A Variational Approach to Estimating Uncertain Parameters in Elliptic Systems

Hans-Werner van Wyk

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Mathematics

Jeff Borggaard, Chair
Martin Day
Terry Herdman
Lizette Zietsman

May 15, 2012
Blacksburg, Virginia

Keywords: Uncertainty quantification, parameter identification, elliptic systems, stochastic collocation methods, variational problems.

A Variational Approach to Estimating Uncertain Parameters in Elliptic Systems

Hans-Werner van Wyk

(ABSTRACT)

As simulation plays an increasingly central role in modern science and engineering research, by supplementing experiments, aiding in the prototyping of engineering systems or informing decisions on safety and reliability, the need to quantify uncertainty in model outputs due to uncertainties in the model parameters becomes critical. However, the statistical characterization of the model parameters is rarely known. In this thesis, we propose a variational approach to solve the stochastic inverse problem of obtaining a statistical description of the diffusion coefficient in an elliptic partial differential equation, based noisy measurements of the model output. We formulate the parameter identification problem as an infinite dimensional constrained optimization problem for which we establish existence of minimizers as well as first order necessary conditions. A spectral approximation of the uncertain observations (via a truncated Karhunen-Lòeve expansion) allows us to estimate the infinite dimensional problem by a smooth, albeit high dimensional, deterministic optimization problem, the so-called 'finite noise' problem, in the space of functions with bounded mixed derivatives. We prove convergence of 'finite noise' minimizers to the appropriate infinite dimensional ones, and devise a gradient based, as well as a sampling based strategy for locating these numerically. Lastly, we illustrate our methods by means of numerical examples.

Acknowledgments

I would like to thank my advisors, Dr. Jeff Borggaard and Dr. Lizette Zietsman, for their unwavering support, guidance and generosity over the past 5 years and for being my family away from home. Thanks to Dr. Max Gunzburger for sparking my interest in uncertainty quantification and suggesting the topic of uncertain parameter estimation. To Dr. Nic van Rensburg for giving me the opportunity to study at Virginia Tech. I would like to thank Dr. Joe Ball for showing me how to get your hands dirty in mathematics, and Dr. John Burns and Dr. George Hagedorn for showing how concrete Functional Analysis can be. My gratitude to the Department of Mathematics at Virginia Tech, to Eileen Shugart for 5 wonderful years of teaching, and to Nicole Sutphin and Ruth Athanson for helping with all the paperwork. I would also like to thank Air Force Office of Scientific Research, for partial funding of my research. To my roommates and many friends who have enriched my life during my stay in Blacksburg: Thomas Weinhart, Raquel Meija, Eun Chang, Kapil Ahuja, Vitor Nunes, Weiwei Hu, Amos Desjardin, Lance Randles, Matt Durham, Xiaojun Wang, Moises Guerra, Edgar Saenz and Idir Mechai. This dissertation is dedicated to my family: Liz van Wyk, Anton van Wyk, Eugene van Wyk, Christof van Wyk and my love Rong Song.

Contents

List of Figures

Notation

General Notation, Deterministic Spaces and Derivatives

D	-	physical domain in $\mathbb{R}^d$, where $d = 1, 2$, or 3.
∂D	-	boundary of domain D
∇	-	gradient in the deterministic variable $x \in D$
α	-	multi-index in $(\mathbb{N} \cup \{0\})^d$; $\lvert\alpha\rvert_1 = \lvert(\alpha_1, ..., \alpha_d)\rvert_1 := \alpha_1 + ... + \alpha_d$
$D_x^\alpha[f(x)]$	-	α^{th} weak derivative with respect to (w.r.t.) the physical variable x
$W^{k,p}(D)$	-	Sobolev space of functions whose k^{th} order weak derivatives are contained in $L^p(D)$
$H^k(D)$	-	Sobolev space $W^{k,2}(D)$
$H_0^1(D)$	-	space of functions in $H^1(D)$ that vanish on the boundary
$H^{-1}(D)$	-	dual space of $H_0^1(D)$
$H(D)$	-	a Hilbert space that embeds continuously in $L^\infty(D)$
$\rightharpoonup, \to$	-	weak convergence and strong convergence, respectively
$\langle \cdot, \cdot \rangle_H$	-	inner product in the Hilbert space H. The underlying space is assumed to be L^2 if no subscript is specified.
$\langle x^*, x \rangle_{\mathscr{X}^* \times \mathscr{X}}$	-	duality pairing representing the action of the dual element $x^* \in \mathscr{X}^*$ on an element $x \in \mathscr{X}$

Probability and Stochastic Spaces

Ω	-	stochastic domain
$\mathcal{F}$	-	a σ-algebra defined on Ω
$\mathbb{P}, d\omega$	-	probability measure on Ω; $d\omega$ denotes integration w.r.t. $\mathbb{P}$.
$\mathbb{E}$	-	expected value with respect to the probability measure $\mathbb{P}$ or $d\omega$
$\mathscr{W}^{k,p}(D)$	-	stochastic Sobolev space $W^{k,p}(D) \otimes L^2(\Omega, d\omega)$
$\mathscr{H}^k(D)$	-	stochastic Sobolev space $\mathscr{W}^{k,2}(D)$

The 'Finite Noise' Approximation

n	-	size of the random vector Y in the 'finite noise' approximation
Y, y	-	a vector of random variables $Y = (Y_1, ..., Y_n)$, while y is the corresponding variable used in integration
Γ^n	-	support of the random vector Y
ρ_n	-	joint density function of the random vector Y
$\mathbf{P}^n$	-	projection operator from $\mathscr{H}_0^1$ onto $H_0^1(D) \otimes \text{span}\{Y_1, ..., Y_n\}$
γ	-	multi-index $\gamma = (\gamma_1, ..., \gamma_n)$ in $(\mathbb{N} \cup \{0\})^n$; $\|\gamma\|_1 = \gamma_1 + ... + \gamma_n$, while $\|\gamma\|_\infty = \max\{\gamma_1, ..., \gamma_n\}$.
D_y^γ	-	γ^{th} derivative with respect to the stochastic variable y, $\gamma \in \mathbb{N}^n$
$\widetilde{W}^{k,p}(D)$	-	'finite noise' stochastic Sobolev space $W^{k,p}(D) \otimes L^2(\Gamma^n, \rho_n)$.
$\widetilde{H}^k(D)$	-	space $\widetilde{W}^{k,2}(D)$.
$H^r_{\text{mix}}(\Gamma^n)$	-	space of bounded mixed derivatives
$\widetilde{H}_{\text{mix}}(D)$	-	tensor product $H^k(D) \otimes H^r_{\text{mix}}(\Gamma^n)$

Optimization and Numerical Implementation

$e(q, u)$	-	equality constraint as a functional equation in $\mathscr{H}_0^1$ or $\widetilde{H}_0^1$
$\tilde{e}(q, u)$	-	equality constraint as a functional equation in $\mathscr{H}^{-1}$ or $\widetilde{H}^{-1}$
$e_n(q, u)$	-	'finite noise' equality constraint formulated in the space $\widetilde{H}_0^1$
$D_q[f(q)]h$	-	Fréchet derivative of f with respect to (w.r.t.) q in direction h
$D_{(q,u)}[f(q,u)](h,v)$	-	Fréchet derivative of f w.r.t. (q, u) in the direction (h, v)
$D_q^2[f(q,u)](h_1,h_2)$	-	second Fréchet derivative $D_q\left[D_q[f(q,u)](h_1)\right](h_2)$
$D_{qu}^2[f(q,u)](h,v)$	-	second Fréchet derivative, $D_u\left[D_q[f(q,u)](h)\right](v)$
$D_{(q,u)}^2[f(q,u)](h,v)^2$	-	second Fréchet derivative, $D_{(q,u)}\left[D_{(q,u)}[f(q,u)](h,v)\right](h,v)$
$y_{\boldsymbol{\ell},\mathbf{j}}$	-	n-dimensional sparse grid point at position $\mathbf{j} = (j_1, ..., j_n)$ and refinement level $\boldsymbol{\ell} = (\ell_1, ..., \ell_n)$
$\psi_{\boldsymbol{\ell},\mathbf{j}}$	-	piecewise multi-linear hat function centered at $y_{\boldsymbol{\ell},\mathbf{j}}$
ϕ_i	-	deterministic finite element basis function
$v^M(x,y)$	-	spatial discretization of function v
$v^{M,N}(x,y)$	-	full discretization of v, in both spatial and stochastic components
$v_+^{M,N}(x,y)$	-	hierarchical surplus of the full discretization $v^{M,N}(x,y)$

Chapter 1

Introduction

An inverse problem can be broadly defined as the process of using observed measurements to infer properties of a physical system of interest. When a mathematical model is used to describe the system, this problem usually amounts to identifying a set of unknown model parameters or initial conditions from measured data. Let $(\mathscr{X}, \|\cdot\|_{\mathscr{X}})$ and $(\mathscr{Y}, \|\cdot\|_{\mathscr{Y}})$ be two normed spaces and $F : \mathscr{X} \to \mathscr{Y}$ be the forward operator, mapping the parameter q to the model output u, which is measured as $\hat{u}$ with some degree of error. The task of inferring information about q from measurements $\hat{u}$ of u then constitutes the inverse problem. Whereas most direct physical models, formulated in the language of ordinary- or partial differential equations, conform to Hadamard's principles of well-posedness, their associated inverse problems are typically non-local as well as non-causal, rendering them ill-posed. Solutions to $F(q) = \hat{u}$ may not exist in the strict sense or may not be unique if F is not bijective, an issue that can be partially resolved by restricting the parameter space $\mathscr{X}$ appropriately and posing the inverse problem in the least squares form

$$\min_{q \in \mathscr{X}} \|F(q) - \hat{u}\|_{\mathscr{Y}}^2. \tag{1.1}$$

A more considerable challenge arises from the possible lack of continuity of the estimated parameter q with respect to the data $\hat{u}$. As a result, a given observation $\hat{u}$, even when measured within only a small margin of error, could have resulted from a wide range of parameter values q. Traditionally, this inherent instability has been addressed through the the use of regularization techniques. These approximate the original inverse problem by one in which the parameter depends continuously on the perturbed data $\hat{u}$, thus ensuring an estimation error in q that is commensurate with the noise level $\delta = \|u - \hat{u}\|_Y$. Tikhonov regularization, the most well-known of these methods, achieves this by adding a penalty term to the least squares cost functional in (1.1), leading to

$$\min_{q \in \mathscr{X}} \frac{1}{2}\|F(q) - \hat{u}\|_{\mathscr{Y}}^2 + \frac{\beta}{2}\|q\|_{\mathscr{X}}^2. \tag{1.2}$$

Under appropriate assumptions on the spaces $\mathscr{X}, \mathscr{Y}$ and the regularity of the forward mapping, it can be shown that the solution q_β^* of (1.2) depends continuously on $\hat{u}$ and that the sequence q_β^* approaches the minimizer q^* of (1.1) as $\beta \to 0^+$ slowly enough (see [18, 28]).

The inverse problem investigated in this dissertation concerns the elliptic system governed by the partial differential equation

$$-\nabla \cdot (q\nabla u) = f \text{ on } D, \quad u = 0 \text{ on } \partial D, \tag{1.3}$$

where $D \subset \mathbb{R}^d$ is the physical domain with piecewise smooth boundary ∂D, q is a spatially varying diffusion coefficient, and f is a given forcing term. Equation (1.3) may describe the steady state flow of fluid with hydraulic head u through a medium with permeability coefficient q or heat conduction across a material with conductivity q and temperature u. The forward problem consists of simulating the hydraulic head (or temperature) u in a material with known diffusion coefficient q and a given forcing term f. Since q cannot be measured directly in practice, it has to be determined indirectly from observations $\hat{u}$ of the model output u, often taking the form of a dataset $\{\hat{u}(x_i)\}_{i=1}^M$ of measurements at given spatial locations $\{x_i\}_{i=1}^M$. In applications, $\hat{u}$ may be treated as a vector or may be interpolated to form a function.

Variational formulations such as (1.2) have been studied extensively for the parameter identification problem related to (1.3). The equation error method, arguably the most direct approach, seeks to obtain q in an admissible set Q_{ad}, by substituting the data $\hat{u}$ for u in (1.3) and solving for q, at least in a least squares sense, i.e.

$$\min_{q \in Q_{\text{ad}}} \frac{1}{2} \left\| -\nabla \cdot (q\nabla \hat{u}) - f \right\|_{H^{-1}}^2 . \tag{1.4}$$

This problem is quadratic in q and can therefore be implemented efficiently, although q might not depend continuously on $\hat{u}$. Output least squares methods offer a more numerically stable alternative, by incorporating the model equation as an equality constraint in the minimization problem

$$\min_{q \in Q_{\text{ad}}} \frac{1}{2} \|u - \hat{u}\|_{H_0^1}^2 + \frac{\beta}{2} R(q), \quad \text{subject to} \qquad -\nabla \cdot (q\nabla u) - f = 0, \tag{1.5}$$

where $R(q)$ is a Tikhonov regularization term. This optimization problem is non-linear in q and is affected considerably by the choice of the function spaces where q is sought and where the equality constraint is imposed. In order to ensure continuity of the parameter-to-output map $q \mapsto u(q)$, q is traditionally assumed to lie in the Sobolev space $H^1(D)$ if the dimension $d = 1$ or $H^2(D)$, if $d = 2, 3$, in which case the regularization term may be taken to be the seminorm $\int_D |\nabla q|^2 \, dx$ or $\int_D |\nabla q|^2 \, dx + \sum_{i_1, i_2} \int_D |D^2_{(x_{i_1}, x_{i_2})} q(x)|^2 \, dx$, respectively [26]. This formulation can also be readily adjusted to incorporate data measured only on a subset of the domain or as a flux on the boundary [33]. A comprehensive treatment of theoretical as well as numerical aspects of the output least squares method is given in the monograph [5], while [16,17,30,54] investigate the stability of (1.5), and [7,19] discuss various regularization techniques.

The augmented Lagrangian method [25,26] is a hybrid approach that combines the equation error method with the output least squares method within a Lagrange multiplier framework.

The algorithm obtains iterates (q_k, u_k) that approximate the output least squares optimal pair (q^*, u^*) by solving a sequence of quadratic minimization problems of the form

$$\min_{(q,u)} \frac{1}{2}\|u - \hat{u}\|^2_{H_0^1} + \frac{\beta}{2}R(q) + \langle \lambda_k, e(q,u)\rangle + \frac{c_k}{2}\|e(q,u)\|^2_{H_0^1}, \tag{1.6}$$

where c_k is a non-decreasing sequence of real numbers, λ_k is a sequence of approximations to the Lagrange multiplier, updated by $\lambda_{k+1} = \lambda_k + c_k e(q_k, u_k)$, and $e(q,u)$ is the equality constraint

$$e(q,u) = (-\Delta)^{-1}(-\nabla \cdot (q\nabla u) - f) \in H_0^1(D), \tag{1.7}$$

interpreted in its weak form. This technique has proven to be numerically efficient and has been continually modified and improved since its inception. Kunisch and Tai [32] investigate the use of operator splitting methods to improve the efficiency of the algorithm, whereas [13, 14] adapt it to the problem of detecting discontinuous diffusion coefficients, modeled as functions in the space of bounded variation. We will use a modified version of the augmented Algorithm in Chapter 4.

Statistical Inverse Problems and Bayesian Methods

When a statistical model for uncertainty in the measurement process or in the dynamical system is available, it may no longer be desirable to treat observations $\hat{u}$ simply as perturbations of the true model output u, but rather to incorporate this additional information into the parameter estimation framework and obtain an approximation not only of the spatial variation of q but also of its probability distribution. Bayesian methods provide a sampling-based approach to statistical inverse problems where the probability distribution of the observations $\hat{u}$ is modeled explicitly.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space. For simplicity, assume the unknown parameter q and the measurement $\hat{u}$, having possibly been discretized or only evaluated at a finite number of points, are realizations of random vectors Q and U in $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively. The presence of uncertainty in the model and/or the measurement process, is incorporated using a random noise vector E in $\mathbb{R}^k$, yielding the observation model $U = F(Q, E)$. For a discussion of the theoretical underpinnings of Bayesian methods in infinite dimensional function spaces, we refer the reader to [60].

Bayesian methods aim to recover the posterior distribution, the probability distribution of the parameter Q conditional on the observed measurement $U = \hat{u}$. Indeed, suppose knowledge of the noise distribution π_e allows us to construct a likelihood function $\pi_{\text{likely}}(u|q)$, a density expressing the likelihood of observing the available data, given a specific value of the parameter. For additive observation models, where $U = f(Q) + E$ and the error E is independent of Q, the likelihood can be easily computed, since $\pi(u|q,e) = \delta(u - F(q) - e)$ and hence formally $\pi_{\text{likely}}(u|q) = \int_{\mathbb{R}^k} \delta(u - F(q) - e)\pi_e(e)\, de = \pi_e(u - F(q))$. Here δ is the Dirac delta function centered at 0. In addition, *assume* that all the information we have

of the variable Q prior to taking measurements can be encoded by a prior density $\pi_{\text{prior}}(q)$. Bayes' Theorem then asserts

$$\pi_{\text{post}}(q|u = \hat{u}) = \frac{\pi_{\text{prior}}(q)\pi_{\text{likely}}(\hat{u}|q)}{\pi(\hat{u})}, \tag{1.8}$$

provided that $\pi(\hat{u}) > 0$. By relating the observation noise in U to the uncertainty associated with the estimated parameter, Bayes' Theorem thus allows us to evaluate the posterior directly through the evaluation of the prior density, assumed given, and the likelihood function, which requires one forward solve in this case. This entirely circumvents the need to solve the classical ill-posed least squares problem. The number $\pi(\hat{u})$ is simply a normalization constant and need not be computed explicitly. The posterior distribution can be used to obtain statistical estimates such as the conditional mean and variance of Q, as well as interval estimates such as confidence bands. Since π_{post} is typically high dimensional, these statistical estimates are usually obtained through sampling schemes such as Monte Carlo methods.

An additive observation model is appropriate when uncertainty in $\hat{u}$ is due solely to variation in the measurement process. However, in many applications, such as the determination of soil porosity from measurements of the pressure head, the measurements may also be spatially correlated. Additionally, uncertainty may be introduced through other spatial parameters, unmodeled multiscale effects, boundary conditions, or forcing terms. In these cases, a more general observation model $U = F(Q, E)$ is necessary, which complicates the construction and evaluation of the likelihood function. Prior distributions are classified according to the biasing effect they have on the posterior distribution. For simplicity, priors that have no great effect are often assumed to be 'conjugate to the posterior,' ensuring that the posterior distribution has the same form as the likelihood function. Informative priors, priors to which the posterior is relatively sensitive, should however be an accurate reflection of the assumptions made on the parameter. Structural information, such as assumptions on the parameter's smoothness, can be encoded through so-called structural priors [27]. Examples include impulse priors, used to model jump discontinuities and Gaussian smoothness priors, although not proper densities, to represent twice continuously differentiable functions.

The trade-off between accuracy and efficiency is reflected in the choice of the observation model leading to the likelihood function, as well as in the choice of the prior. Evaluation of the likelihood can be accelerated by improving the efficiency of forward solves, either through surrogate- or reduced order models [9], or multigrid and multiscale approaches. In [41], prior information of the unknown parameter is used to construct a polynomial chaos approximation of a stochastic forward model, which in turn is used to speed up evaluations of the posterior density directly. Another cost consideration is related to the complexity of the parameter space. When the dimension of the random variable X is tied to pointwise evaluation or to coefficients in a spatial discretization, it can increase considerably with increasing mesh refinement. Attempts to simplify the parameter space include recent work by [42], in which a truncated Karhunen-Loève approximation is computed, based on the prior covariance of the parameter, and by [37] in which a greedy sampling algorithm is used to determine a reduced empirical basis.

Efficient sampling also accelerates the convergence of Bayesian methods by reducing the number of forward solves required to obtain a good estimate of the unknown parameter. In many practical applications, random sampling is achieved by means of Markov chain Monte Carlo (MCMC) techniques, such as the Metropolis-Hastings or Gibbs samplers, that generate a discrete time Markov chain for which π_{post} is the stationary distribution. For given initial value Q_0, a discrete time Markov chain is simply a sequence of random variables $\{Q_n\}_{n=0}^{\infty}$ so that Q_n is independent of $Q_1, ..., Q_{n-2}$ for all $n \in \mathbb{N}$ and the conditional distribution of Q_n, given Q_{n-1}, is specified by the transition probability kernel $P : \mathbb{R}^n \times \mathfrak{B}(\mathbb{R}^n) \to [0, 1]$, which is defined as

$$P(q, A) := \mathbb{P}(Q_n \in A | Q_{n-1} = q) = \int_A \pi_{\text{post}}(q_n | q_{n-1} = q) dq_n$$

for any state $q \in \mathbb{R}^n$ and any Borel set $A \in \mathfrak{B}(\mathbb{R}^n)$. If the transition kernel is aperiodic (preventing the chain from locking into a fixed cycle) and irreducible (allowing the chain to move between any two states with non-zero probability), then it can be shown [46] that the n-step transition probability $P^n(q, A) = \mathbb{P}(Q_n \in A | Q_0 = q)$ approaches the stationary distribution π_{post} in total variation norm and that sample statistics converge almost surely (a.s.) to their respective population statistics, i.e.

$$\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} f(Q_i) = \int_{\mathbb{R}^n} f(q) \pi_{\text{post}}(q) \, dq$$

a.s. on $\mathbb{R}^n$ and for any $f \in L^1(\mathbb{R}^n, \pi_{\text{post}})$. For MCMC methods, the kernel P is unknown and must be estimated as the sample is generated. In the Metropolis-Hastings Algorithm, a candidate generating kernel is chosen initially and the next step in the chain is then either accepted or rejected, based on a computed acceptance ratio [27], whereas the Gibbs sampler uses π_{post} directly to obtain an approximate kernel.

Aperiodicity and irreducibilty of the candidate transition kernel and strict positivity of the acceptance ratio are sufficient conditions for convergence of the Metropolis-Hastings chain, while the Gibbs Algorithm converges if the support of the stationary distribution π_{post} is connected (ruling out certain multimodal densities) and all of its marginals are locally bounded [46]. The rate at which these methods converge, and therefore the size of the sample required to explore π_{post}, is however more difficult to determine [45] and are often studied in terms of the process's distance to stationarity [36]. To avoid undue influence of the initial state Q_0, the chain is usually allowed a certain 'burn in' time before the sample is taken. The central limit Theorem, which guarantees an approximation error of $\frac{\sigma}{\sqrt{N}}$ if the sample $U_n = f(Q_n)$ of size N is independent and identically distributed with variance σ^2, does not apply directly to Markov chains since the process is sequential and hence the sample usually shows some degree of autocorrelation. Parameters such as mixing coefficients play a role in assessing the process's speed of convergence.

Variational Methods for Estimating Uncertain Parameters

Despite the widespread use of the Bayesian methodology, some of the shortcomings mentioned above have led to a continued interest in adapting variational approaches to estimate parameter uncertainty, especially in systems such as (1.3), where parameters are distributed and spatially correlated [4,52,55,63]. Benefits include a well-established infrastructure of existing theory and optimization algorithms, the ability to incorporate structural assumptions on the parameters explicitly into the estimation framework, and clearly defined convergence criteria.

In order to formulate such a least squares identification problem, it is necessary to first describe the propagation of uncertainty in the parameter through the forward model. Suppose the diffusion coefficient q and forcing term f in the stationary elliptic problem (1.3) are random fields $q = q(x,\omega)$ and $f = f(x,\omega)$ associated with a complete probability space $(\Omega, \mathcal{F}, d\omega)$ and satisfying $q(\cdot,\omega) \geq q_{\min} > 0$ and $f(\cdot,\omega) \in H^{-1}(D)$ a.s. on Ω. The corresponding stochastic forward problem consists of finding the random field $u = u(x,\omega)$, so that $u(\cdot,\omega) \in H_0^1(D)$ a.s. on Ω and solves

$$\begin{aligned} -\nabla \cdot (q(x,\omega)\nabla u(x,\omega)) &= f(x,\omega) \qquad && \text{in } D \times \Omega, \\ u(\omega,x) &= 0 && \text{on } \partial D \times \Omega. \end{aligned} \tag{1.9}$$

Equation (1.9) can be interpreted as a family of deterministic forward problems, each corresponding to a certain 'state of the world' ω. Consequently, Monte Carlo sampling methods can be used to obtain a statistical estimate of the random field u.

Stochastic finite element methods [3, 22, 43] provide an efficient alternative to Monte Carlo sampling for estimating u and are based on the weak form of (1.9). If we take u in the stochastic Sobolev space

$$\mathscr{H}_0^1(D) := H_0^1(D) \otimes L^2(\Omega, d\omega),$$

and

$$f \in \mathscr{L}^2(D) := L^2(D) \otimes L^2(\Omega, d\omega),$$

where $L^2(\Omega, d\omega)$ is the set of square integrable functions over Ω, we can formulate the weak form of (1.9), which amounts to finding $u \in \mathscr{H}_0^1(D)$ so that

$$\int_\Omega \int_D q(x,\omega)\nabla u(x,\omega) \cdot \nabla v(x,\omega)\, dx\, d\omega = \int_\Omega \int_D f(x,\omega) v(x,\omega)\, d\omega \quad \forall v \in \mathscr{H}_0^1(D). \tag{1.10}$$

If $q(x,\omega) \geq q_{\min} > 0$ a.s. on $D \times \Omega$ for some non-negative constant $q_{\min}$, then a simple application of the Lax-Milgram Lemma [3] guarantees that a unique solution $u \in \mathscr{H}_0^1(D)$ exists and $\|u\|_{\mathscr{H}_0^1} \leq \frac{1}{q_{\min}} \|f\|_{\mathscr{L}^2}$. Additionally, any solution $u \in \mathscr{H}_0^1$ of (1.9) also solves (1.10).

In numerical approximations of the stochastic forward model, the known parameters $q(x,\omega)$ and $f(x,\omega)$ are usually estimated by functions of a finite number of random variables. If the mean and spatial correlation of the parameters are known, this may be achieved through

spectral expansions such as the Karhunen-Loève (KL) series [38, 56]. Let the mean and covariance functions of a parameter $a(x,\omega) \in L^2(D) \otimes L^2(\Omega, d\omega)$ be given by

$$\mathbb{E}_a(x) = \int_\Omega a(x,\omega)\, d\omega \quad \text{and} \quad C_a(x,x') = \int_\Omega (a(x,\omega) - \mathbb{E}_a(x))(a(x',\omega) - \mathbb{E}_a(x'))\, d\omega$$

respectively. The covariance C_a can now be used to construct a self-adjoint compact kernel operator

$$\mathcal{C}_a : L^2(D) \to L^2(D) | v \in L^2(D) \mapsto \mathcal{C}_a v(x') := \int_D C_a(x,x') \cdot v(x)\, dx$$

with eigenpair sequence $\{(\phi_i, \lambda_i)\}_{i=1}^\infty$, where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \searrow 0$ and $\{\phi_i\}_{i=1}^\infty$ constitutes an orthonormal basis for $L^2(D)$. The KL expansion of the parameter a then takes the form

$$a(x,\omega) = \mathbb{E}_a(x) + \sum_{i=1}^\infty \sqrt{\lambda_i}\phi_i(x) Y_i(\omega), \tag{1.11}$$

where $\{Y_i\}$ is a pairwise uncorrelated sequence of zero mean random variables given by

$$Y_i(\omega) = \frac{1}{\sqrt{\lambda_i}} \int_D (a(x,\omega) - \mathbb{E}_a(x))\phi_i(x)\, dx, \quad i = 1, 2, ...$$

Truncating the series expansion (1.11) yields a best approximation of a in $L^2(D) \otimes L^2(\Omega, d\omega)$.

Suppose that both random fields q and f in (1.10) are approximated by their respective truncated KL expansions and that the random variables $Y_1, Y_2, ..., Y_n$ are jointly distributed with multivariable density function $\rho_n : \Gamma \to [0, \infty)$, where Γ is some hypercube in $\mathbb{R}^n$. It is natural to assume that the random variables $Y_1, ..., Y_i$ associated with q are independent of, and hence uncorrelated with, the random variables $Y_{i+1}, ..., Y_n$ associated with f. The weak form (1.10) can now be expressed in the probability space $(\Omega, \mathcal{F}_n, d\omega)$ where $\mathcal{F}_n = \sigma(\{Y_j\}_{j=1}^n)$ is the sigma algebra generated by $\{Y_1, ..., Y_n\}$. Since the solution u of (1.10) is by default measurable with respect to (w.r.t.) the sigma algebra $\mathcal{F}_n$, the Doob-Dynkin Lemma [8] ensures that $u(x,\omega) = u(x, Y_1(\omega), Y_2(\omega), ..., Y_n(\omega))$, giving rise to the 'finite noise' approximation of the weak form (1.10): Find $u \in \widetilde{H}_0^1(D) := H_0^1(D) \otimes L^2(\Gamma, \rho_n)$ so that

$$\int_\Gamma \int_D q(x,y) \nabla u(x,y) \cdot \nabla v(x,y) \rho_n(y)\, dx\, dy = \int_\Gamma \int_D f(x,y) v(x,y) \rho_n(y)\, dx\, dy \tag{1.12}$$

for all $v \in \widetilde{H}_0^1(D)$, where $y = (y_1, ..., y_n) \in \Gamma$.

In stochastic Galerkin methods, spatial discretization of u is effected by standard finite element methods, while discretization in the stochastic variable y can take a variety of forms, all of which address the potentially high dimension of the support Γ of ρ_n, also known as the 'curse of dimensionality.' Once finite dimensional approximating subspaces, both in

the physical as well stochastic components, are chosen together with their respective sets of basis functions $\{\phi_i\}$ and $\{\psi_j\}$, the solution u may be approximated by

$$u_{IJ}(x,y) = \sum_{i=1}^{I}\sum_{j=1}^{J} c_{ij}\phi_i(x)\psi_j(y).$$

A Galerkin projection of (1.12) onto these subspaces yields an algebraic system that may be solved for the coefficients c_{ij}, $i = 1, ..., I$, $j = 1, ..., J$. If the random variables $Y_1, ..., Y_n$ are independent, their joint density can be written as the product of their marginals $\rho_n(y) = \prod_{i=1}^{n} \rho^{(i)}(y_i)$. This decoupling is exploited by polynomial chaos methods [22, 62], which expand u in terms of a Wiener-Askey chaos basis of mutually orthonormal global polynomials. This approach requires *a priori* computation of various integrals of products of stochastic basis functions and results in a full algebraic system, but also gives a high degree of accuracy. Stochastic collocation methods [47, 48] on the other hand, are based on the fact that the statistical quantities of interest related to the solution u are ultimately of importance and hence the stochastic basis functions are constructed with the aim of approximating the high dimensional integrals appearing both in the variational form (1.12) and in the computation of the statistics. Global interpolating Lagrange polynomials, centered at so-called sparse quadrature points are therefore used as basis functions. If the same quadrature scheme is used to approximate the integrals in (1.12) and to compute statistical quantities of interest related to u, this method is equivalent to evaluating a family of deterministic forward problems, each corresponding to a certain quadrature point, and can therefore be regarded as a sort of sampling scheme. Even though the effectiveness of this basis is not reliant on the independence of $Y_1, ..., Y_n$, the accuracy of the associated numerical quadrature scheme used in the weak form (1.10) as well as in the computation of the statistical quantities of interest, depends on the smoothness of u as a function of the random variables $Y_1, ..., Y_n$. For (1.12), the smoothness of the parameters q and f is sufficient to guarantee the necessary smoothness of the model output u. In cases where such smoothness cannot be guaranteed, it may be more effective to replace the global interpolating Lagrange basis by a high dimensional sparse grid hierarchical finite element type basis with local support, which allows for a certain degree of spatial adaptivity [12, 23, 39].

In this dissertation, we propose an output least squares version of the stochastic parameter identification problem that is based on the weak form (1.10). Suppose we have a statistical model of the measured data $\hat{u}$ and the forcing term f in the form of random fields $\hat{u}(x,\omega) \in \mathscr{H}_0^1(D)$ and $f(x,\omega) \in \mathscr{L}^2(D)$. We then seek to identify the random diffusion coefficient q in terms of $\hat{u}$, as the minimizer of the regularized least squares functional

$$\begin{aligned} &\min_{(q,u)\in\mathscr{H}\times\mathscr{H}_0^1} J(q,u) := \frac{1}{2}\|u-\hat{u}\|^2_{\mathscr{H}_0^1} + \frac{\beta}{2}R(q) \\ &s.t.\ \ q \in Q_{\text{ad}}, \quad e(q,u) = 0, \end{aligned} \tag{P}$$

where the equality constraint $e(q,u) = 0$ represents (1.10) and the Tikhonov regularization term $R(q) > 0$ is added to ensure continuous dependence of the minimizer on the data $\hat{u}$. Here $\mathscr{H} := H(D) \otimes L^2(\Omega)$, where, as in the deterministic case, $H(D)$ is any Hilbert space

that embeds continuously in $L^\infty(D)$. Usually H is taken to be $H^1(D)$ when $d = 1$ or $H^2(D)$ when $d = 2, 3$ (see [26]), and the feasible set Q_{ad} is given by

$$Q_{\text{ad}} = \{q \in \mathscr{H}(D) : 0 < q_{\min} \leq q(x,\omega) \text{ a.s. on } D \times \Omega,\ \|q(\cdot,\omega)\|_H \leq q_{\max} \text{ a.s. on } \Omega\}.$$

While the uniform coercivity condition $0 < q_{\min} \leq q$ ensures that (1.10) has a unique solution, the norm constraint $\|q(\cdot,\omega)\|_H \leq q_{\max}$ contributes broadly to the regularity of the variational problem (P). For instance, under the continuous embedding $H(D) \hookrightarrow L^\infty(D)$ this condition implies $q(x,\omega) \leq q_{\max}$ a.s. on $D \times \Omega$ and hence

$$\begin{aligned} |e(q, u_1 - u_2)v| &= \left| \int_\Omega \int_D q(x,\omega) \nabla \left(u_1(x,\omega) - u_2(x,\omega)\right) \cdot \nabla v \, dx \, d\omega \right| \\ &\leq q_{\max} \int_\Omega \int_D |\nabla \left(u_1(x,\omega) - u_2(x,\omega)\right)| |\nabla v| \, dx \, d\omega \\ &\leq q_{\max} \|u_1 - u_2\|_{\mathscr{H}_0^1} \|v\|_{\mathscr{H}_0^1} \end{aligned}$$

for any two functions $u_1, u_2 \in \mathscr{H}_0^1$ and any test function $v \in \mathscr{H}_0^1$. The equality constraint is therefore continuous in u.

The variational formulation (P) poses a number of theoretical, as well as computational challenges. Whereas $e(q,u)$ is bounded in u, the lack of smoothness of the random field $q = q(x,\omega)$ in its stochastic component ω limits the regularity of $e(q,u)$ as a function of q, making it impossible to use linearization (as in the analogous deterministic case) to establish traditional first order necessary optimality conditions. Likewise, the uniform enforcement of the parameter constraints, while necessary for the well-posedness and regularity of the minimization problem (P), leaves the set of admissible parameters Q_{ad} with an empty interior. Although it is nevertheless possible to establish a saddle point condition for minimizers of (P), based on Hahn-Banach separation arguments, applying standard optimization strategies to locate a minimizer q^* of (P) seems unfeasible.

To deal with this lack of regularity, we approximate the infinite dimensional problem (P) by a more regular 'finite noise' problem. As in the stochastic forward problem, we use methods such as the KL expansion to estimate the input functions, in this case $\hat{u}$ and f, by functions $\hat{u}^n(x, Y_1, ..., Y_i)$ and $f^n(x, Y_{i+1}, ..., Y_n)$ of a finite number of random variables $Y_1, ..., Y_n$ with support Γ^n and joint density ρ_n. The corresponding 'finite noise' discretization of problem (P) then seeks $q = q(x, Y_1, ..., Y_n)$ in terms of $\hat{u}^n$ and f^n as the minimizer of the cost functional.

$$\begin{aligned} &\min_{q \in \widetilde{H}} J^n(q,u) := \frac{1}{2} \int_{\Gamma^n} \|u(\cdot,y) - \hat{u}^n(\cdot,y)\|_{H_0^1}^2 \rho_n(y) dy + \frac{\beta}{2} R_n(q) \\ &s.t.\ q \in Q_{\text{ad}}^n,\ e_n(q,u) = 0, \end{aligned} \tag{P^n}$$

where the admissible set

$$Q_{\text{ad}}^n := \{q(x,y) : 0 < q_{\min} \leq q(x,y) \text{ a.s. on } D \times \Gamma^n,\ \|q(\cdot,y)\|_H \leq q_{\max} \text{ a.s. on } \Gamma^n\},$$

the equality constraint $e_n(q, u)$ represents the finite noise forward problem (1.12), and $R_n(q)$ is an appropriate regularization term. Since q is estimated in terms of $\hat{u}^n$ and f^n, its stochastic complexity is directly tied to that of the measured/known data, which can be quantified explicitly.

As in the deterministic inverse problem (1.5), the space in which q is sought has a strong impact on the well-posedness of (P^n). The lack of smoothness of parameters $q \in H(D) \otimes L^2(\Gamma_n, \rho_n)$ in their stochastic component gives rise to the same complications in regularity as for the infinite noise problem (P). Moreover, the most significant hurdle from a computational point of view is the need to approximate high dimensional integrals when evaluating the cost functional J^n, dealing with the equality constraint (1.10) or computing statistical quantities of interest. Monte Carlo type schemes seem inefficient, especially when compared with Bayesian methods. High-dimensional quadrature techniques are necessary. Many such methods are based on sparse grid interpolation by high order global polynomials, requiring a high degree of smoothness for optimal convergence rates to be attained. Unlike in the 'finite noise' forward model however, the ill-posedness of (P) precludes inference of any smoothness of q in y, based on smoothness properties of the input variables $\hat{u}^n$ and f^n and therefore explicit assumptions on the smoothness are necessary. Ideally, the 'finite noise' parameter space should be chosen so that

1. the 'finite noise' minimizer q_n^* of (P^n) approaches the original minimizer q^* of (P) as $\hat{u}^n \to \hat{u}$ and $f^n \to f$,
2. the resulting regularity of the 'finite noise' optimization problem (P^n) allows for the derivation of more traditional first order necessary optimality conditions and the use of gradient based optimization schemes, and
3. problem (P^n) lends itself to numerical discretizations that scale relatively well with increasing dimension n.

We will make use of the space

$$\widetilde{H}^{r,n}_{\text{mix}}(D) := H(D) \otimes H^r_{mix}(\Gamma_n),$$

where $H^r_{mix}(\Gamma_n)$ is the space of functions with bounded mixed derivatives [12]. We will show in this dissertation that $\widetilde{H}^{r,n}_{\text{mix}}(D)$ indeed fulfills all three criteria mentioned above. Not only does $\widetilde{H}^{r,n}_{\text{mix}}(D)$ embed continuously in $L^\infty(D \times \Omega)$, ensuring that the equality constraint $e_n(q, u)$ is twice Fréchet differentiable and hence allowing for more orthodox optimization approaches, but functions in $H^r_{mix}(\Gamma)$ are also commonly associated with numerical discretization schemes based on hierarchical finite elements, approximations known for their effectiveness in mitigating the 'curse of dimensionality' [10–12, 50]. Hierarchical finite elements also lend themselves well to adaptive quadrature schemes. Their implementation using hierarchical surplusses has been proven effective in reducing the cost of approximating high dimensional integrals [10, 39].

This dissertation discusses various theoretical as well as numerical aspects of the variational parameter estimation problem introduced in (P) and is structured as follows: Chapter 2 introduces the theoretical foundations of the 'infinite noise' problem (P). After a brief overview of stochastic Sobolev spaces and variational forms of the stochastic forward model, we turn to the well-posedness of (P) after which we establish first order necessary optimality conditions in the form of a saddle point condition. Chapter 3 discusses the approximation of the 'infinite noise' problem (P) by the 'finite noise' problem (P^n). The convergence proof makes use of regularization theory for non-linear least squares methods [7] and hinges on the dual role played by the random variables $\{Y_1, ..., Y_n\}$, acting simultaneously as the subset of a complete orthonormal basis as well as a vector of variables in Γ_n. In Chapter 4, existence as well as more traditional first order necessary optimality conditions are established for the 'finite noise' problem (P^n), while Chapter 5 explores various optimization strategies to find a minimizer. Finally, we will illustrate our results by including some numerical experiments in Chapter 6, after which we provide some concluding remarks.

Chapter 2

A Variational Approach to Uncertainty in Parameter Identification

2.1 Introduction

This chapter provides some necessary background material and a preliminary discussion of the variational problem (P). Stochastic Sobolev spaces are tensor product spaces that are convenient for describing the regularity of random fields both in their spatial- as well as their stochastic components. We formulate the weak form (1.10) of the stochastic forward problem (1.9) within this infrastructure, analyze its regularity, and establish its solution's continuous dependence on the parameters q and f in the appropriate norms. The variational formulation (P), in which the forward problem (1.10) appears as the equality constraint $e(q, u) = 0$, is however defined over the Hilbert space $\mathscr{H}(D) \times \mathscr{H}_0^1(D)$, in whose topology $e(q, u)$ is not Fréchet differentiable as a function of the parameter q. We show that (P) nevertheless has minimizers (q^*, u^*) that satisfy the saddle point condition (2.9).

2.2 Partial Differential Equations with Random Coefficients

2.2.1 Stochastic Sobolev Spaces

The quantities of interest encountered in probabilistic descriptions of systems with distributed parameters generally have two distinct components. A deterministic component, stemming from the physical model of the system and a stochastic component associated with the stochastic variability arising from the uncertainty inherent in the model parameters. A natural class of functions with which to describe quantities of this type is the tensor product

of function spaces used in the theory of differential- or partial differential equations, such as Sobolev spaces, with function spaces occurring in the analysis of random variables, most notably the $L^p(\Omega, d\omega)$ spaces. This section offers a brief overview of the elementary theory of tensor products of Hilbert spaces, with special emphasis placed on stochastic Sobolev spaces.

Tensor Products of Hilbert Spaces

Let $(H_1, \langle\cdot,\cdot\rangle_1)$ and $(H_2, \langle\cdot,\cdot\rangle_2)$ be two Hilbert spaces and define for each $\phi_1 \in H_1$, $\phi_2 \in H_2$ the bounded bilinear form $\phi_1 \otimes \phi_2 : H_1 \times H_2 \to \mathbb{C}$, mapping vectors (φ, ψ) to

$$\phi_1 \otimes \phi_2(\varphi, \psi) := \langle \phi_1, \varphi\rangle_1 \langle \phi_2, \psi\rangle_2$$

The collection $\mathscr{E}$ of finite sums $u = \sum_i \varphi_i \otimes \psi_i$ can be equipped with the inner product

$$\langle u, \hat{u}\rangle_{H_1 \otimes H_2} := \sum_{i,j} \langle \varphi_i, \hat{\varphi}_j\rangle_1 \langle \psi_i, \hat{\varphi}_j\rangle_2$$

The Hilbert space tensor product $H_1 \otimes H_2$ is defined to be the completion of $\mathscr{E}$ under this inner product and is therefore also a Hilbert space. It can also be shown [51] that if $\{v_k\}$ and $\{\varphi_j\}$ form orthonormal bases for H_1 and H_2, respectively, then $\{v_k \otimes \varphi_j\}$ forms an orthonormal basis for $H_1 \otimes H_2$ and hence separability is preserved. In the special case where one of the constituent Hilbert spaces is the space $L^2(\Omega, d\omega)$ of square integrable random variables, assumed here to be defined on some underlying probability space $(\Omega, \mathcal{F}, d\omega)$ and to be separable with orthonormal basis $\{v_k\}$, the tensor product takes on a special form. Let H be any separable Hilbert space with orthonormal basis $\{\varphi_j\}$ and consider the space $\mathscr{H} := L^2(\Omega, d\omega; H)$ defined as follows:

Definition 2.2.1. *The space $L^2(\Omega, d\omega; H)$ is the collection of H-valued functions $f : \Omega \to H$ such that*

i) f is measurable, i.e. $\langle y, f(\cdot)\rangle_H$ is measurable w.r.t. the σ-algebra $\mathcal{F}$ for all $y \in H$, and

ii) $\|f\|^2_{\mathscr{H}} := \int_\Omega \|f(\omega)\|^2_H d\omega < \infty$.

Remark 2.2.2. *Condition i) implies that the mappings $\omega \mapsto \|f(\omega)\|^2_H$ and $\omega \mapsto \langle f(\omega), g(\omega)\rangle$ are measurable functions in the traditional sense.*

This space, also known as the space of vector valued Bochner Integrable functions, is known to be unitarily equivalent to the tensor product $L^2(\Omega) \otimes H$ [51].

Proposition 2.2.3. *The space $\mathscr{H} := L^2(\Omega, d\omega; H)$ is a separable Hilbert space with inner product*

$$\langle f, g\rangle_{\mathscr{H}} := \int_\Omega \langle f(\omega), g(\omega)\rangle_H d\omega = \sum_{j=1}^{\infty} \int_\Omega \langle f(\omega), \varphi_j\rangle_H \langle g(\omega), \varphi_j\rangle_H d\omega \tag{2.1}$$

and orthonormal basis $\{v_k(\omega)\varphi_j\}_{j,k=1}^{\infty}$. Moreover, the mapping $U : L^2(\Omega) \otimes H \to \mathscr{H} : f \otimes \varphi \mapsto f(\omega)\varphi$ defines a unitary isomorphism and hence $L^2(\Omega) \otimes H \equiv \mathscr{H}$.

The following examples of tensor products of $L^2(\Omega, d\omega)$ with various other function spaces will feature prominently in future chapters.

Example 2.2.4 (The Product of L^2 Spaces)**.** *If the Hilbert space H is also a separable L^2 space then, in addition to the unitary equivalence described in Proposition 2.2.3, the tensor product $L^2(\Omega_1, d\omega_1) \otimes L^2(\Omega_2, d\omega_2)$ can also be shown to be isometric to the space $L^2(\Omega_1 \times \Omega_2, d\omega_1 \otimes d\omega_2)$ of square integrable multivariate functions with respect to the product measure $d\omega_1 \otimes d\omega_2$ [51], i.e.*

$$L^2(\Omega_1, d\omega_1) \otimes L^2(\Omega_2, d\omega_2) \equiv L^2(\Omega_1, d\omega_1; L^2(\Omega_2, d\omega_2)) \equiv L^2(\Omega_1 \times \Omega_2, d\omega_1 \otimes d\omega_2).$$

The unitary transformations that identify their respective bases are given by $v_k \otimes u_j \mapsto v_k(\omega)u_j$ and $v_k(\omega)u_j \mapsto \{v_k(\omega_1)u_j(\omega_2)$ respectively.

Example 2.2.5 (The space $\mathscr{H}^k(D)$)**.** *Let $D \subset \mathbb{R}^d$ be any bounded domain and suppose that $f : D \times \Omega \to \mathbb{R}$ is a function in $L^2(D \times \Omega, dx \otimes d\omega)$ such that for each $\omega \in \Omega$, the deterministic function $f(\cdot, \omega) : D \to \mathbb{R}$ is weakly differentiable in the sense of distributions. The Hilbert space $\mathscr{H}^k(D) := L^2(\Omega, d\omega; H^k(D))$, $k \in \mathbb{N}$ is defined explicitly as*

$$\mathscr{H}^k(D) = \{v : \Omega \to H^k(D) | v \text{ measurable}, \|v\|_{\mathscr{H}^k} := \sum_{|\alpha| \leq k} \int_\Omega \int_D |D_x^\alpha v(x, \omega)|^2 dx \, d\omega < \infty\},$$

where α denotes the multi-index $\alpha = (\alpha_1, ..., \alpha_d) \in \mathbb{N}^n$, and is equipped with inner product

$$\langle u, v \rangle_{\mathscr{H}^k} := \sum_{|\alpha| \leq k} \int_\Omega \int_D D_x^\alpha u(x, \omega) D_x^\alpha v(x, \omega) dx \, d\omega = \mathbb{E}(\langle u, v \rangle_{H^k}).$$

The subspace $\mathscr{H}_0^1(D) = H_0^1(D) \otimes L^2(\Omega, d\omega) \subset \mathscr{H}^1(D)$, where $H_0^1(D)$ is the H^1-closure of $C_c^\infty(D)$, the space of infinitely differentiable functions with compact support, is defined analogously. Moreover, since the physical domain D is bounded, a direct application of the Poincaré inequality guarantees that the norm $\|\cdot\|_{\mathscr{H}_0^1}$ on $H_0^1(D)$, induced by the inner product

$$\langle u, v \rangle_{\mathscr{H}_0^1} := \int_\Omega \int_D \nabla u(x, \omega) \cdot \nabla v(x, \omega) \, dx \, d\omega,$$

is equivalent to the norm $\|.\|_{\mathscr{H}^1}$ on $\mathscr{H}_0^1(D)$. More generally, for any square integrable weighting function $q(x, \omega) \geq q_{min} > 0$, we can define the subspace

$$\mathscr{H}_{0,q}^1(D) = \left\{ v \in \mathscr{H}_0^1(D) : \|v\|_{\mathscr{H}_{0,q}^1}^2 := \int_D q(x, \omega) |\nabla v(x, \omega)|^2 \, dx \, d\omega < \infty \right\}.$$

This weighted norm is particularly useful in stochastic elliptic problems of the form (1.10), in which the diffusion coefficient $q(x, \omega)$ is an unbounded random field such as a Gaussian or Exponential. Convergence in this weighted norm always implies convergence in the $\|\cdot\|_{\mathscr{H}_0^1}$-norm, whereas the converse only holds when q is uniformly bounded, i.e. $0 < q_{min} \leq q(x, \omega) \leq q_{max} < \infty$ a.s. on $D \times \Omega$, where q_{min}, q_{max} are constants.

More General Stochastic Sobolev Spaces

The definition of stochastic Sobolev spaces can be extended to the non-Hilbert space setting by means of the following construction, which generalizes weak derivatives to the stochastic setting. Let $f \in L^1_{loc}(D \times \Omega, dx \otimes d\omega)$ and $\phi \in C_c^\infty(D)$ be a test function on the physical domain. Then the linear mapping $\Lambda_{\mathbb{E}(f)} : C_c^\infty(D) \to \mathbb{R}$, given by

$$\Lambda_{\mathbb{E}(f)}(\phi) := \int_\Omega \int_D f(x,\omega)\phi(x)dxd\omega, \tag{2.2}$$

can be used to define a distribution. Indeed, since $f(\cdot,\omega)$ gives rise to a regular distribution for each $\omega \in \Omega$, it follows that for any convergent sequence of test functions $\phi_n \to 0$, $\lim_{n\to\infty} \int_D f(x,\omega)\phi_n(x)dx = 0$. Moreover, the sequence ϕ_n is uniformly bounded over some compact subset of D ([53] p.154) and therefore

$$\int_\Omega \sup_n \left| \int_D f(x,\omega)\phi_n(x)dx \right| d\omega < \infty.$$

The convergence $\Lambda_{\mathbb{E}(f)}(\phi_n) \to \Lambda_{\mathbb{E}(f)}(\phi)$ now follows from the Dominated Convergence Theorem. Since Fubini's Theorem allows for interchanging the order of integration in (2.2), it follows that the expected value function $\mathbb{E}(f(x)) = \int_\Omega f(x,\omega)d\omega$ defines a regular distribution on D. Suppose that $f(\cdot,\omega) : D \to \mathbb{R}$ has an α^{th} weak derivative, where $\alpha = (\alpha_1, ...\alpha_d) \in \mathbb{N}^d$. Then

$$\begin{aligned}
\Lambda_{\mathbb{E}(D_x^\alpha f)}(\phi) &:= \int_\Omega \int_D D_x^\alpha f(x,\omega)\phi(x)dxd\omega && \because \text{by (2.2)} \\
&= (-1)^\alpha \int_\Omega \int_D f(x,\omega) D_x^\alpha \phi(x)dxd\omega && \because \text{by definition of } D_x^\alpha f(x) \\
&= (-1)^\alpha \int_D \int_\Omega f(x,\omega) D_x^\alpha \phi(x)d\omega dx =: \Lambda_{D_x^\alpha \mathbb{E}(f)}(\phi) && \because \text{Fubini's Theorem}
\end{aligned}$$

and hence $\mathbb{E}(f(x))$ also has an α^{th} weak derivative and $D_x^\alpha \mathbb{E}(f(x)) = \mathbb{E}(D_x^\alpha f(x))$, where the term on the left can be interpreted in the traditional sense while the term on the right is understood in terms of (2.2).

Example 2.2.6 (The space $\mathscr{W}^{k,p}(D)$)**.** *We now define*

$$\mathscr{W}^{k,p}(D) := L^p(\Omega, d\omega; W^{k,p}(D)), \text{ where } 1 \le p \le \infty,$$

as the space of all measurable functions $v : D \times \Omega \to \mathbb{R}$ so that $v(\cdot,\omega) : D \to \mathbb{R}$ has an α^{th} weak derivative for all $|\alpha| \le k$ a.s. on Ω and for which the norm

$$\begin{aligned}
\|v\|_{\mathscr{W}^{k,p}} &:= \mathbb{E}(\|v\|^p_{\mathscr{W}^{k,p}})^{\frac{1}{p}} = \mathbb{E}\left(\sum_{|\alpha|\le k} \int_D |D_x^\alpha v|^p dx \right)^{\frac{1}{p}}, \qquad 1 \le p < \infty, \\
\|v\|_{\mathscr{W}^{k,\infty}} &:= \max_{|\alpha|\le k} (\operatorname{ess\,sup}_{D\times\Omega} |D_x^\alpha v|)
\end{aligned}$$

is finite. It follows directly from Fubini's Theorem and Jensen's inequality that for any function $f \in \mathscr{W}^{k,p}(D)$, the expectation function $\mathbb{E}(f(x))$ is contained in the corresponding deterministic Sobolev space $W^{k,p}(D)$. In particular, for any $f \in \mathscr{H}^k(D)$,

$$\begin{aligned}\|\mathbb{E}(f)\|_{H^k} &:= \sum_{|\alpha|\leq k} \int_D |D_x^\alpha \mathbb{E}(f(x))|^2 \; dx = \sum_{|\alpha|\leq k} \int_D |\mathbb{E}\left(D_x^\alpha f(x)\right)|^2 \; dx \\ &\leq \sum_{|\alpha|\leq k} \int_D \mathbb{E}\left(|D_x^\alpha f(x)|^2\right) \; dx = \sum_{|\alpha|\leq k} \int_\Omega \int_D |D_x^\alpha f(x)|^2 \; dx \; d\omega \\ &= \|f\|_{\mathscr{H}^k} < \infty\end{aligned}$$

and thus $\mathbb{E}(f) \in H^k(D)$.

2.2.2 Elliptic Boundary Value Problems with Random Coefficients

This section discusses some properties of the stochastic forward model (1.9) that will be useful when incorporating it as an equality constraint in the variational framework (P). For this purpose, it is more convenient to work with the weak form (1.10). Under the assumption

$$0 < q_{\min} \leq q(x,\omega) \leq q_{\max} < \infty \;\text{ a.s. on } D \times \Omega \text{ for some } q_{\min}, q_{\max} \in \mathbb{R}, \tag{A1}$$

the bilinear form $B_q : \mathscr{H}_0^1(D) \times \mathscr{H}_0^1(D) \to \mathbb{R}$, given by

$$(u,v) \mapsto B_q(u,v) := \int_\Omega \int_D q(x,\omega) \nabla u(x,\omega) \cdot \nabla v(x,\omega) \; dx \; d\omega$$

is bounded. Therefore, it defines a bounded linear mapping

$$A_q : \mathscr{H}_0^1(D) \to \mathscr{H}^{-1} |\; u \mapsto A_q(u) := B_q(u,\cdot)$$

and consequently (1.10) can simply be written as the functional equation $A_q(u) = f$ in $\mathscr{H}^{-1}$, where f is understood as the bounded linear functional in $\mathscr{H}^{-1}$, defined as

$$v \in \mathscr{H}_0^1(D) \mapsto \int_\Omega \int_D f(x,\omega) v(x,\omega) \; dx \; d\omega.$$

Moreover, the coercivity condition $0 < q_{\min} \leq q(x,\omega)$ a.s. on $D \times \Omega$ ensures, via the Lax Milgram Theorem [35] that A_q has a bounded inverse, which simply maps f to the unique solution (1.10). For convenience, we will use the notation $-\nabla \cdot (q \nabla u) := A_q(u)$, not to be confused with the classical formulation (1.9). The equality constraint (1.10) can then be written abstractly in $\mathscr{H}^{-1}(D)$ as

$$\widetilde{e}(q,u) := -\nabla \cdot (q \nabla u) - f = 0. \tag{2.3}$$

It may also be useful to write this equality constraint as a functional equation in $\mathscr{H}_0^1(D)$. To this end, we apply the inverse Laplacian $(-\Delta)^{-1} := A_1$ to obtain an equivalent equality constraint

$$e(q,u) := (-\Delta)^{-1} \left(-\nabla \cdot (q \nabla u) - f\right) = 0. \tag{2.4}$$

Remark 2.2.7. *The assumption on the upper bound of the diffusion coefficient $q(x,\omega)$ can also be eliminated at the cost of restricting the solution- and test space of the weak form (1.10) to the weighted Sobolev space $\mathscr{H}^1_{0,q}(D)$ introduced in Example 2.2.5. For the purposes of parameter identification, however, the dependence of the solution's norm on the underlying (unknown) diffusion parameter may lead to unnecessary complications. We will restrict ourselves to parameters that are uniformly bounded and coercive a.s. on Ω.*

Differentiability of $e(q,u)$

The mapping $e(\cdot,\cdot) : \mathscr{H}(D) \times \mathscr{H}^1_0(D) \to \mathscr{H}^1_0(D)$ is readily seen to be affine linear in both arguments. Moreover, as discussed in Chapter 1, for a fixed parameter q satisfying assumption (A1), the map $e(q,\cdot) : \mathscr{H}^1_0(D) \to \mathscr{H}^1_0(D)$ is bounded and therefore Fréchet differentiable in u, with derivative

$$D_q[e(q,u)](v) = (-\Delta)^{-1}(-\nabla\cdot(q\nabla v)) \text{ in direction } v.$$

The mapping $e(\cdot,u) : \mathscr{H} \to \mathscr{H}^1_0(D)$ with fixed u is, however, not as well-behaved when q is assumed to be an L^2-function in the ω-variable. The following simple counter-example shows that it is not even bounded.

Example 2.2.8. *Suppose $D = [0,1]$, $\Omega = [\alpha,1]$ and $D_x u(x,\omega) = \omega$ for $\alpha > 0$ and let $q_n(x,\omega) = \omega^{\frac{1}{2}(-1+\frac{1}{n})}$. Evidently $q_n \to 0$ in L^2. In order for $e(q_n,u) \to 0$ in $\mathscr{H}^{-1}$, the integral*

$$I_n = \int_\alpha^1 \int_0^1 q_n D_x u D_x v dx d\omega$$

is required to approach 0 uniformly in v. By choosing v so that $D_x[v_n(x)] = \frac{q_n}{\sqrt{1-\alpha^{\frac{1}{n}}}}$, we see

$$I_n = \frac{1-\alpha^{1+\frac{1}{n}}}{\sqrt{1-\alpha^{\frac{1}{n}}}} \nrightarrow 0 \quad \text{as } n \to \infty.$$

This endemic lack of smoothness in $q \in \mathscr{H}$ rules out the direct use of linearization techniques for optimization problems in which the elliptic equation enters as an equality constraint. Local minimizers of these optimization problems can thus not be characterized via the traditional Karusch-Kuhn-Tucker conditions [31, 44].

The lack of regularity of the equality constraint in q shown above, relates directly to the choice of the parameter space $\mathscr{H}$. Evidently,

$$\|e(q-\hat{q},u)\|_{\mathscr{H}^1_0} \leq \|q-\hat{q}\|_{\mathscr{L}^\infty}\|u\|_{\mathscr{H}^1_0} \qquad \text{whenever } q,\hat{q} \in \mathscr{L}^\infty(D) := L^\infty(\Omega, L^\infty(D)).$$

Moreover, it was shown in [3] that if, in addition, $q,\hat{q}$ are uniformly coercive and $f,\hat{f} \in \mathscr{H}^{-1}(D)$ are any two forcing terms, the difference in the corresponding solutions u and $\hat{u}$ of Equation (1.10) can be bounded by

$$\|u-\hat{u}\|_{\mathscr{H}^1_0} \leq \frac{C_D}{q_{\min}}\left(\|f-\hat{f}\|_{\mathscr{H}^{-1}} + \frac{1}{q_{\min}}\|q-\hat{q}\|_{L^\infty(\Omega;L^\infty(D))}\|f\|_{\mathscr{H}^{-1}}\right). \tag{2.5}$$

The diffusion coefficient q may also lie in the weaker space $\mathscr{L}^{2p}(D) = L^{2p}(\Omega, L^{2p}(D))$ with $p \in (1,\infty)$, at the cost of certain additional regularity requirements on the solution u. For any $r \in (1,\infty)$ so that $\frac{1}{p} + \frac{1}{r} = 1$, suppose in addition that $u, \hat{u} \in \mathscr{W}^{1,2r}(D) := L^{2r}(\Omega; W^{1,2r}(D))$. Then

$$\|e(q-\hat{q},u)\|_{\mathscr{H}_0^1} \leq \|q-\hat{q}\|_{\mathscr{L}^{2p}} \|u\|_{\mathscr{W}^{1,2r}}$$

and moreover

$$\|u-\hat{u}\|_{\tilde{H}_0^1} \leq \frac{C_D}{q_{\min}} \left(\|f-\hat{f}\|_{\tilde{H}^{-1}} + \frac{1}{q_{\min}} \|q-\hat{q}\|_{L^{2p}(\Omega;L^{2p}(D))} \|u\|_{L^{2r}(\Omega,W^{1,2r}(D)} \right).$$

The additional spatial regularity of u can be obtained by requiring that q and f are smooth enough in the spatial variable x, uniformly over Ω. Since any solution $u \in \mathscr{H}_0^1(D)$ of (1.10) also solves the parametrized family of deterministic boundary value problems (1.9), each of which is assured to have a unique solution under assumption (A1), we are able to address questions regarding the regularity of the model solution u in the spatial and stochastic variables separately. In particular, spatial smoothness conditions on the model parameters that ensure the spatial regularity of the solution for the deterministic problem can be reused for the parameterized problem as long as they are enforced uniformly in the random space. If, for instance, the assumptions

$$\text{The boundary } \partial D \text{ is piecewise } C^2\text{-smooth, and} \tag{A2}$$

$$\exists K < \infty \text{ such that } q(\cdot,\omega) \in C^1(\bar{D}),\ \max_D |\nabla q(x,\omega)| \leq K \text{ a.s. on } \Omega \tag{A3}$$

hold in addition to (A1), then it follows readily from the regularity theory for deterministic second order elliptic partial differential equations [20] that the solution u of (1.10) satisfies $u(\cdot,\omega) \in H^2(D) \cap H_0^1(D)$ a.s on Ω. Moreover, there exists a constant $C > 0$, independent of ω, such that

$$\|u(\cdot,\omega)\|_{H^2(D)} \leq C \|f(\cdot,\omega)\|_{H^{-1}} \quad \text{a.s. on } \Omega. \tag{2.6}$$

This in turn implies that $u \in \mathscr{H}^2(D) \cap \mathscr{H}_0^1(D)$ and $\|u\|_{\mathscr{H}^2} \leq C\|f\|_{\mathscr{H}^{-1}}$. More generally, if we assume that for $m \geq 0$

$$\partial D \text{ is piecewise } C^{m+2} \text{ and } f(\cdot,\omega) \in H^k(D), \text{ and} \tag{$A2^{(m)}$}$$

$$\exists K_\alpha > 0 \text{ so that } q(\cdot,\omega) \in C^{m+1}(D),\ |D_x^\alpha q(\cdot,\omega)| \leq K_\alpha,\ |\alpha| \leq k \text{ a.s. on } \Omega \tag{$A3^{(m)}$}$$

then $u(\cdot,\omega) \in H^{m+2}(D)$ a.s. on Ω ([20] p. 317) and we have the estimate

$$\|u(\cdot,\omega)\|_{H^{m+2}} \leq C\left(\|f(\cdot,\omega)\|_{H^k} + \|u(\cdot,\omega)\|_{\tilde{L}^2}\right) \qquad \text{a.s. on } \Omega.$$

Existence of higher order moments of the solution u can now be deduced directly from the stochastic integrability of f, upon the use of Poincaré's inequality. If $f \in L^p(\Omega, H^k)$, then $u \in L^p(\Omega, H^{m+2})$.

Higher order regularity can also be achieved under slightly milder conditions on q. Indeed suppose assumptions (A2) and (A3) are replaced by

$$\text{The boundary } \partial D \text{ is piecewise smooth with curvature bounded below, and} \tag{A2'}$$

$$q \in \mathscr{W}^{1,p}(D),\ p > d \text{ and } \exists K < \infty \ni \|q(\cdot,\omega)\|_{W^{1,p}} \leq K \text{ a.s. on } \Omega, \tag{A3'}$$

then ([34] pp. 185-189), $u \in \mathscr{H}^2 \cap \mathscr{H}_0^1$ and $\|u\|_{\mathscr{H}^2} \leq C\|f\|_{\mathscr{H}^{-1}}$.

Remark 2.2.9. *The uniform boundedness of the deterministic norm $\|q(\cdot,\omega)\|_{W^{1,p}}$ is essential to ensure higher order regularity of u. If we only assume that $q \in \mathscr{W}^{1,p}(D)$, $p > d$, then there might be some $\omega \in \Omega$ for which $\|q(\cdot,\omega)\|_{W^{1,p}}$ is unbounded. In this case, the deterministic regularity cannot be exploited. The set of ω on which this occurs is, however, arbitrarily small. Indeed, by Chebychev's Theorem, the event $A_\epsilon = \{\omega \in \Omega : \|q(\cdot,\omega)\|_{W^{q,p}} \geq \sqrt{\frac{K}{\epsilon}}\}$ has probability $\mathbb{P}(A_\epsilon) \leq \epsilon$ and on its complement $\Omega \backslash A_\epsilon$, $u(\cdot,\omega) \in H^2(D) \cap H_0^1(D)$.*

Despite the equality constraint's lack of regularity in $q \in \mathscr{H}$, we continue to make use of this parameter space in our analysis, to take advantege of its Hilbert space structure. Other parameter spaces may be used, although the corresponding variational problem will need to be investigated in a general Banach space setting.

2.3 The Parameter Identification Problem

In this section, we verify the existence of a minimizer to problem (P)

$$\begin{aligned} \min_{(q,u)\in\mathscr{H}\times\mathscr{H}_0^1} J(q,u) &:= \frac{1}{2}\|u-\hat{u}\|^2_{\mathscr{H}_0^1} + \frac{\beta}{2}\|q\|^2_{\mathscr{H}} \\ s.t.\ \ q &\in Q_{\text{ad}},\ e(q,u) = 0, \end{aligned} \tag{P}$$

where $Q_{\text{ad}} = \{q \in \mathscr{H}(D) : 0 < q_{\min} \leq q(x,\omega) \text{ a.s. on } D\times\Omega,\ \|q(\cdot,\omega)\|_H \leq q_{\max} \text{ a.s. on } \Omega\}$, and establish first order necessary optimality conditions that are satisfied by minimizers. The lack of smoothness in the equality constraint $e(q,u)$ as a function of $q \in \mathscr{H}$ precludes the use of its linearization and instead we rely on a Hahn-Banach separation argument to establish a saddle point condition satisfied by minimizers of (P). From here on, we assume for the sake of simplicity that the forcing term $f \in L^2(D)$ in (1.10) is deterministic.

2.3.1 Existence of Minimizers

The following lemmas establish the weak compactness of the feasible set Q_{ad}, the continuity of the solution mapping $q \mapsto u(q)$ restricted to Q_{ad}, as well as the weak closedness of its graph in the $\mathscr{H}$ norm. These results will be used in sequel, not only to prove the existence of solutions to (P), but also in showing that (P) can be approximated by a sequence of more tractable 'finite noise' problems (see Chapter 3).

Lemma 2.3.1. *The set Q_{ad} is closed, convex, and hence weakly compact in $\mathscr{H}$.*

Proof. Recall that

$$Q_{\text{ad}} = \{q \in \mathscr{H} : q(x,\omega) \geq q_{\min} > 0, \ \text{ a.s. on } D\times\Omega,\ \|q(\cdot,\omega)\|_H \leq q_{\max} \ \text{ a.s. on } \Omega\}.$$

Convexity is easily verified. In order to show that Q_{ad} is closed, let $\{q_n\}$ be a sequence in Q_{ad} converging to $q \in \mathscr{H}$, i.e.

$$\|q_n - q\|^2_{\mathscr{H}} = \int_\Omega \|q_n(\cdot,\omega) - q(\cdot,\omega)\|^2_H \, d\omega \to 0 \ \text{ as } n \to \infty.$$

Since convergence in $L^2(\Omega, d\omega)$ implies pointwise almost sure convergence of a subsequence $\{q_{n_k}\} \subset Q_{\text{ad}}$ on Ω, it follows that

$$\|q_{n_k}(\cdot,\omega) - q(\cdot,\omega)\|_H \to 0 \quad \text{a.s. on } \Omega, \text{ as } k \to \infty.$$

Additionally, $\|q_{n_k}(\cdot,\omega)\|_H \leq q_{\max}$ a.s. on Ω for $k \in \mathbb{N}$ and therefore q also satisfies this constraint. Finally, $H(D)$ embeds continuously in $L^\infty(D)$, from which it follows that the subsequence $\{q_{n_k}\}$ converges to q pointwise a.s. on $D \times \Omega$, ensuring that q also satisfies the pointwise constraint $q(x,\omega) \geq q_{\min}$ a.s. on $D \times \Omega$. Thus $q \in Q_{\text{ad}}$. □

Lemma 2.3.2. *The mapping $u : q \in Q_{\text{ad}} \mapsto u(q) \in \mathscr{H}_0^1$ defined by (1.10) is continuous.*

Proof. Suppose $q_n \to q$ in Q_{ad} as $n \to \infty$. As in the proof of the previous lemma, there exists a subsequence $q_{n_k} \to q$ pointwise a.s. on $D \times \Omega$. The upper bound (2.5) on the function u ensures that

$$\|u(q_{n_k}) - u(q)\|_{\mathscr{H}_0^1} \leq \left(\frac{C_D \|f\|_{L^2}}{q_{\min}^2}\right) \|q_{n_k} - q\|_{L^\infty(\Omega; L^\infty(D))} \to 0 \quad \text{as } n \to \infty,$$

where C_D is the constant appearing in the Poincaré inequality on D. In order to show $u(q_n) \to u(q)$, suppose to the contrary that there is a subsequence $\{q_{n_k}\}$ of $\{q_n\}$ and a constant $\delta > 0$ for which $\|u(q_{n_k}) - u(q)\|_{\mathscr{H}_0^1} \geq \delta$ for all $k \in \mathbb{N}$. By the arguments given above, this subsequence itself has a further subsequence say $\{q_{n_j}\}$ for which $\lim_{j\to\infty} u(q_{n_j}) = u(q)$, leading to a contradiction. □

Lemma 2.3.3. *The graph $\{(q,u) \in \mathscr{H} \times \mathscr{H}_0^1 : q \in Q_{\text{ad}},\ u = u(q)\}$ of u is weakly closed.*

Proof. Let q_n be a sequence in Q_{ad}, so that $q_n \rightharpoonup q$ in $\mathscr{H}$ and $u(q_n) \rightharpoonup u$ in $\mathscr{H}_0^1$. The weak compactness of Q_{ad} shown in Lemma 2.3.1, directly implies $q \in Q_{\text{ad}}$. It now remains to be shown that $u = u(q)$ or equivalently that u solves $e(q,u) = 0$. Since $u_n = u(q_n)$ and hence $e(q_n, u_n) = 0$ for all $n \geq 1$, i.e.

$$\int_\Omega \int_D q_n \nabla u_n \cdot \nabla v \, dx \, d\omega = \int_\Omega \int_D f \, v \, dx \, d\omega \qquad \text{for all } v \in \mathscr{H}_0^1, \tag{2.7}$$

it suffices to show that there exists a subsequence $\{q_{n_k}\}$ of $\{q_n\}$ so that

$$\lim_{k\to\infty} \int_\Omega \int_D q_{n_k} \nabla u_{n_k} \cdot \nabla v \, dx \, d\omega = \int_\Omega \int_D q \nabla u \cdot \nabla v \, dx \, d\omega \qquad \text{for all } v \in \mathscr{H}_0^1.$$

This would imply

$$\int_\Omega \int_D q \nabla u \cdot \nabla v \, dx \, d\omega = \lim_{k\to\infty} \int_\Omega \int_D q_{n_k} \nabla u_{n_k} \cdot \nabla v \, dx \, d\omega = \int_\Omega \int_D f \, v \, dx \, d\omega$$

for all $v \in \mathscr{H}_0^1$ and hence $e(q,u) = 0$. Now for any $n \geq 1$ and $v \in \mathscr{H}_0^1$,

$$\begin{aligned} \int_\Omega \int_D (q_n \nabla u_n - q^* \nabla u^*) \cdot \nabla v \, dx \, d\omega = &\int_\Omega \int_D (q_n - q^*) \nabla u_n \cdot \nabla v \, dx \, d\omega \\ &+ \int_\Omega \int_D q^* \nabla (u_n - u^*) \cdot \nabla v \, dx \, d\omega. \end{aligned} \tag{2.8}$$

Let $\{q_{n_k}\}$ be the subsequence of $\{q_n\}$ that converges to q pointwise a.s. on $D \times \Omega$, as guaranteed by Lemma 2.3.1. We can then bound the first term by

$$\begin{aligned}
&\left| \int_\Omega \int_D (q_{n_k} - q^*) \nabla u_{n_k} \cdot \nabla v \, dx \, d\omega \right| \\
&\leq \left(\int_\Omega \int_D |q_{n_k} - q^*| |\nabla u_{n_k}|^2 \, dx \, d\omega \right)^{\frac{1}{2}} \left(\int_\Omega \int_D |q_{n_k} - q^*| |\nabla v|^2 \, dx \, d\omega \right)^{\frac{1}{2}} \\
&\leq 2q_{\max} \frac{1}{q_{\min}} \|f\|_{L^2} \left(\int_\Omega \int_D |q_{n_k} - q^*| |\nabla v|^2 \, dx \, d\omega \right)^{\frac{1}{2}} \to 0 \quad \text{as } n_k \to \infty,
\end{aligned}$$

by the Dominated Convergence Theorem, since the integrand is bounded above by $2q_{\max}\|v\|_{\mathscr{H}_0^1}$.

The second term in (2.8) converges to 0 due to the weak convergence $u^n \rightharpoonup u^*$ and the fact that the mapping $\|\cdot\|_{q^*} : u \mapsto \|u\|_{q^*} := \int_\Omega \int_D q^* |\nabla u|^2 \, dx \, d\omega$ defines a norm that is equivalent to $\|\cdot\|_{\mathscr{H}_0^1}$, by virtue of the fact that $0 < q_{\min} \leq q(x, \omega) \leq q_{\max} < \infty$. □

By combining Lemmas 2.3.1, 2.3.2, and 2.3.3, we can show that a solution q^* of the infinite dimensional minimization problem (P) exists for any $\beta \geq 0$.

Theorem 2.3.4 (Existence of Minimizers)**.** *For each $\beta \geq 0$, the problem (P) has a minimizer.*

Proof. Let (q_n, u_n) be a minimizing sequence for the cost functional J over $Q_{\text{ad}} \times \mathscr{H}_0^1$, i.e.

$$\inf_{(q,u) \in Q_{\text{ad}} \times \mathscr{H}_0^1} J(q, u) = \lim_{n \to \infty} J(q_n, u_n) = \lim_{n \to \infty} \frac{1}{2} \|u_n - \hat{u}\|_{\mathscr{H}_0^1}^2 + \frac{\beta}{2} \|q_n\|_{\mathscr{H}}^2.$$

Since u_n satisfies the equality constraint $e(q_n, u_n) = 0$, and consequently $\|u_n\|_{\mathscr{H}_0^1} \leq \frac{1}{q_{\min}} \|f\|_{L^2}$ for all $n \geq 1$ (Lax-Milgram), the Banach-Alaoglu Theorem guarantees the existence of a weakly convergent subsequence $u_{n_k} \rightharpoonup u^* \in \mathscr{H}_0^1(D)$. Moreover, the weak compactness of Q_{ad} established in Lemma 2.3.1 also yields a subsequence $q_{n_k} \rightharpoonup q^*$ as $k \to \infty$, so that $q^* \in Q_{\text{ad}}$. The fact that the infimum of J is attained at the point (q^*, u^*) follows directly from the weak lower semicontinuity of norms [51]. Indeed,

$$\begin{aligned}
J(q^*, u^*) &\leq \liminf_{n_k \to \infty} \frac{1}{2} \|u_{n_k} - \hat{u}\|_{\mathscr{H}_0^1}^2 + \liminf_{n_k \to \infty} \frac{\beta}{2} \|q_{n_k}\|_{\mathscr{H}}^2 \\
&\leq \liminf_{n_k \to \infty} \left(\frac{1}{2} \|u_{n_k} - \hat{u}\|_{\mathscr{H}_0^1}^2 + \frac{\beta}{2} \|q_{n_k}\|_{\mathscr{H}}^2 \right) = \inf_{(q,u) \in Q_{\text{ad}} \times \mathscr{H}_0^1} J(q, u).
\end{aligned}$$

Finally, it follows directly from Lemma 2.3.3 that $u^* = u^*(q^*)$ and hence u^* satisfies the inequality constraint $e(q^*, u^*) = 0$. The regularization term was not required to show the existence of minimizers. □

2.3.2 A Saddle Point Condition

Although solutions to (P) exist, the inherent lack of smoothness of q in the stochastic variable ω complicates the establishment of traditional necessary optimality conditions. This section follows [15] in deriving a weaker saddle point condition for the optimizer (q^*, u^*) of (P) with the help of a Hahn-Banach separation argument.

Let $\langle \cdot, \cdot \rangle$ denote the $L^2(D \times \Omega)$ inner product. For any triple $(q, u, \lambda) \in \mathscr{H} \times \mathscr{H}_0^1 \times \mathscr{H}_0^1$, we define the Langrangian functional by

$$\begin{aligned} L(q, u, \lambda) &= J(q, u) + \langle \widetilde{e}(q, u), \lambda \rangle_{\mathscr{H}^{-1} \times \mathscr{H}_0^1} = J(q, u) + \langle e(q, u), \lambda \rangle_{\mathscr{H}_0^1} \\ &= \frac{1}{2} \|u - \hat{u}\|^2_{\mathscr{H}_0^1} + \frac{\beta}{2} \|q\|^2_{\mathscr{H}} + \langle q \nabla u, \nabla \lambda \rangle - \langle f, \lambda \rangle. \end{aligned}$$

The main theorem of this subsection follows.

Theorem 2.3.5 (Saddle Point Condition)**.** *Let $(q^*, u^*) \in Q_{\text{ad}} \times \mathscr{H}_0^1$ solve problem (P). Then there exists a Lagrange multiplier $\lambda^* \in \mathscr{H}_0^1$ so that the saddle point condition*

$$L(q^*, u^*, \mu) \leq L(q^*, u^*, \lambda^*) \leq L(q, u, \lambda^*) \tag{2.9}$$

holds for all $(q, u, \mu) \in Q_{\text{ad}} \times \mathscr{H}_0^1 \times \mathscr{H}_0^1$.

Proof. Note that the second inequality simply reflects the optimality of (q^*, u^*). To obtain the first inequality, we rely on a Hahn-Banach separation argument. Let

$$S = \{(J(q, u) - J(q^*, u^*) + s, e(q, u)) \in \mathbb{R} \times \mathscr{H}_0^1 : (q, u) \in Q_{\text{ad}} \times \mathscr{H}_0^1, s \geq 0\}$$

and

$$T = \{(-t, 0) \in \mathbb{R} \times \mathscr{H}_0^1 : t > 0\}$$

In the ensuing three lemmas we will show that

1. S and T are convex (Lemma 2.3.6),
2. $S \cap T = \emptyset$ (Lemma 2.3.7), and
3. S has at least one interior point (Lemma 2.3.8).

The Hahn-Banach Theorem thus gives rise to a separating hyperplane, i.e. a pair $(\alpha_0, \lambda_0) \neq (0, 0)$ in $\mathbb{R} \times \mathscr{H}_0^1$, such that

$$\alpha_0(J(q, u) - J(q^*, u^*) + s) + \langle e(q, u), \lambda_0 \rangle_{\mathscr{H}_0^1} \geq -t\alpha_0 \quad \forall t > 0, s \geq 0, (q, u) \in Q_{\text{ad}} \times \mathscr{H}_0^1. \tag{2.10}$$

Letting $s = t = 1$ and $(q, u) = (q^*, u^*)$ readily yields $\alpha_0 \geq 0$. In fact $\alpha_0 > 0$. Suppose to the contrary that $\alpha_0 = 0$. Then by (2.10)

$$\langle e(q, u), \lambda_0 \rangle_{\mathscr{H}^{-1} \times \mathscr{H}_0^1} = \langle q \nabla u, \nabla \lambda_0 \rangle - \langle f, \lambda_0 \rangle \geq 0 \quad \forall (q, u) \in Q_{\text{ad}} \times \mathscr{H}_0^1.$$

If $q = q^*$ and $u \in \mathscr{H}_0^1$ is chosen to satisfy $\langle q^* \nabla u, \nabla \phi \rangle - \langle f - \lambda_0, \phi \rangle = 0 \ \forall \phi \in \mathscr{H}_0^1$, we have

$$\langle q^* \nabla u, \nabla \lambda_0 \rangle - \langle f, \lambda_0 \rangle = \langle f - \lambda_0, \lambda_0 \rangle - \langle f, \lambda_0 \rangle = -\langle \lambda_0, \lambda_0 \rangle \geq 0,$$

which implies that $\lambda_0 = 0$. This contradicts the fact that $(\alpha_0, \lambda_0) \neq (0, 0)$. Dividing (2.10) by α_0 and letting $\lambda^* = \lambda_0/\alpha_0$ yields

$$J(q^*, u^*) \leq J(q, u) + \langle e(q, u), \lambda^* \rangle_{\mathscr{H}_0^1} \ \text{ for all } (q, u) \in Q_{\text{ad}} \times \mathscr{H}_0^1$$

and hence

$$\begin{aligned} L(q^*, u^*, \mu) &= J(q^*, u^*) + \langle e(q^*, u^*), \mu \rangle_{\mathscr{H}_0^1} = J(q^*, u^*) \\ &\leq J(q, u) + \langle e(q, u), \lambda^* \rangle_{\mathscr{H}_0^1} = L(q, u, \lambda^*) \end{aligned}$$

for all $(q, u, \mu) \in Q_{\text{ad}} \times \mathscr{H}_0^1 \times \mathscr{H}_0^1$. □

Lemma 2.3.6. *The sets S and T are convex.*

Proof. Clearly, T is convex. Let $P_1, P_2 \in S$ and consider the convex combination $P_\alpha = \alpha P_1 + (1 - \alpha) P_2$ where $0 \leq \alpha \leq 1$. Hence P_α is of the form $P_\alpha = (p_\alpha, w_\alpha)$ where

$$\begin{aligned} p_\alpha &= \alpha(J(q_1, u_1) - J(q^*, u^*) + s_1) + (1 - \alpha)(J(q_2, u_2) - J(q^*, u^*) + s_2) \\ w_\alpha &= \alpha e(q_1, u_1) + (1 - \alpha) e(q_2, u_2) \end{aligned}$$

with $q_1, q_2 \in Q_{\text{ad}},\ u_1, u_2 \in \mathscr{H}_0^1,$ and $s_1, s_2 \geq 0$. It now remains to show that $w_\alpha = e(q_\alpha, u_\alpha)$ for some $(q_\alpha, u_\alpha) \in Q_{\text{ad}} \times \mathscr{H}_0^1$ and $p_\alpha = J(q_\alpha, u_\alpha) - J(q^*, u^*) + s_\alpha$ for some $s_\alpha \geq 0$. Let $q_\alpha = \alpha q_1 + (1 - \alpha) q_2 \in Q_{\text{ad}}$ and let $u_\alpha \in \mathscr{H}_0^1$ be the unique solution of the variational problem

$$\langle q_\alpha \nabla u_\alpha, \nabla \phi \rangle = \langle \alpha q_1 \nabla u_1 + (1 - \alpha) q_2 \nabla u_2, \nabla \phi \rangle \qquad \forall \phi \in \mathscr{H}_0^1.$$

Therefore

$$\begin{aligned} \langle w_\alpha, \phi \rangle_{\mathscr{H}^{-1}, \mathscr{H}_0^1} &= \langle \alpha q_1 \nabla u_1 + (1 - \alpha) q_2 \nabla u_2, \nabla \phi \rangle - \langle f, \phi \rangle \\ &= \langle q_\alpha \nabla u_\alpha, \nabla \phi \rangle - \langle f, \phi \rangle = \langle e(q_\alpha, u_\alpha), \phi \rangle_{\mathscr{H}_0^1} \qquad \forall \phi \in \mathscr{H}_0^1, \end{aligned}$$

which implies that $w_\alpha = e(q_\alpha, u_\alpha)$. Moreover, it follows readily from the convexity of norms that

$$J(q_\alpha, u_\alpha) \leq \alpha J(q_1, u_2) + (1 - \alpha) J(q_2, u_2)$$

and therefore letting

$$\begin{aligned} s_\alpha &= \alpha J(q_1, u_1) + (1 - \alpha) J(q_2, u_2) - J(q_\alpha, u_\alpha) + \alpha s_1 + (1 - \alpha) s_2 \\ &\geq \alpha s_1 + (1 - \alpha) s_2 \geq 0 \end{aligned}$$

we obtain

$$p_\alpha = J(q_\alpha, u_\alpha) - J(q^*, u^*) + s_\alpha.$$

□

Lemma 2.3.7. *The sets S and T are disjoint.*

Proof. This result follows directly from the fact that $J(q,u) \geq J(q^*,u^*)$ for all (q,u) in $Q_{\text{ad}} \times \mathscr{H}_0^1$. □

Lemma 2.3.8. *The set S has a non-empty interior.*

Proof. Clearly $(s_0, 0) = (J(q^*,u^*) - J(q^*,u^*) + s_0, e(q^*,u^*)) \in S$ for any $s_0 > 0$. For any $\epsilon \in (0,1)$, let (s,w) belong to the ϵ-neighborhood of $(s_0,0)$. In other words $|s-s_0| + \|w\|_{\mathscr{H}_0^1} \leq \epsilon$. Let $q = q^*$ and let u be the solution to the problem

$$\langle q^* \nabla u, \nabla v\rangle = \langle f, v\rangle + \langle \nabla w, \nabla v\rangle, \quad \forall v \in \mathscr{H}_0^1(D). \tag{2.11}$$

Then the functional $\Lambda_w \in \mathscr{H}^{-1}$, defined by $\Lambda_w \phi = \langle \nabla w, \nabla \phi\rangle$, is equal to $e(q^*,u)$

$$\begin{aligned} s' &:= s_0 + J(q^*,u^*) - J(q,u) = s_0 + J(q^*,u^*) - J(q^*,u) \\ &= s_0 + \frac{1}{2}\int_\Omega \int_D |\nabla(u^*(x,\omega) - \hat{u}(x,\omega))|^2 dx d\omega - \frac{1}{2}\int_\Omega \int_D |\nabla(u(x,\omega) - \hat{u}(x,\omega))|^2 \, dx \, d\omega \\ &= s_0 - \frac{1}{2}\int_\Omega \int_D \nabla(u(x,\omega) - u^*(x,\omega)) \cdot \nabla(u(x,\omega) + u^*(x,\omega) - 2\hat{u}(x,\omega)) \, dx \, d\omega. \end{aligned}$$

Now u^* solves $e(q^*,u^*) = 0$ and hence $\|u^*\|_{\mathscr{H}_0^1} \leq \frac{C_D}{q_{\min}}\|f\|_{L^2}$ by Lax-Milgram. Similarly, since u solves (2.11), it follows that $\|u\|_{\mathscr{H}_0^1} \leq \frac{C_D}{q_{\min}}(\|f\|_{L^2} + \|w\|_{\mathscr{H}_0^1}) \leq \frac{C_D}{q_{\min}}(\|f\|_{L^2} + \epsilon)$ and hence $\|u - u^*\|_{\mathscr{H}_0^1} \leq \frac{C_D}{q_{\min}}\epsilon$. We therefore have

$$\begin{aligned} s' &\geq s_0 - \frac{1}{2}\|u-u^*\|_{\mathscr{H}_0^1}(\|u^*\|_{\mathscr{H}_0^1} + \|u\|_{\mathscr{H}_0^1} + 2\|\hat{u}\|_{\mathscr{H}_0^1}) \\ &\geq s_0 - \frac{\epsilon C_D}{2q_{\min}}(\frac{C_D}{q_{\min}}\|f\|_{L^2} + \frac{C_D}{q_{\min}}(\|f\|_{L^2} + \epsilon) + 2\|\hat{u}\|_{\mathscr{H}_0^1}) \\ &\geq s_0 - \frac{\epsilon C_D}{2q_{\min}^2}(2C_D\|f\|_{L^2} + C_D\epsilon + 2q_{\min}\|\hat{u}\|_{\mathscr{H}_0^1}) \geq 0 \end{aligned}$$

for small enough $\epsilon > 0$. Therefore $(s,w) = (J(q^*,u) - J(q^*,u^*) + s', e(q^*,u)) \in S$ for any (s,w) in a small enough ϵ-neighborhood of $(s_0,0)$. □

Chapter 3

Regularization and Approximation by the Finite Noise Problem

Although solutions to (P) exist and satisfy the saddle point condition (2.9), locating these directly by means of optimization strategies that are based on first order approximations, is problematic. In addition to the lack of differentiability of the equality constraint discussed in the previous chapters, the admissible set of parameters Q_{ad}, though physically justifiable, has an empty interior in the $\mathscr{H}$-norm. These difficulties arise as a direct consequence of the inherent lack of smoothness of functions q, u and $\hat{u}$ in the stochastic variable ω. In this chapter, we present an approximation to problem (P) by a 'finite noise' optimization problem (P^n), in which q is permitted to be a smooth, albeit high-dimensional, function in x and a random vector $Y = (Y_1, ..., Y_n)$. As a result, the accompanying equality constraint is twice Fréchet differentiable and the admissible set has a non-empty interior, allowing for the use of more traditional gradient-based optimization approaches. This approximation results from a truncation $\hat{u}^n$ in the Karhunen-Loève series [38,56] of the observed data $\hat{u}$, which may be treated as a perturbation of $\hat{u}$ in the stochastic Sobolev space $\mathscr{H}_0^1$ and can therefore be analyzed conveniently within the setting of regularization methods. A complication arises from the fact that the search space and its associated regularization for problem (P^n) differ from those of related to (P). We exploit of the dual role played by the sequence $\{Y_i\}_{i=1}^n$ of random variables in the KL expansion of $\hat{u}$, acting both as functions in an orthonormal basis for $L^2(\Omega)$ and as a set of variables in the 'finite noise' problem (P^n), to partially resolve this discrepancy by showing that the sequence of minimizers of the 'finite noise' problem have a subsequence converging weakly to a minimizer of (P) as $n \to \infty$. We then show that the search space in the 'finite noise problem' (P^n) satisfies the requisite regularity needed to ensure differentiability of the equality constraint in the parameter q and proceed to establish more traditional, gradient-based first order necessary optimality conditions for minimizers of (P^n). The convergence framework not only informs the choice of numerical discretization schemes used, but also suggests the use of dimension-adaptive methods.

3.1 Stochastic Approximation

According to [56], the Karhunen-Loève expansion given in (1.11) can also be applied in cases where the deterministic component space is a general Hilbert space. In particular, for the measured output $\hat{u} \in \mathscr{H}_0^1(D) = H_0^1(D) \otimes L^2(\Omega, d\omega)$ the expectation and covariance functions

$$\hat{u}_0(x) := \mathbb{E}_{\hat{u}}(\hat{u}(x)), \qquad C_{\hat{u}}(x, x') = \int_\Omega (\hat{u}(x, \omega) - \hat{u}_0(x)) (\hat{u}(x', \omega) - \hat{u}_0(x')) \, d\omega$$

define a self-adjoint, compact covariance operator

$$\mathcal{C}_{\hat{u}} : \mathscr{H}_0^1(D) \to \mathscr{H}_0^1(D) : \phi \mapsto \mathcal{C}_{\hat{u}}(x') = \int_D \nabla_x C_{\hat{u}}(x, x') \cdot \nabla_x \phi(x) \, dx,$$

whose spectral decomposition is given in terms of the eigenpair sequence (λ_i, ϕ_i), where $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n \searrow 0$ and $\{\phi_i\}_{i=1}^\infty$ constitutes an orthonormal basis for $\mathscr{H}_0^1(D)$. This eigenpair sequence can then be used to construct the corresponding KL series of $\hat{u}$

$$\hat{u}(x, \omega) = u_0(x) + \sum_{i=1}^\infty \sqrt{\lambda_i} \phi_i(x) Y_i(\omega), \tag{3.1}$$

where $\{Y_n\}_{n=1}^\infty$ is an uncorrelated (orthonormal) sequence of random variables with zero mean and unit variance, given by

$$Y_i(\omega) = \frac{1}{\sqrt{\lambda_i}} \int_D (\hat{u}(x, \omega) - \mathbb{E}_{\hat{u}}(x)) \phi_i(x) \, dx, \quad i = 1, 2, ... \ .$$

Similar to (1.11), the truncated KL series

$$\hat{u}^n(x, \omega) = u_0(x) + \sum_{i=1}^n \sqrt{\lambda_i} \phi_i(x) Y_i(\omega)$$

gives a best approximation to $\hat{u}$ over finite dimensional subspaces U of $H_0^1(D)$, i.e.

$$\|\hat{u} - \hat{u}^n\|_{\mathscr{H}_0^1}^2 = \inf_{\substack{U \subset H_0^1 \\ \dim(U) = n}} \|\hat{u} - \mathbf{P}_{U \otimes L^2} \hat{u}\|_{\mathscr{H}_0^1}^2 = \sum_{i=n+1}^\infty \lambda_i^2 \to 0 \text{ as } n \to \infty,$$

where $\mathbf{P}_{U \otimes L^2}$ is the projection operator from $\mathscr{H}_0^1$ onto the subspace $U \otimes L^2$.

Remark 3.1.1. *Although the random variables $\{Y_i(\omega)\}_{i=1}^\infty$ form an orthonormal set in $L_0^2(\Omega, d\omega)$, the set of square integrable functions with zero mean, it is not clear* a priori *that they actually span the entire $L_0^2(\Omega, d\omega)$. Since problem (P) aims to identify the unknown parameter q in terms of $\hat{u}$, we assume without loss of generality that the σ-algebra $\mathcal{F}$ is generated by $\hat{u}$ and hence $\{Y_i(\omega)\}_{i=1}^\infty$ forms a basis for $L_0^2(\Omega, d\omega)$. The unknown diffusion parameter $q \in \mathscr{H}(D)$ can then be written in terms of this basis as $q(x, \omega) = q_0(x) + \sum_{i=1}^\infty q_i(x) Y_i(\omega)$.*

In addition, we make the following assumptions on the statistical distribution of the spanning basis $Y_i(\omega)$, that allow for the representation of the probability measure by means of a density function.

Assumption 3.1.2. *Assume that for any n the probability measure of the random vector $Y(\omega) = (Y_1(\omega), ..., Y_n(\omega))$ is absolutely continuous with respect to the Lebesgue measure and hence Y has a joint density function $\rho_n : \Gamma^n \to [0, \infty)$, defined on the hyperrectangle $\Gamma^n \subset \mathbb{R}^n$.*

We can now formulate the least squares parameter estimation problem for the perturbed data function $\hat{u}^n$. Since $\hat{u}^n$ depends on ω only through the intermediary variables $\{Y_i\}_{i=1}^n$, it is reasonable to also estimate the unknown parameter q_n as a function of these intermediary variables, i.e.

$$q_n(x, \omega) = q_n(x, Y_1(\omega), ..., Y_n(\omega)) \in \widetilde{H}(D) := H(D) \otimes L^2(\Gamma^n, \rho_n).$$

The corresponding model output, depending continuously on the parameter q_n, would then also be a function of $Y = (Y_1, ..., Y_n)$, according to the Doob-Dynkin Lemma. The equality constraint now takes the form of the 'finite noise' forward problem (1.12), rewritten here for convenience

$$\int_{\Gamma^n} \int_D q(x, y) \nabla u(x, y) \cdot \nabla v(x, y) \rho_n(y) \, dx \, dy - \int_{\Gamma^n} \int_D f(x, y) v(x, y) \rho_n(y) \, dx \, dy = 0,$$

for all $v \in \widetilde{H}_0^1(D)$ with corresponding strong parametric form

$$\begin{aligned} -\nabla \cdot (q(x, y) \nabla u(x, y)) - f(x) = 0, \quad \text{a.s. on } D \times \Gamma^n, \\ u(x, y) = 0, \quad \text{a.s. on } \partial D \times \Gamma^n, \end{aligned} \tag{3.2}$$

where $y = (y_1, y_2, ..., y_n) \in \Gamma^n$ are the variables of integration corresponding to the random vector $Y = (Y_1, ..., Y_n)$. As in Subsection 2.2.2, we can also express the constraint as an abstract equation in function space, by defining

$$A_q^n : \widetilde{H}_0^1(D) \to \widetilde{H}^{-1}(D) | u \mapsto A_q^n(u)v = \int_{\Gamma^n} \int_D q(x, y) \nabla u(x, y) \cdot \nabla v(x, y) \, \rho_n(y) \, dx \, dy$$

and writing

$$\widetilde{e}_n(q, u) = -\nabla \cdot (q \nabla u) - f = 0,$$

where $-\nabla \cdot (q \nabla u) := A_q^n(u)$. The analogue $e_n(q, u) = 0$ of $e(q, u) = 0$ then takes the form

$$e_n(q, u) = (-\Delta)^{-1}(-\nabla(q \nabla u) - f) = 0,$$

where $\Delta^{-1} = (A_1^n)^{-1}$. The associated 'finite noise' least squares problem can then be formulated as the minimization problem

$$\begin{aligned} \min_{(q,u) \in \widetilde{H} \times \widetilde{H}_0^1} J(q, u) &:= \frac{1}{2} \int_{\Gamma^n} \|u(\cdot, y) - \hat{u}^n(\cdot, y)\|_{H_0^1}^2 \rho_n(y) dy \\ s.t. \quad q \in Q_{\text{ad}}^n, \quad & e_n(q, u) = 0. \end{aligned}$$

We approximate the original set of admissible parameters Q_{ad} by

$$Q_{\text{ad}}^n := \left\{ q \in \widetilde{H}^n : \begin{array}{l} 0 < q_{\min} - \frac{1}{k_n} \leq q(x, y) \text{ a.s. on } D \times \Gamma^n, \\ \|q(\cdot, y)\|_H \leq q_{\max} + \frac{1}{k_n} \text{ a.s. on } \Gamma^n \end{array} \right\},$$

where $k_n \to \infty$ is a monotonically increasing approximation parameter with a rate to be specified later.

This formulation leaves us free to specify the regularity of q_n as a function of the vector $y = (y_1, ..., y_n) \in \Gamma^n \subset \mathbb{R}^n$. At the very least, q_n should be square integrable in y. Moreover, seeing that $\{Y_i\}_{i=1}^{\infty}$ forms a basis for $L_0^2(\Omega)$, the minimizer q^* of the original infinite dimensional problem (P) takes the form

$$q^*(x, \omega) = q_0^*(x) + \sum_{i=1}^{\infty} q_i(x) Y_i(\omega),$$

which is linear in each of the random variables Y_i. Any minimizer q_n^* of (P^n) that approximates q^* is therefore expected to depend relatively smoothly on y when n is large. At low orders of approximation on the other hand, the parameter $q \in Q_{\max}$ that gives rise to the model output $u(q)$ most closely resembling the partial data $\hat{u}^n$ may not exhibit the same degree of smoothness in the first n variables $y = (y_1, ..., y_n)$. Since the accuracy in the approximation of functions and their integrals benefit greatly in high dimensions from a high degree of smoothness [12], this suggests the use of a dimension adaptive strategy in which the smoothness requirement of the parameter is gradually strengthened as the stochastic dimension n increases.

For the sake of our analysis, we seek minimizers q_n^* of (P^n) in the tensor product space $\widetilde{H}_{\text{mix}}(D) := H(D) \otimes H_{\text{mix}}^r(\Gamma^n)$, where $H_{\text{mix}}^r(\Gamma^n)$ is the space of functions with bounded mixed derivatives of order r [61]. For any function $v \in \widetilde{H}_{\text{mix}}(D) \subset L^2(D \times \Gamma^n)$, the norm

$$\|v\|_{\widetilde{H}_{\text{mix}}}^2 := \sum_{|\gamma|_\infty \le r} \sum_{|\alpha|_1 \le s_d} \int_D \int_{\Gamma^n} \left| D_y^\gamma D_x^\alpha v(x, y) \right|^2 \rho_n(y) dy dx < \infty \tag{3.3}$$

where $\gamma = (\gamma_1, ..., \gamma_n) \in \mathbb{N}^n$ and $\alpha = (\alpha_1, ..., \alpha_d) \in \mathbb{N}^d$ are multi-indices, with $|\gamma|_\infty = \max\{\gamma_1, ..., \gamma_n\}$, $|\alpha|_1 = \alpha_1 + ... + \alpha_d$ and $s_d = 1$ when $d = 1$ or $s_d = 2$ when $d = 2, 3$. Posing the minimization problem (P^n) in this space not only guarantees that the right hand side $e_n(q, u)$ of the 'finite noise' equality constraint is twice Fréchet differentiable in q (see Section 3.2), but also allows for the use of numerical discretization schemes based on hierarchical finite elements, approximations known for their effectiveness in mitigating the curse of dimensionality [12].

In addition to using the space $\widetilde{H}_{\text{mix}}(D)$, we further modify the finite noise problem by adding a regularization term to ensure the continuous dependence of the optimizer q_n^* on the data $\hat{u}^n$. In summary, (P^n) is written as:

$$\begin{aligned} &\min_{(q,u) \in \widetilde{H}_{\text{mix}} \times \widetilde{H}_0^1} J^n(q, u) := \frac{1}{2} \|u - \hat{u}^n\|_{\widetilde{H}_0^1}^2 + \frac{\beta_n}{2} \|q\|_{\widetilde{H}_{\text{mix}}}^2 \\ &s.t. \;\; q \in Q_{\text{ad}}^n, \quad e_n(q, u) = 0. \end{aligned} \tag{P^n}$$

We will justify the use of this approximation scheme by demonstrating that as $n \to \infty$ and $\beta_n \to 0$, the sequence of minimizers q_n^* of problem (P^n) has a weakly convergent subsequence and that the limit of all convergent subsequences minimizes the infinite dimensional problem (P). Note that for any $v \in \widetilde{H}_0^1(D)$,

$$\|v\|_{\mathscr{H}_0^1}^2 = \int_\Omega \|v(\cdot,\omega)\|_{H_0^1}^2 \, d\omega = \int_{\Gamma^n} \|v(\cdot,y)\|_{H_0^1}^2 \rho_n(y) \, dy = \|v\|_{\widetilde{H}_0^1}^2.$$

Similarly $\|h\|_{\mathscr{H}} = \|h\|_{\widetilde{H}}$ for any $h \in \widetilde{H}(D)$ and therefore it is relatively straightforward to shift back and forth between these functions and their finite noise approximations. The norms $\|\cdot\|_{\mathscr{H}}$ and $\|\cdot\|_{\widetilde{H}_{\text{mix}}}$ are quite different on the other hand. In order to mediate between the minimizer q_n^* of the finite noise problem (P^n), formulated in the $\widetilde{H}_{\text{mix}}$ norm, and that of the infinite dimensional problem, whose minimizer q^* is measured in the $\mathscr{H}$ norm, we make use of the projection of q^* onto the first n basis vectors $\{Y_1, ..., Y_n\}$:

$$\mathbf{P}^n q^* = q_0^*(x) + \sum_{i=1}^n q_i(x) Y_i(\omega).$$

It is evident that $\mathbf{P}^n q^* \to q^*$ as $n \to \infty$ in $\mathscr{H}$. Moreover, seeing that $\mathbf{P}^n q^*$ is linear in y, we can bound the norm in $\widetilde{H}_{\text{mix}}$ in terms of its norm in $\mathscr{H}$ as the following computation shows:

Lemma 3.1.3.

$$\|\mathbf{P}^n q^*\|_{\widetilde{H}_{\text{mix}}} \leq \sqrt{2} \|\mathbf{P}^n q^*\|_{\mathscr{H}}.$$

Proof. Let e_i be the i^{th} standard basis vector for $\mathbb{N}^n$. We now apply formula (3.3) to $\mathbf{P}^n q^*$ to obtain

$$\begin{aligned}
\|\mathbf{P}^n q^*\|_{\widetilde{H}_{\text{mix}}}^2 &:= \sum_{|\gamma|_\infty \leq r} \sum_{|\alpha|_1 \leq s_d} \int_D \int_{\Gamma^n} \left| D_y^\gamma D_x^\alpha \left[q_0(x) + \sum_{i=1}^n q_i(x) y_i \right] \right|^2 \rho_n(y) dy dx \\
&= \sum_{|\alpha|_1 \leq s_d} \int_D \int_{\Gamma^n} \left| D_y^0 D_x^\alpha \left[q_0(x) + \sum_{i=1}^n q_i(x) y_i \right] \right|^2 \rho_n(y) dy dx \\
&\quad + \sum_{i=1}^n \sum_{|\alpha|_1 \leq s_d} \int_D \int_{\Gamma^n} \left| D_y^{e_i} D_x^\alpha \left[\sum_{i=1}^n q_i(x) y_i \right] \right|^2 \rho_n(y) dy dx \\
&= \int_{\Gamma^n} \|\mathbf{P}^n q^*(\cdot,\omega)\|_H^2 \rho_n(y) dy + \sum_{i=1}^n \sum_{|\alpha|_1 \leq s_d} \int_D \int_{\Gamma^n} |D_x^\alpha q_i(x)|^2 \rho_n(y) dy dx \\
&= \|\mathbf{P}^n q^*\|_{\mathscr{H}}^2 + \sum_{i=1}^n \|q_i\|_H^2 = 2 \sum_{i=0}^n \|q_i\|_H^2 - \|q_0\|_H^2 \leq 2 \|\mathbf{P}^n q^*\|_{\mathscr{H}}^2.
\end{aligned}$$

The second and third equalities follow from

$$D_y^\gamma \left[\sum_{i=1}^n q_i(x) y_i \right] = \begin{cases} \sum_{i=1}^n q_i(x) y_i, & \text{if } \gamma = 0 \\ q_i(x), & \text{if } \gamma = e_i \\ 0, & \text{otherwise} \end{cases}$$

and the fact that the random variables $\{Y_i\}_{i=1}^n$ are orthonormal and have zero mean. □

The following lemma addresses the feasibility of $\mathbf{P}^n q^*$. Although $\mathbf{P}^n q^*$ does not necessarily lie in the approximate feasible region Q^n_{ad}, the set on which $\mathbf{P}^n q^* \notin Q^n_{\text{ad}}$ can be made arbitrarily small as $n \to \infty$. Let $\mathscr{A}_n$ be the event that $\mathbf{P}^n q^*$ lies inside the approximate feasible region Q^n_{ad}, i.e.

$$\mathscr{A}_n := \left\{ \omega \in \Omega : 0 < q_{\min} + \frac{1}{k_n} \leq \mathbf{P}^n q^*(x,\omega) \text{ a.s. on } D,\ \|\mathbf{P}^n q^*(\cdot,\omega)\|_H \leq q_{\max} + \frac{1}{k_n} \right\}.$$

Lemma 3.1.4. *There is a monotonically increasing sequence $k_n \to \infty$ so that $\mathbb{P}(\Omega \backslash \mathscr{A}_n) \leq \frac{1}{k_n}$ for all $n \in \mathbb{N}$.*

Proof. For any $n \geq 1$, let k_n satisfy $\|\mathbf{P}^n q^* - q^*\|^2_{\mathscr{H}} = \frac{1}{C^2 k_n^3}$, where $C \geq 1$ is the embedding constant for $H(D) \hookrightarrow L^\infty(D)$. Clearly $k_n \to \infty$ as $n \to \infty$. Define the event

$$\mathscr{B}_n = \left\{ \omega \in \Omega : \|\mathbf{P}^n q^*(\cdot,\omega) - q^*(\cdot,\omega)\|_H \leq \frac{1}{C k_n} \right\}.$$

For any $\omega \in \mathscr{B}_n$,

$$\left| \|\mathbf{P}^n q^*(\cdot,\omega)\|_H - \|q^*(\cdot,\omega)\|_H \right| \leq \|\mathbf{P}^n q^*(\cdot,\omega) - q^*(\cdot,\omega)\|_H \leq \frac{1}{C k_n} \leq \frac{1}{k_n}$$

and

$$\|\mathbf{P}^n q^*(\cdot,\omega) - q^*(\cdot,\omega)\|_{L^\infty} \leq C \|\mathbf{P}^n q^*(\cdot,\omega) - q^*(\cdot,\omega)\|_H \leq \frac{1}{k_n},$$

which implies $\mathscr{B}_n \subset \mathscr{A}_n$. Moreover, according to Chebychev's inequality

$$\begin{aligned} \mathbb{P}(\Omega \backslash \mathscr{A}_n) \leq \mathbb{P}(\Omega \backslash \mathscr{B}_n) &\leq C^2 k_n^2 \int_\Omega \|\mathbf{P}^n q^*(\cdot,\omega) - q^*(\cdot,\omega)\|^2_H d\omega \\ &= C^2 k_n^2 \|\mathbf{P}^n q^* - q^*\|^2_{\mathscr{H}} \leq \frac{1}{k_n}. \end{aligned}$$

□

In order to ensure strict adherence to the inequality constraints of (P^n) for every n, we modify $\mathbf{P}^n q^*(\cdot,\omega)$ on $\Omega \backslash \mathscr{A}_n$.

Definition 3.1.5. *For all $n \in \mathbb{N}$, let $\hat{q}^*_n \in \widetilde{H}_{\text{mix}} \subset \mathscr{H}$ be defined as follows:*

$$\hat{q}^*_n := \begin{cases} \mathbf{P}^n q^*, & \omega \in \mathscr{A}_n \\ q^*_n, & \omega \notin \mathscr{A}_n \end{cases}. \tag{3.4}$$

Evidently $\hat{q}^*_n \in Q_{\text{ad}} \cap \widetilde{H}_{\text{mix}}$ and in light of Lemma 3.1.4, it is reasonable to expect $\hat{q}^*_n \approx \mathbf{P}^n q^*$ for large n, except on sets of negligible measure. Thus, we show

Lemma 3.1.6. *$\hat{q}_n^* \to q^*$ in $\mathscr{H}$ as $n \to \infty$.*

Proof.

$$\begin{aligned}\|\hat{q}_n^* - q^*\|_{\mathscr{H}} &= \int_{\mathscr{A}_n} \|\mathbf{P}^n q^*(\cdot,\omega) - q^*(\cdot,\omega)\|_H^2 d\omega + \int_{\Omega\backslash\mathscr{A}_n} \|q_n^*(\cdot,\omega) - q^*(\cdot,\omega)\|_H^2 d\omega \\ &\leq \|\mathbf{P}^n q^* - q^*\|_{\mathscr{H}}^2 + \mathbb{P}(\Omega\backslash\mathscr{A}_n) \sup_{\omega\in\Omega} \|q_n^*(\cdot,\omega) - q^*(\cdot,\omega)\|_H^2 \\ &\leq \|\mathbf{P}^n q^* - q^*\|_{\mathscr{H}}^2 + \frac{1}{k_n} 4 \left(q_{\max} + \frac{1}{k_1}\right)^2 \to 0 \text{ as } n \to \infty.\end{aligned}$$

□

We are now in a position to prove the main theorem of this section. For its proof we will make use of the fact that, due to the lower semicontinuity of norms

$$x_n \rightharpoonup x \text{ and } \limsup_{n\to\infty} \|x_n\| \leq \|x\| \;\Rightarrow\; x_n \to x \tag{3.5}$$

for any sequence $\{x_n\}$ in a Hilbert space.

Theorem 3.1.7. *Let $\delta_n := \|\hat{u} - \hat{u}^n\|_{\mathscr{H}_0^1} \to 0$ and $\beta_n \to 0$ as $n \to \infty$. Then the sequence of minimizers q_n^* of (P^n) has a subsequence converging weakly to a minimizer of the infinite dimensional problem (P) and the limit of every weakly convergent subsequence is a minimizer of (P). The corresponding model outputs converge strongly to the infinite dimensional minimizer's model output.*

Proof. Since q_n^* is optimal for (P^n), we have

$$\|u(q_n^*) - \hat{u}^n\|_{\widetilde{H}_0^1}^2 + \beta_n \|q_n^*\|_{\widetilde{H}_{\text{mix}}}^2 \leq \|u(\hat{q}_n^*) - \hat{u}^n\|_{\widetilde{H}_0^1}^2 + \beta_n \|\hat{q}_n^*\|_{\widetilde{H}_{\text{mix}}}^2. \tag{3.6}$$

Moreover, by definition $\hat{q}_n^*(\cdot, Y(\omega)) = q_n^*(\cdot, Y(\omega))$ for all $Y \in Y(\Omega\backslash\mathscr{A}_n)$ and hence

$$\begin{aligned}&\|\hat{q}_n^*\|_{\widetilde{H}}^2 - \|q_n^*\|_{\widetilde{H}}^2 \\ &= \sum_{|\gamma|_\infty \leq 1} \sum_{|\alpha|_1 \leq s_d} \left(\int_{Y(\mathscr{A}_n)} \int_D \left| D_y^\gamma D_x^\alpha \mathbf{P}^n q^* \right|^2 \rho_n(y) dx dy - \int_{Y(\mathscr{A}_n)} \int_D \left| D_y^\gamma D_x^\alpha q_n^* \right|^2 \rho_n(y) dx dy \right) \\ &\leq \sum_{|\gamma|_\infty \leq 1} \sum_{|\alpha|_1 \leq s_d} \left(\int_{Y(\mathscr{A}_n)} \int_D \left| D_y^\gamma D_x^\alpha \mathbf{P}^n q^* \right|^2 \rho_n(y) dx dy \right) \leq \|\mathbf{P}^n q^*\|_{\widetilde{H}_{\text{mix}}}^2 \leq 2\|\mathbf{P}^n q^*\|_{\widetilde{H}}^2\end{aligned}$$

from which it follows that

$$\begin{aligned}\|u(q_n^*) - \hat{u}^n\|_{\widetilde{H}_0^1}^2 &\leq \|u(\hat{q}_n^*) - \hat{u}^n\|_{\widetilde{H}_0^1}^2 + \beta_n \|\hat{q}_n^*\|_{\widetilde{H}_{\text{mix}}}^2 - \beta_n \|q_n^*\|_{\widetilde{H}_{\text{mix}}}^2 \\ &\leq \|u(\hat{q}_n^*) - \hat{u}^n\|_{\widetilde{H}_0^1}^2 + 2\beta_n \|\mathbf{P}^n q^*\|_{\widetilde{H}}^2.\end{aligned}$$

By Lemmas 3.1.6 and 2.3.2

$$\limsup_{n\to\infty} \|u(q_n^*) - \hat{u}^n\|_{\widetilde{H}_0^1}^2 \leq \lim_{n\to\infty} \|u(\hat{q}_n^*) - \hat{u}^n\|_{\widetilde{H}_0^1}^2 + \beta_n \|\mathbf{P}^n q^*\|_{\widetilde{H}}^2 = \|u(q^*) - \hat{u}\|_{\widetilde{H}_0^1}^2,$$

which, together with the Banach Alaoglu Theorem, guarantees the existence of a subsequence $u(q^*_{n_j})$ converging weakly to some $u_0 \in \mathscr{H}_0^1$. Since feasible sets $\{Q^n_{\text{ad}}\}_{n=1}^{\infty}$ form a nested sequence, all functions $q^*_n \in Q^n_{\text{ad}} \subset Q^1_{\text{ad}}$, which is weakly compact (Lemma 2.3.1). The sequence $q^*_n \in Q_{\text{ad}}$ therefore has a subsequence, $q^*_{n_j} \rightharpoonup q_0 \in Q^1_{\text{ad}}$ in $\mathscr{H}$. Additionally, since Q^n_{ad} is nested and the graph of u is weakly closed (Lemma 2.3.3) we have $q_0 \in \cap_{n=1}^{\infty} Q^n_{\text{ad}} = Q_{\text{ad}}$ and $u_0 = u(q_0)$. Therefore

$$\begin{aligned}
\|u(q_0) - \hat{u}\|^2_{\widetilde{H}_0^1} &= \lim_{j\to\infty} \langle u(q^*_{n_j}) - \hat{u}^{n_j}, u(q_0) - \hat{u}\rangle_{\widetilde{H}_0^1} && (3.7)\\
&\leq \liminf_{j\to\infty} \|u(q^*_{n_j}) - \hat{u}^{n_j}\|_{\widetilde{H}_0^1} \|u(q_0) - \hat{u}\|_{\widetilde{H}_0^1} && (3.8)\\
&\leq \limsup_{j\to\infty} \|u(q^*_{n_j}) - \hat{u}^{n_j}\|_{\widetilde{H}_0^1} \|u(q_0) - \hat{u}\|_{\widetilde{H}_0^1} && (3.9)\\
&\leq \|u(q^*) - \hat{u}\|_{\widetilde{H}_0^1} \|u(q_0) - \hat{u}\|_{\widetilde{H}_0^1}, && (3.10)
\end{aligned}$$

which implies $\|u(q_0) - \hat{u}\|_{\mathscr{H}_0^1} \leq \|u(q^*) - \hat{u}\|_{\mathscr{H}_0^1}$ and hence $q_0 \in Q_{\text{ad}}$ is a minimizer for (P). Inequalities (3.8) and (3.9) further lead to

$$\lim_{j\to\infty} \|u(q^*_{n_j}) - \hat{u}^{n_j}\|_{\widetilde{H}_0^1} = \|u(q_0) - \hat{u}\|_{\widetilde{H}_0^1}$$

which, together with the weak convergence $u(q^*_{n_j}) - \hat{u}^{n_j} \rightharpoonup u(q_0) - \hat{u}$, implies that $u(q^*_{n_j}) - \hat{u}^{n_j} \to u(q_0) - \hat{u}$ due to (3.5). In addition, the fact that $\hat{u}^{n_j} \to \hat{u}$ implies that $u(q_{n_j}) \to u(q_0)$. This argument holds for any convergent subsequence of $\{q^*_n\}$ and hence the Theorem is proved. □

3.1.1 Extensions and Related Results

Theorem 3.1.7 only guarantees weak convergence of a subsequence of 'finite noise' minimizers q^*_n to the appropriate infinite noise minimizer q^*. Within the framework outlined above (see also [7]), the proof for strong convergence relies heavily on the result (3.5)

$$x_n \rightharpoonup x, \text{ and } \limsup_{n\to\infty} \|x_n\| \leq \|x\| \Rightarrow x_n \to x.$$

While weak convergence $q^*_n \rightharpoonup q_0$ in $\mathscr{H}$ follows trivially from the weak compactness of the admissible set, finding the necessary bound $\limsup_{n\to\infty} \|q^*_n\|_{\mathscr{H}} \leq \|q^*\|_{\mathscr{H}}$ is complicated by the fact that problems (P) and (P^n) are posed in different norms.

Indeed, if we choose to approximate problem (P) instead by the 'finite noise' problem with regularization term $R(q) = \frac{\beta}{2}\|q\|^2_{\widetilde{H}}$, i.e.

$$\begin{aligned}
&\min_{(q,u)\in\widetilde{H}\times\widetilde{H}_0^1} J^n(q,u) := \frac{1}{2}\|u - \hat{u}^n\|^2_{\widetilde{H}_0^1} + \frac{\beta_n}{2}\|q\|^2_{\widetilde{H}} \\
&s.t.\ \ q \in Q_{\text{ad}}, \quad e_n(q,u) = 0,
\end{aligned} \tag{$P^n_{\widetilde{H}}$}$$

where $Q_{\text{ad}} = \{q \in \widetilde{H} : q \geq q_{\min} > 0,\ \|q(\cdot, y)\|_H \leq q_{\max} \text{ a.s. on } \Gamma^n\}$ and $e_n(q,u) = 0$ is the 'finite noise' equality constraint, we can readily establish the necessary condition $\limsup_{n\to\infty} \|q^*_n\|_{\mathscr{H}} \leq \|q^*\|_{\mathscr{H}}$ within the given framework.

Theorem 3.1.8. *Let $\delta_n := \|\hat{u} - \hat{u}^n\|_{\mathscr{H}_0^1} \to 0$ and $\beta_n \to 0$ as $n \to \infty$ so that $\frac{\delta_n}{\beta_n}$ is bounded. Then the sequence of minimizers q_n^* of $(P_{\widetilde{H}}^n)$ has a subsequence converging* strongly *to a minimizer of the infinite dimensional problem (P) and the limit of every convergent subsequence is a minimizer of (P) with $\beta = 0$. The corresponding model outputs converge strongly to the infinite dimensional minimizer's model output.*

Proof. The proof of this theorem is similar to that of Theorem 3.1.7. Let q_n^* be the minimizer of the 'finite noise' problem $(P_{\widetilde{H}}^n)$. Then

$$\|u(q_n^*) - \hat{u}^n\|_{\widetilde{H}_0^1}^2 + \beta_n \|q_n^*\|_{\widetilde{H}}^2 \leq \|u(q^*) - \hat{u}^n\|_{\widetilde{H}_0^1}^2 + \beta_n \|q^*\|_{\widetilde{H}}^2, \tag{3.11}$$

which, by virtue of the fact that $\beta_n(\|q^*\|_{\widetilde{H}}^2 - \|q_n^*\|_{\widetilde{H}}^2) \leq \beta_n \|q^*\|_{\widetilde{H}} \to 0$ as $n \to \infty$, implies

$$\limsup_{n\to\infty} \|u(q_n^*) - \hat{u}\|_{\widetilde{H}_0^1}^2 \leq \|u(q^*) - \hat{u}\|_{\widetilde{H}_0^1}^2. \tag{3.12}$$

As in Theorem 3.1.7, this bound, together with the weak compactness of Q_{ad} (Lemma 2.3.1) and the fact that the graph of u is weakly closed (Lemma 2.3.3), give rise to a subsequence $q_{n_j}^*$ of q_n^* so that $q_{n_j}^* \rightharpoonup q_0$ and $u(q_{n_j}^*) \rightharpoonup u(q_0)$ for some function $q_0 \in Q_{\mathrm{ad}}$. The weak convergence, together with the upper bound (3.12) lead directly to the chain of inequalities (3.7)-(3.10). As a result, q_0 is a minimizer of (P) with $\beta = 0$ and

$$\lim_{j\to\infty} \|u(q_{n_j}^*) - \hat{u}^{n_j}\|_{\widetilde{H}_0^1} = \|u(q_0) - \hat{u}\|_{\widetilde{H}_0^1},$$

which implies $\lim_{j\to\infty} u(q_{n_j}^*) = u(q_0)$, by virtue of (3.5). In addition,

$$\begin{aligned}
\beta_{n_j}\|q_{n_j}^*\|_{\mathscr{H}}^2 &\leq \overbrace{\|u(q_{n_j}^*) - \hat{u}\|_{\mathscr{H}_0^1}^2 - \|u(q_0) - \hat{u}\|_{\mathscr{H}_0^1}^2}^{\geq 0} + \beta_{n_j}\|q_{n_j}^*\|_{\mathscr{H}}^2 \\
&= \|u(q_{n_j}^*) - \hat{u}^n\|_{\mathscr{H}_0^1}^2 + \beta_{n_j}\|q_{n_j}^*\|_{\mathscr{H}}^2 - \|u(q_0) - \hat{u}^{n_j}\|_{\mathscr{H}_0^1}^2 \\
&\quad + 2\langle u(q_{n_j}) - u(q_0), \hat{u}^{n_j} - \hat{u}\rangle_{\widetilde{H}_0^1} \\
&\leq \beta_{n_j}\|q_0\|_{\mathscr{H}}^2 + 2\langle u(q_{n_j}^*) - u(q_0), \hat{u}^{n_j} - \hat{u}\rangle_{\widetilde{H}_0^1}
\end{aligned}$$

and hence

$$\limsup_{j\to\infty} \|q_{n_j}^*\|_{\mathscr{H}} \leq \|q_0\|_{\mathscr{H}}^2 + 2 \lim_{j\to\infty} \|u(q_{n_j}^*) - u(q_0)\|_{\widetilde{H}_0^1} \frac{\delta_{n_j}}{\beta_{n_j}} = \|q_0\|_{\mathscr{H}}^2.$$

Inequality (3.5) then guarantees that $q_{n_j} \to q_0$ as $j \to \infty$. □

3.2 Necessary Optimality in the Finite Noise Problem

An immediate benefit of approximating the problem in $\widetilde{H}_{\mathrm{mix}}$ results from its continuous embedding in $L^\infty(D \times \Gamma^n)$. This ensures that the equality constraint is twice continuously Fréchet differentiable in q, which in turn enables the derivation of more traditional first

and second order optimality conditions. These can be used as the basis of the optimization strategies, discussed in the following chapter. By virtue of the tensor product structure of $H^r_{\text{mix}}(\Gamma^n)$ we may consider Sobolev regularity component-wise, which, in conjunction with the compact embedding of $H^r(\Gamma_i)$ in $L^\infty(\Gamma_i)$ for $r \geq 1$, gives rise to this property as the following lemma shows.

Lemma 3.2.1. *The space $\widetilde{H}_{\text{mix}}(D)$ embeds continuously in $L^\infty(D \times \Gamma^n)$ for all $n \in \mathbb{N}$.*

Proof. For any fixed value y_0 of the random component y and any multi-index $\gamma \in \mathbb{N}^n$, the function $D_y^\gamma q(\cdot, y_0) \in H^{s_d}(D)$ whenever $|\gamma|_\infty \leq r$, where $s_d = 1$ is the physical dimension $d = 1$ and $s_d = 2$ for $d = 1, 2$. Similarly, if both spatial variable x and all but the i^{th} component y_i of the stochastic variable y are fixed at x_0 and $y_0^1, ..., y_0^{i-1}, y_0^{i+1}, ...y_0^n$ respectively, and $\alpha \in \mathbb{N}^d$, $\gamma_i^* := (\gamma_1, ..., \gamma_{i-1}, 0, \gamma_{i+1}, ..., \gamma_n) \in \mathbb{N}^n$ are multi-indices satisfying $|\alpha|_1 \leq s_d$, $|\gamma_i^*|_\infty \leq r$, then

$$D_x^\alpha D_{\mathbf{y}}^{\gamma_i^*} q(x_0, y_0^1, ..., y_0^{i-1}, \cdot, y_0^{i+1}, ..., y_0^n) \in H^r(\Gamma_i) \hookrightarrow L^\infty(\Gamma_i).$$

By repeated application of the one-dimensional Sobolev Imbedding Theorem [1], it follows that

$$\begin{aligned}
\|q\|_{L^\infty(D\times\Gamma)} &= \max_{x\in D, y\in\Gamma^n} |q(x,y)| = \max_{y\in\Gamma^n} \|q(\cdot,y)\|_{L^\infty(D)} \leq C \max_{(y_1,...,y_n)\in\Gamma^n} \|D_x^\alpha q(\cdot,y)\|_{H^1(D)} \\
&\leq C \max_{(y_1,...,y_{n-1})\in\Gamma^{n-1}} \left(\sum_{|\alpha|_1\leq s_d} \int_D \left(\max_{y_n\in\Gamma_n} |D_x^\alpha q(x,y_1,...,y_n)| \right)^2 dx \right)^{\frac{1}{2}} \\
&\leq C C_{\Gamma_n} \max_{(y_1,...,y_{n-1})\in\Gamma^{n-1}} \left(\sum_{|\alpha|_1\leq s_d} \sum_{\gamma_n=0}^{r} \int_D \int_{\Gamma_n} |D_x^\alpha D_{y_n}^{\gamma_n} q(x,y_1,...,y_n)|^2 d\omega\, dx \right)^{\frac{1}{2}} \\
&\leq ... \\
&\leq C \prod_{i=1}^n C_{\Gamma_i} \left(\sum_{|\alpha|_1\leq s_d} \sum_{|\gamma|_\infty\leq r} \int_D \int_{\Gamma^n} |D_x^\alpha D_y^\gamma q(x,y)|^2 \rho_n(y)\, dy\, dx \right)^{\frac{1}{2}} = C_n \|q\|_{\widetilde{H}_{\text{mix}}}
\end{aligned}$$

for some constant $C_n > 0$, independent of q, but possibly dependent on the total dimension $d_{\text{tot}} = d + n$.

□

3.2.1 First- and Second Order Optimality

Recall that the equality constraint $e_n(\cdot,\cdot) : \widetilde{H}_{\text{mix}}(D) \times \widetilde{H}_0^1(D) \to \widetilde{H}_0^1(D)$ satisfies

$$\langle e_n(q,u), v\rangle_{\widetilde{H}_0^1} := \int_D \int_{\Gamma^n} q\nabla u \cdot \nabla v\, \rho_n(y)\, dy\, dx - \int_D \int_{\Gamma^n} f\, v\, \rho_n(y)\, dy\, dx$$

for all $v \in \widetilde{H}_{\text{mix}}(D)$. Since $e_n(q,u)$ is affine linear in both arguments, its Fréchet differentiability with respect to either q or u follows directly from its boundedness in these variables.

Continuity in u can be shown in precisely the same way as in the infinite dimensional case $e(q,u)=0$. For any $u,\tilde{u}\in\widetilde{H}_0^1$,

$$\left|\langle e_n(q,u-\tilde{u}),v\rangle_{\widetilde{H}_0^1}\right| = \left|\int_D\int_{\Gamma^n} q\nabla(u-\tilde{u})\cdot\nabla v\ \rho_n dy\ dx\right| \le q_{\max}\|u-\tilde{u}\|_{\widetilde{H}_0^1}\|v\|_{\widetilde{H}_0^1} \quad \forall\phi\in\widetilde{H}_0^1$$

and hence $\|e_n(q,u-\tilde{u})\|_{\widetilde{H}_0^1}\le q_{\max}\|u-\tilde{u}\|_{\widetilde{H}_0^1}$. In order to establish the continuity of $e_n(q,u)$ with respect to $q\in\widetilde{H}_{\mathrm{mix}}$, we make use of the embedding from Lemma 3.2.1. Let $q,\tilde{q}\in\widetilde{H}_{\mathrm{mix}}$ and $v\in\widetilde{H}_0^1$ be arbitrary. Then

$$\begin{aligned}\left|\int_D\int_{\Gamma^n}(q-\tilde{q})\nabla u\cdot\nabla v\ \rho_n(y)\ dy\ dx\right| &\le \|q-\tilde{q}\|_{L^\infty(D\times\Gamma)}\|u\|_{\widetilde{H}_0^1}\|v\|_{\widetilde{H}_0^1}\\ &\le \tilde{C}_n\|q-\tilde{q}\|_{\widetilde{H}_{\mathrm{mix}}}\|u\|_{\widetilde{H}_0^1}\|v\|_{\widetilde{H}_0^1},\end{aligned}$$

from which it readily follows that $\|e_n(q-\tilde{q},u)\|_{\mathscr{H}^{-1}}\le\tilde{C}_n\|u\|_{\widetilde{H}_0^1}\|q-\tilde{q}\|_{\widetilde{H}_{\mathrm{mix}}}$ and therefore e_n is twice continuously Fréchet differentiable.

A simple calculation then reveals that the first derivative of e_n in the direction $(h,v)\in\widetilde{H}_{\mathrm{mix}}\times\widetilde{H}_0^1$ is given by:

$$D_{(q,u)}[e_n(q,u)](h,v)=D_q[e_n(q,u)]h+D_u[e_n(q,u)]v\in\widetilde{H}_0^1,$$

where

$$\begin{aligned}\langle D_q[e_n(q,u)]h,\phi\rangle_{\widetilde{H}_0^1} &= \int_D\int_{\Gamma^n} h\nabla u\cdot\nabla\phi\ \rho_n\ dy\ dx \quad \text{and}\\ \langle D_u[e_n(q,u)]v,\phi\rangle_{\widetilde{H}_0^1} &= \int_D\int_{\Gamma^n} q\nabla v\cdot\nabla\phi\ \rho_n\ dy\ dx \quad \text{for all } \phi\in\widetilde{H}_0^1.\end{aligned}$$

Similarly, the second derivative $D^2_{(q,u)}[e_n(q,u)](h_1,v_1)(h_2,v_2)$ in directions $h_1,h_2\in\widetilde{H}$ and $v_1,v_2\in\widetilde{H}_0^1$ can be computed as

$$D^2_{(q,u)}[e_n(q,u)](h_1,v_1)(h_2,v_2)=D^2_{qu}[e_n(q,u)](h_1,v_2)+D^2_{uq}[e_n(q,u)](v_1,h_2)\in\widetilde{H}_0^1(D),$$

where

$$\begin{aligned}\langle D^2_{qu}[e_n(q,u)](h_1,v_2),w\rangle_{\widetilde{H}_0^1} &= \int_D\int_{\Gamma^n} h_1\nabla v_2\cdot\nabla w\ \rho_n\ dy\ dx\\ \langle D^2_{uq}[e_n(q,u)](v_1,h_2),w\rangle_{\widetilde{H}_0^1} &= \int_D\int_{\Gamma^n} h_2\nabla v_1\cdot\nabla w\ \rho_n\ dy\ dx\end{aligned}$$

We can now derive more traditional first order necessary optimality conditions

Theorem 3.2.2 (Existence of Lagrange Multipliers)**.** *Let (q_n^*,u_n^*) be a minimizer for problem (P^n). Then there exists a unique Lagrange multiplier $\lambda_n^*\in\widetilde{H}_0^1$ for which the Lagrange functional*

$$L(q,u;\lambda)=J^n(q,u)+\langle\lambda,e_n(q,u)\rangle_{\widetilde{H}_0^1}$$

satisfies

$$D_{(q,u)}\left[L(q_n^*, u_n^*; \lambda_n^*)\right](h,v) \geq 0 \;\; \textit{for all } (h,v) \in C(q_n^*) \times \widetilde{H}_0^1$$

where

$$C(q_n^*) = \{b(c - q^*) : c \in Q_{\mathrm{ad}},\; b \geq 0\}.$$

In particular, the adjoint equation and complementary condition hold

$$\begin{aligned} \langle q_n^* \nabla \lambda_n^*, \nabla v\rangle =& \langle \nabla(u_n^* - \hat{u}^n), \nabla v\rangle && \textit{for all } v \in \widetilde{H}_0^1 && (3.13)\\ \beta\langle q_n^*, q - q_n^*\rangle_{\widetilde{H}_{\mathrm{mix}}} - & \langle \nabla\lambda^*, (q - q_n^*)\nabla u^*\rangle \geq 0 && \textit{for all } q \in Q_{\mathrm{ad}}. && (3.14)\end{aligned}$$

Proof. Let (q_n^*, u_n^*) be a minimizer of problem (P^n). We will show that (q_n^*, u_n^*) satisfies the regular point condition

$$D_{(q,u)}[e_n(q_n^*, u_n^*)](C(q^*) \times \widetilde{H}_0^1) = \widetilde{H}, \tag{3.15}$$

from which the existence of the Lagrange multiplier follows directly by [44]. To this end, let $w \in \widetilde{H}_0^1$ be arbitrary. Regularity of (q_n^*, u_n^*) is then equivalent to the solvability of the equation

$$D_q[e_n(q^*, u^*)]h + D_u[e_n(q^*, u^*)]v = w \quad \text{for } h \in C(q^*) \text{ and } v \in \widetilde{H}_0^1.$$

Since $0 \in C(q^*)$ and the elliptic equation $\langle q_n^* \nabla v, \nabla \cdot\,\rangle_{L^2} = \langle w, \cdot\,\rangle_{\widetilde{H}_0^1}$ is solvable for any $w \in \widetilde{H}_0^1$, condition (3.15) holds and hence there exists a Lagrange multiplier $\lambda_n^* \in \widetilde{H}_0^1$ such that

$$D_{(q,u)}[J(q_n^*, u_n^*)](h,v) + \langle \lambda_n^*, D_{(q,u)}[e_n(q_n^*, u_n^*)](h,v)\rangle_{\widetilde{H}_0^1} \geq 0 \quad \text{for all } (h,v) \in C(q^*) \times \widetilde{H}_0^1,$$

which simplifies to

$$\begin{aligned} 0 &\leq \langle u_n^* - \hat{u}^n, v\rangle_{\widetilde{H}_0^1} + \beta\langle q_n^*, h\rangle_{\widetilde{H}_{\mathrm{mix}}} + \langle D_q[e_n(q_n^*, u_n^*)]h + D_u[e_n(q_n^*, u_n^*)]v, \lambda_n^*\rangle_{\widetilde{H}_0^1} \\ &= \langle \nabla(u_n^* - \hat{u}^n), \nabla v\rangle + \beta\langle q_n^*, h\rangle_{\widetilde{H}_{\mathrm{mix}}} - \langle h\nabla u_n^* + q_n^* \nabla v, \nabla \lambda_n^*\rangle \end{aligned} \tag{3.16}$$

for all $(h,v) \in C(q_n^*) \times \widetilde{H}_0^1$. In particular, choosing $h = 0$, we obtain

$$\langle \nabla\lambda_n^*, q_n^* \nabla v\rangle = \langle \nabla(u_n^* - \hat{u}^n), \nabla v\rangle \qquad \text{for all } v \in \widetilde{H}_0^1,$$

which yields the adjoint equation (3.13). The uniqueness of λ_n^* now follows directly from the uniqueness of the solution to the elliptic equation (3.13). Finally, setting $v = 0$ in (3.16) yields the complementary condition (3.14)

$$\beta\langle q_n^*, q - q_n^*\rangle_{\widetilde{H}_{\mathrm{mix}}} - \langle \nabla\lambda_n^*, (q - q_n^*)\nabla u_n^*\rangle \geq 0 \qquad \text{for all } q \in Q_{\mathrm{ad}}.$$

□

Remark 3.2.3. *The following remarks apply to Theorem 3.2.2*

1. *These results concerning Fréchet differentiability of the equality constraint and first order optimality of minimizers can readily be modified to accommodate the case when* $e_n(q,u) = 0$ *in* $\widetilde{H}_0^1$ *is replaced by finite noise equivalent* $\tilde{e}_n(q,u) = 0$ *in* $\widetilde{H}^{-1}$ *of the constraint* $\tilde{e}(q,u) = 0$ *in* $\widetilde{H}^{-1}$*, given in equation* (2.3)*.*

2. *Note that, according to the Lax Milgram Lemma, the solution to the adjoint equation* (3.13) *satisfies the bound*

$$\|\lambda_n^*\|_{\widetilde{H}_0^1} \leq \frac{1}{q_{\min}} \|u_n^* - \hat{u}^n\|_{\widetilde{H}_0^1},$$

which will be use in the proof of the coercivity of the second Fréchet derivative below.

The second Fréchet derivative of the Lagrangian is given by

$$\begin{aligned} D^2_{(q,u)}[L(q,u,\lambda)](h,v) &= \|v\|^2_{\widetilde{H}_0^1} + \beta\|q\|^2_{\widetilde{H}_{\text{mix}}} + 2\langle \lambda, D_{(q,u)}[e_n(q,u)](h,v)\rangle_{\widetilde{H}_0^1} \\ &= \|v\|^2_{\widetilde{H}_0^1} + \beta\|q\|^2_{\widetilde{H}_{\text{mix}}} + 2\langle \nabla\lambda, h\nabla v\rangle. \end{aligned}$$

Under the additional assumption

$$\frac{C_n}{q_{\min}} \|u_n^* - \hat{u}^n\| \leq \sqrt{\beta}, \tag{3.17}$$

where C_n is the embedding constant of $\widetilde{H}(D) \hookrightarrow L^\infty(D \times \Gamma^n)$, we can also derive second order optimality conditions that will prove useful in the next chapter. The following theorem has been adapted from [32]. Its proof is included here for completeness.

Theorem 3.2.4. *Under assumption* (3.17)*, the second order Fréchet derivative,* $D^2_{(q,u)}[L(q^*,u^*;\lambda^*)]$*, satisfies*

$$D^2_{(q,u)}[L(q^*,u^*,\lambda^*)](h,v)^2 \geq \tau(\|h\|^2_{\widetilde{H}_{\text{mix}}} + \|v\|^2_{\widetilde{H}_0^1}) \quad \forall (h,v) \in \widetilde{H}_{\text{mix}}(D) \times \widetilde{H}_0^1(D) \tag{3.18}$$

for some $\tau = \tau(\beta) > 0$

Proof. By the continuous embedding $\|\cdot\|_{L^\infty} \leq C_n \|\cdot\|_{\widetilde{H}_{\text{mix}}}$, we have for any real $\alpha \neq 0$

$$\begin{aligned} D^2_{(q,u)}[L(q_n^*,u_n^*,\lambda_n^*)](h,v)^2 &= \|v\|^2_{\widetilde{H}} + \beta\|h\|^2_{\widetilde{H}_{\text{mix}}} + 2\langle \lambda_n^*, h\nabla v\rangle \\ &\geq \|v\|^2_{\widetilde{H}} + \beta\|h\|^2_{\widetilde{H}_{\text{mix}}} - 2\|\lambda_n^*\|_{\widetilde{H}_0^1}\|v\|_{\widetilde{H}_0^1}\|h\|_{L^\infty} \\ &\geq \|v\|^2_{\widetilde{H}} + \beta\|h\|^2_{\widetilde{H}_{\text{mix}}} - 2\frac{C_n}{q_{\min}}\|u_n^* - \hat{u}^n\|_{\widetilde{H}_0^1}\|v\|_{\widetilde{H}_0^1}\|h\|_{\widetilde{H}_{\text{mix}}}. \end{aligned}$$

Let $c = \frac{C_n}{q_{\min}}\|u_n^* - \hat{u}^n\|_{\widetilde{H}_0^1}$ and $\alpha \neq 0$ be arbitrary. Then the simple quadratic inequality

$$0 \leq \left(\alpha\|h\|_{\widetilde{H}_{\text{mix}}} + \frac{c}{2\alpha}\|v\|_{\widetilde{H}_0^1}\right)^2 = \alpha^2\|h\|^2_{\widetilde{H}_{\text{mix}}} - c\|h\|_{\widetilde{H}_{\text{mix}}}\|v\|_{\widetilde{H}_0^1} + \frac{c^2}{4\alpha^2}\|v\|_{\widetilde{H}_0^1}$$

implies

$$\begin{aligned} D^2_{(q,u)}[L(q_n^*, u_n^*, \lambda_n^*)](h,v)^2 &\geq \|v\|^2_{\widetilde{H}} + \beta\|h\|^2_{\widetilde{H}_{\mathrm{mix}}} - c\|v\|_{\widetilde{H}^1_0}\|h\|_{\widetilde{H}_{\mathrm{mix}}} \\ &\geq (1-\alpha^2)\|h\|^2_{\widetilde{H}_{\mathrm{mix}}} + \left(1 - \frac{c^2}{4\alpha^2}\right)\|v\|_{\widetilde{H}^1_0}. \end{aligned}$$

It now suffices to show that there exists an $\alpha \neq 0$ for which $\tau := \min\{\beta - \alpha^2, 1 - \frac{c}{4\alpha^2}\} > 0$. Under assumption (3.17), $c^2 < 4\beta$ and therefore the strict positivity of τ can be guaranteed by choosing

$$\alpha := \begin{cases} \frac{1}{2}(\sqrt{(1-\beta)^2 + c^2} - (1-\beta)), & \text{if } \beta \in (0,1] \\ \frac{1}{2}(\sqrt{(\beta-1)^2 + c^2} - (\beta-1)), & \text{if } \beta \in (1,\infty) \end{cases}.$$

□

Chapter 4

Optimization Strategies

4.1 Introduction

The finite noise problem (P^n) lends itself readily to analysis. First and second order necessary optimality conditions can be proved completely analogously to the deterministic case, as was shown at the end of the previous chapter. Moreover, many existing optimization methods [25] designed for identification and control of systems governed by partial differential equations are formulated and analyzed within a general Hilbert space setting, endowing these algorithms with sufficient generality to be applied to the stochastic problem (P^n). The primary challenge, however, lies in implementing these algorithms efficiently and obtaining accurate approximations of q, even in the presence of the relatively high dimension of the underlying stochastic domain Γ^n and moderate assumptions on the smoothness of $q(x, y)$ as a function of $y \in \Gamma^n$. This chapter investigates various optimization-based numerical strategies for the parameter identification problem (P^n) that reduce computational cost either through the choice of discretization space or by exploiting the parallel structure inherent in the problem.

Finite dimensional discretizations of problem (P^n) should ideally lead to accurate approximations of both the forward and adjoint systems as well as of the high-dimensional integrals that are required for evaluating the cost functional and the associated gradient. The tensor product structure of the underlying function spaces allows us to employ different approximation spaces in the physical and stochastic domains. Since the effect of spatial approximation on the solvability of deterministic variational problems is relatively well understood [5,13,14], we use standard finite element methods to discretize the spatial domain D and focus instead on discretization in the stochastic domain Γ^n. Stochastic collocation methods [2,47,48] provide a means of efficiently computing statistical quantities of interest related to solutions of the model or adjoint equations, by solving their parametric versions in parallel. As an example, suppose a candidate parameter $q(x, y)$ is given and consider the stochastic forward

problem (1.12)

$$\int_{\Gamma^n}\int_D q(x,y)\nabla u(x,y)\cdot\nabla v(x,y)\ \rho_n(y)\ dx\ dy = \int_{\Gamma^n}\int_D f(x,y)v(x,y)\ \rho_n(y)\ dx\ dy$$

for all $v \in \widetilde{H}_0^1(D)$. Let V_I be an I-dimensional subspace of $H_0^1(D)$ with basis functions $\{\phi_1,...,\phi_I\}$ and Z_J be a J-dimensional subspace of $L^2(\Gamma^n)$ with basis functions $\{\psi_1,...,\psi_J\}$. The projection of the model output $u \in \widetilde{H}_0^1(D)$ onto $V_I \otimes Z_J$ then takes the form

$$u_{IJ}(x,y) = \sum_{i=1}^{I}\sum_{j=1}^{J} c_{ij}\phi_i(x)\psi_j(y),$$

which can be used to formulate a Galerkin-type approximation to the forward problem (1.12)

$$\begin{aligned}&\int_{\Gamma^n}\int_D q(x,y)\left(\sum_{i=1}^{I}\sum_{j=1}^{J} c_{ij}\nabla\phi_i(x)\psi_j(y)\right)\cdot\nabla\phi_{i'}(x)\psi_{j'}(y)\rho_n(y)\ dx\ dy\\ =&\int_{\Gamma^n}\int_D f(x,y)\phi_{i'}(x)\psi_{j'}(y)\rho_n(y)\ dx\ dy \quad \text{for } i'=1,2,...,I \text{ and } j'=1,2,...,J.\end{aligned} \tag{4.1}$$

Solving (4.1) for the unknown coefficients c_{ij} then yields an approximation of u in terms of the given basis. Once an approximation of u is obtained, it can be used to compute statistical quantities of interest (qoi), such as moments, probabilities or stochastic Sobolev norms, all of which require the evaluation of integrals of the form

$$G_{\text{qoi}}(x) = \int_{\Gamma^n} G(u(x,y),x,y)\rho_n(y)\ dy, \tag{4.2}$$

where G is an appropriately defined functional. The functional G may take the form $G(u(x,y)) = u(x,y)^k$ when $G_{\text{qoi}}(x) = \mathbb{E}(u(x,\cdot)^k)$ is the k^{th} moment, $G(u(x,y)) = \chi_{\{u(x,y)\le\alpha\}}(y)$ when $G_{\text{qoi}}(x) = \mathbb{P}(u(x,\cdot)\le\alpha)$ is a cumulative probability, or $G(u(\cdot,y)) = \|u(\cdot,y)\|_{H_0^1}^2$ when the quantity of interest is the stochastic Sobolev norm $G_{\text{qoi}} = \|u\|_{\widetilde{H}_0^1}$. In general, the integrals over Γ^n appearing in (1.12) and (4.2) must be estimated by a numerical quadrature scheme, specified by a sequence of quadrature weights $\{w_k\}_{k=1}^K$ and nodes $\{y_k\}_{k=1}^K$. If the basis functions $\{\psi_j\}_{j=1}^J$ are chosen to be a nodal basis of Z_J centered at these quadrature points, i.e. $J=K$ and $\psi_j(y_k)=\delta_{jk}$ for $j,k=1,...,K$, then the system (4.1) can be approximated by

$$\begin{aligned}&\sum_{k=1}^{K} w_k\int_D q(x,y_k)\left(\sum_{i=1}^{I}\sum_{j=1}^{J} c_{ij}\nabla\phi_i(x)\psi_j(y_k)\right)\cdot\nabla\phi_{i'}(x)\psi_{j'}(y_k)\rho_n(y_k)\ dx\\ &= \sum_{k=1}^{K} w_k\int_D f(x,y_k)\phi_{i'}(x)\psi_{j'}(y_k)\rho_n(y_k)\ dx\\ &\text{for } i'=1,2,...,I \text{ and } j'=1,2,...,J\\ \Rightarrow\quad &\sum_{i=1}^{I} c_{ik}\int_D q(x,y_k)\nabla\phi_i(x)\cdot\nabla\phi_{i'}(x)\ dx = \int_D f(x,y_k)\phi_{i'}(x)\ dx\\ &\text{for } i'=1,...,I, \text{ and } k=1,...,K.\end{aligned} \tag{4.3}$$

This decoupling of the forward problem allows us to estimate the solution of (4.1) by computing solutions $u(x, y_k)$ of a family of deterministic forward problems, each corresponding to a node y_k, that can then be used as interpolants for approximating statistical quantities of interest in (4.2).

4.2 Approximation by Sparse Hierarchical Finite Elements

The choice of the type of nodal basis used to discretize (1.12) (or the adjoint system (3.13)) depends on the smoothness of the stochastic integrands in (4.1) and (4.2) as functions of y. Under certain smoothness conditions on the parameter $q(x, y)$, which are readily satisfied if q is written in terms of its KL expansion, the model output $u(x, y)$ can be shown to be analytic in y, warranting the use of global interpolating basis functions such as Lagrange polynomials [2]. In our case $q(x, y)$ is written in terms of the random variables in the KL expansion of the measured data $\hat{u}$ and hence such smoothness conditions may no longer hold. Consequently, neither the model output u, nor the Lagrange multiplier λ, characterized by the adjoint equation, are guaranteed to exhibit the requisite smoothness as functions of y to allow for their approximation by a global polynomial basis. Here we make use of an interpolating basis of piecewise smooth, multi-linear hat functions.

Assume, without loss of generality, that the stochastic domain $\Gamma^n = [0, 1]^n$. While much is known about interpolation formulas on one-dimensional domains, the problem of computing efficient and accurate multi-dimensional interpolants remains a challenge. Sparse grid methods [6,21,49,59] efficiently combine one-dimensional interpolation schemes to obtain accurate interpolants in higher dimensions with only a moderate number of grid points. Suppose Γ^n is subdivided along each dimension into one-dimensional grids X^{ℓ_t}, $t = 1, 2, ..., n$ of equally spaced points, where the multi-index $\ell = (\ell_1, ..., \ell_n) \in \mathbb{N}^n$ denotes the level of refinement in each direction. In particular, each grid X^{ℓ_t} consists of nodes $\{y_{\ell_t, j_t}\}_{j_t=0}^{m_t^\ell}$, where

$$m_t^\ell = \begin{cases} 1, & \text{if } \ell_t = 1 \\ 2^{\ell_t}, & \text{if } \ell_t > 1 \end{cases} \quad \text{and} \quad y_{\ell_t, j_t} = \begin{cases} 0.5, & \text{if } \ell_t = 1, j_t = 1 \\ 2^{-\ell_t} j_t, & \text{if } \ell_t > 1, \text{ for } j_t = 0, 1, ..., m_t^\ell \end{cases}.$$

For convenience, we define $\mathbf{m}^\ell := (m_1^\ell, ..., m_n^\ell)$ and take $\mathbf{j} \leq \mathbf{m}^\ell$ to mean $j_t \leq m_t^\ell$ for all $t = 1, .., n$. The full tensor product grid X^ℓ on Γ^n, given by

$$X^\ell := X^{\ell_1} \times \cdots \times X^{\ell_n},$$

thus consists of the points $\{y_{\ell, \mathbf{j}}\}_{\mathbf{j} \leq \mathbf{m}^\ell}$. Let $\{\psi_{\ell_t, j_t}\}_{j_1=0}^{m_t^\ell}$ denote a set of one-dimensional, nodal interpolating basis functions centered at the grid points $\{y_{\ell_t, j_t}\}_{j_t=0}^{m_t^\ell}$ of each one-dimensional grid X^{ℓ_t}, $t = 1, ..., n$. We use bases of one-dimensional piecewise linear hat functions, defined by $\psi_{\ell_t, j_t}(y) := 1$ when $\ell_t = 1$, and

$$\psi_{\ell_t, j_t}(y_t) := \psi\left(m_t^\ell \left(y_t - \frac{j_t}{m_t^\ell}\right)\right), \qquad \psi(z) := \begin{cases} 1 - |z|, & \text{if } -1 \leq z \leq 1 \\ 0, & \text{otherwise} \end{cases},$$

when $\ell_t > 1$. A basis function $\psi_{\ell,\mathbf{j}}$ centered at a node $y_{\ell,\mathbf{j}} = (y_{\ell_1,j_1}, ..., y_{\ell_n,j_n})$ in the multi-dimensional grid $X^{\ell_1} \times ... \times X^{\ell_n} \subset [0,1]^n$ can then be obtained by taking the product of the appropriate univariate nodal basis functions, i.e.

$$\psi_{\ell,\mathbf{j}}(y_1, ..., y_n) = \psi_{\ell_1,j_1} \otimes \cdots \otimes \psi_{\ell_n,j_n} := \prod_{t=1}^{n} \psi_{\ell_t,j_t}(y_t),$$

for any $y = (y_1, ..., y_n) \in [0,1]^n$. Note that the one-dimensional grids are nested, i.e. $X^1 \subset X^2 \subset ... \subset X^{\ell_t}$ for any $\ell_t \in \mathbb{N}$. As a result, the subspaces spanned by one-dimensional interpolating basis functions are also nested and hence it is relatively straightforward to compare the accuracy of one-dimensional grids with various refinement levels ℓ_t. A multi-dimensional interpolation formula with refinement level l_t in the t^{th} direction can be obtained by combining the one-dimensional interpolation formulas

$$U^{\ell_t}(v) = \sum_{j_t=0}^{m_t} v(y_{\ell_t,j_t}) \psi_{\ell_t,j_t}$$

to form the full tensor multi-variate interpolant

$$\left(U^{\ell_1} \otimes \cdots \otimes U^{\ell_n}\right)(v) = \sum_{\mathbf{j} \leq \mathbf{m}^{\ell}} v(y_{\ell,\mathbf{j}}) \psi_{\ell,\mathbf{j}}.$$

The number of grid points needed to construct this interpolant is $2^{\ell_1} \cdot 2^{\ell_2} \cdots 2^{\ell_n}$, which scales exponentially as the dimension n of the space increases. In particular, if we require the same refinement level L in each direction, we obtain the L-level full tensor interpolant

$$\left(U^{L} \otimes \cdots \otimes U^{L}\right)(v) = \sum_{\mathbf{j} \leq \mathbf{m}^{L}} v(y_{\ell,\mathbf{j}}) \psi_{\ell,\mathbf{j}},$$

for which 2^{nL} grid points are needed. Sparse grid methods form L-level interpolants $A(L,n)v$ from linear combinations of certain lower order full tensor interpolants

$$A(L,n)(v) = \sum_{L-n+1 \leq |\ell|_1 \leq L} (-1)^{N-|\ell|_1} \begin{pmatrix} n-1 \\ L-|\ell|_1 \end{pmatrix} \left(U^{\ell_1} \otimes \cdots \otimes U^{\ell_n}\right)(v), \tag{4.4}$$

where $L \geq n \in \mathbb{N}$ is called the interpolation level of $A(L,n)$. This formula, although seemingly more costly than the full tensor product formula, is in fact much more economical in its use of grid points. Whereas the full tensor product interpolant $U^L \otimes \cdots \otimes U^L(v)$ requires function values at all nodes on the full grid

$$X^L \times ... \times X^L = \bigcup_{|\ell|_\infty \leq L} X^{\ell_1} \times ... \times X^{\ell_n},$$

$A(L,n)v$ effectively requires function values only at points on the sparse grid

$$H(L,n) = \bigcup_{|\ell|_1 \leq L} X^{\ell_1} \times ... \times X^{\ell_n}.$$

A comparison of the full tensor grid $X^L \times \cdots X^L$, used for full tensor interpolation, and the sparse grid $H(L,n)$ is given in Figure 4.2. This reduction in computational effort comes at the cost of a slightly lower level of accuracy. Indeed, if

$$Z^\ell := \text{span}\{\psi_{\ell,\mathbf{j}}, \mathbf{0} \le \mathbf{j} \le \mathbf{m}^\ell\},$$

then the full tensor product interpolant is exact for functions in the direct sum $\bigoplus_{|\ell|_\infty \le L} Z^\ell$, while the sparse grid interpolant is exact on the smaller subspace $\bigoplus_{|\ell|_1 \le L} Z^\ell$.

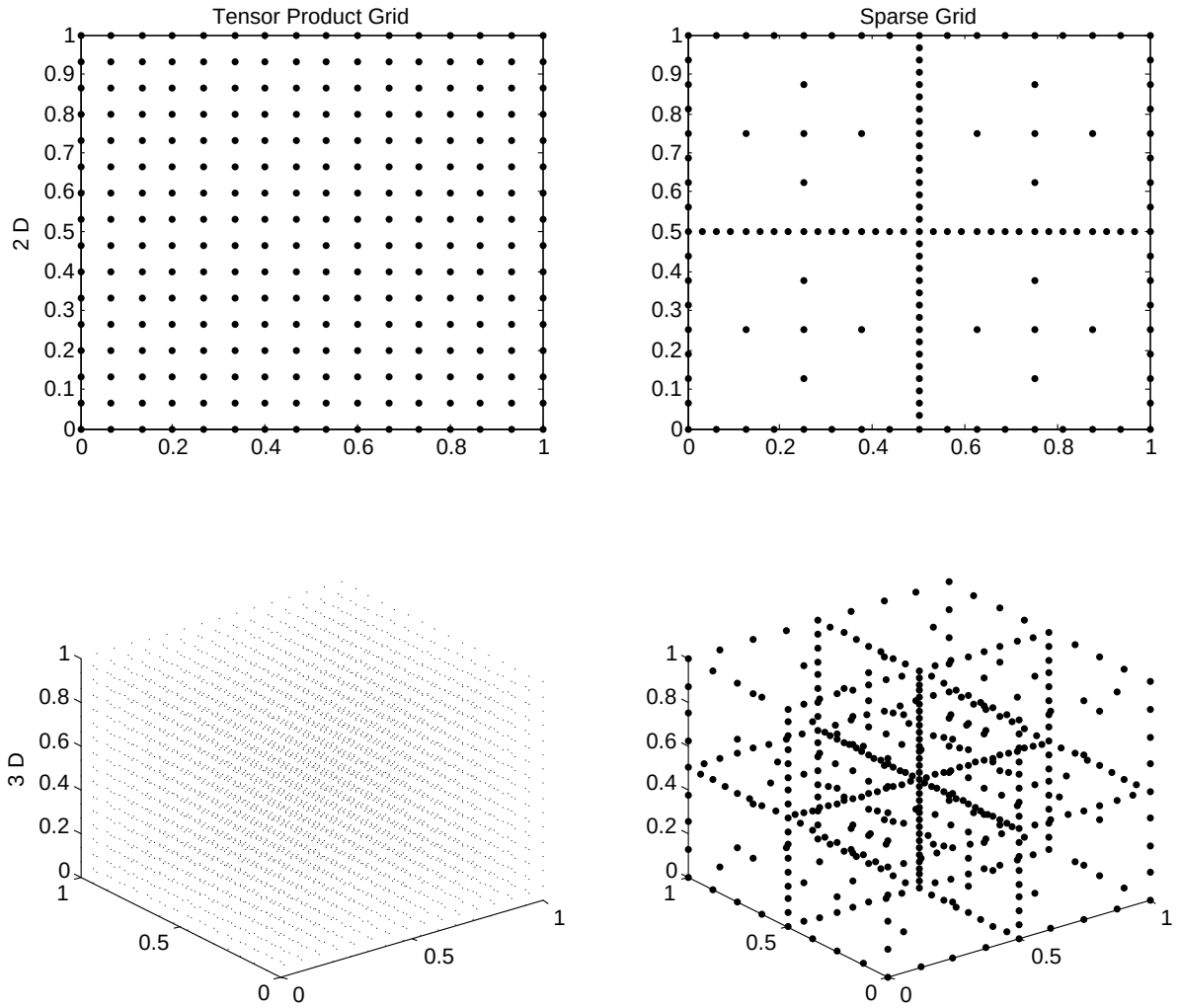

Figure 4.1: Comparison of the 2D and 3D full and sparse grids at interpolation level $L = 4$.

In practice, formula (4.4) is not used directly to construct interpolants. Instead, higher order interpolants are constructed recursively from lower order ones by adding corrections on the appropriately refined grid. Indeed, let $\Delta^{\ell_t} = U^{\ell_t} - U^{\ell_t - 1}$ for $\ell_t \ge 1$ and let $U^0 = 0$. Then it can be shown that for any $L \ge n$

$$A(L,n)v = A(L-1,n)v + \Delta A(L,n),$$

where

$$\Delta A(L,n) = \sum_{|\ell|_1=L} \sum_{\mathbf{j}\leq \mathbf{m}^\ell} \underbrace{[v(y_{\ell,\mathbf{j}}) - A(L-1,n)v(y_{\ell,\mathbf{j}})]}_{w_{\ell,\mathbf{j}}} \cdot \psi_{\ell,\mathbf{j}}(y).$$

The coefficients $w_{\ell,\mathbf{j}} = v(y_{\ell,\mathbf{j}}) - A(L-1,n)v(y_{\ell,\mathbf{j}})$ appearing in the update $\Delta A(L,n)$, also known as hierarchical surpluses, represent the discrepancy between the function v and the $L-1$ level interpolant $A(L-1,n)v$ at the new gridpoints $H(L,n)\backslash H(L-1,n)$. Hierarchical surpluses $w_{\mathbf{j},\ell}$ provide useful *a posteriori* error estimates that can readily be employed by an adaptive scheme to identify the regions in which the grid should be refined [10, 39, 40]. Unfortunately, it is difficult to incorporate adaptive approximation seamlessly into the high-dimensional gradient-based optimization methods developed in Chapter 3. Since the functions q_k, u_k and λ_k are changing at each iteration of the optimization algorithm, the adaptive refinement scheme would have to be adjusted throughout the duration of the algorithm. This can be costly, especially in light of the fact that the relevant bilinear and trilinear forms would have to be updated after each adaptive refinement or coarsening step. In Section 4.5, adaptivity is revisited in the context of a collocation-based method known as the inverse sampling method.

4.3 A Steepest Descent Algorithm

An equivalent formulation of problem (P^n) in which the cost functional depends exclusively on q, is given by

$$\min_{q\in\widetilde{H}_{\text{mix}}} \hat{J}^n(q) = \frac{1}{2}\|u(q) - \hat{u}^n\|^2_{\widetilde{H}^1_0} + \frac{\beta}{2}\|q\|^2_{\widetilde{H}_{\text{mix}}} \quad \text{subject to } q \in Q_{\text{ad}}, \tag{4.5}$$

where $u(q) \in \widetilde{H}^1_0(D)$ is the unique solution of $\tilde{e}(q,u) = 0$. Steepest descent methods obtain iterates approximating the minimizer by moving from a previous iterate in the direction in which the cost function descends most rapidly, i.e. in the direction of the negative gradient. Evidently $\hat{J}^n(q) = J^n(q,u(q))$ and hence its derivative is computed by means of the chain rule

$$\begin{aligned} D_q[\hat{J}^n(q)]h &= \langle D_u[J^n(q,u(q))], D_q[u(q)]h\rangle_{\widetilde{H}^{-1}\times\widetilde{H}^1_0} + \langle D_q[J^n(q,u(q))], h\rangle_{\widetilde{H}^*_{\text{mix}}\times\widetilde{H}_{\text{mix}}} \\ &= \langle D_q[u(q)]^* D_u[J^n(q,u(q))], h\rangle_{\widetilde{H}^*_{\text{mix}}\times\widetilde{H}_{\text{mix}}} + \beta\langle q, h\rangle_{\widetilde{H}_{\text{mix}}}, \end{aligned}$$

where $\widetilde{H}^*_{\text{mix}}$ denotes the dual of $\widetilde{H}_{\text{mix}}$ and $D_q[u(q)]^*$ is the adjoint of $D_q[u(q)]$. The term $D_q[u(q)]^* D_u[J^n(q,u(q))]$ can be obtained by differentiating the equality constraint $\tilde{e}(q,u(q)) = 0$ with respect to q, yielding

$$\langle D_q[\tilde{e}_n(q,u(q))], h\rangle_{\widetilde{H}^*_{\text{mix}}\times\widetilde{H}_{\text{mix}}} + \langle D_u[\tilde{e}_n(q,u(q))], D_q[u(q)]h\rangle_{\widetilde{H}^{-1}\times\widetilde{H}^1_0} = 0 \quad \forall h \in \widetilde{H}_{\text{mix}}$$

and hence

$$\begin{aligned} &D_q[\tilde{e}_n(q,u(q))]^* + D_q[u(q)]^* D_u[\tilde{e}_n(q,u(q))]^* = 0 \text{ in } \widetilde{H}^*_{\text{mix}} \text{ or} \\ &D_q[u(q)]^* = -D_q[\tilde{e}_n(q,u(q))]^* D_u[\tilde{e}_n(q,u(q)))]^{-*}. \end{aligned}$$

Therefore $D_q[u(q)]^* D_u[J^n(q,u(q))] = D_q[\tilde{e}_n(q,u(q))]p$, where $p \in \widetilde{H}_0^1$ solves the adjoint equation

$$D_u[\tilde{e}_n(q,u(q))]^* p = -D_u[J^n(q,u(q))].$$

Since $D_u[\tilde{e}_n(q,u)]^* : \widetilde{H}^{-1} \to \widetilde{H}_0^1 : v \mapsto \langle q\nabla v, \nabla\cdot\rangle$, the adjoint equation takes the form

$$\langle q\nabla p, \nabla\cdot\rangle = -\langle \nabla u - \nabla\hat{u}, \cdot\rangle,$$

which is the same as equation (3.13), derived in Chapter 3. Finally, $D_q[\tilde{e}_n(q,u)]^* : \widetilde{H}_0^1 \to \widetilde{H}_{\text{mix}}^*$ maps $p \mapsto \langle \cdot\nabla p, \nabla u\rangle$ and hence

$$\langle \nabla\hat{J}^n(q), h\rangle := D_q[\hat{J}^n(q)]h = \langle h\nabla p, \nabla u\rangle + \beta\langle q, h\rangle_{\widetilde{H}_{\text{mix}}}. \tag{4.6}$$

The presence of additional constraints Q_{ad} on q complicates the procedure somewhat, since moving in directions of steepest descent may lead to infeasible iterates. The fact that the admissible set Q_{ad} is closed and convex in $\widetilde{H}_{\text{mix}}(D)$, however, allows us to consider set projections of the form

$$\mathbf{P}_{Q_{\text{ad}}} : \widetilde{H}_{\text{mix}}(D) \to Q_{\text{ad}} | s \mapsto \mathbf{P}_{Q_{\text{ad}}}(s) \in Q_{\text{ad}}, \quad \|\mathbf{P}_{Q_{\text{ad}}}(s) - s\| = \min_{q\in Q_{\text{ad}}} \|s - q\|_{\widetilde{H}_{\text{mix}}}.$$

Algorithms such as the projected Armijo Algorithm [24] provide a (theoretically) simple means to find the optimal q. It can be shown [24] that a step size σ_k satisfying the projected Armijo rule always exists and leads to a first order globally convergent optimization algorithm.

Algorithm 1 The Projected Armijo Algorithm

Choose $q_0 \in \widetilde{H}_{\text{mix}}(D)$, $k = 0$, $\gamma \in (0,1)$.
while not converged **do**
  Set $s_k = -\nabla\hat{J}^n(q_k)$.
  Choose σ_k to be the $\max\{1, 1/2, 1/4, ...\}$ so that

$$\hat{J}^n(\mathbf{P}_{Q_{\text{ad}}}(q_k + \sigma_k s_k)) - \hat{J}^n(q_k) \leq -\frac{\sigma_k}{\gamma}\|\mathbf{P}_{Q_{\text{ad}}}(q_k + \sigma_k s_k) - q_k\|_{\widetilde{H}_{\text{mix}}}. \tag{4.7}$$

  Set $q_{k+1} := \mathbf{P}_{Q_{\text{ad}}}(q_k + \sigma_k s_k)$.
  Set $k = k + 1$ and test for convergence.
end while

4.4 An Augmented Lagrangian Algorithm

The augmented Lagrangian method may be viewed as a modified penalty method. Quadratic penalty methods avoid explicit enforcement of the equality constraint $e_n(q,u) = 0$, by including an additional term, that penalizes violations of the constraint, in the cost functional. For (P^n), this algorithm would require solving a sequence of subproblems of the form

$$\min_{(q,u)\in\widetilde{H}_{\text{mix}}\times\widetilde{H}_0^1} J^n(q,u) + \frac{c_k}{2}\|e_n(q,u)\|^2_{\widetilde{H}_0^1}$$
$$\text{subject to } q \in Q_{\text{ad}},$$

where the sequence $\{c_k\}$ increases steadily as $k \to \infty$. Convergence of this class of methods to a minimizer q_n^* of (P^n) requires the sequence $c_k \to \infty$ as $k \to \infty$, leading to a progressive deterioration in the conditioning of the subproblems. Augmented Lagrangian methods consider instead the augmented problem

$$\begin{aligned} &\min_{(q,u)\in \widetilde{H}_{\mathrm{mix}}\times \widetilde{H}_0^1} J_c^n(q,u) := J^n(q,u) + \frac{c}{2}\|e(q,u)\|^2_{\widetilde{H}_0^1} \\ &\text{subject to } q \in Q_{\mathrm{ad}}, \quad e_n(q,u) = 0. \end{aligned} \tag{P_c^n}$$

It can readily be verified that the minimizers (q_n^*, u_n^*) of (P^n) coincide with those of the augmented problem (P_c^n). Using the same arguments as in Theorem 3.2.2, we can also derive first order necessary optimality conditions, i.e. there exists a Lagrange multiplier μ_n^* so that

$$D_{(q,u)}\left[L_c(q_n^*, u_n^*, \mu_n^*)\right](h,v) \geq 0 \text{ for all } (h,v) \in C(q^*) \times \widetilde{H}_0^1$$

where the augmented Lagrangian $L_c : \widetilde{H}_{\mathrm{mix}} \times \widetilde{H}_0^1 \times \widetilde{H}_0^1 \to \mathbb{R}$ is given by

$$L_c(q,u,\mu) = \frac{1}{2}\|u - \hat{u}^n\|^2_{\widetilde{H}_0^1} + \frac{\beta}{2}\|q\|^2_{\widetilde{H}_{\mathrm{mix}}} + \langle \mu, e_n(q,u)\rangle_{\widetilde{H}_0^1} + \frac{c}{2}\|e_n(q,u)\|^2_{\widetilde{H}_0^1} \tag{4.8}$$

and

$$C(q^*) = \{l(c - q^*) : c \in Q_{\mathrm{ad}},\ l \geq 0\}.$$

In particular, the adjoint equation takes the form

$$\langle q\nabla(\mu + ce_n(q,u)), \nabla v\rangle = -\langle u - \hat{u}^n, v\rangle_{\widetilde{H}_0^1} \text{ for all } v \in \widetilde{H}_0^1. \tag{4.9}$$

Upon comparing (4.9) with (3.13), we are led by the Lax Milgram Lemma to conclude that the solution μ of (4.9) is related to the solution λ of (3.13) by $\lambda = \mu + ce_n(q,u)$. Moreover, the feasibility of the minimizers (q_n^*, u_n^*) of (P_c^n) implies the equality constraint $e_n(q_n^*, u_n^*) = 0$ and hence $\mu_n^* = \lambda_n^*$.

Under certain coercivity conditions on the second Fréchet derivative of L_c [26], it can be shown that minimizers of the constrained problem (P_c^n) can also be found by minimizing

$$\min_{(q,u)\in \widetilde{H}_{\mathrm{mix}}\times \widetilde{H}_0^1} L_c(q,u,\lambda_n^*), \quad \text{subject to } q \in Q_{\mathrm{ad}} \tag{4.10}$$

for c in some interval. Since the exact Lagrange multiplier λ_n^* is unknown in general, it must be approximated. The augmented Lagrangian algorithm, described below, estimates λ_n^* iteratively, together with q_n^* and u_n^*, by solving a series of auxiliary minimization subproblems (P_{aux}) with no equality constraint and fixed $\lambda = \lambda_{n,k}$ and using the minimizers $(q_{n,k}^*, u_{n,k}^*)$ to update the Lagrange multiplier via the correction $\lambda_{n,k+1} = \lambda_{n,k} + c_k e_n(q_{n,k}^*, u_{n,k}^*)$. Note that the sequence $\{c_k\}$ below is not required to converge to ∞.

Algorithm 2 The Augmented Lagrangian Algorithm

Choose $\lambda_{n,0} \in \widetilde{H}_0^1(D)$, and non-decreasing sequence $\{c_k\}$ with $c_0 > 0$.
Set $k = 0$.
while not converged **do**
Obtain minimizers $(q_{n,k}^*, u_{n,k}^*)$ by solving the auxiliary problem

$$\min_{(q,u)\in \widetilde{H}_{\mathrm{mix}} \times \widetilde{H}_0^1} L_{c_k}(q, u, \lambda_{n,k}) \text{ subject to } q \in Q_{\mathrm{ad}} \tag{P_{aux}}$$

Set $\lambda_{n,k+1} := \lambda_{n,k} + c_k e_n(q_{n,k}^*, u_{n,k}^*)$.
Set $k = k + 1$ and test for convergence.
end while

While the coercivity condition on the second Fréchet derivative of L_c mentioned earlier is not necessary for the convergence of Algorithm 2, it is required [32] that the following second order sufficient optimality condition holds at the minimizer (q_n^*, u_n^*).

Assumption 4.4.1. *Assume there exists a constant $\tau = \tau(\beta) > 0$ so that*

$$D^2_{(q,u)}[L(q_n^*, u_n^*, \lambda_n^*)](h, v)^2 \geq \tau(\|h\|^2_{\widetilde{H}_{\mathrm{mix}}} + \|v\|^2_{\widetilde{H}_0^1}) \quad \textit{for all } (h, v) \in \widetilde{H}_{\mathrm{mix}} \times \widetilde{H}_0^1.$$

Under this assumption, the following convergence theorem can be shown to hold. Its original proof, formulated in a general Hilbert space setting, carries over directly to our problem. We refer the interested reader to the cited references.

Theorem 4.4.2 (Convergence, [26, 32])**.** *The auxiliary problem (P_{aux}) has a solution for each $k \in \mathbb{N} \cup \{0\}$. If Assumption 4.4.1 also holds then the iterates of Algorithm 2 satisfy*

$$\frac{\tau}{2}\left(\|q_{n,k}^* - q_n^*\|^2_{\widetilde{H}_{\mathrm{mix}}} + \|u_{n,k}^* - u_n^*\|^2_{\widetilde{H}_0^1}\right) + \frac{1}{2c_k}\|\lambda_{n,k}^* - \lambda_n^*\|^2_{\widetilde{H}_0^1} \leq \frac{1}{2c_k}\|\lambda_{n,k-1}^* - \lambda_n^*\|^2_{\widetilde{H}_0^1}.$$

Consequently,

$$\sum_{k=0}^{\infty}\left(\|q_{n,k}^* - q_n^*\|^2_{\widetilde{H}_{\mathrm{mix}}} + \|u_{n,k}^* - u_n^*\|^2_{\widetilde{H}_0^1}\right) \leq \frac{1}{2\tau c_0}\|\lambda_{n,0} - \lambda_{n,k}^*\|^2_{\widetilde{H}_0^1}$$

and hence

$$(q_{n,k}^*, u_{n,k}^*) \to (q_n^*, u_n^*) \textit{ as } k \to \infty \textit{ and } \|\lambda_{n,k}^* - \lambda_n^*\|_{\widetilde{H}_0^1} \leq \|\lambda_{n,0} - \lambda_n^*\|_{\widetilde{H}_0^1} \quad \textit{for all } k = 1, 2, \ldots .$$

If in addition $q_n^* \in \mathrm{int}\ (Q_{\mathrm{ad}})$, then there exists an index $k_0 \in \mathbb{N} \cup \{0\}$ and a constant C so that

$$\|q_n^* - q^*\|_{\widetilde{H}_{\mathrm{mix}}} + \|u_n^* - u^*\|_{\widetilde{H}_0^1} + \|\lambda_{n,k}^* - \lambda_n^*\|_{\widetilde{H}_0^1} \leq \frac{C}{\sqrt{c_k}}\|\lambda_{n,k-1}^* - \lambda_n^*\|_{\widetilde{H}_0^1} \textit{ for } k \geq k_0,$$

from which it follows that $(q_{n,k}^*, u_{n,k}^*, \lambda_{n,k}^*) \to (q_n^*, u_n^*, \lambda_n^*)$ in $\widetilde{H}_{\mathrm{mix}} \times \widetilde{H}_0^1 \times \widetilde{H}_0^1$ as $k \to \infty$.

Note that the cost functional L_c is quadratic in q for fixed u and λ and quadratic in u for fixed q and λ, suggesting the use of sequential splitting methods to speed up the solution of the auxiliary subproblem. An alternative to Algorithm 2, replaces the subproblem (P_{aux}) with a pair of optimization problems for q and u.

Algorithm 3 The Augmented Lagrangian Algorithm with Sequential Splitting

Choose $\lambda_{n,0} \in \widetilde{H}_0^1(D)$, and non-decreasing sequence $\{c_k\}$ with $c_0 > 0$.
Set $k = 0$.
while not converged **do**
  Solve the auxiliary problem sequentially:
  For iterates $q^*_{n,k}$ and $u^*_{n,k}$ and c_k
  Obtain $q^*_{n,k+1}$ by solving the minimization problem

$$\min_{q \in Q_{\text{ad}}} L_{c_k}(q, u^*_{n,k}, \lambda^*_{n,k}).$$

  Obtain $u^*_{n,k+1}$ by solving the minimization problem

$$\min_{u \in \widetilde{H}_0^1} L_{c_k}(q_{n,k+1}, u, \lambda^*_{n,k}).$$

  Set $\lambda_{n,k+1} := \lambda_{n,k} + c_k e_n(q^*_{n,k+1}, u^*_{n,k+1})$.
  Set $k = k + 1$ and test for convergence.
end while

4.5 An Inverse Sampling Method

The optimization-based approaches outlined above simultaneously approximate the spatial as well as stochastic components of q. During the execution of these algorithms, the inherent parallel structure in the forward and adjoint problems can be exploited by collocation-based discretizations. Thus, the computation of the cost functional and gradient can be sped up considerably. In this section, we develop a sampling-based method for the inverse problem (P) that lends itself more fully to parallelization, comparable to the level of parallelization achieved for the forward problem or for Bayesian methods. As an illustration of how naturally this method arises in an experimental or industrial setting, consider the problem of determining the statistical distribution of the conductivity coefficient q in a collection of plates that are manufactured from an inhomogeneous material in an experimental or industrial setting. An intuitively appealing procedure to obtain a statistical estimate of q is to subject a random sample of plates to a heat source under similar conditions and record the temperature distribution for each plate. Upon estimating the conductivity coefficient deterministically for each plate in the sample, we can statistically aggregate the results. It therefore seems like a natural idea to treat the inverse problem itself as a sampling problem.

Within the setting of the variational problem (P), our approach amounts to finding a parameter $q_p \in \mathscr{H}$ together with its corresponding model output $u_p := u(q_p)$ that satisfy the following family of deterministic problems pointwise a.s. on Ω:

$$\begin{aligned} \min_{(q,u)\in H\times H_0^1} & J^\omega(q,u) = \frac{1}{2}\|u-\hat{u}(\cdot,\omega)\|_{H_0^1}^2 + \frac{\beta}{2}\|q\|_H^2 \\ \text{subject to: } & \int_D q\nabla u\cdot\nabla v(x)\,dx = \int_D f(x)v(x)\,dx \ \text{ for all } v\in H_0^1 \\ & q\in H,\ \|q\|_H \le q_{\max},\ 0<q_{\min}\le q(x) \text{ a.e. on } D. \end{aligned} \tag{P^ω}$$

The function q_p is therefore defined pointwise a.s. on Ω by $q_p(\cdot,\omega) := q_\omega^*$, where $(q_\omega^*, u(q_\omega^*))$ are minimizers of the problem (P^ω). Since $\|q_\omega^*\|_H \le q_{\max}$ a.s. on Ω, $q_p \in \mathscr{H}$. Solving (P^ω) repeatedly for different realizations ω_i then yields an interpolant of q_p, which can be used to approximate statistical quantities of interest via numerical quadrature. Clearly (q_p, u_p), defined as the pointwise minimizer of (P^ω), also solves the original variational problem (P). Indeed, for any minimizer (q^*, u^*) of (P), the pointwise optimality of $(q_p, u(q_p))$ yields the pointwise inequality

$$J(q_p(\cdot,\omega), u_p(\cdot,\omega)) \le J(q^*(\cdot,\omega), u^*(\cdot,\omega)) \quad \text{for all } \omega\in\Omega,$$

and hence (q_p, u_p) minimizes the cost functional J. Moreover, it is evident that q_p satisfies the pointwise bounds constituting the admissibility criteria, while (q_p, u_p) satisfies the equality constraint $e(q,u)=0$, by virtue of the relation between (1.9) and (1.10). This approach therefore solves (P) by embedding a deterministic parameter identification algorithm within a sampling scheme, whereas the gradient-based optimization methods discussed in the previous sections employ sampling-type methods within each iteration loop of an optimization algorithm. Bayesian methods offer yet another approach in which only the forward problem is embedded within a sampling scheme.

Although this method involves multiple deterministic optimization problems, these can be solved completely in parallel and each deterministic 'inversion' is relatively small, compared with the optimization problem in the gradient-based approach. Moreover, the regularization term $\frac{\beta}{2}\|q\|_H^2$ ensures that proximal realizations of $\hat{u}$ give rise to similar deterministic optimization problems, a property which may readily be exploited by the optimization algorithms used for deterministic inversion. This approach also leads to a greater flexibility in the resolution of spatial features in the inverse problem, allowing for the use of spatially adaptive optimization algorithms to locate discontinuities in q for example. The samples obtained by this type of inversion also have an immediate physical significance, in contrast to the samples obtained by Bayesian method.

One drawback of this approach is that each sample is relatively costly to obtain. If the parameter q is statistically complex, then obtaining a representative sample may become prohibitively expensive. However, since the statistical complexity of q is directly linked to that of $\hat{u}$ by virtue of Theorem 3.1.8, we can obtain a simple approximation of q in the

statistical sense, by considering the 'finite noise' discretization $(P^n_{\widetilde{H}})$ of (P). The 'finite noise' analogue of (P^ω) can then be written as

$$\begin{aligned} \min_{(q,u)\in H\times H_0^1} & J^{y,n}(q,u) = \frac{1}{2}\|u-\hat{u}^n(\cdot,y)\|^2_{H_0^1} + \frac{\beta}{2}\|q\|^2_H \\ \text{subject to: } & \int_D q\nabla u\cdot\nabla v(x)\,dx = \int_D f(x)v(x)\,dx \ \text{ for all } v\in H_0^1 \\ & q\in H,\ \|q\|_H\le q_{\max},\ 0<q_{\min}\le q(x) \text{ a.e. on } D. \end{aligned} \tag{P^n_y}$$

Clearly the 'finite noise' form of this inverse sampling method has the same useful properties as (P^ω), with the added benefit that q is now only estimated in terms of the finite set of random variables $\{Y_1,...,Y_n\}$ that contribute most prominently to the statistical variation in $\hat{u}^n$. In addition, the 'finite noise' problem allows for the estimation of q by a variety of interpolation schemes, including stochastic collocation schemes based on sparse grid hierarchical finite element methods. Since inverse sampling methods construct q directly by interpolation, adaptive interpolation/quadrature methods can readily be implemented.

Chapter 5

Numerical Experiments

5.1 Introduction

This chapter centers on the numerical implementation of the variational methods discussed in this dissertation, used to identify the uncertain diffusion parameter q in the elliptic system (1.10). The statistical estimation of the parameter q can be achieved in three distinct, but related steps. The pre-processing phase consists of statistically analyzing the measured data $\hat{u}$ to obtain its 'finite noise,' or truncated Karhunen-Loève representation as well as a reasonable estimate of the joint density function of the associated random vector $Y = (Y_1, ..., Y_n)$. Once the 'finite noise' problem is specified, one of the optimization strategies introduced in Chapter 4 can be used to locate the random field q. This q gives rise to a model output most closely resembling the 'finite noise' data, subject to additional regularity requirements. Finally, the estimate obtained can be used to derive related statistical quantities of interest, such as the mean, variance or probabilities of failure. Section 5.2 discusses how the truncated KL expansion and the various optimization algorithms can be carried out under numerical discretizations of the functions $q, u, \hat{u}$ and f. Section 5.3 addresses three numerical examples that illustrate the use of both the gradient-based strategies, in which the entire q is estimated within an optimization algorithm, as as well as the inverse sampling method, in which q is obtained by sampling the minimizers of a family of deterministic inverse problems.

5.2 Numerical Discretization

Let $V^M \subset \mathscr{H}^1(D)$ be a finite element space over the physical domain D, spanned by a basis $\{\phi_1, ..., \phi_M\}$ of piecewise continuous polynomial functions. Furthermore, let V_0^M be the subset of V^M that approximates functions in $\mathscr{H}_0^1(D)$. For our one-dimensional examples, we assume $V_0^M = \text{span}\{\phi_2, \ldots, \phi_{M-1}\}$. Thus, any functions $v \in \mathscr{H}^1(D)$ and $v_0 \in \mathscr{H}_0^1(D)$

can be approximated spatially by

$$v^M(x,\omega) = \sum_{i=1}^{M} v(x_i,\omega)\phi_i(x) \quad \text{and} \quad v_0^M(x,\omega) = \sum_{j=2}^{M-1} v(x_i,\omega)\phi_i(x),$$

respectively, for every $\omega \in \Omega$. We denote by $\mathbf{v}^M(\omega) = [v(x_1,\omega),...,v(x_M,\omega)]$ the spatial coordinatization of v^M and define the bilinear forms

$$[A]_{i_1,i_2} = \langle \phi_{i_1}, \phi_{i_2} \rangle \quad \text{and} \quad [A_x]_{i_1,i_2} = \langle \nabla\phi_{i_1}, \nabla\phi_{i_2} \rangle, \quad \text{for } i_1, i_2 = 1,\ldots,M.$$

The inner products in $L^2(D\times\Omega)$ or $\widetilde{H}_0^1(D)$ of any two functions $v, w \in V^M$ with respective coordinatizations $\mathbf{v}$ and $\mathbf{w}$ are therefore given by

$$\begin{aligned} \langle v, w \rangle_{\tilde{L}^2} &= \mathbf{v}^T A\ \mathbf{w}, \quad \text{and} \\ \langle v, w \rangle_{\widetilde{H}_0^1} &= \langle \nabla v, \nabla w \rangle = \mathbf{v}^T A_x\ \mathbf{w}. \end{aligned}$$

The first and last rows and columns of the bilinear forms A, A_x can be deleted for the finite element approximations $v_0^M \in V_0^M$ of functions in H_0^1. We denote these modified bilinear forms by A^0 and A_x^0.

5.2.1 The Karhunen-Loève Expansion

In order to reduce our variational problem (P) to its 'finite noise' approximation (P^n), we must first compute the truncated KL expansion of the measured data $\hat{u} \in \mathscr{H}_0^1$. This requires the spectral decomposition of the compact covariance operator $\mathcal{C}_{\hat{u}} : \widetilde{H}_0^1 \to \widetilde{H}_0^1$, defined in terms of its covariance kernel

$$\begin{aligned} C_{\hat{u}}(x,x') &= \mathbb{E}[(\hat{u}(x') - \hat{u}_0(x'))(\hat{u}(x) - \hat{u}_0(x))] \\ v \in \mathscr{H}_0^1 \mapsto \mathcal{C}_{\hat{u}} v(x') &= \int_D \nabla_x C(x,x') \cdot \nabla v(x)\, dx \in \mathscr{H}_0^1, \end{aligned}$$

where $\hat{u}_0(x) = \mathbb{E}[\hat{u}(x,\cdot)]$. In practice, $\hat{u}$ commonly occurs in the form of a random sample $\{\hat{u}(x,\omega_j\}_{j=1}^{N^s}$. An approximate KL expansion can be obtained by discretizing $\hat{u}(\cdot,\omega_j)$ in space for each $j = 1,...,N^s$, i.e. $\hat{u}^M(x,\omega_j) = \sum_{i=1}^M \hat{u}(x_i,\omega_j)\phi_i(x)$ for $j = 1,...,N^s$. We can then form approximate sample estimates of the mean $\hat{u}_0^h(x)$ and the covariance kernel $C_{\hat{u}}^h(x,x')$ by computing

$$\begin{aligned} \hat{u}_0^M(x) &:= \frac{1}{N^s} \sum_{j=1}^{N^s} \hat{u}^M(x,\omega_j) \quad \text{and} \\ C_{\hat{u}}^M(x,x') &:= \frac{1}{N^s} \sum_{j=1}^{N^s} \left(\hat{u}^M(x,\omega_i) - \hat{u}_0^M(x)\right)\left(\hat{u}^M(x',\omega_i) - \hat{u}_0^M(x')\right), \end{aligned}$$

It is also convenient to represent the spatially discretized sample $\{\hat{u}^M(x,\omega_j)\}_{j=1}^{N^s}$ as an $N^s \times M$ matrix $\hat{\mathbf{U}}^M$ of basis coefficients, where $\hat{\mathbf{U}}^M_{i,j} = [\hat{u}(x_j,\omega_i)]$ with $i = 1,...,N^s$ and $j = 1,...,M$. In this notation

$$\hat{\mathbf{u}}_0^M := [\hat{u}_0^M(x_1),\cdots,\hat{u}_0^M(x_M)] = \frac{1}{N^s}\left[\sum_{i=1}^{N^s}\hat{\mathbf{U}}^M_{i,1},\cdots,\sum_{i=1}^{N^s}\hat{\mathbf{U}}^M_{i,M}\right]$$

$$\mathbf{C}^M_{\hat{u}} := \left[C^M_{\hat{u}}(x_{j_1},x_{j_2})\right]_{j_1,j_2=1}^M = \frac{1}{N^s}\left(\hat{\mathbf{U}}^M - \begin{bmatrix}\hat{\mathbf{u}}_0^M\\ \vdots \\ \hat{\mathbf{u}}_0^M\end{bmatrix}\right)\left(\hat{\mathbf{U}}^M - \begin{bmatrix}\hat{\mathbf{u}}_0^M\\ \vdots \\ \hat{\mathbf{u}}_0^M\end{bmatrix}\right)^T.$$

The covariance operator $\mathcal{C}_{\hat{u}} : \mathscr{H}_0^1 \to \mathscr{H}_0^1$ can therefore be approximated by the finite dimensional operator $\mathcal{C}^M_{\hat{u}} : V^M \to V^M$ that maps $v^M = \sum_{j=1}^M v_j\phi(x) \in V^M$ to V^M,

$$\begin{aligned} w^M(x') \in V^M := \mathcal{C}^M_{\hat{u}} \circ v^M(x') &= \int_D \nabla_x C^M_{\hat{u}}(x,x')\cdot\nabla_x v^M(x)\,dx \\ &= \sum_{j_1=1}^M\sum_{j_2=1}^M C^M_{\hat{u}}(x_{j_1},x')v_{j_2}\int_D \nabla\phi_{j_1}(x)\cdot\nabla\phi_{j_2}(x)\,dx \\ \Rightarrow \quad \mathbf{w}^M &= \mathbf{C}^M_{\hat{u}}A_x\mathbf{v}^M. \end{aligned}$$

The spectral decomposition of the approximate covariance operator $\mathcal{C}^M_{\hat{u}}$ therefore amounts to finding the eigenpairs $(\lambda,\mathbf{v})$ satisfying

$$\mathbf{C}^M_{\hat{u}}A_x\mathbf{v} = \lambda\mathbf{v} \quad \Leftrightarrow \quad A_x\mathbf{C}^M_{\hat{u}}A_x\mathbf{v} = \lambda A_x\mathbf{v}.$$

Note that since the covariance matrix is positive definite, the eigenvectors may be chosen to be orthonormal. Moreover, while the countable sequence $\{\lambda_i, v_i\}$ of eigenpairs of the infinite dimensional operator satisfy $\lambda_i \to 0$ as $i \to \infty$, the finite dimensional operator only has a finite rank and therefore a finite dimensional eigenspace. The decay in the eigenvalues allows us to determine a suitable truncation level $n \leq M$ and hence we obtain the truncated, discretized KL expansion

$$\hat{u}^M(x,\omega_i) = \hat{u}_0^M(x) + \sum_{k=1}^n \sqrt{\lambda_k}v_k^M(x)Y_k(\omega_i),$$

sampled at the points $\{\omega_i\}_{i=1}^{N^s}$, where $\{Y_k(\omega_i)\}_{k=1}^n$ are realizations of is an uncorrelated sequence of zero mean random variables $\{Y_k\}_{k=1}^n$, so that

$$Y_k(\omega_i) = \frac{1}{\sqrt{\lambda_k}}\int_D \left(\hat{u}^M(x,\omega_i) - \hat{u}_0^M(x)\right)v_k^M(x)\,dx \tag{5.1}$$

for $k = 1,...,n$ and $i = 1,...,N^s$.

Estimation of the Joint Density Function

The formulation of the 'finite noise' problem (P^n) requires the estimation of the density function ρ_n of the random vector $Y = [Y_1,\cdots,Y_n]$ appearing in the truncated KL expansion

$\hat{u}^n$ of $\hat{u}$. Let $\{Y(\omega_j)\}_{j=1}^{N^s}$ be the sample realizations of the random vector Y, obtained in (5.1), from which a joint distribution must be estimated. Multi-dimensional density estimation is a vast and active field of statistical research, well beyond the scope of this dissertation, and the reader is referred to the books [29, 57], as well as the survey article [58], for a more exhaustive treatment of the subject. There are two broad approaches to distribution fitting. Parametric methods assume that the underlying parametric form of the distribution from which the sample has been drawn, is known and hence these approaches make use the data only to fit the necessary parameters. For example, if the random vector Y is assumed to be multivariate normal, i.e. $Y \sim MVN(\mu, \Sigma)$, then the general form of ρ_n is known and the data $\{Y(\omega_j)\}_{j=1}^{N_s}$ is used to obtain statistics of the population mean μ and covariance matrix Σ. Nonparametric methods, on the other hand, make no assumptions on the type of distribution from which the data is drawn and therefore must take into account the geometric properties of the data, such as clustering, skewness and modality, as well as assumptions on the densities smoothness, to obtain an empirical multivariate density describing the pertinent qualitative aspects of the data. The random vectors encountered in the following two examples are only of moderate size and we make use of kernel density estimators to approximate the relevant densities empirically.

Discretization in the Stochastic Component

As discussed in Chapter 4, we use an interpolating sparse grid hierarchical finite element basis $\{\psi_{\ell,i}\}$, where $L - n + 1 \leq |\ell|_1 \leq L$ and $\mathbf{i} \leq \mathbf{m}^\ell$, to effect the discretization of the functions $q \in \widetilde{H}_{\text{mix}}(D)$ and $u, \hat{u} \in \widetilde{H}_0^1(D)$ in their stochastic components. Suppose $i = 1, ..., N$ is an enumeration of the sparse grid points, i.e.

$$\{y_i\}_{i=1}^N = \{y_{\ell,\mathbf{i}} : L - n + 1 \leq |\ell|_1 \leq L,\ \mathbf{i} \leq \mathbf{m}^\ell\}.$$

Generally, it is more convenient to use this simple indexing scheme, although it may also be necessary to use the multi-index notation established in Chapter 4. For the sake of simplicity, we will use these two indexing systems interchangeably in the sequel. For a function $v : D \times \Gamma^n \to \mathbb{R}$ with spatial discretization

$$v^M(x, y) = \sum_{j=1}^{M} v(x_j, y)\phi_j(x),$$

let $\{y_i\}_{i=1}^N$ be the set of sparse grid points at a given level and $\{v_+(x_j, y_i)\}_{i=1}^N$ be the corresponding collection of hierarchical surpluses of $v(x_j, \cdot)$. Then the full discretization $v^{M,N}(x, y)$ of $v(x, y)$ takes the form

$$v_+^{M,N}(x, y) = \sum_{i=1}^{N} v_+^M(x, y_i)\psi_i(y) = \sum_{i=1}^{N}\sum_{j=1}^{M} v_+(x_j, y_i)\phi_j(x)\psi_i(y).$$

Note that the fully discretized function $v^{M,N}$ can also be represented by the matrix of interpolants $[\mathbf{V}^{M,N}]_{ij} = v(x_j, y_i)$ or equivalently, using the matrix of hierarchical surpluses

$$\left[\mathbf{V}_+^{M,N}\right]_{ij} = v_+(x_j, y_i), \quad \text{where} \quad i = 1, ..., N \quad \text{and} \quad j = 1, ..., M.$$

Approximation of the Cost Functional and Gradient

Both cost functionals

$$\hat{J}(q) = \frac{1}{2}\|u(q) - \hat{u}^n\|^2_{\widetilde{H}^1_0} + \frac{\beta}{2}\|q\|^2_{\widetilde{H}_{\text{mix}}} \quad \text{and}$$
$$L_c(q,u,\lambda) = \frac{1}{2}\|u - \hat{u}^n\|^2_{\widetilde{H}^1_0} + \frac{\beta}{2}\|q\|^2_{\widetilde{H}_{\text{mix}}} + \langle q\nabla u, \nabla\lambda\rangle + \frac{c}{2}\|e_n(q,u)\|^2_{\widetilde{H}^1_0}$$

appearing in optimization problems (4.5) and (4.8), respectively, require the evaluation of high-dimensional stochastic integrals. We approximate these integrals with the help of the weighted bilinear forms S_ρ and S^r_ρ and the trilinear form T_ρ, given by

$$[S_\rho]_{i_1,i_2} = \int_{\Gamma^n} \psi_{i_1}(y)\psi_{i_2}(y)\ \rho_n(y)\ dy,$$
$$\left[S^r_\rho\right]_{i_1,i_2} = \sum_{|\gamma|_\infty \le r} \int_{\Gamma^n} D^\gamma_y \psi_{i_1}(y) D^\gamma_y \psi_{i_2}(y)\ \rho_n(y)\ dy, \quad \text{and}$$
$$[T_\rho]_{i_1,i_2,i_3} = \int_{\Gamma^n} \psi_{i_1}(y)\psi_{i_2}(y)\psi_{i_3}(y)\ \rho_n(y)\ dy$$

as $i_1, i_2, i_3 = 1, \ldots, n$. The evaluation of these multi-dimensional integrals for any given density function ρ_n is a challenging task in general. Note that, whereas each basis function $\psi_i(y)$ is of the form

$$\psi_{\mathbf{l},\mathbf{i}}(y_{\mathbf{l},\mathbf{i}}) = \prod_{t=1}^{n} \psi_{l_t,i_t}(y_{l_t,i_t}) \tag{5.2}$$

for the appropriate multi-indices $\mathbf{l}$ and $\mathbf{i}$, the density function ρ_n cannot necessarily be decomposed into the product of its marginals. This prevents an effective decoupling of the integral into the product of integrals over one-dimensional domains. In the case of the bilinear form S_ρ, we circumvent this difficulty by approximating the density function ρ_n with its sparse grid hierarchical interpolant ρ^N_n, which leads to

$$[S_\rho]_{i_1,i_2} \approx \left[S_{\rho^N_n}\right]_{i_1,i_2} = \sum_{i_3=1}^{N} \rho^N_{n,z}(y_{i_3}) \int_{\Gamma^n} \psi_{i_1}(y)\psi_{i_2}(y)\psi_{i_3}(y)\ dy \quad \text{or} \quad S_{\rho^N_n} = \sum_{i_3=1}^{N} \rho^N_{n,z}(y_{i_3}) T_{\cdot,\cdot,i_3},$$

where T is the unweighted trilinear form $[T]_{i_1,i_2,i_3} = \int_{\Gamma^n} \psi_{i_1}(y)\psi_{i_2}(y)\psi_{i_3}(y)\ dy$ for indices $i_1, i_2, i_3 = 1, ..., N$. The array T is relatively simple to compute, in comparison to T_ρ, since the multi-dimensional integral required to compute T readily decomposes into the product of one-dimensional integrals. Indeed, suppose the indices i_1, i_2 and i_3 correspond to pairs of multi-indices $\{\mathbf{l}, \mathbf{i}\}$, $\{\mathbf{l}', \mathbf{i}'\}$ and $\{\mathbf{l}'', \mathbf{i}''\}$ respectively. Then the $(i_1, i_2, i_3)^{th}$ entry of T can readily be calculated by

$$[T]_{i_1,i_2,i_3} = \int_{\Gamma^n} \psi_{\mathbf{l},\mathbf{i}}(y)\psi_{\mathbf{l}',\mathbf{i}'}(y)\psi_{\mathbf{l}'',\mathbf{i}''}(y)\ dy = \prod_{t=1}^{n} \int_{\Gamma_t} \psi_{l_t,i_t}(y)\psi_{l'_t,i'_t}(y)\psi_{l''_t,i''_t}(y)\ dy.$$

A similar approximation can be made for S^r_ρ through the use of the unweighted trilinear form

$$[T^r]_{i_1,i_2,i_3} = \sum_{|\gamma|_\infty \le r} \int_{\Gamma^n} D_y^\gamma \psi_{\mathbf{l},\mathbf{i}}(y) D_y^\gamma \psi_{\mathbf{l}',\mathbf{i}'}(y) \psi_{\mathbf{l}'',\mathbf{i}''}(y)\, dy.$$

To approximate the weighted trilinear form T_ρ, we make use of the unweighted quadrilinear form $[R^\gamma]_{i_1,i_2,i_3,i_4} = \int_{\Gamma^n} \psi_{i_1}(y)\psi_{i_2}(y)\psi_{i_3}(y)\psi_{i_4}(y)\, dy$ and the sparse grid hierarchical surplus $\rho^N_{n,z}$ of ρ_n.

The variational form of the finite noise equality constraint $e_n(q,u) = 0$ is also written in terms of a stochastic integral and it is therefore natural to consider the possibility of using these bilinear and trilinear forms to represent its Galerkin approximation (4.1). The resulting $MN \times MN$-dimensional linear system is solved directly, without resorting to a quadrature rule to approximate the stochastic integrals. The matrices $S_{\rho^N_n}$ and $S^r_{\rho^N_n}$ and the array $T_{\rho^N_n}$ are generally not sparse, as shown in Figure 5.2.1. The linear system, while being a more accurate approximation of (1.12) than (4.3), would nevertheless be more difficult to solve, possibly requiring the use of appropriate preconditioners.

In our numerical examples, we therefore solve both the forward and adjoint systems using the fully decoupled approximation, given in (4.3). Note that function $e_n(q,u) \in \widetilde{H}^1_0(D)$ appearing in the penalty term of L_c, is the unique solution of the variational problem

$$\int_{\Gamma^n}\int_D \nabla e_n \cdot \nabla v \rho_n(y)\, dx\, dy = \int_{\Gamma^n}\int_D q\nabla u \cdot \nabla v \rho_n(y)\, dx\, dy - \int_{\Gamma^n}\int_D f(x) v \rho_n(y)\, dx\, dy \quad (5.3)$$

for all $v \in \widetilde{H}^1_0(D)$. By arguments analogous to those used to derive (4.3), we approximate this problem by the decoupled system

$$\int_D \nabla e_n(y_i) \cdot \nabla v\, dx = -\int_D q(x,y_i)\nabla u(x,y_i) \cdot \nabla v(x)\, dx - \int_D f(x)v(x)\, dx \quad \forall v \in H^1_0,$$

for $i = 1, ..., N$.

We are now in a position to derive the form of the cost functional when the functions $q, u, \hat{u}^n$ and λ are approximated by $q^{M,N}, u^{M,N}, \hat{u}^{M,N}$ and $\lambda^{M,N}$, respectively. Indeed, the first term of $J(q,u)$ becomes

$$\begin{aligned}
&\|u^{M,N} - \hat{u}^{M,N}\|^2_{\widetilde{H}^1_0} = \int_{\Gamma^n}\int_D |\nabla u^{M,N}(x,y) - \nabla \hat{u}^{M,N}(x,y)|^2\, \rho_n(y)\, dx \\
&= \sum_{j,j'=1}^{M}\sum_{i,i'=1}^{N} d_+(x_j,y_i) d_+(x_{j'},y_{i'}) \left(\int_D \nabla\phi_i(x)\cdot\nabla\phi_{i'}(x)\, dx\right)\left(\int_{\Gamma^n} \psi_i(y)\psi_{i'}(y)\rho_n(y)\, dy\right) \\
&= \sum_{i,i'=1}^{N} \mathbf{d}_+(y_i)^T A_x \mathbf{d}_+(y_{i'}) \left(\int_{\Gamma^n} \psi_i(y)\psi_{i'}(y)\rho_{n_s}(y)\, dy\right) = \mathbf{d}_+^T \left(S_{\rho^N_n} \otimes A_x\right) \mathbf{d}_+,
\end{aligned}$$

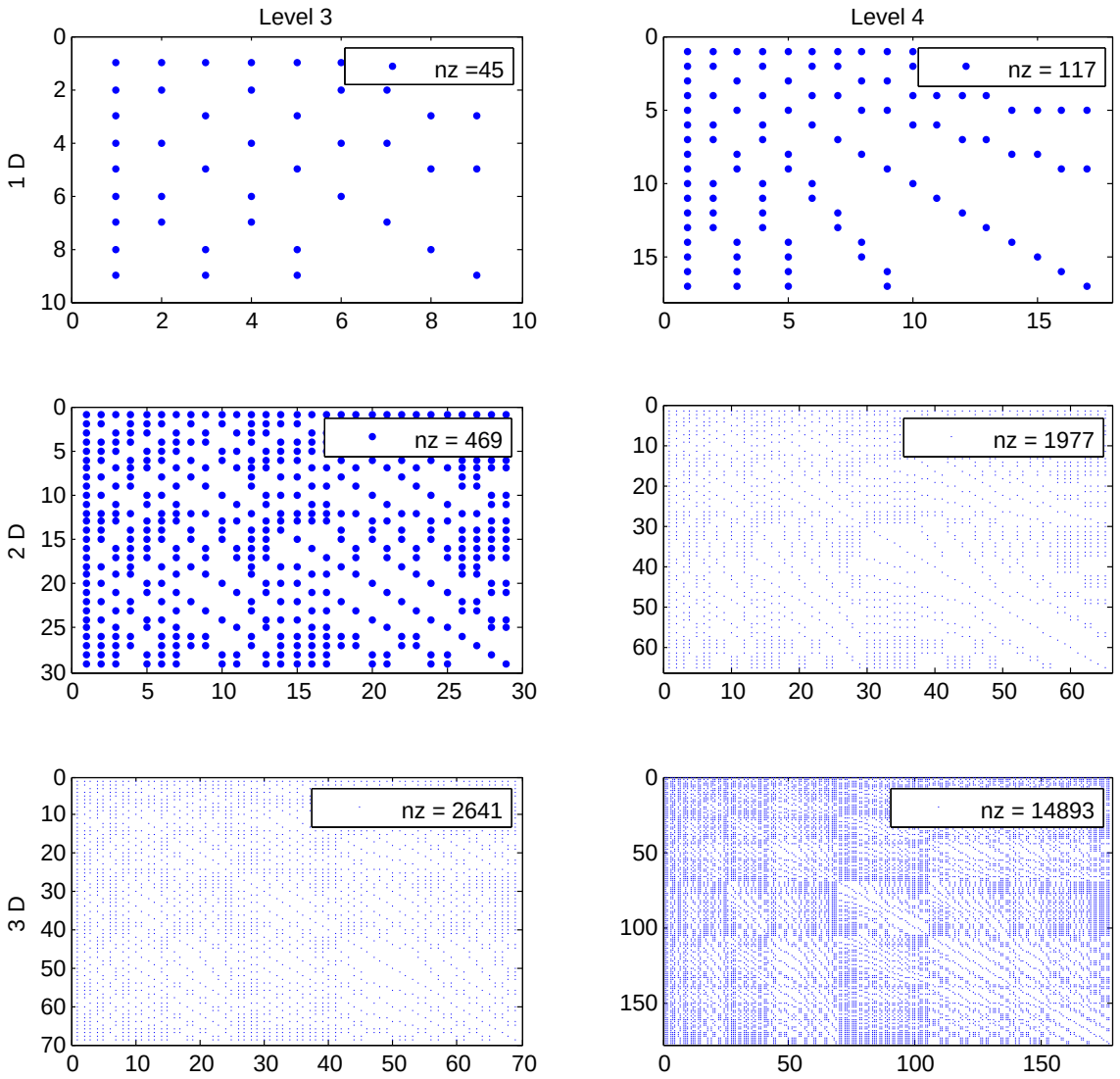

Figure 5.1: Sparsity pattern of unweighted bilinear forms S for varying dimensions and levels of interpolation. The quantity nz represents the number of non-zero entries of S.

where $d_+(x_j, y_i) = u_+(x_j, y_i) - \hat{u}_+(x_j, y_i)$, $\mathbf{d}_+(y_i) = [d_+(x_1, y_i), ..., d_+(x_M, y_i)]^T \in \mathbb{R}^M$, $\mathbf{d}_+ = [\mathbf{d}_+(y_1), ..., \mathbf{d}_+(y_N)]^T \in \mathbb{R}^{MN}$, and $\otimes$ denotes the Kronecker product. Similarly

$$\|q^{M,N}\|^2_{\widetilde{H}_{\text{mix}}} = \mathbf{q}_+^T \left(S^r_{\rho_n^N} \otimes Ax \right) \mathbf{q}_+.$$

The term $\langle q \nabla u, \nabla \lambda \rangle$ on the other hand requires the use of the weighted trilinear form $T_\rho \approx T_{\rho_n^N}$. Indeed

$$\begin{aligned}
&\langle q^{M,N} \nabla u^{M,N}, \nabla \lambda^{M,N} \rangle = \int_{\Gamma^n} \int_D q^{M,N}(x,y) \nabla u^{M,N}(x,y) \cdot \nabla \lambda^{M,N}(x,y) \; \rho_n(y) \; dx \; dy \\
&= \sum_{j=1}^{M} \sum_{i=1}^{N} q_+(x_j, y_i) \int_{\Gamma^n} \int_D \phi_j(x) \psi_i(y) \nabla u^{M,N}(x,y) \cdot \nabla \lambda^{M,N}(x,y) \; \rho_n(y) \; dx \; dy \\
&= \sum_{j',j''=1}^{M} \sum_{i',i''=1}^{N} u_+(x_{j'}, y_{i'})^T S_{\rho_n^N, q} \; \lambda_+(x_{j''}, y_{i''}) = \boldsymbol{u}_+^T \left(T_{\rho_n^N, q^{M,N}} \right) \boldsymbol{\lambda}_+
\end{aligned}$$

where

$$T_{\rho_n^N, q^{M,N}} = \left[\sum_{j=1}^{M} \sum_{i=1}^{N} q_+(x_j, y_i) \left(\int_{\Gamma^n} \psi_i \psi_{i'} \psi_{i''} \; \rho_n \; dy \right) \left(\int_D \phi_j \nabla \phi_{j'} \cdot \nabla \phi_{j''} \; dx \right) \right]_{\substack{i',i''=1,...,N \\ j',j''=1,...,M}}.$$

Combining these terms yields an expression for the discretized cost functionals

$$\begin{aligned}
J(q^{M,N}, u^{M,N}) &= \frac{1}{2} \mathbf{d}_+^T \left(S_{\rho_n^N} \otimes A_x \right) \mathbf{d}_+ + \frac{\beta}{2} \mathbf{q}_+^T \left(S^r_{\rho_n^N} \otimes Ax \right) \mathbf{q}_+ \\
L_c(q^{M,N}, u^{M,N}) &= J(q^{M,N}, u^{M,N}) + \mathbf{u}_+^T \left(T_{\rho_n^N, q^{M,N}} \right) \boldsymbol{\lambda}_+ + \frac{c}{2} \mathbf{e}_{n,+}^T \left(S_{\rho_n^N} \otimes A_x \right) \mathbf{e}_{n,+}.
\end{aligned}$$

The gradient vector can be obtained in a similar fashion. Recall that for the cost functional

$$\hat{J}(q) := J(q, u(q)) = \frac{1}{2} \|u - \hat{u}\|^2_{\widetilde{H}_0^1} + \frac{\beta}{2} \|q\|^2_{\widetilde{H}_{\text{mix}}},$$

we have

$$D_q \left[\hat{J}(q) \right] h = \langle h \nabla u(q), \nabla p \rangle_{\widetilde{H}_0^1} + \beta \langle q, h \rangle_{\widetilde{H}_{\text{mix}}},$$

where $u(q)$ is the model output and p is the solution of the adjoint problem (3.13), both of which may be obtained by solving the decoupled, parameterized versions of (1.12) and (3.13), respectively. The gradient corresponding to the augmented Lagrangian $L_c(u, q; \lambda)$ is given by

$$D_{(q,u)} \left[L_c(q, u; \lambda) \right] (h, v) = \langle u - \hat{u}^n, v \rangle_{\widetilde{H}_0^1} + \beta \langle q, h \rangle_{\widetilde{H}_{\text{mix}}} + \langle h \nabla v, \nabla \lambda \rangle + c \langle \nabla e_n(q, u), h \nabla v \rangle,$$

where $e_n(q, u)$ is the equality constraint satisfying (5.3).

5.3 Numerical Examples

5.3.1 Example 1: Spatially Constant, Random Parameters

The first example illustrates the variational method for a problem in which the diffusion coefficient is constant in space. In particular, we study the problem of estimating the random variable $q(\omega) = 1 + Y(\omega)$, from measurements $\hat{u}$ of the model output

$$u(x, Y(\omega)) = x(1 - x) + Y(\omega)\sin(\pi x),$$

when the forcing term f is given by

$$f(x, Y(\omega)) := -\nabla \cdot [q(Y(\omega)\nabla u(x, Y(\omega))] = 2 + 2Y(\omega) + \pi^2 Y(\omega)\sin(\pi x) + \pi^2 Y(\omega)^2 \sin(\pi x).$$

The KL expansion of the of measured output $\hat{u}$ has a single mode whose random variable $\tilde{Z}$ has empirical density function given in Figure 5.2.

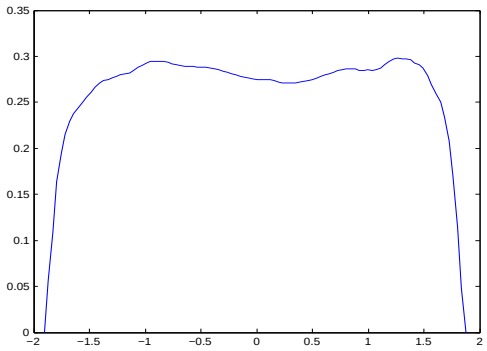

Figure 5.2: Density function of random variable associated with the measured output $\hat{u}$

The estimated parameter q^* written as a function of this random variable closely resembles the original parameter q, as shown in Figure 5.3

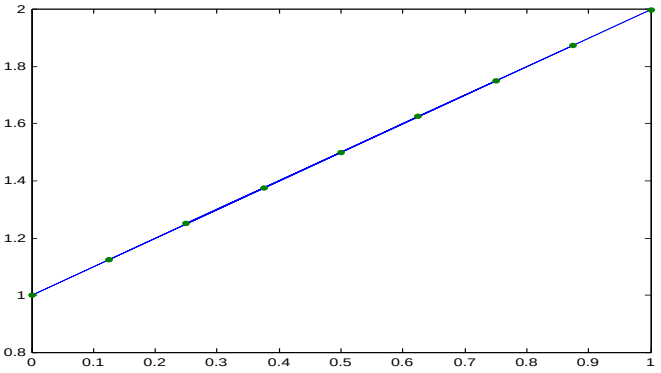

Figure 5.3: The original parameter q as a function of random variable Z (solid line), as well as its estimated value q^* (dotted line).

5.3.2 Example 2: Distributed Random Parameters

For the second example, we employ the augmented Lagrangian method to estimate the spatially varying random diffusion coefficient

$$q(x,\omega) = 3 + x^2 + Y_1(\omega)\sin(\pi x) + Y_2(\omega)\sin(2\pi x),$$

where the random vector $(Y_1, Y_2) \sim \text{UNIF}([0,1]^2)$ is uniformly distributed over the cube $[0,1]^2$ and the deterministic forcing term f is given by

$$f(x) = -6 + 18x - 8x^2 + 12x^3.$$

The model data $\hat{u}$ was computed from the data generated by q and f and expanded in the same random basis (Y_1, Y_2). We used continuous quadratic basis functions on 16 elements, i.e. with 33 degrees of freedom, to effect spatial discretization of the functions $q, u, \hat{u}$ and f. The sparse grid hierarchical finite element basis introduced in Section 4.2 with interpolation level $L = 3$ is used to discretize the stochastic variable $y = (y_1, y_2)$. The resulting discretized optimization problem therefore had 957 degrees of freedom in total. Figures 5.4 and 5.5 illustrate the strength of the approximation of the exact diffusion coefficient q_{exact} by the estimated function q_{approx}. These figures also compare both the sample paths associated with the appropriate sparse grid collocation points and the first nine moments of the respective random fields. The first nine moments of the estimate and the exact solution are comparable.

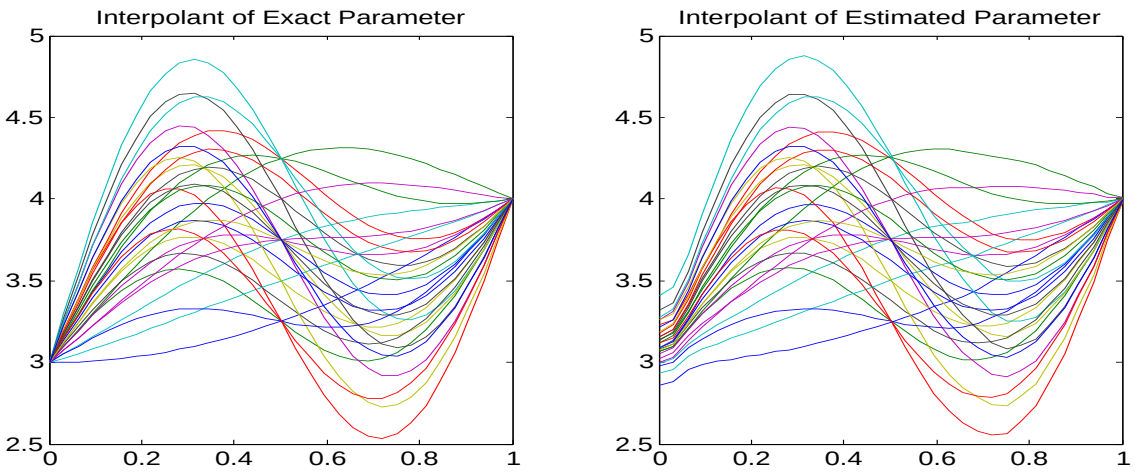

Figure 5.4: Interpolants of q_{exact} and q_{approx} at the sparse grid collocation points.

5.3.3 Example 3: The Inverse Sampling Method

The third example examines the use of the inverse sampling method introduced in Section 4.5 to statistically estimate the spatially varying random parameter q, when the underlying

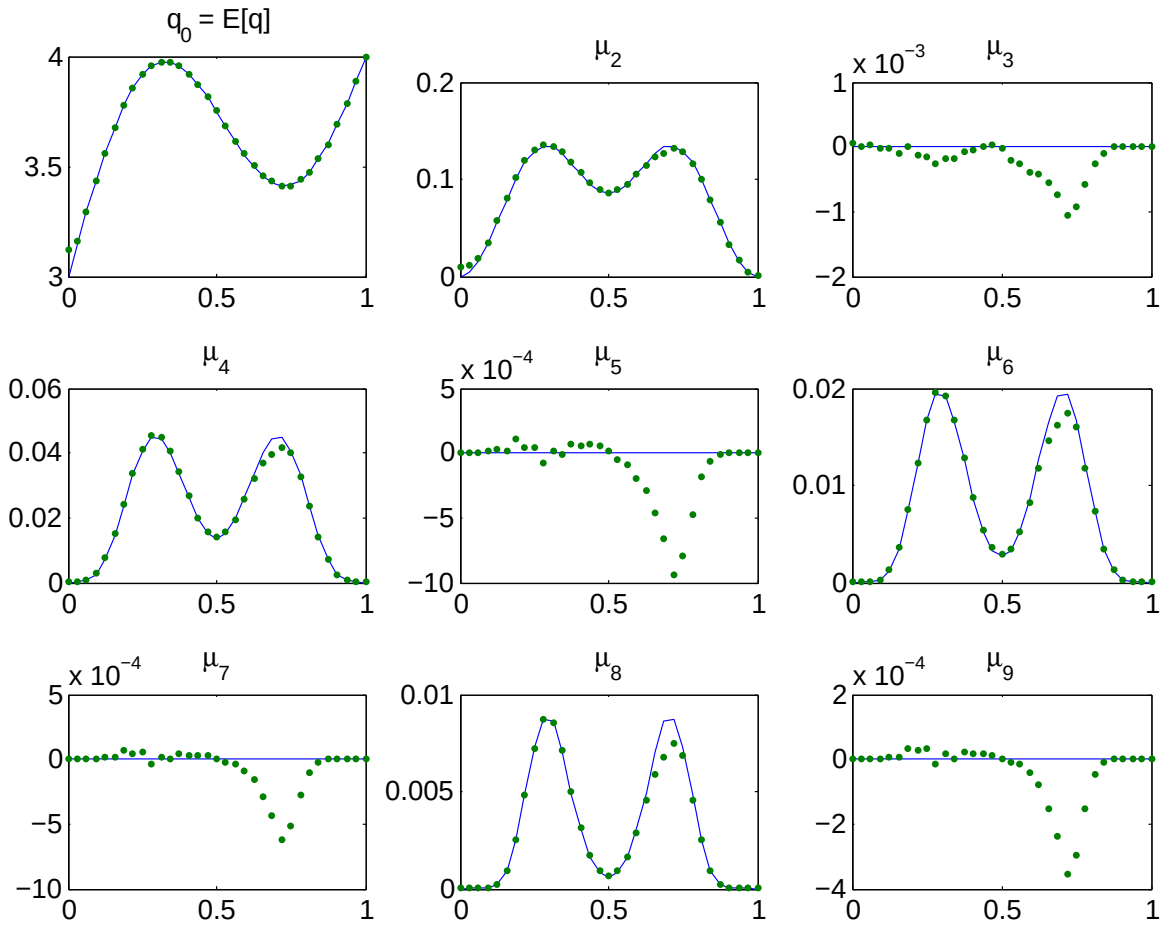

Figure 5.5: First nine central moments $\mu_k = E\left[(q - q_0)^k\right]$ of the exact (solid line) and the estimated (dotted line) parameters.

distribution is non-standard. In particular, we wish to estimate the diffusion parameter $q \in H^1([0,1]) \otimes \widetilde{H}_{\text{mix}}(D)$ where

$$
\begin{aligned}
q(x, Y_q(\omega)) &= 1 + x^3 + \sin(2\pi x) Y_q(\omega), \quad y_q \sim \rho_q && \text{(true parameter)} \\
u(x, Y_u(\omega)) &= x(1 - x^2) Y_u(\omega), \quad Y_u(\omega) \sim \rho_u && \text{(true model output)} \\
f(x, Y_q(\omega), Y_u(\omega)) &= -\nabla \left(q(x, Y_q(\omega)) \nabla u(x, Y_u(\omega)) \right) && \text{(fixed forcing term)}
\end{aligned}
$$

The random variables Y_q and Y_u are independently distributed with density functions

$$
\rho_q(y_q) = \begin{cases} \cos(4\pi(y_q + \frac{3}{4})) + 1, & y_q \in [-1, -\frac{1}{2}] \\ \cos(4\pi(y_q - \frac{3}{4})) + 1, & y_q \in [\frac{1}{2}, 1] \\ 0, & \text{otherwise} \end{cases}
$$

$$
\rho_u(y_u) = \begin{cases} \frac{1}{2}, & y_u \in [-1, 1] \\ 0, & \text{otherwise} \end{cases} .
$$

Figures 5.6 and 5.7 depict 200 sample paths for each of the three functions and the density function of $Y_q(\omega)$ respectively. The bimodal behavior of the density function related to

the parameter q is evident in the concentration of sample paths. The density function,

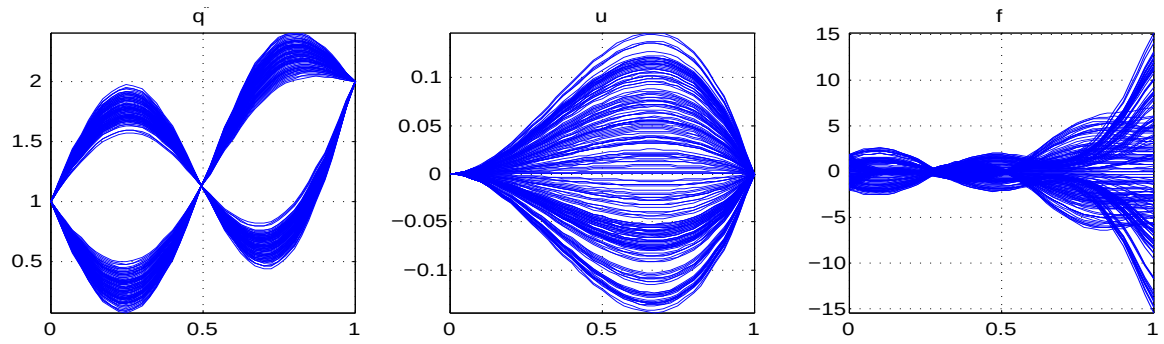

Figure 5.6: Sample paths of true parameter q, model output u and forcing term f.

cumulative and inverse cumulative distribution for Y_q are given in Figure 5.7.

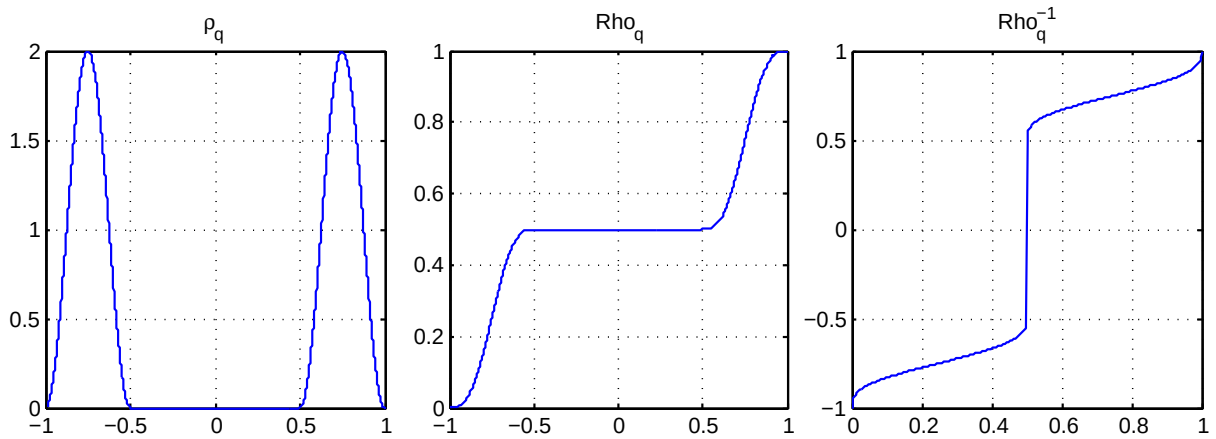

Figure 5.7: Density function, distribution function, and inverse distribution function of Y_q.

Statistical Pre-processing

In contrast to the earlier sections, the statistical variation of the model output u in this example results not only from uncertainty in the diffusion parameter q, but also from that of the forcing term f. In order to obtain a statistical estimate of q in terms of $\hat{u}^n$, it is therefore necessary to also control for the influence of f. To this end, we generate an independent sample of trajectories of both the forcing term and model output and corrupt the sample with white noise to simulate measurement noise. Upon analyzing the data sets of $\hat{u}$ and f,

we obtain their truncated KL expansions

$$\hat{u}^n(x,\omega) = \hat{u}_0(x) + \phi_1^u(x)Y_1^u(\omega) \quad \text{and}$$
$$\hat{f} = \hat{f}_0(x) + \phi_1^f(x)Y_1^f(\omega) + \phi_2^f(x)Y_2^f(\omega).$$

The graphs of the relevant functions in the expansion are given in the figures below.

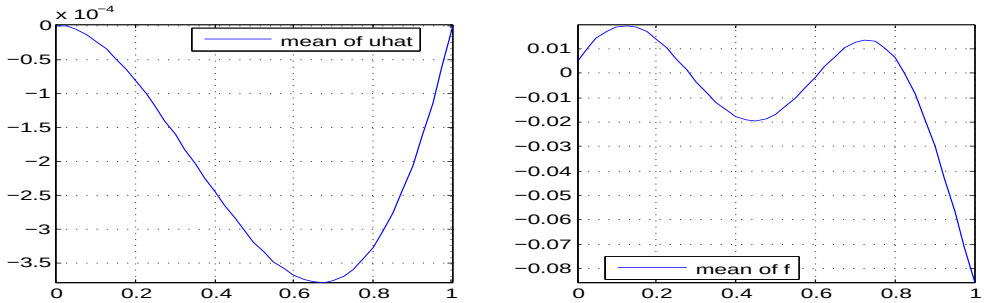

Figure 5.8: Estimated mean of $\hat{u}$ and f.

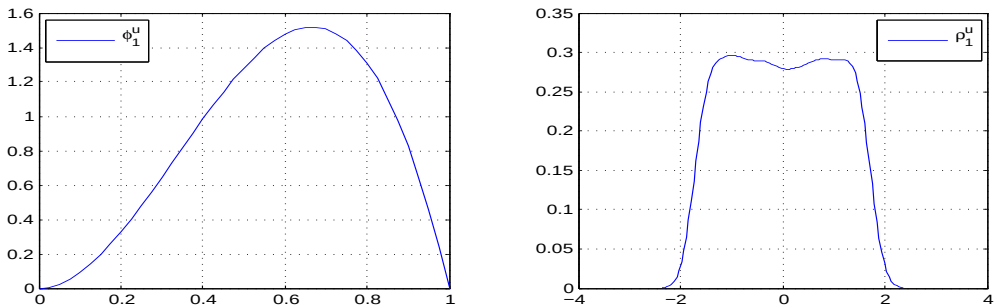

Figure 5.9: Significant mode and associated density function in KL expansion of $\hat{u}$.

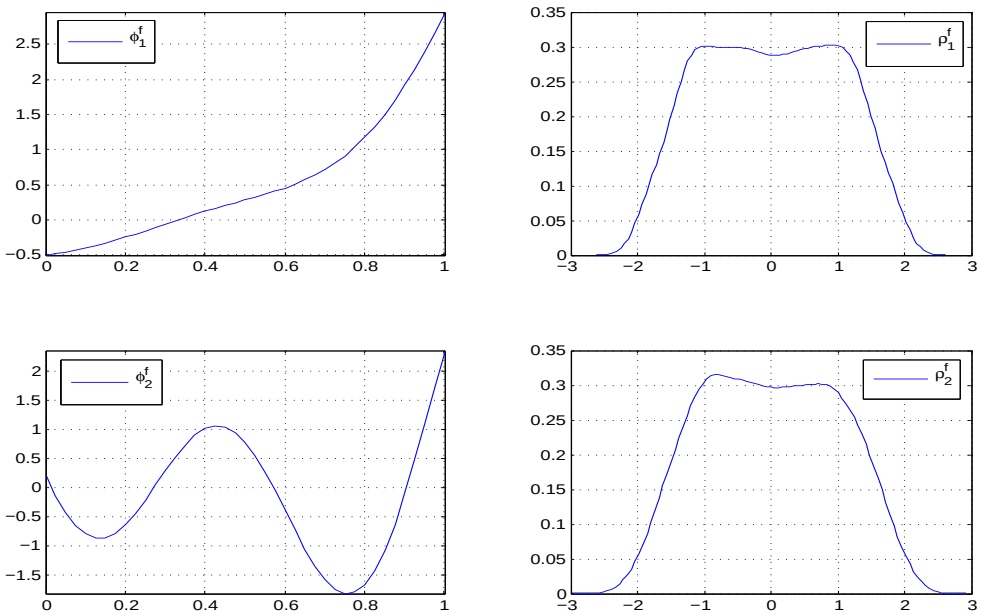

Figure 5.10: Modes and associated density functions in the KL expansion of f.

The parameter q is therefore estimated as a function of the spatial variable x as well as all the random variables Y_1^u, Y_1^f, Y_2^f given above.

We can now proceed to estimate the joint density function of the random vector $Y = (Y_1^u, Y_1^f, Y_2^f)$. The three-dimensional scatter plot, shown in Figure 5.11, reveals that the data set lies almost entirely on a two-dimensional subspace of $\mathbb{R}^3$ thus reducing the effective dimension of the sample. This subspace may be identified and the data set projected onto it by means of principle component analysis. This entails computing the singular value decomposition of the sample correlation matrix $\hat{\Sigma}$, i.e. $\hat{\Sigma} = V\Lambda V^T$ and projecting the data onto the left singular vectors (columns of V) corresponding to significant singular values.

Here

$$\hat{\Sigma} = \begin{bmatrix} 1 & 0.9899 & -0.1419 \\ 0.9899 & 1 & 0 \\ 0.1419 & 0 & 1 \end{bmatrix},$$

which reveals a strong correlation (and hence redundancy) between Y_1^u and Y_1^f. Furthermore

$$V = \begin{bmatrix} -0.7071 & 0.0000 & -0.7071 \\ -0.6999 & 0.1419 & 0.6999 \\ 0.1004 & 0.9899 & -0.1004 \end{bmatrix}, \quad \text{and} \quad \Lambda = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

indicating that only the subspace consisting of the first two left singular vectors is significant. We thus transform the data set corresponding to the random vector $Y = [Y_1^u, Y_1^f, Y_2^f]$ to that corresponding to the random vector $Z = [Z_1, Z_2, Z_3]$ by left multiplication of V, i.e. $Z = YV$. Plotting the transformed data (Figure 5.12) reveals that the third coordinate is

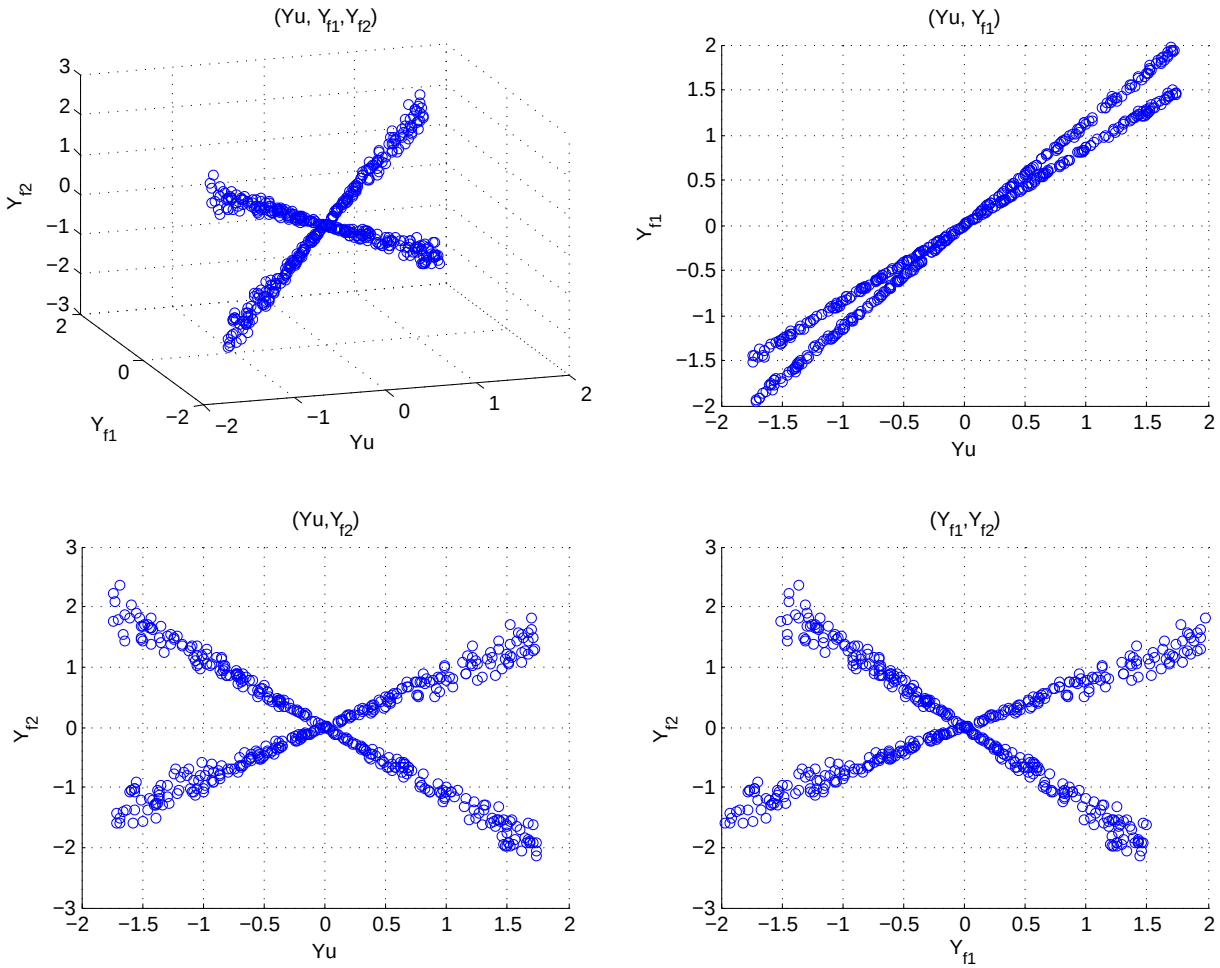

Figure 5.11: Scatter plots showing dependencies among random variables.

effectively zero and may therefore be left out without loss of too much accuracy, i.e.

$$[Z_1, Z_2] = [Y_1, Y_2, Y_3] \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \\ v_{31} & v_{32} \end{bmatrix}.$$

The Karhunen-Loève expansion of f and $\hat{u}$ under this change of variables is given by

$$\hat{u}(x, z_1, z_2) = \hat{u}_0(x) + \left([z_1, z_2] \begin{bmatrix} v_{11} \\ v_{12} \end{bmatrix} \right) \sqrt{\lambda_1^u} \phi_1^u(x),$$

$$\hat{f}(x, z_1, z_2) = \hat{f}_0(x) + \left([z_1, z_2] \begin{bmatrix} v_{21} & v_{31} \\ v_{22} & v_{32} \end{bmatrix} \right) \begin{bmatrix} \sqrt{\lambda_1^f} & 0 \\ 0 & \sqrt{\lambda_2^f} \end{bmatrix} \begin{bmatrix} \phi_1^f(x) \\ \phi_2^f(x) \end{bmatrix}.$$

Note that the transformation V is an orthogonal projection and hence has a Jacobian equal

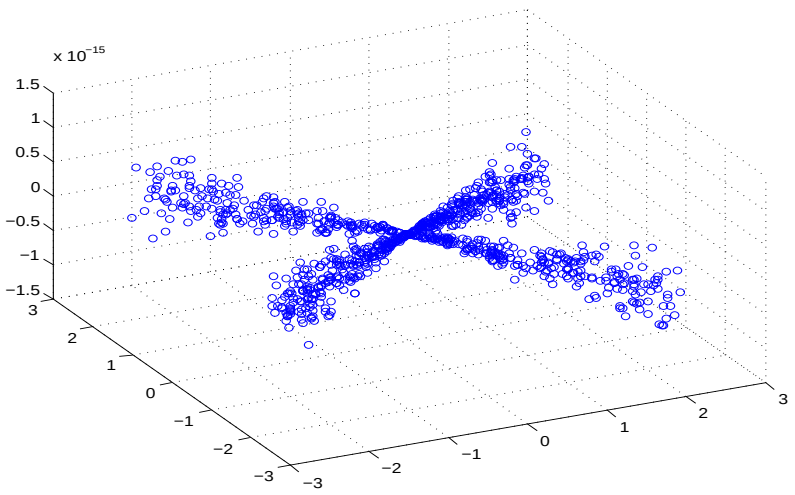

Figure 5.12: Scatter plot of the data set corresponding to random vector Z.

to one.

Once the data is projected onto the (z_1, z_2)-plane, we can estimate the joint density function ρ_{z_1,z_2}. A straightforward approximation can be achieved by first constructing a histogram to estimate the marginal density ρ_{z_1}, then dividing the cross-sectional data z_2 according to bins in the first component and hence estimating the conditional density function $\rho_{z_2|z_1}$ over each cross-sectional bin, using one-dimensional kernel density estimates. We may also use multivariate kernel density estimates directly. Figures 5.13 and 5.14 show the densities obtained by these two methods. Figure 5.15 illustrates the accuracy of the density estimate by comparing the first three moments computed with the help of the density ρ_{z_1,z_2} with those obtained through Monte Carlo sampling.

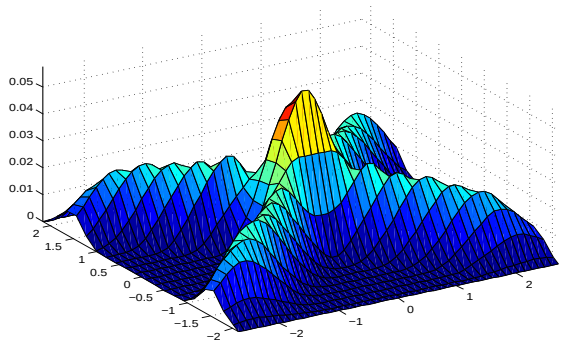

Figure 5.13: Joint density function estimated using a histogram.

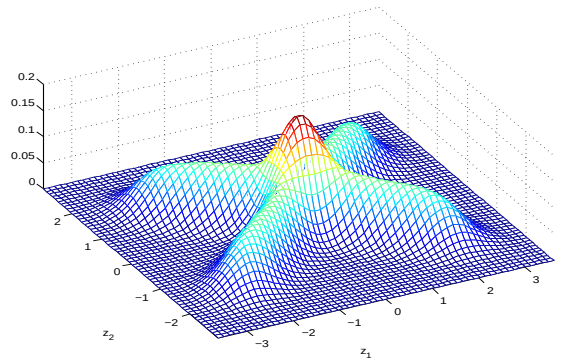

Figure 5.14: Kernel density estimate of joint density function.

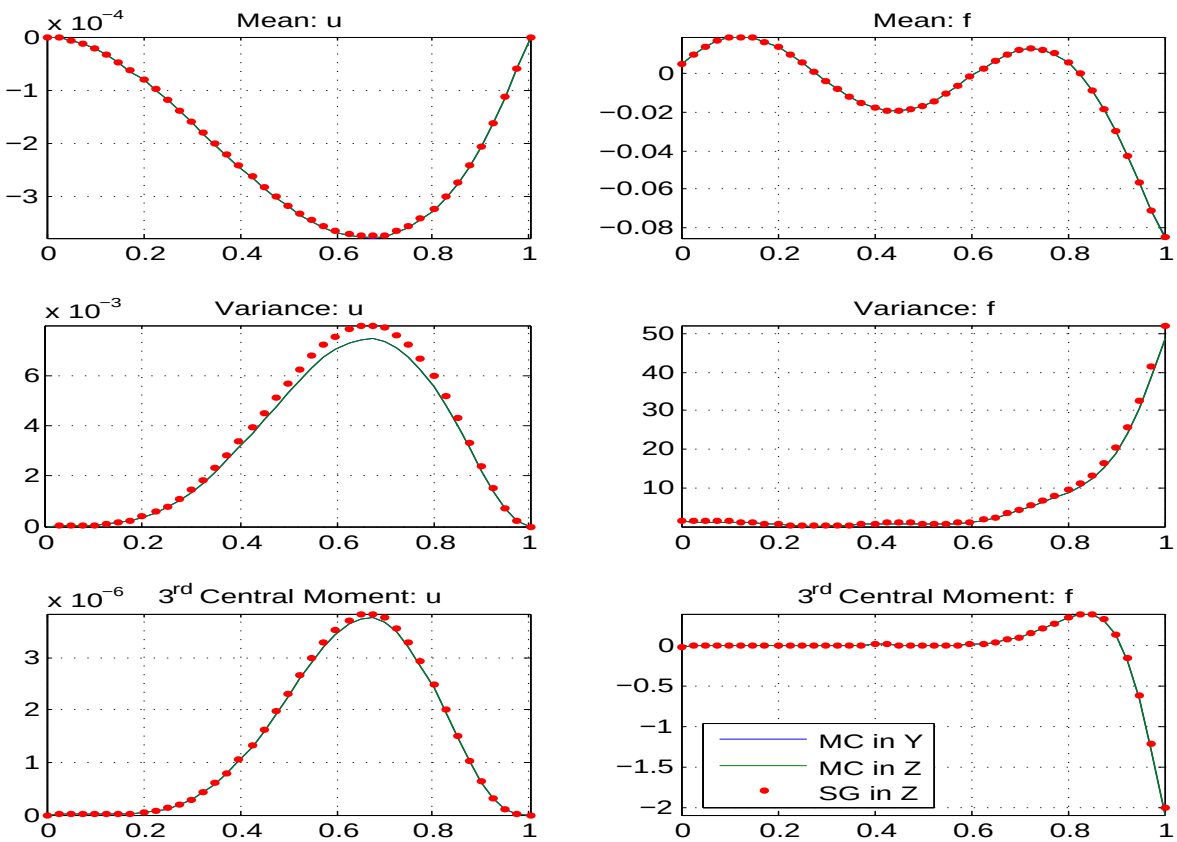

Figure 5.15: Comparison of the moments of the Karhunen-Lòeve Expansions of $\hat{u}^n$ and f, using the estimated density function, with those obtained by Monte Carlo methods.

Inverse Sampling

To estimate the random field q through inverse sampling, we generate a small collection of pairs of sample paths $\{(\hat{u}(\cdot, Z(\omega_i)), f(\cdot, Z(\omega_i))\}_{i=1}^{N_s}$ and invert each such pair deterministically to obtain a sample $\{q^*(\cdot, Z(\omega_i))\}_{i=1}^{N_s}$ of paths of the unknown optimal field $q^*(x, Z)$. Here we sample directly from the Karhunen-Loève expansion of u and f to obtain a sample of 100 pairs $\{q^*(\cdot, Z(\omega_i))\}_{i=1}^{N_s}$. Since the output pairs were directly obtained from the sampled values of q, all optimization runs were completed within a short period of time and we obtain an accurate representation of the unknown q. Figure 5.16 below gives the sample paths of the estimated function q^*, which compares favorably with that given in Figure 5.6. The first

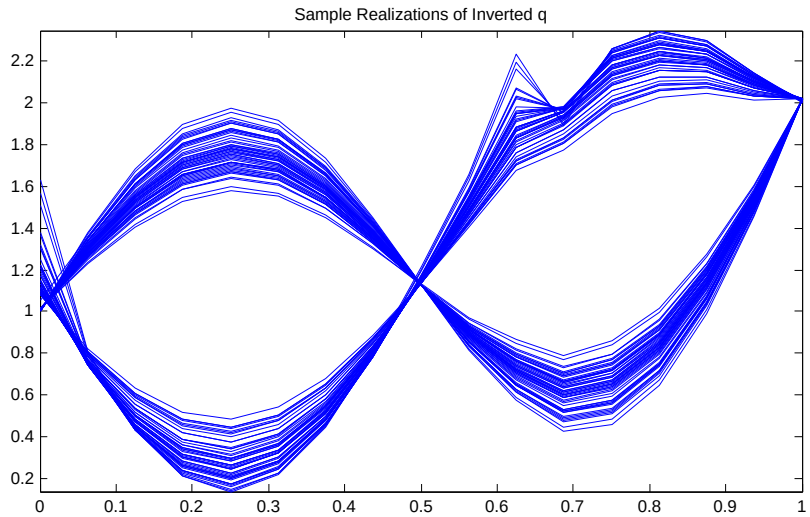

Figure 5.16: Sample paths of the parameter estimate q^*.

four moments of the estimated parameter q^* also match the corresponding moments of the true parameter q, as shown in Figure 5.17.

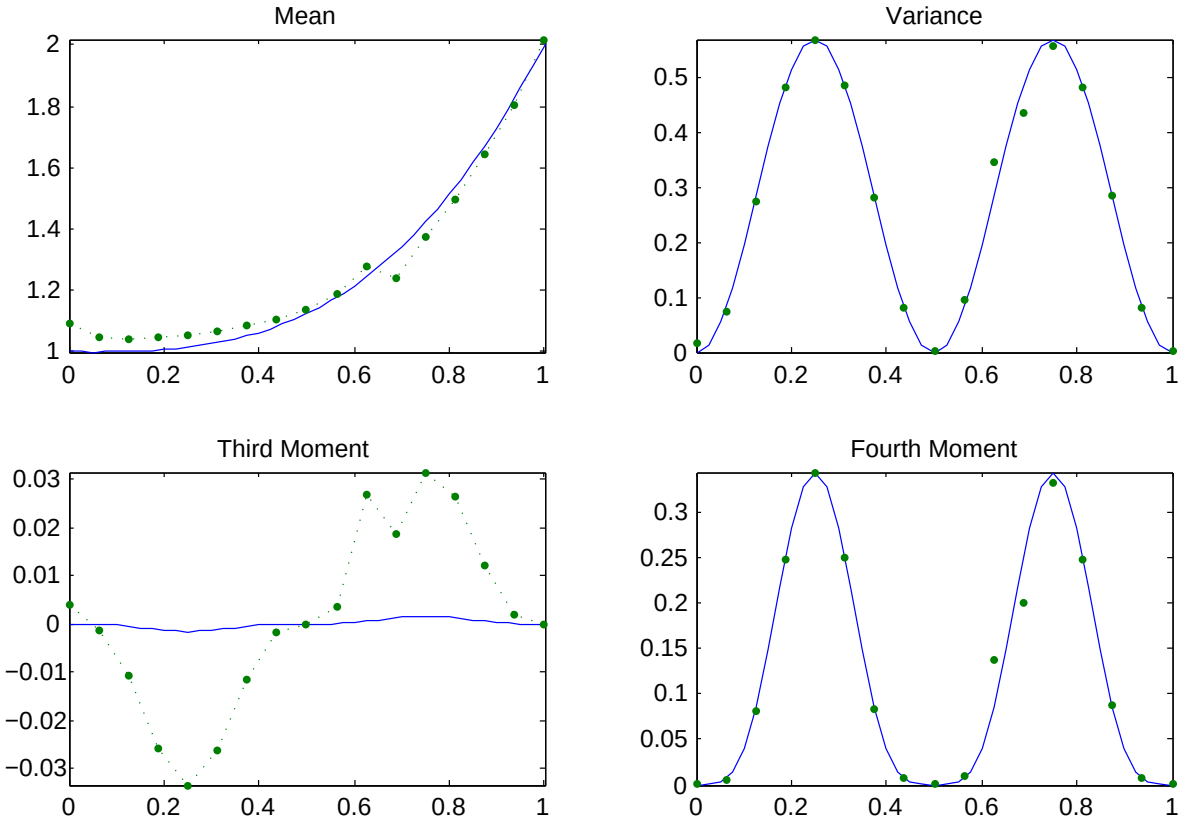

Figure 5.17: Comparison of the first four moments of the original parameter (solid line) and its estimate q^* (dotted line).

Finally, we compare the first mode in the respective KL expansions of the original q and its estimate q^*. Unfortunately, the KL expansion of the estimated parameter q^* has four modes, whereas that of the original parameter q only has one. Incredibly, not only do the first deterministic mode, but also the density function of the first random variable in the KL expansion correspond, see Figures 5.18 and 5.19

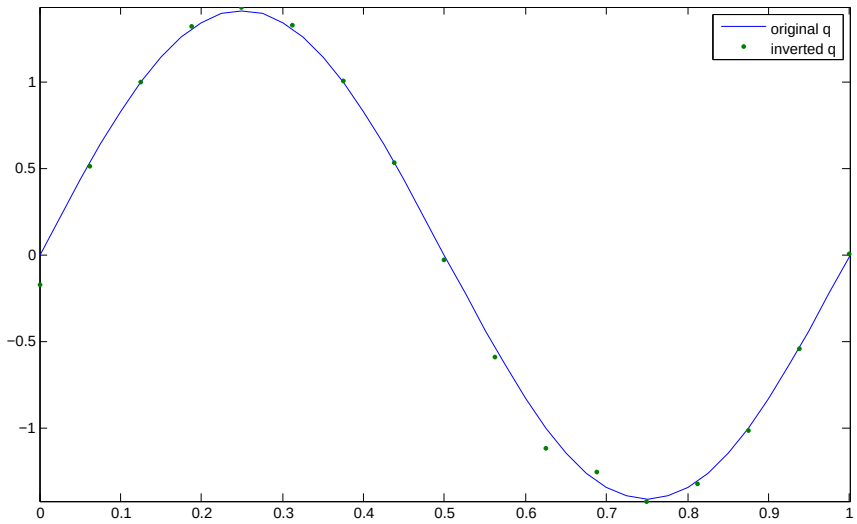

Figure 5.18: First deterministic mode in the KL expansion of q.

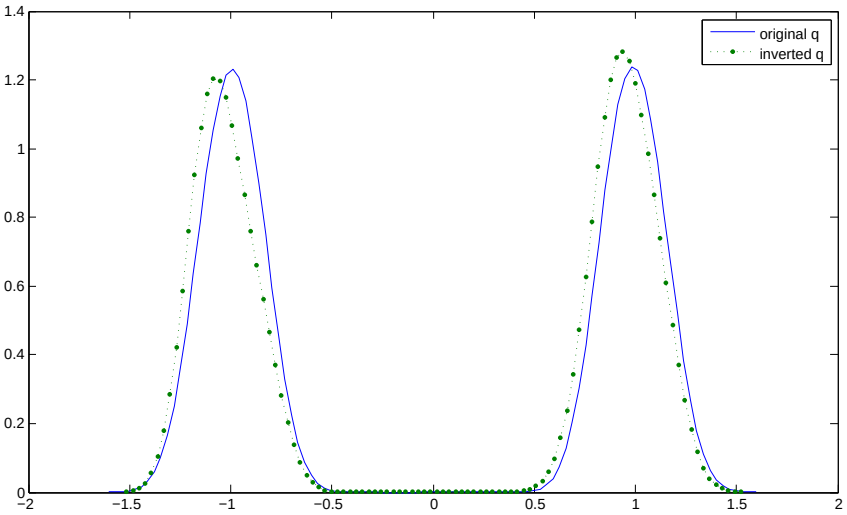

Figure 5.19: Actual and estimated density functions of the first random variable in the KL expansion of q.

An outstanding issue in the context of the inverse sampling method relates to its robustness in light of possible statistical mischaracterizations of the measured data. In practice, the random behavior of the forcing term f is often independent of q and the statistical model for f may be derived independently from that of the model output $\hat{u}^n$. If, during the pre-processing stage there is an error in the statistical characterization of the measured data $\hat{u}^n$, or forcing term f, the sampling process may yield pairs $(\hat{u}(\cdot, Z(\omega_i)), f(\cdot, Z(\omega_i))$ for which

the corresponding diffusion coefficient $q^*(x, Z(\omega))$ does not exist or is difficult to estimate numerically. For the last example, we therefore use the transformed bivariate data set Z to construct a t-copula from which it is easy to sample, but which may not be an accurate density estimate. The t-copula is obtained by first transforming the marginal random variables Z_1 and Z_2 to uniform random variables and estimating their dependence structure via linear or rank based correlation, which is enough to specify the t-copula. Figure 5.20 highlights the distortion of the density, brought about by the copula estimate. In Figure 5.20, we compare the t-copula density estimate with the more accurate kernel density estimate. This loss of

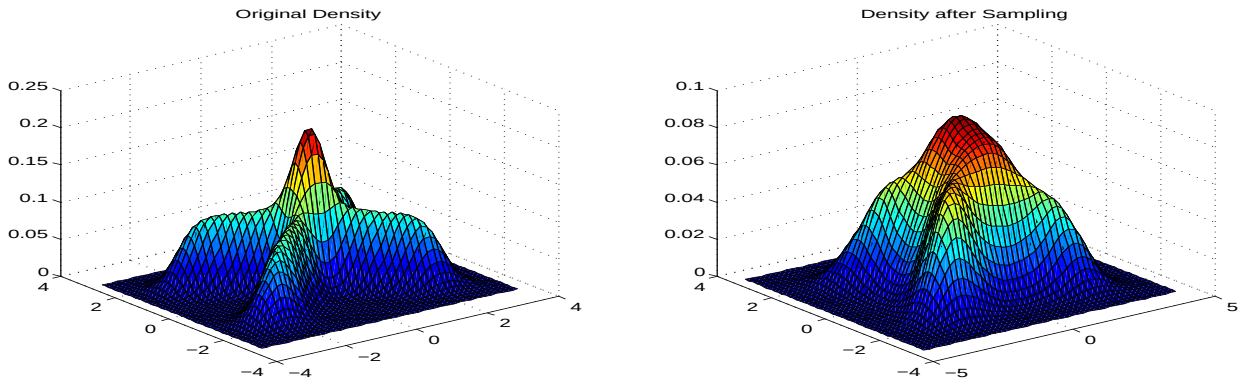

Figure 5.20: Comparison of the original kernel density estimate and t-copula estimate of the joint density function of Z.

accuracy is further illustrated by the scatter diagrams in Figures 5.21 and 5.22 of a sample generated from the t-copula and the original sample, respectively. Evidently, the distinct cross-like shape is blurred in the sample generated from the t-copula. This is due to the fact that traditional linear measures of dependence used to determine the copula (such as correlation) are unable to capture the non-standard dependence relation between Z_1 and Z_2. Despite these differences in the exact shape of these two densities however, there seems to be a relatively good correspondence in the first two moments of the Karhunen-Loève expansions of $\hat{u}^n$ and $\hat{f}$, constructed from the original sample Z and from the copula sample, respectively.

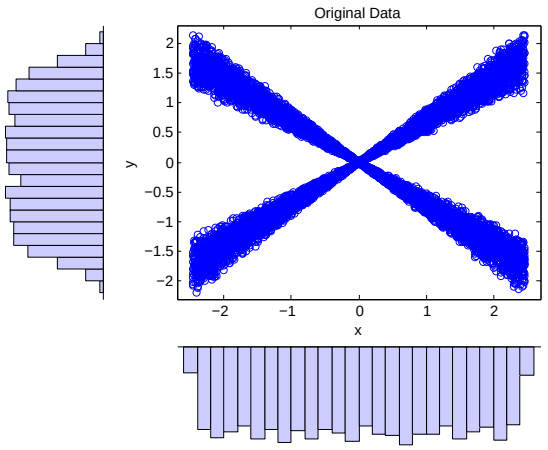

Figure 5.21: Scatter diagram of the original data.

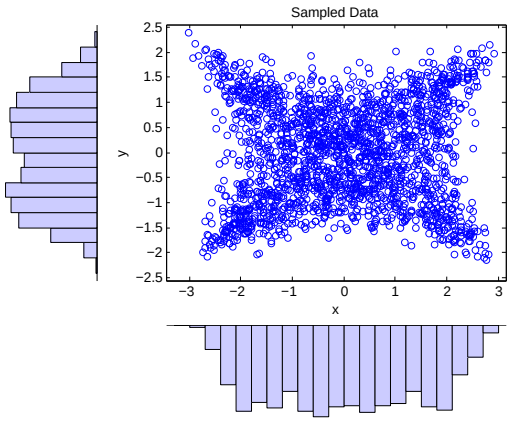

Figure 5.22: Scatter diagram of the sample generated by the copula.

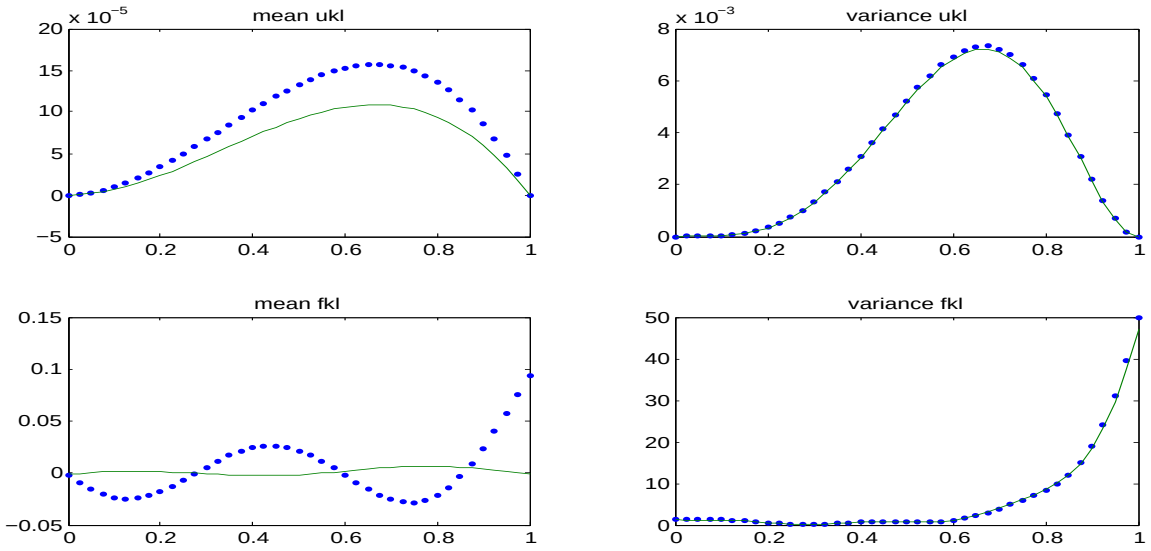

Figure 5.23: Comparison of the mean and variance of the model output u and the forcing term f, using different random samples. The dotted line represents the moments obtained from the copula sample.

Since the support of the t-copula includes points that are not in the support of the original density function, it is possible that the sample pairs $(\hat{u}(\cdot, Z(\omega_i)), f(\cdot, Z(\omega_i))$ generated by the copula are no longer related to each other via the given model, leading to a lack of identifiability of the parameter $q^*(\cdot, Z(\omega_i))$ for certain realizations. Indeed, upon using a random sample of size 100 generated by the t-copula, we observed that the optimization algorithm fails to identify q^* correctly for some pairs $(\hat{u}(\cdot, Z(\omega_i)), f(\cdot, Z(\omega_i))$, and at times

does not even converge. The 'failure rate' of the optimization algorithm can be visualized by recording the approximate relative H_0^1-error in the model output in the Box Plot in Figure 5.24.

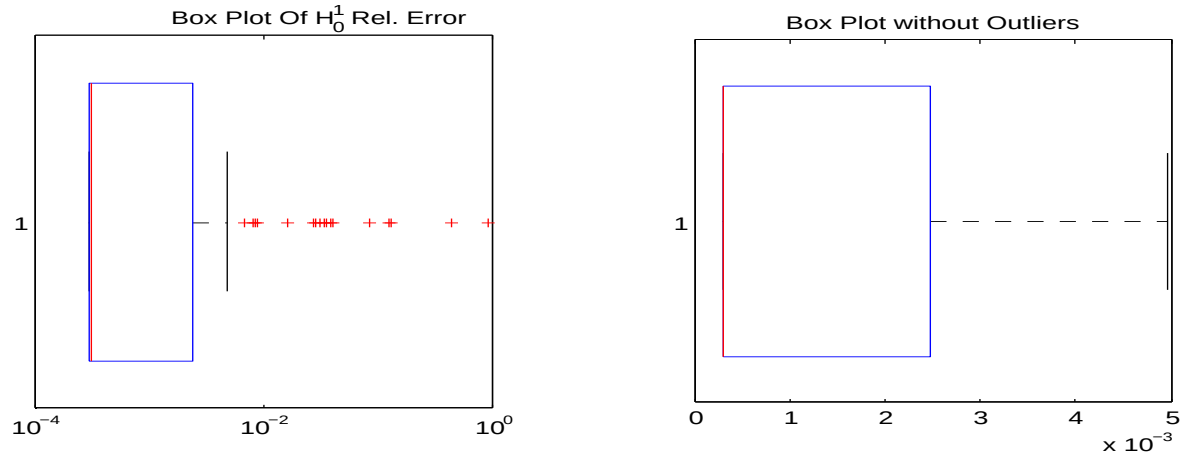

Figure 5.24: Box Plot of the relative H_0^1 discrepancy of the model output from the data (Median = 3.1246e-04).

We use the error estimate to filter out the realizations $q^*(\cdot, Z(\omega))$ whose model output differs from the measured data by more than a given threshold (in our case 5×10^{-3}) and obtain a reduced sample of size 66. The sample paths of the reduced sample are given in Figure 5.25.

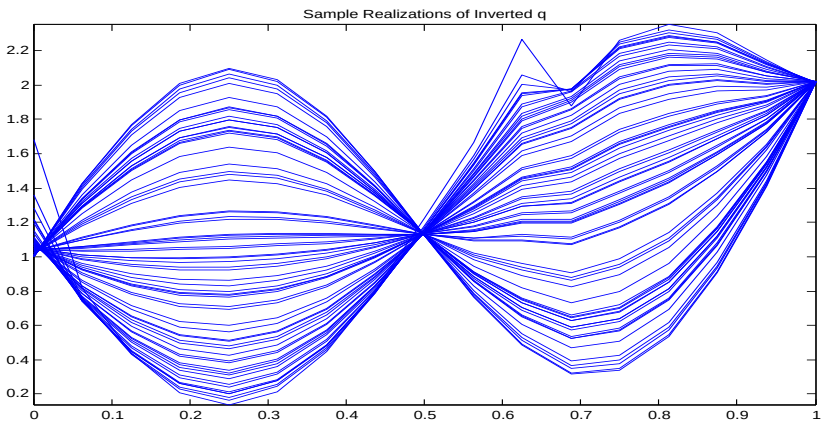

Figure 5.25: Sample paths of the estimate q^* of the parameter q obtained by inverse sampling.

Clearly, the bimodal behavior, which wasn't accurately preserved in sampling from the joint density of the output, is also not strongly evident here. Figure 5.26 compares the first four moments of the original parameter q and its approximation q^*. A possible reason for the understatement of the variance and fourth moment by the estimated parameter q^* relates

to the possible ill-posedness of the deterministic problem when q is low. Sample outputs corresponding to realizations of q^* whose minimum is uncomfortably close to zero may lead to ill-posed inverse problems, most of the results of which were filtered out. The mean is approximated quite accurately, however.

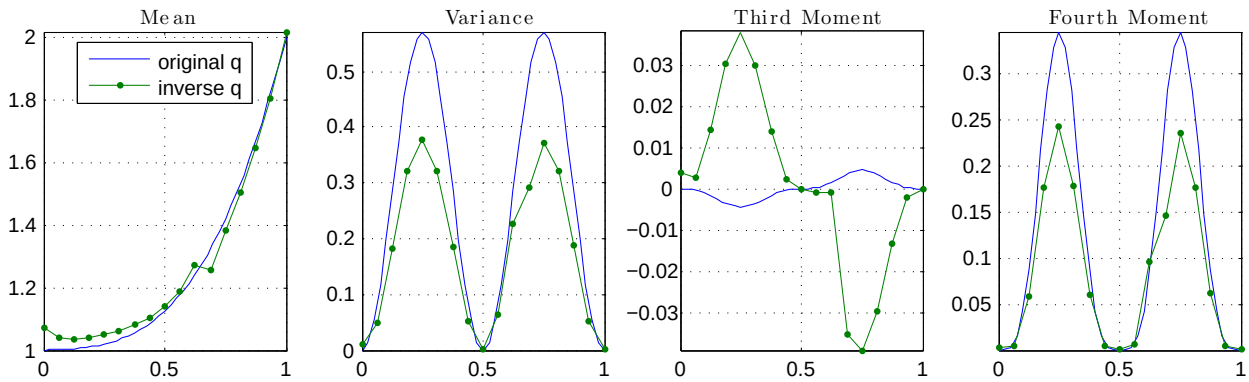

Figure 5.26: Comparison of the first four moments of the estimated parameter q^* with those of the original parameter q.

Finally we compare the Karhunen–Loève expansions of the original parameter q and its estimate q^*. The estimate q^* has four significant modes (according to our cutoff criteria) whereas the original parameter q only has one significant mode. We can compare the first mode (Figure 5.28), however, which seems to correspond well (up to a sign).

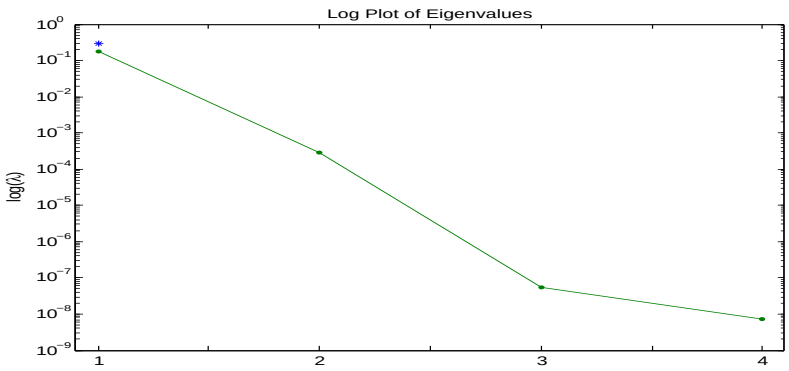

Figure 5.27: Eigenvalue decay in KL expansion. * represents the original eigenvalue

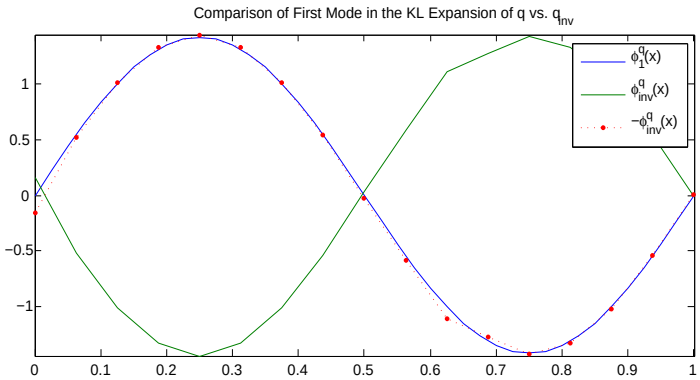

Figure 5.28: Comparison of the first mode in the KL expansion for the original parameter and its estimate q^*.

Chapter 6

Conclusions and Future Work

The quantification of uncertainty in the simulation output of computational models relies on an accurate statistical description of the underlying uncertain model parameters. In this dissertation, we proposed a variational method for estimating the uncertain diffusion coefficient q in stationary elliptic systems, based on measurements of the associated model output $\hat{u}$. We assumed that $\hat{u}$ is a given random field in the stochastic Sobolev space $\mathscr{H}_0^1(D)$ and posed the parameter identification problem as an infinite dimensional least squares optimization problem (P). This optimization problem was posed subject to the equality constraint $e(q,u)=0$, representing the stochastic forward problem, alongside the general admissibility constraint $q \in Q_{\text{ad}}$. In future studies, it would be reasonable to explore the possibility of allowing for other types of observation in this variational framework, such as measurements on certain parts of the domain, or of fluxes on the boundary.

For our formulations, we showed that minimizers of (P) exist and satisfy the appropriate saddle point conditions. The n-term truncation $\hat{u}^n$ of the Karhunen-Loève expansion of the measured output $\hat{u}$ leads to a 'finite noise' approximation (P^n) of (P), in which the parameter q is estimated as a relatively smooth, albeit high-dimensional function of the spatial variable x, as well as a the random variables $Y_1, ..., Y_n$ occurring in the truncated KL expansion of $\hat{u}^n$. Since the smoothness of q as a function of $Y_1, ..., Y_n$ is not known *a priori*, we impose a certain degree of smoothness on q by choosing the parameter space $\widetilde{H}_{\text{mix}}$, which not only improves the regularity of problem (P^n), but also allows for approximations by sparse grid hierarchical finite elements, a discretization scheme well-known for its effectiveness in mitigating the 'curse of dimensionality.' We proved the existence of minimizers of the 'finite noise' problem (P^n), established Karush-Kuhn-Tucker necessary optimality conditions, and showed how classical Tikhonov regularization theory could be generalized to prove that 'finite noise' minimizers converge to minimizers of (P) as the accuracy in the approximation $\hat{u}^n$ improves.

The 'finite noise' problem lends itself readily to a variety of existing optimization approaches, such as the steepest descent and augmented Lagrangian methods. The evaluation of the appropriate cost functional or gradient vector requires the estimation of high dimensional integrals, which can be achieved by means of sparse grid quadrature methods. The cost

of every step in the optimization procedure is dominated by the cost of forming the cost functional and the gradient, which requires the solution of the stochastic adjoint equation. We applied sparse grid stochastic collocation methods to resolve the equality constraint as well as the adjoint system. A promising avenue for future research would be to further explore alternative implementations of the given gradient-based optimization algorithms. In particular, methods that make greater use of sparsity or exploit the problem's underlying parallel structure would be required to apply this approach to more complex problems.

As an alternative to gradient-based optimization methods, we introduced an inverse sampling method, in which a stochastic interpolant of the estimated parameter q was constructed by solving a collection of deterministic inverse problems, essentially interchanging the operations of integration and optimization. The interpolant could then be used to compute statistical quantities of interest related q, such as its mean and variance. This method is relatively straightforward to implement in parallel and makes no *a priori* assumptions on the smoothness the parameter q as a function of the stochastic component. Its cost is, however, governed by the size of the interpolant needed to accurately approximate the integral that is used to compute the desired statistical quantity of interest. Since discontinuities in q have a considerable influence on the accuracy of numerical quadrature methods, the smoothness of q therefore still plays an indirect role in the number of required stochastic quadrature points. In the future, we plan to investigate the use of adaptive quadrature scheme in conjunction with the inverse sampling methods. It is also not clear whether inverse sampling methods are ultimately more efficient than gradient-based full optimization strategies in general and a more in-depth comparative study will be necessary to establish this.

Finally, we plan to apply the variational approach developed in this dissertation to more complex physical systems. Of particular interest are systems that exhibit time dependence, nonlinear operators, or material interfaces.

Bibliography

[1] R. A. Adams and J. J. F. Fournier. *Sobolev spaces*, volume 140 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam, second edition, 2003.

[2] I. Babuška, F. Nobile, and R. Tempone. A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM J. Numer. Anal.*, 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 J. Numer. Anal.*, 42(2):800–825, 2004.

[4] H. T. Banks and K. L. Bihari. Modelling and estimating uncertainty in parameter estimation. *Inverse Problems*, 17(1):95–111, 2001.

[5] H. T. Banks and K. Kunisch. *Estimation techniques for distributed parameter systems*, volume 1 of *Systems & Control: Foundations & Applications*. Birkhäuser Boston Inc., Boston, MA, 1989.

[6] V. Barthelmann, E. Novak, and K. Ritter. High dimensional polynomial interpolation on sparse grids. *Adv. Comput. Math.*, 12(4):273–288, 2000. Multivariate polynomial interpolation.

[7] A. Binder, H. W. Engl, C. W. Groetsch, A. Neubauer, and O. Scherzer. Weakly closed nonlinear operators and parameter identification in parabolic equations by Tikhonov regularization. *Appl. Anal.*, 55(3-4):215–234, 1994.

[8] Z. Brzeźniak and T. Zastawniak. *Basic stochastic processes*. Springer Undergraduate Mathematics Series. Springer-Verlag London Ltd., London, 1999. A course through exercises.

[9] T. Bui-Thanh, K. Willcox, and O. Ghattas. Model reduction for large-scale systems with high-dimensional parametric input space. *SIAM J. Sci. Comput.*, 30(6):3270–3288, 2008.

[10] H.-J. Bungartz and S. Dirnstorfer. Multivariate quadrature on adaptive sparse grids. *Computing*, 71(1):89–114, 2003.

[11] H.-J. Bungartz and M. Griebel. A note on the complexity of solving Poisson's equation for spaces of bounded mixed derivatives. *J. Complexity*, 15(2):167–199, 1999.

[12] H.-J. Bungartz and M. Griebel. Sparse grids. *Acta Numer.*, 13:147–269, 2004.

[13] T. F. Chan and X.-C. Tai. Identification of discontinuous coefficients in elliptic problems using total variation regularization. *SIAM J. Sci. Comput.*, 25(3):881–904, 2003.

[14] T. F. Chan and X.-C. Tai. Level set and total variation regularization for elliptic inverse problems with discontinuous coefficients. *J. Comput. Phys.*, 193(1):40–66, 2004.

[15] Z. Chen and J. Zou. An augmented Lagrangian method for identifying discontinuous parameters in elliptic systems. *SIAM J. Control Optim.*, 37(3):892–910, 1999.

[16] F. Colonius and K. Kunisch. Stability for parameter estimation in two-point boundary value problems. *J. Reine Angew. Math.*, 370:1–29, 1986.

[17] F. Colonius and K. Kunisch. Output least squares stability in elliptic systems. *Appl. Math. Optim.*, 19(1):33–63, 1989.

[18] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*, volume 375 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1996.

[19] H. W. Engl, K. Kunisch, and A. Neubauer. Convergence rates for Tikhonov regularisation of nonlinear ill-posed problems. *Inverse Problems*, 5(4):523–540, 1989.

[20] L. C. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1998.

[21] T. Gerstner and M. Griebel. Numerical integration using sparse grids. *Numer. Algorithms*, 18(3-4):209–232, 1998.

[22] R. G. Ghanem and P. D. Spanos. *Stochastic finite elements: a spectral approach*. Springer-Verlag, New York, 1991.

[23] M. Griebel. Adaptive sparse grid multilevel methods for elliptic PDEs based on finite differences. *Computing*, 61(2):151–179, 1998.

[24] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich. *Optimization with PDE constraints*, volume 23 of *Mathematical Modelling: Theory and Applications*. Springer, New York, 2009.

[25] K. Ito and K. Kunisch. The augmented Lagrangian method for equality and inequality constraints in Hilbert spaces. *Math. Programming*, 46(3, (Ser. A)):341–360, 1990.

[26] K. Ito and K. Kunisch. The augmented Lagrangian method for parameter estimation in elliptic systems. *SIAM J. Control Optim.*, 28(1):113–136, 1990.

[27] J. Kaipio and E. Somersalo. *Statistical and computational inverse problems*, volume 160 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2005.

[28] A. Kirsch. *An introduction to the mathematical theory of inverse problems*, volume 120 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1996.

[29] J. Klemelä. *Smoothing of multivariate data: Density estimation and visualization*. Wiley Series in Probability and Statistics. John Wiley & Sons Inc., Hoboken, NJ, 2009.

[30] C. Kravaris and J. H. Seinfeld. Identification of parameters in distributed parameter systems by regularization. *SIAM J. Control Optim.*, 23(2):217–241, 1985.

[31] H. W. Kuhn and A. W. Tucker. Nonlinear programming. In *Proc. of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1950*, pages 481–492, Berkeley and Los Angeles, 1951. University of California Press.

[32] K. Kunisch and X.-C. Tai. Sequential and parallel splitting methods for bilinear control problems in Hilbert spaces. *SIAM J. Numer. Anal.*, 34(1):91–118, 1997.

[33] K. Kunisch and L. White. Parameter estimation for elliptic equations in multidimensional domains with point and flux observations. *Nonlinear Anal.*, 10(2):121–146, 1986.

[34] O. A. Ladyzhenskaya and N. N. Ural'tseva. *Linear and quasilinear elliptic equations*. Translated from the Russian by Scripta Technica, Inc. Translation editor: Leon Ehrenpreis. Academic Press, New York, 1968.

[35] P. D. Lax and A. N. Milgram. Parabolic equations. In *Contributions to the theory of partial differential equations*, Annals of Mathematics Studies, no. 33, pages 167–190. Princeton University Press, Princeton, N. J., 1954.

[36] D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2009. With a chapter by James G. Propp and David B. Wilson.

[37] C. Lieberman, K. Willcox, and O. Ghattas. Parameter and state model reduction for large-scale statistical inverse problems. *SIAM J. Sci. Comput.*, 32(5):2523–2542, 2010.

[38] M. Loève. *Probability theory. II*. Springer-Verlag, New York, fourth edition, 1978. Graduate Texts in Mathematics, Vol. 46.

[39] X. Ma and N. Zabaras. An adaptive hierarchical sparse grid collocation algorithm for the solution of stochastic differential equations. *J. Comput. Phys.*, 228(8):3084–3113, 2009.

[40] X. Ma and N. Zabaras. An adaptive high-dimensional stochastic model representation technique for the solution of stochastic partial differential equations. *J. Comput. Phys.*, 229(10):3884–3915, 2010.

[41] Y. Marzouk and D. Xiu. A stochastic collocation approach to Bayesian inference in inverse problems. *Commun. Comput. Phys.*, 6(4):826–847, 2009.

[42] Y. M. Marzouk and H. N. Najm. Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems. *J. Comput. Phys.*, 228(6):1862–1902, 2009.

[43] H. G. Matthies and A. Keese. Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations. *Comput. Methods Appl. Mech. Engrg.*, 194(12-16):1295–1331, 2005.

[44] H. Maurer and J. Zowe. First and second order necessary and sufficient optimality conditions for infinite-dimensional programming problems. *Math. Programming*, 16(1):98–110, 1979.

[45] K. L. Mengersen and R. L. Tweedie. Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.*, 24(1):101–121, 1996.

[46] S. Meyn and R. L. Tweedie. *Markov chains and stochastic stability*. Cambridge University Press, Cambridge, second edition, 2009. With a prologue by Peter W. Glynn.

[47] F. Nobile, R. Tempone, and C. G. Webster. An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM J. Numer. Anal.*, 46(5):2411–2442, 2008.

[48] F. Nobile, R. Tempone, and C. G. Webster. A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM J. Numer. Anal.*, 46(5):2309–2345, 2008.

[49] E. Novak and K. Ritter. High-dimensional integration of smooth functions over cubes. *Numer. Math.*, 75(1):79–97, 1996.

[50] D. Pflüger, B. Peherstorfer, and H.-J. Bungartz. Spatially adaptive sparse grids for high-dimensional data-driven problems. *J. Complexity*, 26(5):508–522, 2010.

[51] M. Reed and B. Simon. *Methods of modern mathematical physics. I. Functional analysis*. Academic Press, New York, 1972.

[52] R. T. Rockafellar. Optimization under uncertainty. Lecture Notes, 2001.

[53] W. Rudin. *Functional analysis*. McGraw-Hill Book Co., New York, 1973. McGraw-Hill Series in Higher Mathematics.

[54] D. L. Russell. Some remarks on numerical aspects of coefficient identification in elliptic systems. In *Optimal control of partial differential equations (Oberwolfach, 1982)*, volume 68 of *Internat. Schriftenreihe Numer. Math.*, pages 210–228. Birkhäuser, Basel, 1984.

[55] A. Sandu. *Solution of Inverse Problems using Discrete ODE Adjoints*, pages 345–365. John Wiley and Sons, Ltd., 2010.

[56] C. Schwab and R. A. Todor. Karhunen-Loève approximation of random fields by generalized fast multipole methods. *J. Comput. Phys.*, 217(1):100–122, 2006.

[57] D. W. Scott. *Multivariate density estimation, theory, practice, and visualization*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Inc., New York, 1992.

[58] D. W. Scott and S. R. Sain. Multidimensional density estimation. *Handbook of Statistics*, 24:229–261, 2005.

[59] S. A. Smolyak. Quadrature and interpolation formulas for tensor products of certain classes of functions. *Dokl. Akad. Nauk SSSR*, 4:240–243, 1963.

[60] A. M. Stuart. Inverse problems: a Bayesian perspective. *Acta Numer.*, 19:451–559, 2010.

[61] V. N. Temlyakov. Approximation of functions with bounded mixed derivative. *Proc. Steklov Inst. Math.*, 1(178):vi+121, 1989. A translation of Trudy Mat. Inst. Steklov **1**78 (1986), Translated by H. H. McFaden.

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

[63] N. Zabaras and B. Ganapathysubramanian. A scalable framework for the solution of stochastic inverse problems using a sparse grid collocation approach. *Journal of Computational Physics*, 227(9):4697–4735, 2008.