QUASI-MONTE CARLO METHODS WITH APPLICATIONS TO PARTIAL DIFFERENTIAL EQUATIONS WITH RANDOM COEFFICIENTS

A thesis submitted for the degree of
Doctor of Philosophy

By
James Nichols
B.Sci.(Hons)

School of Mathematics,
The University of New South Wales.

May 2014

Abstract

This thesis provides the theoretical foundation for the component-by-component (CBC) construction of randomly shifted lattice rules that are tailored to integrals over $\mathbb{R}^s$ arising from Darcy-flow PDE problems where the permeability coefficient is given by a log-normal random field. We focus on the problem of computing the expected value of linear functionals of the solution of the PDE, which gives rise to integrals of the form $\int_{\mathbb{R}^s} f(\boldsymbol{y}) \prod_{j=1}^{s} \phi(y_j) \, \mathrm{d}\boldsymbol{y}$ with a univariate probability density ϕ. Our general strategy is to first map the integral into the unit cube $[0,1]^s$ using the inverse of the cumulative distribution function of ϕ, and then apply quasi-Monte Carlo (QMC) methods. However, the transformed integrand in the unit cube does not fall within the standard QMC settings from the literature. Therefore, a non-standard function space setting for integrands over $\mathbb{R}^s$ is required for the analysis. Such spaces were previously considered in [39], however due to the needs of the PDE problem, we must extend the theory of the aforementioned paper in several nontrivial directions, including a new error analysis for the CBC construction of lattice rules with general non-product weights, the introduction of an unanchored weighted space for the setting, the use of coordinate-dependent weight functions in the norm, and the strategy for fast CBC construction with POD ("product and order dependent") weights.

Our method of numerical approximation of this problem includes piecewise linear finite element approximation in physical space, the truncation of the parameterised expansion of the random field, and QMC quadrature rules for computing integrals over parameterised probability space which define the expected values. We give a rigorous error analysis for the effect of all three of these types of approximation. We show, using the non-standard function space setting developed in the thesis, that the quadrature error decays with $\mathcal{O}(n^{-1+\delta})$ with respect to the number of quadrature points n, where $\delta > 0$ is arbitrarily small and where the implied constant in the asymptotic error bound is independent of the dimension of the domain of integration.

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis.

I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.

James Nichols

Acknowledgements

I can not thank my supervisors Ian and Frances enough. Your kindness and humour has made the PhD entertaining and invigorating. There is no gratitude that can repay the opportunities you gave me in allowing me to join you in your endeavours. I must thank Ivan Graham, Rob Schiechl, and Christoph Schwab for being great teachers, even over great distances, for the chance to work with you on these projects, and for accommodating me on my visits abroad.

I'd hope I've shown the gratitude my friends and family deserve. The privilege of undertaking this degree would not be possible without the support that you all have given whether knowingly or not, and should not be underestimated. Particular thanks to my family Jennifer, John and Alex, and to Pia, thank you for the inspiration you give me daily. To my housemates past and present at The Barn, thank you for a great chapter of my life. And to the UNSW Maths and Stats family, which has proven to be such a cohesive, driven, exciting and unpredictable bunch, thank you for to opportunity to join your ranks as a friend and peer. And one last thanks to the coffee gang.

Contents

Chapter 1

Introduction and motivation

1.1 The subject of this thesis

Quasi-Monte Carlo (QMC) methods have proven to be very effective at tackling high dimensional integration and approximation problems. QMC methods involve n-point quadrature to approximate an integral over the unit cube $[0,1]^s$, where the quadrature points are chosen deterministically from the $[0,1]^s$. There are countless examples of problems to which QMC provides an invaluable tool, including, but not limited to, mathematical finance, ray-tracing in computer graphics, agent-based modelling and statistical maximum-likelihood problems. In this thesis we explore the application of QMC methods of PDEs with random coefficients.

We are concerned with a particular problem of single phase fluid flow in a saturated porous medium. We call this the porous-flow problem. Such a model can be used to simulate the flow of fluids through ground-rock or soils, for the purpose of aquifer management or the study of pollutant spread. To model this we consider the following second-order elliptic PDE

$$-\nabla \cdot (a(\boldsymbol{x})\nabla u(\boldsymbol{x})) \;=\; f(\boldsymbol{x})\,,$$

which we solve in some bounded domain $D \subset \mathbb{R}^d$ for $d = 1$, 2 or 3. We wish to solve for u, the residual pressure field. The coefficient a is provided as the permeability of the underlying medium. We are interested in generalising this to let the coefficient a be a random field. This step reflects the roughness of the permeability field for rock formations, as well as the fact that, to some degree, it is impossible to have complete knowledge of the permeability of a given region in the earth's crust, especially at small scales. Thus we write $a(\boldsymbol{x}, \boldsymbol{y})$, for $\boldsymbol{y} \in \mathbb{R}^{\mathbb{N}}$, where $\mathbb{R}^{\mathbb{N}}$ is a (parameterised) probability space equipped with a probability measure ρ. The solution $u(\boldsymbol{x}, \boldsymbol{y})$ must now also depend on $\boldsymbol{y}$.

We are thus interested in finding statistical answers to our questions, for example "what is the expected time for a pollutant particle to cross the domain?". In asking for an expected value of a quantity, the answer must involve an integration over the probability space $\mathbb{R}^{\mathbb{N}}$. More specifically, if for example the functional $\mathcal{G}(u)$ provides us with the crossing time for a particle, where u is a solution of the PDE with a particular realisation of the field a, then we seek the expectation $\mathbb{E}^{\boldsymbol{y}}[\mathcal{G}(u)] = \int_{\mathbb{R}^{\mathbb{N}}} \mathcal{G}(u)\,\mathrm{d}\rho(\boldsymbol{y})$.

There are many techniques in the literature to solve this problem, including variants of stochastic collocation and stochastic Galerkin methods. Here however, we take the following approach of three approximations. The first approximation is to be able to solve

the PDE for a given field realisation. This we do with piecewise continuous finite element methods (FEM). The second is to truncate $\boldsymbol{y}$, that is, reduce out parameterised probability space to a finite dimensional space $\mathbb{R}^s$, making the integral above finite (but possibly quite high) dimensional. And finally we apply QMC methods to perform the numerical integration on the truncated probability space. Here we analyse the contribution towards the approximation error of all three elements.

We pay particular attention to the analysis of the QMC error. We apply shifted lattice rules, a type of QMC point sets. Shifted lattice rules may be constructed for integrands belonging to certain weighted high-dimensional functions spaces, using the *component-by-component* (CBC) algorithm, a greedy algorithm. We can prove that the error of the lattice rule converges fast in these function spaces. Unfortunately however, because of the unbounded nature of the domain of integration $\mathbb{R}^s$, the integrand for the PDE problem does not belong to to the usual Sobolev type function spaces used throughout the literature of lattice rules. Thus we open up a new problem in the realm of shifted lattice rules, where we must consider a function space suitable for our PDE problem, then prove that lattice rules obtain good convergence in these spaces.

The success of this technique relies on the ability of the QMC method to perform better than the alternatives, particularly Monte Carlo (MC) methods, while being easy to use, and readily accessible to practitioners. Throughout the paper we tackle many technical results, eventually proving the results we desired, which is certainly beyond the scope of a practitioner. However the outcome is beautiful in its simplicity – the generating vector is no more than a vector of integers, and use of lattice rules is simple. We conclude the thesis with numerical experiments whereby we construct these lattice rules for our custom weighted function space that is tailored fit to the PDE, and demonstrate that they perform well over a wide variety of parameters.

1.2 The main achievements of this thesis

Some key achievements in this thesis can be summarised as follows.

1. A novel weighted, unanchored, and unbounded space has been derived, and the corresponding reproducing kernel derived.
2. A proof that lattice rules built using the CBC construction, for the unanchored space with general weights and coordinate dependent weight functions, can achieve good or even optimal $\mathcal{O}(n^{-1+\delta})$ theoretical convergence.
3. We have described the details of implementing the fast CBC algorithm for the unanchored space with POD weights, including the necessary procedures to use matrix storage and FFT methods to speed up calculation, and have outlined techniques to avoid numerical issues to do with POD weights with factorial order dependence.
4. We present results for the analysis of the truncation and finite element approximation error of the PDE with parameterised random coefficients has been provided.

5. It has been demonstrated that the linear functionals of the solution u of the PDE, under the right conditions, exists in the weighted unanchored space derived earlier, thus the QMC results of good lattice rules apply to this problem.
6. The CBC algorithm for these novel weight unbounded spaces as well as a QMC-finite element solver have been implemented, and tested, and have been shown to perform very well against MC approximation.

Many of these achievements arose from work with Ian Sloan, Frances Kuo, Ivan Graham, Rob Scheichl, and Christoph Schwab, with the results of our collaborations in [46] and [24] submitted for publication.

1.3 The outline of this thesis

In Chapter 2 we begin with a survey of QMC methods and lattice rules. We review the notions of discrepancy and worst-case error and other measurements of "good" lattice rules. Following this is a survey of reproducing kernel Hilbert spaces, with their applications to QMC. We then review weighted spaces, including the weighted Korobov and weighted Sobolev spaces. We describe the CBC algorithm, and provide results from the literature that demonstrates that the algorithm provides good lattice rules.

In Chapter 3 we make the necessary generalisations of the theory from the preceding chapter to be able to apply our QMC methods to the porous-flow problem. This includes the derivation of the weighted unanchored and unbounded function space. We also review the anchored space that is covered in earlier publications. We prove here that lattice rules constructed with the CBC algorithm in the weighted unanchored space, with general weights, can show optimal convergence.

In Chapter 4 we take a turn towards the Darcy-flow PDE problem. We carefully specify the PDE and the random field, and make our assumptions on the smoothness of the random field and hence the regularity of the problem. Then we specify the integral problem that we wish to approximate as the expectation of the functional of the pressure field over the underlying probability space. We then start by providing results for the finite element as well as truncation errors. Following this we analyse the behaviour of the partial derivatives of the integrand with respect to the stochastic variables (rather than the spatial variables), bounding the partial derivatives, such that we can show that the integrand belongs to our unanchored weighted space. Finally we state our convergence results for this QMC-FE method.

In Chapter 5 we examine the implementation and numerical results of our theoretical work in this thesis. We outline the details of the fast CBC algorithm for the unanchored space, including a method to cope with POD weights with strong order-dependent growth. We also propose a method to use the CBC algorithm on the anchored space, which is otherwise difficult for technical reasons. Also involved in the CBC algorithm is the ability to calculate the shift-average kernel, this must be done with one-dimensional numerical integration, which we discuss. Then present some worst-case error results for a few model problems. Finally we tackle the PDE problem, defining a model problem to apply our

QMC quadrature. We define a range of parameters and perform the CBC algorithm, tailored to the PDE. We present the worst-case error results, as well as the quadrature error results. Included in this is a discussion of unstable parameters, and difficulties we had in setting various parameters of the weighted unanchored space parameters to fit the PDE problem.

Chapter 2

QMC methods and shifted lattice rules

We are concerned here with the standard problem of numerical integration on the s-dimensional hypercube $[0,1]^s$,

$$I_s(f) := \int_{[0,1]^s} f(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y},$$

using an n-point approximation, or *quadrature rule*,

$$Q_n(\mathcal{P};\, f) := \frac{1}{n}\sum_{k=1}^{n} f(\boldsymbol{t}^{(k)}),$$

where the quadrature points $\mathcal{P} = \{\boldsymbol{t}^{(k)}\}_{k=1}^{n}$ are a set of well-distributed points in $[0,1]^s$. While many practical settings for numerical integration may be on domains other than $[0,1]^s$, our theory here easily applies to any compact domain in $\mathbb{R}^s$ that can readily be mapped back to $[0,1]^s$ with no singularities in the mapping.

There are many techniques of choosing quadrature points $\boldsymbol{t}^{(i)}$ available to us. Perhaps the most commonly used technique is Monte Carlo quadrature, where the points are chosen randomly within $[0,1]^s$, usually with a uniform distribution. In addition to this there are quasi-Monte Carlo (QMC) methods, where the quadrature points are chosen deterministically but to perform better than their random counterparts. Some QMC point sets include Sobol′, Halton and Faure sequences, and lattice points. A review of these QMC methods and more can be found in [49].

An essential part of estimating $I_s(f)$ with numerical quadrature is the ability to estimate the error of our approximation, that is to be able to estimate the difference between the our target integral and the approximation, $|I_s(f) - Q_n(\mathcal{P};\, f)|$. Ideally we would like to find bound of the error with respect to the number of quadrature points n, for example a bound of the form Cn^{-p} for some $p > 0$, where the constant C may or may not depend on s. For example, we know that for Monte Carlo point sets, the root-mean-square error is $\mathcal{O}(n^{-1/2})$, while for certain shifted lattice rules, optimal bounds are of the form $\mathcal{O}(n^{-1+\delta})$ for some small $\delta > 0$.

In particular we are interested in the *worst-case error* for a given rule $\mathcal{P}$ for a given function space $\mathcal{F}$ which we take to be a Banach space, but often will be a Hilbert space.

We define the worst-case error as follows.

$$e_{s,n}(\mathcal{P}, \mathcal{F}) := \sup_{f \in \mathcal{F},\, \|f\|_{\mathcal{F}} \leq 1} |I_s(f) - Q_n(\mathcal{P};\, f)|. \tag{2.1}$$

Clearly we have that for any $f \in \mathcal{F}$

$$|I_s(f) - Q_n(\mathcal{P};\, f)| \leq e_{s,n}(\mathcal{F}, \mathcal{P})\, \|f\|_{\mathcal{F}}, \tag{2.2}$$

meaning we can take the worst-case error to be a measure of the "quality" of point sets that is independent of any specific function. Choosing the function space $\mathcal{F}$ of integrands allows us to specify what properties we wish the integrands $f \in \mathcal{F}$ have, for example we may wish them to have square integrable first-order derivatives or Fourier series with a certain polynomial decay. This choice, and in particular what functions it then includes in the unit ball, affects the size of the worst-case error. Later in this chapter we shall encounter settings in which it is possible to derive closed-form expressions of this quantity.

Another notion of "quality" of the point set $\mathcal{P}$ is the *discrepancy*. For $\boldsymbol{y} \in [0,1]^s$, we define the *local discrepancy function* as

$$\mathrm{discr}_{\mathcal{P}}(\boldsymbol{y}) = \frac{\#\{\boldsymbol{t}^{(i)} : \boldsymbol{t}^{(i)} \in [0, \boldsymbol{y})\}}{n} - \prod_{k=1}^{s} y_k,$$

where $\#A$ is the counting-measure, or the number of points in a discrete set, and as a shorthand we have used the notation $[0, \boldsymbol{y})$, where $[0, \boldsymbol{y}) = [0, y_1) \times \ldots \times [0, y_s)$. The local discrepancy function $\mathrm{discr}_{\mathcal{P}}$ represents the difference between the proportion of quadrature points in the hypercube $[0, \boldsymbol{y})$ and the volume of $[0, \boldsymbol{y})$, which intuitively is an indicator of the quality of the spread of $\mathcal{P}$. In absolute value it is smaller (closer to 0) when the points $\boldsymbol{t}^{(i)}$ are evenly spread, such that the proportion of points in any given hypercube is reasonably close to its volume, and larger in absolute value in areas where there are holes or clusters in $\mathcal{P}$.

It is instructive to derive the following results in one dimension, i.e., the unit-interval, $[0, 1]$. In one dimension we see that

$$\mathrm{discr}_{\mathcal{P}}(y) = \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{[0,y)}(t^{(k)}) - y = \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{(t^{(k)},1]}(y) - y$$

where $\mathbf{1}_A(y)$ is the indicator function of a set A. We can now demonstrate the following,

$$\begin{aligned}\int_0^1 f'(y)\,\mathrm{discr}_{\mathcal{P}}(y)\,\mathrm{d}y &= \int_0^1 f'(y)\left(\frac{1}{n}\sum_{k=1}^{n}\mathbf{1}_{(t^{(k)},1)}(y) - y\right)\mathrm{d}y \\ &= \frac{1}{n}\sum_{k=1}^{n}\int_{t^{(k)}}^1 f'(y)\,\mathrm{d}y - \int_0^1 yf'(y)\,\mathrm{d}y \\ &= \frac{1}{n}\sum_{k=1}^{n}\Big(f(1) - f(t^{(k)})\Big) - \left([yf(y)]_0^1 - \int_0^1 f(y)\,\mathrm{d}y\right) \\ &= \int_0^1 f(y)\,\mathrm{d}y - \frac{1}{n}\sum_{k=1}^{n} f(t^{(k)}) = I_s(f) - Q_n(\mathcal{P};\,f),\end{aligned}$$

which gives us the simplified 1-dimensional *Zaremba identity*,

$$I_s(f) - Q_n(\mathcal{P};\,f) = \int_0^1 f'(y)\,\mathrm{discr}_{\mathcal{P}}(y)\,\mathrm{d}y. \tag{2.3}$$

Now take the absolute value of (2.3) and apply Hölder's innequality to obtain

$$\begin{aligned}|I_s(f) - Q_n(\mathcal{P};\,f)| &\le \int_0^1 |f'(y)\,\mathrm{discr}_{\mathcal{P}}(y)|\,\mathrm{d}y \\ &\le \left(\int_0^1 |f'(y)|^p\right)^{1/p}\left(\int_0^1 |\mathrm{discr}_{\mathcal{P}}(y)|^q\right)^{1/q},\end{aligned} \tag{2.4}$$

where $1/p + 1/q = 1$. There are two special cases, $(p, q) = (1, \infty)$ and $(2, 2)$, from which we obtain

$$|I_s(f) - Q_n(\mathcal{P};\,f)| \le |f|_1 D_\infty(\mathcal{P}), \tag{2.5}$$

and

$$|I_s(f) - Q_n(\mathcal{P};\,f)| \le |f|_2 D_2(\mathcal{P}),$$

where the semi-norms $|\cdot|_1$ and $|\cdot|_2$ are $|f|_1 = \int_0^1 |f'(x)|\,\mathrm{d}x$ and $|f|_2 = (\int_0^1 |f'(x)|^2\,\mathrm{d}x)^{1/2}$, and we have labelled the two notions of discrepancy, the *star discrepancy*

$$D_\infty(\mathcal{P}) = \sup_{y\in[0,1]} |\mathrm{discr}_{\mathcal{P}}(y)|,$$

and the *L_2-discrepancy*

$$D_2(\mathcal{P}) = \left(\int_{[0,1]} |\mathrm{discr}_{\mathcal{P}}(y)|^2\,\mathrm{d}y\right)^{1/2}. \tag{2.6}$$

Evidently both these quantities are small when the points are evenly distributed over the unit cube. The inequality in (2.5) is a simplified form of the original Koksma-Hlawka inequality (see e.g. [32]), which, similar to (2.2), neatly bounds the error by a property of

the integrand (the semi-norm, $|\cdot|_1$) and the quality of the point set $\mathcal{P}$ (the star discrepancy, $D_\infty(\mathcal{P})$).

Before proceeding further, we look briefly at the extension to higher dimensions $s \geq 1$, i.e. $[0,1]^s$. The Zaremba identity then becomes

$$I_s(f) - Q_n(\mathcal{P};\, f) = \sum_{\emptyset \neq \mathfrak{u} \subseteq \{1:s\}} (-1)^{|\mathfrak{u}|+1} \int_{[0,1]^{|\mathfrak{u}|}} \frac{\partial^{|\mathfrak{u}|} f}{\partial \boldsymbol{y}_\mathfrak{u}}(\boldsymbol{y}_\mathfrak{u}; \mathbf{1}) \,\mathrm{discr}_\mathcal{P}(\boldsymbol{y}_\mathfrak{u}; \mathbf{1})\, \mathrm{d}\boldsymbol{y}_\mathfrak{u}, \tag{2.7}$$

where $\mathfrak{u} \subset \{1, 2, ..., s\}$ is a sub-collection of the indices of the dimensions, evidently there are 2^s such collections (note that we use the Fraktur font for $\mathfrak{u}$, which is reserved for this set notation), $\frac{\partial^{|\mathfrak{u}|} f}{\partial \boldsymbol{y}_\mathfrak{u}}$ denotes the mixed partial derivative of f in the coordinates x_i provided $i \in \mathfrak{u}$ (note that it is a mixed partial *first* derivative), and $(\boldsymbol{y}_\mathfrak{u}; \mathbf{1}) \in [0,1]^s$ is the vector whose i-th components are x_i if $i \in \mathfrak{u}$, and 1 if $i \notin \mathfrak{u}$. Also note that $\mathfrak{u}$ can contain any dimension coordinate at most once (this will be important when we later examine multi-index notation). A corresponding generalisation of (2.4), where once again we use the Hölder's inequality, is

$$|I_s(f) - Q_n(\mathcal{P};\, f)| \leq \left(\sum_{\emptyset \neq \mathfrak{u} \subseteq \{1:s\}} \left\| \frac{\partial^{|\mathfrak{u}|} f}{\partial \boldsymbol{y}_\mathfrak{u}}(\boldsymbol{y}_\mathfrak{u}; \mathbf{1}) \right\|_{L^p}^p \right)^{1/p} \left(\sum_{\emptyset \neq \mathfrak{u} \subseteq \{1:s\}} \|\mathrm{discr}_\mathcal{P}(\boldsymbol{y}_\mathfrak{u}; \mathbf{1})\|_{L^q}^q \right)^{1/q}. \tag{2.8}$$

2.1 Tractibility

As well as providing useful upper bounds for quadrature error for a given n, it is natural to ask a related question - given some $\varepsilon > 0$ and $s \geq 1$, what is the minimum number of quadrature points n that is required to achieve an error less than ε? The result ideally is some function $n(s, \varepsilon)$, a direct relationship between the number of dimensions and desired error and the number of points needed, for a given class of integrands. This concept applies to the much wider class of *approximation problems*, which includes the n point quadrature that we investigate in this thesis. The books [51, 52] contain a detailed account of the field of tractability.

There are settings for which we may want to consider integration on an *infinite dimensional* domain, e.g. $[0,1]^\mathbb{N}$. In practice we consider finite dimensional integrals that are approximations, by virtue of being *truncated*, of infinite dimensional integrals. To be precise

$$\lim_{s \to \infty} \int_{[0,1]^s} f(\boldsymbol{y}; \mathbf{0})\, \mathrm{d}\boldsymbol{y} \;=\; \int_{[0,1]^\mathbb{N}} f(\boldsymbol{y})\, \mathrm{d}\boldsymbol{y}$$

where we have written $(\boldsymbol{y}; \mathbf{0})$ to indicate the sequence $(y_1, \ldots, y_s, 0, 0, \ldots)$. Such is the case in the porous-flow problem that we shall discuss later. In this setting we need to know how our error bounds behave as we let $s \to \infty$, and whether it will be necessary to consider more quadrature points as we increase the number of dimensions.

Hence we say that a problem is *tractable* if the function $n(s, \epsilon)$ of a given class of integrands can be, roughly speaking, bounded by polynomial expressions of ε and s. If

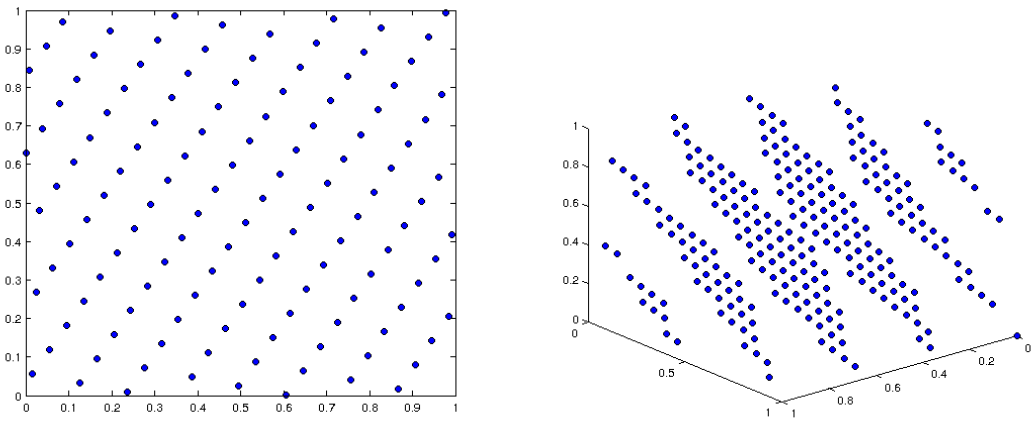

Figure 2.1: Two examples of lattice rules, to the left, 127 points in $[0,1]^2$ with generator $(1,27)$ and shift $(0.15,0.67)$, to the right, 257 points in $[0,1]^3$ with generator $(1,76,113)$.

we can drop the dependence on s, that is we get a function $n(s,\epsilon)$ that is independent of the dimensionality, then we say that the problem is *strongly tractable*.

2.2 Lattice rules and classical theory

In much of the theory that follows shortly, we shall be considering QMC algorithms that are *shifted rank-1 lattice rules*. We start by defining the following set

$$\mathcal{Z}_n := \{1 \leq z \leq n : \gcd(z,n) = 1\}. \tag{2.9}$$

Then for any $z \in \mathcal{Z}_n$ we can see that, for $k = 1,\ldots,n$, we can generate the entire set of integers $1 \ldots n$ with $kz (\mathrm{mod} n)$. Also note that if n is prime, then $\mathcal{Z}_n = \{1,\ldots,n-1\}$. Now, we can define the shifted rank-1 lattice rule,

$$Q_n(\boldsymbol{z},\boldsymbol{\Delta};\, f) = \frac{1}{n}\sum_{k=1}^{n} f\left(\left\{\frac{k}{n}\boldsymbol{z} + \boldsymbol{\Delta}\right\}\right), \tag{2.10}$$

where $\boldsymbol{z} \in \mathcal{Z}_n^s$ is said to be the *generating vector*, and $\boldsymbol{\Delta} \in [0,1]^s$ is the *shift*, and we have written $\{\cdot\}$ to represent taking only the fractional part of the components of a number or vector, i.e. $\{x\} = x - \lfloor x \rfloor$. The shift is usually taken to be a random vector in $[0,1]^s$, and as we shall see, our error bounds will be in terms of an expectation over this shift. If there is no shift, i.e., if $\boldsymbol{\Delta} = \boldsymbol{0}$, then we write $Q_n(\boldsymbol{z};f) = Q_n(\boldsymbol{z},\boldsymbol{0};f)$ for the regular *rank-1 lattice rule*. As each $z_j \in \mathcal{Z}_n$, we have that each 1-dimensional projection of the lattice rule covers $[0,1]$ at n equally spaced points. Higher dimensional projections are not guaranteed to cover their respective subspaces though.

The theory of "good lattice points" originates from the works of Korobov, see [31], as well as Hlawka, see [30]. A good summary of the field can be found in [64]. Throughout the rest of this chapter and the next, we present results on the existence and methods of construction of $\boldsymbol{z}$ that produce "good" lattice rules. That is, lattice rules that provide

fast-converging upper bounds on the quadrature error $|I_s(f) - Q_n(\boldsymbol{z}, \boldsymbol{\Delta}; f)|$, which we demonstrate via bounding worst-case errors, (2.1).

Classical theory centres around measurement of goodness of lattice rules for functions from the *Korobov class*, $E_\alpha(c)$. A periodic function f is in $E_\alpha(c)$ if

$$\left|\hat{f}(\boldsymbol{h})\right| \leq \frac{c}{(\bar{h}_1\bar{h}_2\cdots\bar{h}_s)^\alpha}, \tag{2.11}$$

where $\hat{f}(\boldsymbol{h})$ is the $\boldsymbol{h}$-th Fourier-series component, and $\bar{h} = \max(1, |h|)$. This is essentially a requirement of the smoothness of functions f in $E_\alpha(c)$, and indeed the larger α is the faster the Fourier components must decay, and hence the smoother the function is.

Let us write $a \equiv_n b$ if $a = b \,(\bmod n)$. If $\mathcal{L}$ is our set of rank-1 lattice points $\boldsymbol{t}^{(k)} = \left\{\frac{k}{n}\boldsymbol{z}\right\}$, then we define the *dual lattice*

$$\mathcal{L}^\perp = \{\boldsymbol{h} \in \mathbb{Z}^s : \boldsymbol{z}\cdot\boldsymbol{h} \equiv_n 0\}$$

The dual lattice has the following important property.

Identity 1

$$Q_n(\boldsymbol{z}; \exp(2\pi i\boldsymbol{h}\cdot\boldsymbol{x})) \;=\; \frac{1}{n}\sum_{k=1}^{n}\exp(2\pi ik\boldsymbol{h}\cdot\boldsymbol{z}/n) \;=\; \begin{cases} 1 & \text{if } \boldsymbol{h}\in\mathcal{L}^\perp, \\ 0 & \text{otherwise.} \end{cases} \tag{2.12}$$

This result is trivial, however we note that it also applies to *general* lattices, a full discussion of which can be found in [64]. This property allows us to represent the error for functions in the class $E_\alpha(c)$.

Theorem 2 *If f has an absolutely convergent Fourier series, then*

$$Q_n(\boldsymbol{z};\, f) - I_s(f) = \sum_{\substack{\boldsymbol{h}\in\mathcal{L}^\perp \\ \boldsymbol{h}\neq\boldsymbol{0}}} \hat{f}(\boldsymbol{h}) \tag{2.13}$$

Proof. Applying to the Fourier series representation

$$Q_n(\boldsymbol{z};\, f) \;=\; Q_n\left(\boldsymbol{z};\, \sum_{\boldsymbol{h}\in\mathbb{Z}^s}\hat{f}(\boldsymbol{h})\right) \;=\; \sum_{\boldsymbol{h}\in\mathbb{Z}^s} Q_n(\boldsymbol{z};\, \hat{f}(\boldsymbol{h})),$$

noting that $I_s(f) = \hat{f}(\boldsymbol{0})$, and using Identity 1 gives us the result. □

Hence by combining (2.13) and the definition of $E_\alpha(c)$, (2.11), we get the following bound that applies to all $f \in E_\alpha(c)$,

$$\begin{aligned} |I_s(f) - Q_n(\boldsymbol{z};\, f)| &\leq c\sum_{\boldsymbol{h}\in\mathcal{L}^\perp}{}' \frac{1}{(\bar{h}_1\bar{h}_2\cdots\bar{h}_s)^\alpha} \\ &= cP_\alpha(\boldsymbol{z}, n) \end{aligned} \tag{2.14}$$

Where we have re-labelled the sum in the second half of the above equation as $P_\alpha(\boldsymbol{z}, n)$. This quantity $P_\alpha(\boldsymbol{z}, n)$ is independent of the function f which we are trying to integrate, and entirely dependent on the choice of generating vector $\boldsymbol{z}$ and smoothness α of the Korobov class. Hence we have another way of bounding our quadrature with a quantity that is independent of f, in similar vein to (2.8). In some of the classic literature of lattice rules, such as [21, 47, 48, 50], work is undertaken to find good bounds on $P_\alpha(\boldsymbol{z}, n)$ and other related quantities. However, from here we will base our work on Hilbert space techniques rather than these classes, though we note that they rely on similar notions such as dual lattices, and measurements of goodness that are akin to $P_\alpha(\boldsymbol{z}, n)$.

2.3 Reproducing kernel Hilbert spaces

We consider here a Hilbert space $\mathcal{H}$ of functions defined on a domain D (which typically will be $[0,1]^s$) that is equipped with a *reproducing kernel*, which we define shortly. The use of these spaces has been important in the theory of QMC, as it facilitates analysis of worst-case error. In fact, as we shall see shortly, there are useful Hilbert space settings where closed-form expressions exist for the worst-case error.

In the following subsections we look at some examples of spaces that are useful to study, and corresponding kernels for those spaces. Typically spaces that are used are of Korobov type, where successive Fourier components are bounded, or Sobolev type, where the integral of mixed derivatives are bounded. Later we shall continue on to consider the *weighted* spaces, and examine the challenges the weighted setting allow us to overcome.

Here we present a brief introduction to the topic in the abstract setting, presenting necessary results for the work that follows in later chapters, and in some cases presenting a sketch of the relevant proof. For a survey of the topic of reproducing kernel Hilbert spaces in an abstract setting, we refer the reader to [3].

Definition 3 (Reproducing kernel Hilbert space) *We say that a Hilbert space $\mathcal{H}$ of functions $f : D \to \mathbb{R}$, with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$, is a* reproducing kernel Hilbert space *if it is equipped with a function $K : D \times D \to \mathbb{R}$ such that*

1. *for any fixed $\boldsymbol{y} \in D$, $K(\cdot, \boldsymbol{y}) \in \mathcal{H}$, and*
2. *$K(\boldsymbol{x}, \boldsymbol{y})$ obeys the* reproducing property, *$\langle f, K(\cdot, \boldsymbol{y}) \rangle_{\mathcal{H}} = f(\boldsymbol{y})$, for all $f \in \mathcal{H}$ and $\boldsymbol{y} \in D$.*

Some important properties of reproducing kernels follow.

Proposition 4 *A reproducing kernel K in a Hilbert space $\mathcal{H}$ that satisfies the properties in Definition 3 is*

1. Symmetric*: $K(\boldsymbol{x}, \boldsymbol{y}) = K(\boldsymbol{y}, \boldsymbol{x})$ for all $\boldsymbol{x}, \boldsymbol{y} \in D$.*
2. Positive semi-definite*: $\sum_{i,j=1}^{n} a_i a_j K(\boldsymbol{y}_i, \boldsymbol{y}_j) \geq 0$ for all finite collections $\{a_i\}_{i=1}^n \subset \mathbb{R}$ and $\{\boldsymbol{y}_i\}_{i=1}^n \subset D$.*
3. Unique*: For any other function $\widetilde{K}(\boldsymbol{x}, \boldsymbol{y})$, $\widetilde{K}(\boldsymbol{x}, \boldsymbol{y}) = K(\boldsymbol{x}, \boldsymbol{y})$*
4. *And in particular $K(\boldsymbol{x}, \boldsymbol{x}) \geq 0$ for all $\boldsymbol{x} \in D$.*

Proof. Details of the proofs can be found in [3]. □

The following result outlines the conditions for which a Hilbert space can be equipped with such kernels

Proposition 5 *For a reproducing kernel $K(\boldsymbol{x}, \boldsymbol{y})$ to exist in a Hilbert space $\mathcal{H}$ of real-valued functions on D, it is a necessary and sufficient condition that function evaluations be a continuous linear functional in $\mathcal{H}$. That is for any $\boldsymbol{y} \in D$, the functional $E_{\boldsymbol{y}}(f) = f(\boldsymbol{y})$ is continuous in f.*

Proof. The proof follows as a consequence of the Riesz representation theorem. See [3] for details. □

We note that the conditions in Proposition 5, for the existence of a reproducing kernel, are reasonably straightforward, and that for some of these Hilbert spaces a corresponding reproducing kernel can notionally be found. In [67] the reproducing kernel is derived for various examples of Hilbert spaces for which Proposition 5 applies. In the following sections we shall present well known spaces and provide their kernels as a given, however in Chapter 3 we shall be considering a new space for which we must derive the kernel.

Conversely, any given function $K(\boldsymbol{x}, \boldsymbol{y})$ that is symmetric and positive semi-definite defines a unique Hilbert space, equipped with an inner product $\langle \cdot, \cdot \rangle_K$, where $K(\boldsymbol{x}, \boldsymbol{y})$ has the reproducing property. We state this in the following.

Proposition 6 *Suppose $K : D \times D \to \mathbb{R}$ is a symmetric, positive semi-definite function, then there is a unique Hilbert space of real-valued functions on D such that K is a reproducing kernel.*

Proof. Consider functions of the form $\sum_{k=1}^{n} a_k K(\boldsymbol{x}_k, \cdot)$ where $\{a_k\}_{k=1}^{n} \subset \mathbb{R}$ and $\{\boldsymbol{x}_k\}_{k=1}^{n} \subset D$. The inner product defined by

$$\left\langle \sum_{k=1}^{m} b_k K(\boldsymbol{y}_k, \cdot), \sum_{k=1}^{n} a_k K(\boldsymbol{x}_k, \cdot) \right\rangle := \sum_{j=1}^{n} \sum_{k=1}^{m} a_j b_k K(\boldsymbol{y}_k, \boldsymbol{x}_j),$$

can be shown to satisfy the criteria in Definition 3. The proof then involves defining a Hilbert space in terms of the completion this subspace of functions. A detailed proof can be found in [3]. □

We note from the uniqueness property of Proposition 4, as well as Proposition 5, that most useful Hilbert spaces will have a reproducing kernel that is unique for that space. Thus there is a "one-to-one" correspondence between the norm of the space and the reproducing kernel. This means that characteristics of the norm must be reflected in the kernel, if any parameter appears in the norm, for example weights, must necessarily appear in the reproducing kernel.

The following lemma shows us in fact that we can express any bounded linear functional in an alternative form in terms of the reproducing kernel. This is an important

aspect of the theory, as this means both integration and numerical quadrature can be expressed in this form, leading to a representation of quadrature error in terms of the kernel.

Lemma 7 *For any bounded linear functional on a reproducing kernel Hilbert space* $T : \mathcal{H} \to \mathbb{R}$*, we have for every* $f \in \mathcal{H}$

$$T(f) \;=\; \langle f, g \rangle \,,$$

where $g(\boldsymbol{x}) = T(K(\cdot, \boldsymbol{x}))$*. We call* $T(K(\cdot, \boldsymbol{x}))$ *the* representer *of* T*.*

Proof. From the Riesz representation theorem we know that there exists a unique $t \in \mathcal{H}$ such that $T(f) = \langle f, t \rangle$ for all $f \in \mathcal{H}$. Using this and the reproducing property we also have

$$t(\boldsymbol{y}) = \langle t, K(\cdot, \boldsymbol{y}) \rangle \;=\; \langle K(\cdot, \boldsymbol{y}), t \rangle \;=\; T(K(\cdot, \boldsymbol{y})) \;=\; T(K(\boldsymbol{y}, \cdot)) \,.$$

So we see that we have

$$T\left(\langle f, K(\cdot, \boldsymbol{y}) \rangle\right) \;=\; T(f) \;=\; \langle f, t \rangle \;=\; \langle f, T(K(\cdot, \boldsymbol{y})) \rangle \;.$$

□

2.3.1 QMC in reproducing kernel Hilbert spaces

We now investigate the integration and n-point quadrature operators and apply Lemma 7 to find a representer for them in terms of the reproducing kernel in our Hilbert space $\mathcal{H}$. We shall see that this enables us to express $e_{s,n}(\mathcal{P}, \mathcal{H})$, defined (2.1), in closed form in terms of the kernel K.

First we investigate integration, which we assume to be a bounded linear operator, that is, we assume $\mathcal{H}$ to be embedded in L_1. From the reproducing property and Lemma 7, we have that

$$I_s(f) = \int_{[0,1]^s} f(\boldsymbol{y}) \,\mathrm{d}\boldsymbol{y} = \left\langle f, \int_{[0,1]^s} K(\cdot, \boldsymbol{y}) \,\mathrm{d}\boldsymbol{y} \right\rangle_{\mathcal{H}} .$$

Furthermore, we can define the *initial error* of integration and use the above to derive the following,

$$e_{s,0}(0, \mathcal{H}) \;=\; \|I\|_{\mathcal{H}} \;=\; \sup_{\|f\| \le 1} \left\langle f, \int_{[0,1]^s} K(\cdot, \boldsymbol{y}) \,\mathrm{d}\boldsymbol{y} \right\rangle_{\mathcal{H}} .$$

We can derive a similar expression for the n-point quadrature,

$$Q_n(\mathcal{P};\, f) \;=\; \frac{1}{n} \sum_{i=1}^{n} f(\boldsymbol{t}^{(i)}) \;=\; \frac{1}{n} \sum_{i=1}^{n} \left\langle f, K(\cdot, \boldsymbol{t}^{(i)}) \right\rangle_{\mathcal{H}} \;=\; \left\langle f, \frac{1}{n} \sum_{i=1}^{n} K(\cdot, \boldsymbol{t}^{(i)}) \right\rangle_{\mathcal{H}} .$$

Finally we can use these representations to derive the following for the quadrature error,

$$\begin{aligned}|I_s(f) - Q_n(\mathcal{P};\, f)| \;&=\; \left|\left\langle f, \int_D K(\cdot, \boldsymbol{y})\,\mathrm{d}\boldsymbol{y}\right\rangle_{\mathcal{H}} - \left\langle f, \frac{1}{n}\sum_{i=1}^{n} K(\cdot, \boldsymbol{t}^{(i)})\right\rangle_{\mathcal{H}}\right| \\ &= \left|\left\langle f, \int_D K(\cdot, \boldsymbol{y})\,\mathrm{d}\boldsymbol{y} - \frac{1}{n}\sum_{i=1}^{n} K(\cdot, \boldsymbol{t}^{(i)})\right\rangle_{\mathcal{H}}\right| \\ &= \left|\langle f, h\rangle_{\mathcal{H}}\right|, \end{aligned} \tag{2.15}$$

where we have introduced the *representer of the quadrature error*

$$h(\boldsymbol{y}) \;:=\; \int_D K(\boldsymbol{y}, \boldsymbol{x})\,\mathrm{d}\boldsymbol{x} - \frac{1}{n}\sum_{i=1}^{n} K(\boldsymbol{y}, \boldsymbol{t}^{(i)}). \tag{2.16}$$

We can also write $g(\boldsymbol{y}) := \int_D K(\boldsymbol{y}, \boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$ for the representer of integration. From the Cauchy-Schwarz inequality we get a bound on the error

$$|I_s(f) - Q_n(\mathcal{P};\, f)| = |\langle f, h\rangle_{\mathcal{H}}| \le \|f\|_{\mathcal{H}}\, \|h\|_{\mathcal{H}}. \tag{2.17}$$

We can now revisit the worst-case error, (2.1). From (2.17) we have

$$e_{s,n}(\mathcal{P}, \mathcal{H}) \;:=\; \sup_{f\in\mathcal{H},\, \|f\|_{\mathcal{H}}\le 1} |I_s(f) - Q_n(\mathcal{P};\, f)| \;=\; \sup_{f\in\mathcal{H},\, \|f\|_{\mathcal{H}}\le 1} |\langle f, h\rangle_{\mathcal{H}}| \;=\; \|h\|_{\mathcal{H}},$$

as clearly the supremum is obtained when $f = h/\|h\|_{\mathcal{H}}$ (ensuring $\|f\|_{\mathcal{H}} = 1$). If we take the square of $e_{s,n}(\mathcal{P}, \mathcal{H})$ and then expand using the definition of h,

$$\begin{aligned} e^2_{s,n}(\mathcal{P}, \mathcal{H}) = \langle h, h\rangle_{\mathcal{H}} &= \left\langle \int_D K(\cdot, \boldsymbol{x})\,\mathrm{d}\boldsymbol{x} - \frac{1}{n}\sum_{i=1}^{n} K(\cdot, \boldsymbol{t}^{(i)}),\, \int_D K(\cdot, \boldsymbol{y})\,\mathrm{d}\boldsymbol{y} - \frac{1}{n}\sum_{j=1}^{n} K(\cdot, \boldsymbol{t}^{(j)})\right\rangle_{\mathcal{H}} \\ &= \int_D\int_D K(\boldsymbol{x}, \boldsymbol{y})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y} - \frac{2}{n}\sum_{i=1}^{n}\int_D K(\boldsymbol{x}, \boldsymbol{t}^{(i)})\,\mathrm{d}\boldsymbol{x} + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} K(\boldsymbol{t}^{(i)}, \boldsymbol{t}^{(j)}). \end{aligned} \tag{2.18}$$

The last line of the above equation is derived from the linearity of the inner product, along with Lemma 7 and the reproducing property. Similar steps can be used to show that we have the following expression for the initial error,

$$\begin{aligned} e^2_{s,0}(0, \mathcal{H}) \;=\; \|I\|^2_{\mathcal{H}} \;&=\; \left\langle \int_D K(\cdot, \boldsymbol{x})\,\mathrm{d}\boldsymbol{x},\, \int_D K(\cdot, \boldsymbol{y})\,\mathrm{d}\boldsymbol{y}\right\rangle_{\mathcal{H}} \\ &= \int_D\int_D K(\boldsymbol{x}, \boldsymbol{y})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y} \end{aligned} \tag{2.19}$$

Note that the results in this section, particularly (2.18), are independent of the type of quadrature points used in the QMC quadrature, and applies to any deterministic point set. For some cases of point sets $\mathcal{P}$, most notably lattice rules, and for particular kernels, (2.18)

simplifies considerably and can be dealt with in closed form, whereas notions from previous sections such as discrepancy $D_q(\mathcal{P})$, for evaluating Koksma-Hlawka style inequalities, have no simple closed form simplifications.

2.3.2 Korobov spaces

Our first example of a useful Hilbert space is the Korobov space $\mathcal{K}_\alpha$ of functions on the unit interval with absolutely convergent Fourier series. Here we present the details of the space, but we will not examine the reproducing kernel, which we save later for the discussion of weighted spaces. For $\alpha \geq 0$ we define the inner product

$$\langle f, g\rangle_{\mathcal{K}_\alpha} \;=\; \hat{f}(0)\,\overline{\hat{g}(0)} + \sum_{h\in\mathbb{Z}\setminus\{0\}} \hat{f}(h)\,\overline{\hat{g}(h)}\,h^\alpha,$$

where again $\hat{f}(h)$ represents the h-th Fourier component of f. As usual the norm is given by $\|f\|_{\mathcal{K}_\alpha} = \langle f, f\rangle_{\mathcal{K}_\alpha}^{1/2}$.

Heuristically speaking, this norm measures the smoothness of f as functions with small Fourier components for high h will tend to be smoother. Furthermore, for the norm of f to remain finite, we evidently require greater than $\mathcal{O}(h^{-\alpha})$ convergence of $\hat{f}(h)$. Also note that for some $\alpha \leq \beta$, if we have $f \in \mathcal{K}_\beta$, then $\|f\|_{\mathcal{K}_\alpha} \leq \|f\|_{\mathcal{K}_\beta}$.

Now for the higher dimensional generalisation. We define the Korobov space of s-dimensional functions with the inner product

$$\langle f, g\rangle_{\mathcal{K}_{\alpha,s}} \;=\; \hat{f}(\mathbf{0})\,\overline{\hat{g}(\mathbf{0})} \;+\; \sum_{\emptyset\neq\mathfrak{u}\subseteq\{1:s\}} \sum_{\boldsymbol{h}_\mathfrak{u}\in(\mathbb{Z}\setminus\{0\})^{|\mathfrak{u}|}} \hat{f}(\boldsymbol{h}_\mathfrak{u};\mathbf{0})\,\overline{\hat{g}(\boldsymbol{h}_\mathfrak{u};\mathbf{0})} \prod_{j\in\mathfrak{u}} |h_j|^\alpha\,. \tag{2.20}$$

where we have used notation similar to (2.7), that is, $(\boldsymbol{y}_\mathfrak{u};\mathbf{0}) \in [0,1]^s$ is the vector whose i-th components are x_i if $i \in \mathfrak{u}$, and 0 if $i \notin \mathfrak{u}$.

2.3.3 Sobolev spaces

We explore Sobolev space of functions on $[0,1]$ with square-integrable mixed first derivatives. In the literature of QMC we consider two main flavours of Sobolev space, *anchored* and *unanchored*, which are well known to be Hilbert spaces. Consider first the anchored Sobolev space $\mathcal{H}_c$ of absolutely continuous functions with square-integrable first derivatives, with inner product

$$\langle f, g\rangle_{\mathcal{H}_c} = f(c)\,g(c) + \int_0^1 f'(y)\,g'(y)\,\mathrm{d}y\,,$$

where $c \in [0,1]$ is our *anchor*. We require the evaluation at the anchor to make the quantity $\langle f, f\rangle_{\mathcal{H}_c}^{1/2}$ a norm, rather than a mere semi-norm. For the unanchored space $\mathcal{H}$ of absolutely continuous functions, we define the inner product as follows.

$$\langle f, g\rangle_{\mathcal{H}} = \int_0^1 f(y)\,\mathrm{d}y \int_0^1 g(y)\,\mathrm{d}y + \int_0^1 f'(y)\,g'(y)\,\mathrm{d}y\,.$$

We can generalise the anchored space to s-dimensions as follows.

$$\langle f, g\rangle_{\mathcal{H}_{c,s}} = \sum_{\mathfrak{u}\subseteq\{1:s\}} \int_{[0,1]^s} \frac{\partial^{|\mathfrak{u}|} f}{\partial \boldsymbol{y}_{\mathfrak{u}}}(\boldsymbol{y}_{\mathfrak{u}}; \boldsymbol{c}_{-\mathfrak{u}}) \frac{\partial^{|\mathfrak{u}|} g}{\partial \boldsymbol{y}_{\mathfrak{u}}}(\boldsymbol{y}_{\mathfrak{u}}; \boldsymbol{c}_{-\mathfrak{u}})\, \mathrm{d}\boldsymbol{y}_{\mathfrak{u}}\,, \tag{2.21}$$

where the vector $\boldsymbol{c} \in [0,1]^s$ is again called the anchor, and similarly to the Korobov space we have written $(\boldsymbol{y}_{\mathfrak{u}}; \boldsymbol{c}_{-\mathfrak{u}})$ to represent the vector in $[0,1]^s$ with components equal to y_j if $j \in \mathfrak{u}$ and the remaining components set to c_j if $j \notin \mathfrak{u}$, that is, we write $-\mathfrak{u}$ to represent the complement of $\mathfrak{u}$, i.e. $-\mathfrak{u} = \{1:s\} \setminus \mathfrak{u}$. Common choices for $\boldsymbol{c}$ include $(0,\dots,0)$, $(1/2,\dots,1/2)$ and $(1,\dots,1)$. We do not present the s-dimensional generalisation of the unanchored space as for the rest of this chapter we only present results in the anchored space.

2.4 Weighted spaces

It can be observed in many practical high-dimensional problems that some coordinates make the more important contribution towards a function in some way. In terms of the integration problem, we may say that these coordinates are more "important" or "difficult" than others. Weights were introduced in [62] to address exactly this.

We make this more precise. Consider the s-dimensional generalisations of the Korobov and Sobolev spaces, (2.20) and (2.21) respectively. We found their norms both to be of the form

$$\|f\|^2 = \sum_{\mathfrak{u}\subseteq\{1:s\}} \|f\|_{\mathfrak{u}}^2\,,$$

where, as an example, in the Sobolev space we have

$$\|f\|_{\mathfrak{u}}^2 = \int_{[0,1]^s} \left(\frac{\partial^{|\mathfrak{u}|} f}{\partial \boldsymbol{y}_{\mathfrak{u}}}(\boldsymbol{y}_{\mathfrak{u}}; \boldsymbol{c}_{-\mathfrak{u}})\right)^2 \mathrm{d}\boldsymbol{y}_{\mathfrak{u}}\,.$$

Some sets $\mathfrak{u}$ may make most of the contribution to this sum. In practical applications this is often observed to be the case for $\mathfrak{u}$ that include earlier rather than later coordinates, as well as the $\mathfrak{u}$ with smaller rather than larger cardinalities.

Now we consider a collection positive numbers $\boldsymbol{\gamma} = \{\gamma_{s,\mathfrak{u}} : \mathfrak{u} \in \{1:s\}\}$ which we call the weights. Now we can make a *weighted norm*

$$\|f\|_{\boldsymbol{\gamma}}^2 = \sum_{\mathfrak{u}\subseteq\{1:s\}} \gamma_{s,\mathfrak{u}}^{-1}\, \|f\|_{\mathfrak{u}}^2\,,$$

We observe that the weighted norm modifies the unit ball in the sense that if $\gamma_{s,\mathfrak{u}}$ is small, then $\|f\|_{\mathfrak{u}}$ is forced to be small if f is to remain in the unit ball. Thus the weights are in proportion to the measurement of "difficulty" that we allow in each coordinate collection. In this setting we adopt the following convention: if $\gamma_{s,\mathfrak{u}} = 0$, then also if $\|f\|_{\mathfrak{u}} = 0$ we say that $\gamma_{s,\mathfrak{u}}^{-1}\|f\|_{\mathfrak{u}} = 0$. That is, we adopt the convention that $0/0 = 0$ in the case of the weighted norm. For the Sobolev space example, in the extreme case where $\gamma_{s,\mathfrak{u}} = 0$,

quadrature in the $\mathfrak{u}$ coordinates requires at most one point as f will have to be constant in these coordinates.

These weights may come in many different flavours. We present a few examples here.

- Weights were originally introduced in *product* form, that is for a sequence $\gamma_1, \gamma_2, \ldots, \gamma_s$ of positive numbers, then for $\mathfrak{u} \subseteq \{1:s\}$ the weights can be written as

$$\gamma_{s,\mathfrak{u}} = \gamma_{\mathfrak{u}} = \prod_{j \in \mathfrak{u}} \gamma_j \,. \tag{2.22}$$

  In writing $\gamma_{s,\mathfrak{u}} = \gamma_{\mathfrak{u}}$, we have emphasised that there is no inherent dependence on the dimension of the problem s.
- The weights are said to be *order-dependent* if we have

$$\gamma_{s,\mathfrak{u}} = \Gamma_{s,|\mathfrak{u}|}$$

  where $\Gamma_{s,0}, \Gamma_{s,1}, \ldots, \Gamma_{s,s}$ are a set of non-negative numbers, i.e. there is a dependency on cardinality of $\mathfrak{u}$, but not to what $\mathfrak{u}$ may contain.
- We say the weights are *finite-order* if for some integer q we have

$$\gamma_{s,\mathfrak{u}} = 0 \qquad \text{for all } s \text{ and } \mathfrak{u} \text{ with } |\mathfrak{u}| > q.$$

  We say the finite-order weights are of order q^*, if q^* is the smallest possible integer that satisfies the property above.
- If we have two sets of non-negative $\gamma_1, \ldots, \gamma_s$ and $\Gamma_1, \ldots, \Gamma_s$, then we say the weights are of *product and order dependent (POD)* type if we can write

$$\gamma_{s,\mathfrak{u}} = \gamma_{\mathfrak{u}} = \Gamma_{|\mathfrak{u}|} \prod_{j \in \mathfrak{u}} \gamma_j \tag{2.23}$$

  This "hybrid" weight, relatively new in the literature, will be important in chapters to follow.

When the weights are assumed to have no specific form, we say that they are *general weights*, a term that is usually used to distance ourselves from the product weights setting, which is the most common setting in the literature. While weights were originally introduced in product form, general weights were introduced in [70].

Weighted spaces were originally introduced in [62] to demonstrate the conditions under which multivariate integration is strongly tractable. That is, to demonstrate the conditions under which there exists a QMC rule for which the error bound can be found to be independent of the dimensionality s of the integrand. These results were considered in the setting of *tensor product* Hilbert spaces, that is spaces where the weights will necessarily have to take the product form (2.22). In this setting it was shown in [62] that

the multivariate integration is strongly tractable in weighted Korobov or Sobolev spaces if and only if

$$\sum_{j=1}^{\infty} \gamma_j \, < \, \infty \, .$$

Weights will prove to be useful not just in infinite dimensional integration or tractability results. As the weights are in the norm, they must also make an appearance in the reproducing kernel. Thus we find that the weights appear in the worst-case error, as a consequence of (2.18). Hence we find that the weights to be a useful parameter in minimising the upper bound on quadrature error as expressed in (2.17), as both the worst-case error and the norm depend on the weights. We note that this is a relatively new approach to the use of weighted spaces, it was first discussed in [41] and [19]. Classical literature tended to take the point of view that the weights were given, rather than being parameters that can be chosen to best fit a specific problem.

2.5 Weighted Korobov spaces

Here we consider the weighted Korobov spaces, that is we make a weighted generalisation of (2.20). We shall present formulae for the kernel K in this space, and with lattice rules as our points sets, we will be able to develop (2.18) and (2.17) further to present convergence results. In this approach we skip the usual tensor-product construction of the space $\mathcal{K}_{s,\alpha,\boldsymbol{\gamma}}$, as is often done in the literature (e.g. [63]), and define our s-dimensional function spaces directly. This approach follows the presentation in [20] and [18].

First we introduce the following useful identity.

Identity 8 *For a sequence of numbers a_j,*

$$\prod_{j=1}^{s}(1+a_j) \, = \, \sum_{\mathfrak{u}\subseteq\{1:s\}} \prod_{j\in\mathfrak{u}} a_j \, = \, 1 + \sum_{\emptyset\neq\mathfrak{u}\subseteq\{1:s\}} \prod_{j\in\mathfrak{u}} a_j \, ,$$

noting the implicit assumption that we take the summand for $\mathfrak{u} = \emptyset$ to be $\prod_{j\in\emptyset} a_j = 1$.

Now for the higher dimensional generalisation we consider a set of weights $\gamma_{s,\mathfrak{u}} \geq 0$. We define the weighted Korobov space of functions on $[0,1]^s$ with absolutely convergent Fourier series, with the inner product

$$\langle f, g\rangle_{\mathcal{K}_{s,\alpha,\boldsymbol{\gamma}}} \, = \, \hat{f}(\mathbf{0})\,\overline{\hat{g}(\mathbf{0})} \, + \sum_{\emptyset\neq\mathfrak{u}\subseteq\{1:s\}} \gamma_{s,\mathfrak{u}}^{-1} \sum_{\boldsymbol{h}_\mathfrak{u}\in(\mathbb{Z}\setminus\{0\})^{|\mathfrak{u}|}} \hat{f}(\boldsymbol{h}_\mathfrak{u};\mathbf{0})\,\overline{\hat{g}(\boldsymbol{h}_\mathfrak{u};\mathbf{0})} \prod_{j\in\mathfrak{u}} |h_j|^\alpha \, ,$$

where we have written $(\boldsymbol{h}_\mathfrak{u};\mathbf{0})$ to express the vector in $\mathbb{Z}^s$ for which the components are h_j if $j \in \mathfrak{u}$ and the remaining components are set to 0.

In this setting the kernel is given by

$$K_{s,\alpha,\boldsymbol{\gamma}} \, = \, \sum_{\mathfrak{u}\subseteq\{1:s\}} \gamma_{s,\mathfrak{u}}\, K_{\mathfrak{u},\alpha}(\boldsymbol{x}_\mathfrak{u}, \boldsymbol{y}_\mathfrak{u}) \, , \tag{2.24}$$

where

$$K_{\mathfrak{u},\alpha}(\boldsymbol{x}_{\mathfrak{u}}, \boldsymbol{y}_{\mathfrak{u}}) = \prod_{j\in\mathfrak{u}} \sum_{h\in\mathbb{Z}\setminus\{0\}} \frac{\exp(2\pi i h(x_j - y_j))}{|h|^\alpha}.$$

Note that, in line with conventions taken thus far, the summand for $\mathfrak{u} = \emptyset$ is equal to 1, that is we take $\gamma_{\emptyset,s} K_{\emptyset,\alpha} = 1$.

Now we can proceed with directly applying our formula for the worst-case error in a reproducing kernel Hilbert space for a rule $\mathcal{P}$, (2.18), to this example,

$$e^2_{s,n}(\mathcal{P}, \mathcal{K}_{s,\alpha,\boldsymbol{\gamma}}) = \frac{1}{n^2} \sum_{k=1}^{n} \sum_{\ell=1}^{n} \sum_{\emptyset\neq\mathfrak{u}\subseteq\{1:s\}} \gamma_{s,\mathfrak{u}} \prod_{j\in\mathfrak{u}} \sum_{h\in\mathbb{Z}\setminus\{0\}} \frac{\exp(2\pi i h(t_j^{(k)} - t_j^{(\ell)}))}{|h|^\alpha}.$$

If we take our point set to be a lattice rule, as defined in (2.10), we note that the shift makes no difference in the expression above, and that this expression simplifies to

$$e^2_{s,n}(\boldsymbol{z}, \mathcal{K}_{s,\alpha,\boldsymbol{\gamma}}) = \frac{1}{n} \sum_{k=1}^{n} \sum_{\emptyset\neq\mathfrak{u}\subseteq\{1:s\}} \gamma_{s,\mathfrak{u}} \prod_{j\in\mathfrak{u}} \sum_{h\in\mathbb{Z}\setminus\{0\}} \frac{\exp(2\pi i h k z_j/n)}{|h|^\alpha}. \tag{2.25}$$

We would like to find convenient bounds of $e_{s,n}(\boldsymbol{z}, \mathcal{K}_{s,\alpha,\boldsymbol{\gamma}})$ the form $\mathcal{O}(n^{-r})$ for some $r > 0$ for a "good" lattice rule, which we can then substitute in to (2.2).

We will show the existence of a good lattice rule using an averaging argument. Assume for the rest of this section that n is a prime number. We write the mean of the squared worst-case error for all possible generating vectors $\boldsymbol{z} \in \mathcal{Z}_n^s$,

$$M_{s,n}(\alpha) := \frac{1}{(n-1)^s} \sum_{\boldsymbol{z}\in\mathcal{Z}_n^s} e^2_{s,n}(\boldsymbol{z}). \tag{2.26}$$

Now, let $\zeta(x) := \sum_{h=1}^{\infty} h^{-x}$ for $x > 1$ denote the Riemann zeta function. Note that if n is prime, the set $\mathcal{Z}_n$ contains all numbers from 1 to $n-1$ inclusive, thus contains $n-1$ elements. As it turns out, it is possible to obtain an explicit formula for this quantity. We present this in the following.

Theorem 9 *If n is prime and $\alpha > 1$, with $M_{s,n}(\alpha)$ defined as in (2.26), we have*

$$M_{s,n}(\alpha) \leq \frac{1}{n} \sum_{\emptyset\neq\mathfrak{u}\subseteq\{1:s\}} \gamma_{s,\mathfrak{u}} \, (2\zeta(\alpha))^{|\mathfrak{u}|}, \tag{2.27}$$

thus there exists a generating vector $\boldsymbol{z}_* \in \mathcal{Z}_n^s$ such that

$$e_{s,n}(\boldsymbol{z}_*, \mathcal{K}_{s,\alpha,\boldsymbol{\gamma}}) \leq \frac{1}{\sqrt{n}} \left(\sum_{\emptyset\neq\mathfrak{u}\subseteq\{1:s\}} \gamma_{s,\mathfrak{u}} (2\zeta(\alpha))^{|\mathfrak{u}|} \right)^{1/2} \tag{2.28}$$

Proof. The proof of (2.27) follows similarly to [63, Theorem 3] upon setting $\beta_j = 1$ and using Identity 8. Also we refer the reader to the proof of [20, Theorem 1], which takes

in to consideration the same proof for general weights. For the second part we use the principle that there is always at least one choice that is as good as or better than average. Thus there must be a generating vector $\boldsymbol{z}$ that beats the mean, and thus also beat the estimate (2.27). □

Theorem 9 presents an upper bound for the worst-case error of a lattice rule only in terms of the number of points n and the weights $\gamma_{s,\mathfrak{u}}$. As it turns out we can demonstrate much better convergence, but first we recall a variant of Jensen's inequality. If we have some non-negative sequence a_j then

$$\sum_j a_j \;\le\; \left(\sum_j a_j^r\right)^{1/r} \qquad \text{for any } 0 < r \le 1. \tag{2.29}$$

Theorem 10 *If n is prime and $\alpha > 1$ then there exists a generating vector $\boldsymbol{z}_* \in \mathcal{Z}_n^s$ such that for all $\lambda \in (1/\alpha, 1]$ we have*

$$e_{s,n}(\boldsymbol{z}_*, \mathcal{K}_{s,\alpha,\boldsymbol{\gamma}}) \;\le\; \frac{1}{n^{1/(2\lambda)}} \left(\sum_{\emptyset \ne \mathfrak{u} \subseteq \{1:s\}} \gamma_{s,\mathfrak{u}}^{\lambda}\, (2\zeta(\alpha\lambda))^{|\mathfrak{u}|}\right)^{1/(2\lambda)} \tag{2.30}$$

Proof. We refer the reader to [33] for the proof. □

Thus we have a result that bounds the error for arbitrarily large s. It is not surprising that the rate of convergence of $e_{s,n}$, with respect to n, is directly linked to the parameter α, the larger α, the smaller we can take λ, and hence the faster our convergence rate. Recall that α characterises the "smoothness" of the integrands f that are allowable in the space $\mathcal{K}_{s,\alpha,\boldsymbol{\gamma}}$, it seems to makes sense that quadrature error might converge faster for a smooth function.

One problem of this result is that we have not provided a proof that is constructive. There is no insight in *how* to find the generating vector, we have merely stated that there is some vector, out of all the possibilities in $\mathcal{Z}_n^s$, that satisfies the bound. Shortly we shall be investigating a method of construction of good generating vectors $\boldsymbol{z}_*$, and then prove that vectors constructed this way satisfy similar bounds.

2.6 Weighted Sobolev spaces

Now we turn our attention to weighted Sobolev spaces. We define the *weighted anchored Sobolev space* $\mathcal{H}_{s,\boldsymbol{c},\boldsymbol{\gamma}}$ to be the s-dimensional tensor product space of the 1-dimensional spaces $\mathcal{H}_c$, but we define the inner product as

$$\langle f, g\rangle_{\mathcal{H}_{s,\boldsymbol{c},\boldsymbol{\gamma}}} \;=\; \sum_{\mathfrak{u} \subseteq \{1:s\}} \gamma_{s,\mathfrak{u}}^{-1} \int_{[0,1]^s} \frac{\partial^{|\mathfrak{u}|} f}{\partial \boldsymbol{y}_{\mathfrak{u}}}(\boldsymbol{y}_{\mathfrak{u}}; \boldsymbol{c}_{-\mathfrak{u}}) \frac{\partial^{|\mathfrak{u}|} g}{\partial \boldsymbol{y}_{\mathfrak{u}}}(\boldsymbol{y}_{\mathfrak{u}}; \boldsymbol{c}_{-\mathfrak{u}})\, \mathrm{d}\boldsymbol{y}_{\mathfrak{u}}\,, \tag{2.31}$$

where we have used the same notation in §2.3.3. These weighted Sobolev spaces have been considered in [28], [29] and [59].

In this setting the reproducing kernel is

$$K_{\boldsymbol{c},\boldsymbol{\gamma}}(\boldsymbol{x},\boldsymbol{y}) \;=\; \sum_{\mathfrak{u}\subseteq\{1:s\}} \gamma_{s,\mathfrak{u}} \prod_{j\in\mathfrak{u}} \mu_{c_j}(x_j,y_j)\,,$$

where

$$\mu_c(x,y) \;=\; \begin{cases} \min(|x-c|,|y-c|) & \text{if } x,\, y \ge c \text{ or } x,\, y < c \\ 0 & \text{otherwise.} \end{cases}$$

A quantity that will come up shortly in this setting is the initial error, $e_{s,0}(0,\mathcal{H}_{c,\boldsymbol{\gamma}})$, which from (2.19) is given by

$$\begin{aligned} e^2_{s,0}(0,\mathcal{H}_{s,\boldsymbol{c},\boldsymbol{\gamma}}) \;&=\; \sum_{\mathfrak{u}\subseteq\{1:s\}} \gamma_{s,\mathfrak{u}} \prod_{j\in\mathfrak{u}} \int_{[0,1]^2} \mu_{c_j}(x_j,y_j)\,\mathrm{d}x_j\,\mathrm{d}y_j \\ &=\; \sum_{\mathfrak{u}\subseteq\{1:s\}} \gamma_{s,\mathfrak{u}} \prod_{j\in\mathfrak{u}} (c_j^2 - c_j + 1/3)\,. \end{aligned} \tag{2.32}$$

From here we could proceed as in §2.5, and derive an expression for the worst-case error from (2.18) for this setting, however this does not prove to be fruitful. As it turns out, the way to proceed lies in the random shift, $\boldsymbol{\Delta}$. First let us define the *shift invariant kernel* K^{sh}, associated with any kernel K by

$$K^{\mathrm{sh}}(\boldsymbol{x},\boldsymbol{y}) \;=\; \int_{[0,1]^s} K(\{\boldsymbol{x}+\boldsymbol{\Delta}\},\{\boldsymbol{y}+\boldsymbol{\Delta}\})\,\mathrm{d}\boldsymbol{\Delta}\,. \tag{2.33}$$

Now let us express, for a moment, the worst case error as a function of associated kernel, that is we write $e_{s,n}(\boldsymbol{z},K)$ for the quantity given in (2.18). We consider mean-square over all possible random shifts of the worst-case error for a shifted lattice rule, given that the $\boldsymbol{\Delta}$ is uniformly distributed on $[0,1]^s$. It is shown in [29] that the *shift averaged worst-case error* is equal to the worst-case error with the shift invariant kernel, that is,

$$[e^{\mathrm{sh}}_{s,n}(\boldsymbol{z},K)]^2 \;:=\; \mathbb{E}^{\boldsymbol{\Delta}}[e^2_{s,n}(\boldsymbol{z},\boldsymbol{\Delta},K)] \;=\; \int_{[0,1]^s} e^2_{s,n}(\boldsymbol{z},\boldsymbol{\Delta},K)\,\mathrm{d}\boldsymbol{\Delta} \;=\; e^2_{s,n}(\boldsymbol{z},K^{\mathrm{sh}}),$$

For the Sobolev space we also write $e^{\mathrm{sh}}_{s,n}(\boldsymbol{z},\mathcal{H}_{s,\boldsymbol{c},\boldsymbol{\gamma}}) = e^{\mathrm{sh}}_{s,n}(\boldsymbol{z},K_{\boldsymbol{c},\boldsymbol{\gamma}})$, where K is assumed to be the appropriate kernel. We consider the shift invariant kernel for the weighted Sobolev space, which we can compute to be

$$K^{\mathrm{sh}}_{\boldsymbol{c},\boldsymbol{\gamma}}(\boldsymbol{x},\boldsymbol{y}) \;=\; \sum_{\mathfrak{u}\subseteq\{1:s\}} \gamma_{s,\mathfrak{u}} \prod_{j\in\mathfrak{u}} (B_2(|x_j-y_j|)+m_j)$$

where $B_2(x) = x^2 - x + 1/6$ is the Bernoulli polynomial of degree 2 and $m_j = c_j^2 - c_j + 1/3$. We can also compute an expression for the shift averaged worst-case error,

$$[e^{\mathrm{sh}}_{s,n}(\boldsymbol{z},\mathcal{H}_{s,\boldsymbol{c},\boldsymbol{\gamma}})]^2 \;=\; \sum_{\emptyset\ne\mathfrak{u}\subseteq\{1:s\}} \gamma_{s,\mathfrak{u}} \left(\frac{1}{n} \sum_{k=0}^{n-1} \prod_{j\in\mathfrak{u}} \left(B_2\left(\left\{ \frac{k z_j}{n} \right\} \right) + m_j \right) - \prod_{j\in\mathfrak{u}} m_j \right)\,. \tag{2.34}$$

Now, for simplicity of further exposition, we assume that our weights are of *product type*, that is $\gamma_{s,\mathfrak{u}} = \prod_{j\in\mathfrak{u}} \gamma_j$. Under this assumption, and using Identity 8, we can re-write the shift invariant kernel as

$$\begin{aligned} K^{\mathrm{sh}}_{\boldsymbol{c},\boldsymbol{\gamma}}(\boldsymbol{x},\boldsymbol{y}) &= \prod_{j=1}^{s} (1 + \gamma_j (B_2(|x_j - y_j|) + m_j)) \\ &= e^2_{s,0}(0,\mathcal{H}_{s,\boldsymbol{c},\boldsymbol{\gamma}}) \prod_{j=1}^{s} \left(1 + \hat{\gamma}_j \sum_{h\in\mathbb{Z}\setminus\{0\}} \frac{\exp(2\pi i(x_j - y_j))}{h^2}\right) \\ &= e^2_{s,0}(0,\mathcal{H}_{s,\boldsymbol{c},\boldsymbol{\gamma}}) \sum_{\mathfrak{u}\subseteq\{1:s\}} \prod_{j\in\mathfrak{u}} \hat{\gamma}_j \sum_{h\in\mathbb{Z}\setminus\{0\}} \frac{\exp(2\pi i(x_j - y_j))}{h^2}, \end{aligned}$$

where we have used the well known Fourier series expansion of B_2,

$$B_2(x) = \frac{1}{2\pi^2} \sum_{h\in\mathbb{Z}\setminus\{0\}} \frac{\exp(2\pi i x)}{h^2}.$$

as well as (2.32) and the substitution

$$\hat{\gamma}_j = \frac{\gamma_j}{2\pi^2(1+\gamma_j m_j)}.$$

Thus we see that $K^{\mathrm{sh}}_{\boldsymbol{c},\boldsymbol{\gamma}}$ is same (ignoring the $\|I\|_{\mathcal{H}_{\boldsymbol{c},\boldsymbol{\gamma}}}$ factor in front) as the kernel $K_{\alpha,\boldsymbol{\gamma}}$ from the Korobov space in (2.24) with $\alpha = 2$, thus we can apply Theorem 10 to $e_{s,n}(\boldsymbol{z}, K^{\mathrm{sh}}_{\boldsymbol{c},\boldsymbol{\gamma}})$. Thus in the weighted anchored Sobolev space we obtain the following,

Theorem 11 *Let n be prime and assume the weights $\gamma_{s,\mathfrak{u}}$ are of product form, then there exists a generating vector $\boldsymbol{z}_0 \in \mathcal{Z}_n^s$ such that for any $1/2 > \delta > 0$ we have*

$$e^{\mathrm{sh}}_{s,n}(\boldsymbol{z}_0,\mathcal{H}_{\boldsymbol{c},\boldsymbol{\gamma}}) \le \frac{\|I\|_{\mathcal{H}_{\boldsymbol{c},\boldsymbol{\gamma}}}}{n^{1-\delta}} \left(\prod_{j=1}^{s} \left(1 + 2\hat{\gamma}_j^{1/(2-2\delta)} \zeta\left(\frac{1}{1-\delta}\right)\right)\right)^{1-\delta}. \tag{2.35}$$

Proof. We apply Theorem 10, noting that as $\alpha = 2$, we can take λ to be arbitrarily close to $1/2$, hence the result follows from the choice $\lambda = \frac{1}{2(1-\delta)}$ for some small $\delta > 0$. □

Once again the result is not constructive, it is merely an existence result. Also note that the result applies to the shift averaged worst-case error, not the original worst-case error of (2.1). This means two things: Firstly this implies the existence of some shift $\boldsymbol{\Delta}^*$ for which the bound holds, again using the principle that there is always a choice that is at least as good as the average. Secondly, in using shifted lattice rules on integrands in $\mathcal{H}_{\boldsymbol{c},\boldsymbol{\gamma}}$, we need to average the quadrature we obtain over a number of shifts (usually a reasonably small number can be used, of the order of 10) to be able to confidently apply this result. Often random shifts are employed with QMC rules anyhow, as the random shifts make $Q_n(\mathcal{P}; f)$ an unbiased estimator of $I_s(f)$, and furthermore allows us to calculate the standard error of our estimator. Hence, the fact that the theory depends

on the random shifts is not an inconvenience. We refer the reader to [58, Theorem 2.2] for a proof that the shifted lattice rule is an unbiased estimator.

We mention briefly that it is possible to allow the weights to be of general non-product type and still yield a result akin to that of Theorem 11, however we have not done so here for two reasons. Firstly, to do so is non-trivial and not particularly instructive. Secondly, we shall be presenting the results that allow for general weights in the unbounded setting, which involves use of these auxiliary weights, in Chapter 3. The general weights case in this Sobolev setting is examined in [37].

Finally we consider the implications of allowing the number of coordinates to grow, that is, letting $s \to \infty$. We see that, for $\lambda \in (1/2, 1]$, if we have

$$\sum_{j=1}^{\infty} \hat{\gamma}_j^{\lambda} \, < \, \infty,$$

then, as shown in [63], we can deduce that $\prod_{j=1}^{\infty} (1 + 2\hat{\gamma}_j^{\lambda} \zeta(2\lambda)) < \infty$, and hence we can bound $e^{\mathrm{sh}}_{s,n}(\boldsymbol{z}_0, \mathcal{H}_{\boldsymbol{c},\boldsymbol{\gamma}})$ independently of s. This implies strong tractability.

A variant of the Sobolev setting introduced in this section is the weighted *unanchored* Sobolev space, which is an s-dimensional generalisation of the unanchored space introduced in §2.3.3. For simplicity, and also as we will be introducing unanchored spaces in the unbounded setting in great detail for the next chapter, we opt not to consider the unanchored space here.

2.6.1 Relationships with discrepancy

We have seen in these last few sections a common thread with the material introduced at the start of the chapter, of discrepancies, and the associated Koksma-Hlawka type inequalities. Here we show a direct relationship in the Sobolev space with anchor 1. For simplicity we look only at the one-dimensional unweighted space, $\mathcal{H}_{1,1}$. In this setting we can write the kernel as

$$K_{1,1}(x, y) = 1 + \min(1 - x, 1 - y) = 1 + (1 - x)\mathbf{1}_x(y) + (1 - y)\mathbf{1}_y(x)\,. \tag{2.36}$$

Now, recalling (2.17), we have that

$$\begin{aligned} I_s(f) - Q_n(\mathcal{P}; f) \; &= \; \langle f, h \rangle_{\mathcal{H}_{1,1}} \\ &= \; f(1)h(1) + \int_0^1 f'(y)h'(y)\,\mathrm{d}y\,. \end{aligned} \tag{2.37}$$

From (2.36) we can show that $\partial K_{1,1}(x,y)/\partial y = \mathbf{1}_{[0,y)}(x)$, hence we can derive the following,

$$\begin{aligned}\frac{\mathrm{d}h(y)}{\mathrm{d}y} &= \int_0^1 \frac{\partial}{\partial y} K_{1,1}(x,y)\mathrm{d}x - \frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial y} K_{1,1}(t^{(i)},y) \\ &= -\int_0^1 \mathbf{1}_{[0,y)}(x)\mathrm{d}x + \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{[0,y)}(t^{(i)}) \\ &= -y + \frac{\#\{t^{(i)} : t^{(i)} \in [0,y)\}}{n} \;=\; \mathrm{discr}_{\mathcal{P}}(y).\end{aligned}$$

Evidently we can also have that $K_{1,1}(x,1) = K_{1,1}(1,x) = 1$, hence $h(1) = 0$. Thus from (2.37) we find that

$$I_s(f) - Q_n(\mathcal{P}; f) \;=\; \int_0^1 f'(y)\,\mathrm{discr}_{\mathcal{P}}(y)\,\mathrm{d}y\,,$$

which is precisely the Zaremba identity of (2.3). Furthermore we see that the worst-case error is in fact equal to the L_2-discrepancy of (2.6), which we can show as follows.

$$e_{1,n}^2(\mathcal{P},\mathcal{H}_1) \;=\; \langle h,h\rangle_{\mathcal{H}_{1,1}} \;=\; \int_0^1 |\mathrm{discr}_{\mathcal{P}}(y)|^2\;\mathrm{d}y \;=\; [D_2(\mathcal{P})]^2,$$

In fact this generalises to s dimensions, that is, $e_{s,n}(\mathcal{P},\mathcal{H}_{s,1}) = D_2(\mathcal{P})$.

2.7 Component-by-component construction

Sections 2.5 and 2.6 explore optimal bounds on worst-case errors for (shifted) lattice rules. As discussed, however, these results are not constructive, they present no scheme for finding good generating vectors that satisfy our bounds. A method of construction known as the *component-by-component algorithm* that dates back to the work of Korobov, see e.g. [31], was also further developed in [35, 58, 59, 60]. This method proposes a simple "greedy" approach where the values for each component of the generating vector are chosen separately to minimise the worst-case error.

Algorithm 12 (CBC Algorithm) *For any prime n and $s \in \mathbb{N}$*
1. Set $z_1 = 1$
2. For each $d = 2,3,\ldots,s$ with $z_1,\ldots,z_{d-1}$ fixed, choose $z_d \in \mathcal{Z}_n$ such that $e_{n,d}^{\mathrm{sh}}(\{z_1,\ldots,z_d\},\mathcal{H}_{\boldsymbol{c},\boldsymbol{\gamma}})$ is minimised.

The advantage of using the CBC algorithm is that it reduces the search space for a good vector quite considerably. A brute force approach to finding $\boldsymbol{z}$ that minimises (2.34) might consider all $(n-1)^s$ possibilities and evaluate the worst-case error at each. Evidently this is exponentially expensive to compute, even when considering a few possible reductions that different symmetries might afford us. However, the CBC method enables

us to consider n possibilities for each component, thus we only have to evaluate the worst-case error sn times.

It is not immegiately obvious that the vector $\boldsymbol{z}$ constructed in the CBC algorithm would satisfy the sorts of bounds in Theorems 10 or 11. However it was shown in [60] that lattice rules constructed this way have worst-case errors that converge as $\mathcal{O}(n^{-1/2})$ for the Korobov class of functions, and a similar result was subsequently shown in [59, 58] for weighted Sobolev spaces. In [59] the principle of the CBC algorithm is also applied to $\boldsymbol{\Delta}$, that is, Δ_d is chosen after z_d, with $\Delta_1, \ldots, \Delta_{d-1}$ and $z_1, \ldots, z_d$ fixed. It can be shown that this construction for $\boldsymbol{\Delta}$ and $\boldsymbol{z}$ leads to a good shifted lattice rule with $\mathcal{O}(n^{-1/2})$ error convergence.

However, these results do not match the optimal rates of convergence we have in Theorems 10 and 11, where, under the right conditions, $\mathcal{O}(n^{-1+\delta})$ convergence can be observed. While numerical experiments in [58] suggest that generating vectors do exhibit these improved orders of convergence, it was not until [33] and [16] that optimal convergence was proven for generating vectors produced by the CBC algorithm. We state the result here.

Theorem 13 *Let n be prime and assume the weights $\gamma_{s,\mathfrak{u}}$ are of product form, and let $\boldsymbol{z}_*$ be constructed using the component-by-component algorithm, then for any $1/2 > \delta > 0$ we have*

$$e^{\mathrm{sh}}_{s,n}(\boldsymbol{z}_*, \mathcal{H}_{\boldsymbol{c},\boldsymbol{\gamma}}) \leq \frac{\|I\|_{\mathcal{H}_{\boldsymbol{c},\boldsymbol{\gamma}}}}{n^{1-\delta}} \left(\prod_{j=1}^{s} \left(1 + 2\hat{\gamma}_j^{1/(2-2\delta)} \zeta\left(\frac{1}{1-\delta} \right) \right) \right)^{1-\delta}. \tag{2.38}$$

Proof. The proof uses an inductive argument, assuming that (2.38) holds for some $\boldsymbol{z} \in \mathcal{Z}_n^s$, then it can be shown that the natural extension of (2.38) holds for $(\boldsymbol{z}, z_{s+1})$ as well, where $\boldsymbol{z}$ is fixed and z_{s+1} is chosen by the minimisation criteria in step 2 of the algorithm. For details we refer to [33, Theorem 5 and 8], noting that the proof technique again makes use of the connection between Korobov and Sobolev spaces. We do not present the details here as a full proof of a similar result will be shown in the next chapter, where we will allow for further generality including unbounded integrands. □

Chapter 3

Unbounded functions with general weights

In this chapter we now consider integrals of functions that are defined on unbounded regions, for example $\mathbb{R}^s$, that is, integrals of the form

$$\int_{\mathbb{R}^s} f(\boldsymbol{y}) \prod_{j=1}^{s} \phi(y_j)\,\mathrm{d}\boldsymbol{y}\,,$$

where ϕ is a univariate probability density function on $\mathbb{R}$, frequently a Gaussian density. These integrals arise, most prominently, in Darcy-flow fluid modeling in porous media problems, which is our main motivation for the extensions of theory in this chapter, and the main focus of Chapter 4. Other examples of problems for which unbounded integration arises include, but by no means are limited to, integrals arising from option pricing problems in finance, see for example [6, 1, 42, 23, 26, 27], and maximum likelihood problems in statistics, see for example [14, 15, 72, 34, 57]. The aim of this chapter is to provide the theoretical foundation for the fast CBC construction of randomly shifted lattice rules that are tailored to these practical integrals, and demonstrate that these lattice rules obtain "good" error convergence rates.

The natural first step in applying QMC methods to an integral formulated over $\mathbb{R}^s$ is to transform the integral into the unit cube $[0,1]^s$, such that we may use the point sets explored in the previous chapter. After some appropriate manipulations, an integral over $\mathbb{R}^s$ can be rewritten in the form

$$\int_{\mathbb{R}^s} f(\boldsymbol{y}) \prod_{j=1}^{s} \phi(y_j)\,\mathrm{d}\boldsymbol{y} \;=\; \int_{[0,1]^s} f(\Phi^{-1}(\boldsymbol{u}))\,\mathrm{d}\boldsymbol{u}, \tag{3.1}$$

where Φ^{-1} denotes the inverse of the cumulative distribution function, corresponding to ϕ, which, when applied to vectors, is done component-wise. A QMC method with points $\boldsymbol{t}^{(1)}, \ldots, \boldsymbol{t}^{(n)} \in [0,1]^s$ then approximates the last integral in (3.1) by

$$\frac{1}{n} \sum_{k=1}^{n} f(\Phi^{-1}(\boldsymbol{t}^{(k)})). \tag{3.2}$$

However, the transformation in (3.1) often results in an integrand $f(\Phi^{-1}(\cdot))$ that is either unbounded at the boundary of the unit cube, or has unbounded derivatives near the boundary. In those cases, the standard QMC theory that we have seen thus far cannot be

applied, as the transformed integrand will not have a finite norm in any of the appropriate functions spaces examined thus far.

Various methods have been proposed in the literature of tackling this issue. To this end an anchored function space over $\mathbb{R}^s$ was considered in [39], see also [68, 69, 40]. It was proposed that, in the norm or inner product, the integrand be multiplied by a *weight function* ψ that decays quickly at infinity, as a way of controlling the behaviour of the integrand at the edges. It was shown that randomly shifted lattice rules can be obtained using a CBC construction to achieve close to the optimal convergence rate in this non-standard setting.

Here we make use of this weight function in a space reminiscent of the weighted Sobolev spaces of Chapter 2. To understand the setting it is instructive to examine the details of the norm of the space of functions in one dimension, $\mathcal{F}_1$, which we define as

$$\|f\|_{\mathcal{F}_1}^2 \;=\; [f(0)]^2 + \frac{1}{\gamma}\int_{\mathbb{R}} [f'(y)\psi(y)]^2 \,\mathrm{d}y. \tag{3.3}$$

The weight function ψ allows us to control the limiting behaviour of functions in $\mathcal{F}_1$, and either make it very "small" or very "big". For example, if we took $\psi(y) = \exp(-\alpha|y|)$, then we can see that $\mathcal{F}_1$ would contain any polynomial, as the norm would be finite. Even the function $\exp(\beta|y|)$ would have a finite norm, provided $\beta < \alpha$. Clearly in this example the weight γ does little to the space except scale the norm, however we include it here as we shall be considering weighted spaces later in the chapter, in the spirit of the weighted Sobolev and Korobov spaces from the last chapter. Although ϕ does not make an explicit difference to the space $\mathcal{F}_1$, as it is not present in the inner product, our integration operator $I_{s,\phi}$ is defined in terms of it, hence ϕ is central to the results that follow.

This chapter provides a number of important extensions to the theory in [39], all motivated by the needs of the application to porous flow problems. This work is original research, and has been submitted for publication in [46]. Firstly, we remove the assumption that the weight parameters $\gamma_{\mathfrak{u}}$ take the product form $\gamma_{\mathfrak{u}} = \prod_{j\in\mathfrak{u}} \gamma_j$. In the previous chapter we allowed weights to be of general form in the standard setting on the unit cube, except in some results relevant to the weighted Sobolev spaces. Similarly, in the literature general weights were introduced in [70] and have been considered in [20, 61]. However, allowing for general weights is novel territory for the this unbounded setting.

This generalisation was prompted by, and is essential to, the porous-flow problem that this thesis is concerned with. In the following chapter, as well as [36, 38], it is demonstrated that the overall error bound of our QMC method is minimised when the weight parameters take the form of POD weights of (2.23), which we recall to be of the form

$$\gamma_{\mathfrak{u}} = \Gamma_{|\mathfrak{u}|} \prod_{j\in\mathfrak{u}} \gamma_j,$$

The weight parameters are determined by the choice of two sequences $\Gamma_0 = \Gamma_1 = 1, \Gamma_2, \ldots$ and $\gamma_1, \gamma_2, \ldots$. Allowing for these weights, or any other general forms of weights, means that the theoretical basis for the CBC construction of randomly shifted lattice rules in this non-standard setting needs to be proved anew, and this non-trivial result is the main theorem of this chapter.

As highlighted in [17] for the standard setting, the CBC construction with general non-product weights in the anchored setting has an issue with the computational cost due to the need to work with some "auxiliary weights". The same issue holds for the non-standard setting considered here. Our second major advance in this chapter is to introduce an "unanchored" version of the function space over $\mathbb{R}^s$. We provide the complete theory for the CBC construction in this unanchored variant, and discuss the computational strategies for implementing fast CBC construction with POD weights.

In addition to the two major extensions, we also make other generalisations. We allow the weight parameters $\gamma_{\mathfrak{u}}$ to depend on the dimension s: later we write, more explicitly, $\gamma_{s,\mathfrak{u}}$. This is natural for the maximum likelihood problems [57] since all model parameters depend on the dimension s. This is also useful from the point of view of linking the theory between anchored and unanchored settings. We allow the weight function ψ to be coordinate dependent, that is, we have a weight function ψ_j for each coordinate. This turns out to be a crucial step in modelling the PDE applications in the next chapter. In our analysis we also allow the integration domain to be more general, to cater for other potential future applications.

The outline of this chapter is as follows. In §3.1 we discuss two additional practical applications that motivate the theoretical developments of this chapter, and outline how the theory can be applied in each case. In §3.2 we introduce the function space settings of this chapter, briefly discussing relevant extensions to reproducing kernel Hilbert spaces, and introduce the expression for the shift-averaged worst-case errors for randomly lattice rules in this setting. In §3.2.2 we review known results in the anchored setting, and then in §3.2.3 we derive various results in the new unanchored setting. In §3.3 we present the CBC algorithm for constructing a good generating vector of randomly shifted lattice rules, and then prove the main convergence results of the shift-averaged worst-case error for both the anchored and unanchored spaces with general non-product weights. We save the discussion of implementation and numerical results for Chapter 5.

3.1 Motivating applications

Integrals over $\mathbb{R}^s$ often arise from practical applications in the form of multivariate expected values

$$\mathbb{E}_\rho[q] \;=\; \int_{\mathbb{R}^s} q(\boldsymbol{y})\, \rho(\boldsymbol{y})\, \mathrm{d}\boldsymbol{y}, \tag{3.4}$$

where q is some quantity of interest which depends on a vector $\boldsymbol{y} = (y_1, \ldots, y_s)$ of parameters or variables in s dimensions, and ρ is some multivariate probability density function, not necessarily a product of univariate functions, describing the distribution of $\boldsymbol{y}$. We

have discussed briefly our major application, explored further in Chapter 4, of the porous flow problem. Below we discuss two additional motivating applications.

3.1.1 Application to option pricing problems

Following the Black-Scholes model, integrals arising from option pricing problems take the general form of (3.4), with

$$q(\boldsymbol{y}) = \max(\mu(\boldsymbol{y}), 0) \quad \text{and} \quad \rho(\boldsymbol{y}) = \frac{\exp(-\frac{1}{2}\boldsymbol{y}^{\mathsf{T}}\Sigma^{-1}\boldsymbol{y})}{\sqrt{(2\pi)^s \det(\Sigma)}},$$

where the variables $\boldsymbol{y} = (y_1, \ldots, y_s)^{\mathsf{T}}$ correspond to a discretization of the underlying Brownian motion over a time interval $[0, T]$, and the covariance matrix has entries $\Sigma_{ij} = (T/s)\min(i, j)$. Here $q(\cdot)$ is called the *payoff function*, and typically is of the form $q(\boldsymbol{y}) = \max(\mu(\boldsymbol{y}), 0)$, where $\mu(\boldsymbol{y})$ is a smooth function of the asset price $S_t(\boldsymbol{y})$.

Typically one takes a factorization $\Sigma = AA^{\mathsf{T}}$ and applies a change of variables $\boldsymbol{y} = A\boldsymbol{y}'$, leading us to an integral of the form (3.1), with $f(\boldsymbol{y}') = q(A\boldsymbol{y}')$, and ϕ being the standard normal density. The choice of factorization therefore determines the function f. If A is obtained via the Cholesky factorization, then it is called the "standard construction". The "Brownian bridge construction" [1] yields a different matrix A, while the matrix A obtained from the eigenvalue decomposition of Σ is known by the QMC community as the "principal components construction" [6].

The success of QMC for option pricing cannot be explained by the standard theory, however analysis of the integrand, as seen in [27], suggests that all ANOVA terms of f are smooth, with the exception of the highest order term, $f_{\{1:s\}}$. This suggests, noting that ANOVA terms are orthogonal in the new unanchored function space setting of this paper, that the analysis on [26] can potentially be adapted to show that the function $f - f_{\{1:s\}}$ belongs to the unanchored setting of this chapter.

3.1.2 Application to maximum likelihood problems

Another source of inspiration towards the non-standard setting in this paper is a class of generalized response models in statistics, as examined in [34, 39, 57]. A specific example of the time series Poisson likelihood model considered in these papers involves an integral of the form (3.4), with

$$q(\boldsymbol{y}) = \prod_{j=1}^{s} \frac{\exp(\tau_j(\beta + y_j) - e^{\beta + y_j})}{\tau_j!} \quad \text{and} \quad \rho(\boldsymbol{y}) = \frac{\exp(-\frac{1}{2}\boldsymbol{y}^{\mathsf{T}}\Sigma^{-1}\boldsymbol{y})}{\sqrt{(2\pi)^s \det(\Sigma)}}.$$

Here $\beta \in \mathbb{R}$ is a model parameter, $\tau_1, \ldots, \tau_s \in \{0, 1, \ldots\}$ are the count data, and Σ is a Toeplitz covariance matrix with $\Sigma_{ij} = \sigma^2\kappa^{|i-j|}/(1-\kappa^2)$, where σ^2 is the variance and $\kappa \in (-1, 1)$ is the autoregression coefficient. An obvious way to rewrite this integral in the form (3.1) is to factorize Σ as discussed above for the option pricing applications, but this yields unacceptable integrands f. Instead the strategy developed in [34] recentres and rescales the exponent of the integrand $q(\boldsymbol{y})\rho(\boldsymbol{y}) = \exp(F(\boldsymbol{y}))$. Using the results from

this current paper and following the strategy for choosing weight parameters in [36, 38], the recent paper [57] provides careful estimates of the norm of the resulting integrand f corresponding to three different choices of density ϕ, with the weight function taken as $\psi \equiv 1$, and gives the formula for the weight parameters $\gamma_{\mathfrak{u}}$ that minimise the overall error bound.

3.2 Function space setting

3.2.1 General framework of reproducing kernel Hilbert spaces

The previous chapter surveyed the literature of numerical integration on $[0,1]^s$. We wish to extend this theory to unbounded domains. First we define our problem somewhat more precisely.

Suppose that our domain is $D := \overline{(a,b)}$, allowing unbounded intervals such as $\mathbb{R}$. Let ϕ be a univariate probability density function on D, that is, $\phi(y) > 0$ for all $y \in D$ and $\int_a^b \phi(y)\,\mathrm{d}y = 1$. For $s \geq 1$, we define the cumulative distribution function $\Phi : D \to [0,1]$ as

$$\Phi(y) := \int_a^y \phi(t)\,\mathrm{d}t,$$

and denote its inverse by $\Phi^{-1} : [0,1] \to D$. Given a vector $\boldsymbol{v} \in [0,1]^s$ we apply Φ^{-1} component-wise, that is we write $\Phi^{-1}(\boldsymbol{v}) = (\Phi^{-1}(v_1), \dots, \Phi^{-1}(v_s))$. We are interested in the integral of a function $f : D^s \to \mathbb{R}$ with respect to the product probability density, that is

$$I_{s,\phi}(f) := \int_{D^s} f(\boldsymbol{y}) \prod_{j=1}^{s} \phi(y_j)\,\mathrm{d}\boldsymbol{y}.$$

Ultimately, our approximation of this integral amounts to using the n-point randomly shifted rank-1 lattice rule on the transformed integrand $f \circ \Phi^{-1}$, that is

$$Q_{s,n}(\boldsymbol{\Delta}; f) := \frac{1}{n} \sum_{i=1}^{n} f\left(\Phi^{-1}\left(\left\{\frac{i\,\boldsymbol{z}}{n} + \boldsymbol{\Delta}\right\}\right)\right). \tag{3.5}$$

We assume that the integrand f belongs to a weighted reproducing kernel Hilbert space (RKHS) $\mathcal{F}$ of real valued functions on D^s, which are (at least) integrable with respect to the s-fold tensor product of the density ϕ. In this setting we can continue to use the machinery of §2.3, except that here we are concerned with the unbounded integration operator $I_{s,\phi}$, and that to use QMC methods, we need to transform the integrand to the unit cube. We derive results similar to those found in §2.3.1 for this specific setting.

To be able to apply QMC methods, we must map the integrand from the Hilbert space $\mathcal{F}$ of functions on D^s to a Hilbert space $\mathcal{G}$ of functions on $[0,1]^s$, where we use the isometry

$$f \in \mathcal{F} \iff g = f(\Phi^{-1}(\cdot)) \in \mathcal{G}, \quad \text{with} \quad \|f\|_{\mathcal{F}} = \|g\|_{\mathcal{G}}.$$

It is important to note that the integral remains the same under this isometry,

$$I_{s,\phi}(f) = I_s(g) := \int_{[0,1]^s} g(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}.$$

Conveniently, it can be shown that the space $\mathcal{G}$ is also a RKHS, where the kernel is

$$K_{\mathcal{G}}(\boldsymbol{u},\boldsymbol{v}) = K_{\mathcal{F}}(\Phi^{-1}(\boldsymbol{u}),\Phi^{-1}(\boldsymbol{v})), \quad \boldsymbol{u},\,\boldsymbol{v}\in[0,1]^s. \tag{3.6}$$

Once again we will make use of the shift-invariant kernel, as defined in (2.33),

$$\begin{aligned} K^{\mathrm{sh}}_{\mathcal{G}}(\boldsymbol{u},\boldsymbol{v}) &:= \int_{[0,1]^s} K_{\mathcal{G}}(\{\boldsymbol{u}+\boldsymbol{\Delta}\},\{\boldsymbol{v}+\boldsymbol{\Delta}\})\,\mathrm{d}\boldsymbol{\Delta} \\ &= \int_{[0,1]^s} K_{\mathcal{F}}(\Phi^{-1}(\{\boldsymbol{u}+\boldsymbol{\Delta}\}),\Phi^{-1}(\{\boldsymbol{v}+\boldsymbol{\Delta}\}))\,\mathrm{d}\boldsymbol{\Delta}, \end{aligned} \tag{3.7}$$

Note that the shift-invariant kernel is only dependent on the difference of the two points $\boldsymbol{u}$ and $\boldsymbol{v}$. With a slight abuse of notation we write

$$K^{\mathrm{sh}}_{\mathcal{G}}(\boldsymbol{u},\boldsymbol{v}) = K^{\mathrm{sh}}_{\mathcal{G}}(\{\boldsymbol{u}-\boldsymbol{v}\},\boldsymbol{0}) = K^{\mathrm{sh}}_{\mathcal{G}}(\{\boldsymbol{u}-\boldsymbol{v}\}).$$

Hence we approximate the integral $I_s(g) = I_{s,\phi}(f)$ by a QMC rule

$$Q_{s,n}(\mathcal{P};g) \;=\; \frac{1}{n}\sum_{k=1}^{n} g(\boldsymbol{t}^{(k)}),$$

with points $\boldsymbol{t}^{(1)},\ldots,\boldsymbol{t}^{(n)}\in[0,1]^s$, and it is in this space $\mathcal{G}$ that we can study the *worst-case error*, defined as

$$e_{s,n}(\mathcal{P},\mathcal{G}) \;=\; \sup_{\|g\|_{\mathcal{G}}\le 1} |I_s(g) - Q_{s,n}(\mathcal{P};g)|.$$

Then it is straightforward to relate this to the original integration problem for $f\in\mathcal{F}$,

$$\begin{aligned} |I_{s,\phi}(f) - Q_{s,n}(\mathcal{P};f\circ\Phi^{-1})| \;&=\; |I_s(g) - Q_{s,n}(\mathcal{P};g)| \\ &\le\; e_{s,n}(\mathcal{P},\mathcal{G})\,\|g\|_{\mathcal{G}} \;=\; e_{s,n}(\mathcal{P},\mathcal{G})\,\|f\|_{\mathcal{F}}. \end{aligned} \tag{3.8}$$

The last expression illustrates the fact that while we study the worst-case error in $\mathcal{G}$, for which we have explicit expressions, we can keep the analysis of the norm of f in the original space $\mathcal{F}$, which is more convenient.

Once again both integration and QMC quadrature are linear functionals on both $\mathcal{F}$ and $\mathcal{G}$, so we can write

$$|I_s(g) - Q_n(\mathcal{P};\,g)| \;=\; \left|\langle g,h\rangle_{\mathcal{G}}\right|,$$

and we see that our representer of quadrature error, $h \in \mathcal{G}$, should be defined here as

$$\begin{aligned} h(\boldsymbol{u}) &:= \int_{[0,1]^s} K_{\mathcal{G}}(\boldsymbol{u},\boldsymbol{v})\,\mathrm{d}\boldsymbol{v} - \frac{1}{n}\sum_{i=1}^{n} K_{\mathcal{G}}(\boldsymbol{u},\boldsymbol{t}^{(i)}) \\ &= \int_{D^s} K_{\mathcal{F}}(\Phi^{-1}(\boldsymbol{u}),\boldsymbol{y}) \prod_{j=1}^{s} \phi(y_j)\,\mathrm{d}\boldsymbol{y} - \frac{1}{n}\sum_{i=1}^{n} K_{\mathcal{G}}(\boldsymbol{u},\boldsymbol{t}^{(i)})\,. \end{aligned}$$

which can be derived using the same steps as in (2.15).

The initial error of integration in $\mathcal{G}$ is the same as in the original space $\mathcal{F}$,

$$e_s(0,\mathcal{G}) \;=\; \sup_{\|g\|_{\mathcal{G}}\le 1} |I_s(g)| \;=\; \sup_{\|f\|_{\mathcal{F}}\le 1} |I_{s,\phi}(f)| \;=\; e_s(0,\mathcal{F}).$$

which we require to be finite, and can be calculated to be

$$[e_s(0,\mathcal{F})]^2 \;=\; \int_{D^s}\int_{D^s} K_{\mathcal{F}}(\boldsymbol{x},\boldsymbol{y}) \prod_{j=1}^{s} \left(\phi(x_j)\,\phi(y_j)\right)\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y} < \infty. \tag{3.9}$$

To ensure the embedding of $\mathcal{F}$ in $L_{2,\phi}(D^s)$, which is required in our later analysis, we further assume that

$$\int_{D^s} K_{\mathcal{F}}(\boldsymbol{y},\boldsymbol{y}) \prod_{j=1}^{s} \phi(y_j)\,\mathrm{d}\boldsymbol{y} < \infty. \tag{3.10}$$

Once again we are interested in evaluating the worst-case errors for shifted lattice rules. Since the set of points that make up the shifted lattice rule $Q_{s,n}$ are dependent only on the vectors $\boldsymbol{z}$ and $\boldsymbol{\Delta}$, for short hand notation we write the worst-case error as $e_{s,n}(\mathcal{P};\mathcal{G}) = e_{s,n}(\boldsymbol{z},\boldsymbol{\Delta})$.

As we consider *randomly* shifted lattice rules, we are interested in the shift-average worst-case error, which now only depends on $\boldsymbol{z}$ and is well-known (see e.g., [58]) to reduce to

$$\begin{aligned} [e^{\mathrm{sh}}_{s,n}(\boldsymbol{z})]^2 := \int_{[0,1]^s} [e_{s,n}(\mathcal{P};\mathcal{G})]^2\,\mathrm{d}\boldsymbol{\Delta} &= \int_{[0,1]^s} [e_{s,n}(\boldsymbol{z},\boldsymbol{\Delta})]^2\,\mathrm{d}\boldsymbol{\Delta} \\ &= -\int_{[0,1]^s}\int_{[0,1]^s} K_{\mathcal{G}}(\boldsymbol{u},\boldsymbol{v})\,\mathrm{d}\boldsymbol{u}\,\mathrm{d}\boldsymbol{v} + \frac{1}{n}\sum_{k=1}^{n} K^{\mathrm{sh}}_{\mathcal{G}}\left(\left\{\frac{k\boldsymbol{z}}{n}\right\}\right). \end{aligned} \tag{3.11}$$

3.2.2 Anchored spaces

Here we review the *weighted anchored spaces* as studied in [40, 39], but with some new developments. Given an anchor $c \in D$, a set of *weight parameters* $\gamma_{s,\mathrm{u}} > 0$ (or "weights" for short) and a set of *weight functions* $\psi_j : D \to \mathbb{R}$, the space $\mathcal{F}$ is the Hilbert space of functions from D^s to $\mathbb{R}$, for which $I_{s,\phi}(f)$ is bounded, with the inner product

$$\langle f, g\rangle_{\mathcal{F}} = \sum_{\mathrm{u}\subseteq\{1:s\}} \frac{1}{\gamma_{s,\mathrm{u}}} \int_{D^{|\mathrm{u}|}} \frac{\partial^{|\mathrm{u}|} f}{\partial \boldsymbol{y}_{\mathrm{u}}}(\boldsymbol{y}_{\mathrm{u}};\boldsymbol{c}_{-\mathrm{u}})\, \frac{\partial^{|\mathrm{u}|} g}{\partial \boldsymbol{y}_{\mathrm{u}}}(\boldsymbol{y}_{\mathrm{u}};\boldsymbol{c}_{-\mathrm{u}}) \prod_{j\in\mathrm{u}} \psi_j^2(y_j)\,\mathrm{d}\boldsymbol{y}_{\mathrm{u}}, \tag{3.12}$$

where the same notation is used as in (2.31). We take $\gamma_{s,\emptyset} = 1$. As usual the corresponding norm is $\|f\|_{\mathcal{F}} = \langle f, f\rangle_{\mathcal{F}}^{1/2}$. The generalisations in (3.12), as compared to similar work in [39, 40, 68, 69], include the allowance for general (non-product) weights $\gamma_{s,\mathfrak{u}}$, which may depend on the dimension s, as well as for coordinate-dependent weight functions ψ_j. This function space setting has three key ingredients:

- The *univariate probability density* $\phi : \mathbb{R} \to \mathbb{R} \setminus \mathbb{R}^-$ in (3.1) controls the mapping from $\mathbb{R}^s$ to the unit cube $[0,1]^s$. Although ϕ does not affect the norm (3.3), it determines the transformed integrand $f(\Phi^{-1}(\cdot))$ over the unit cube. Hence it will affect the integration, and in particular the worst-case and initial errors.
- The *weight function* $\psi : \mathbb{R} \to \mathbb{R}^+$ in the norm (3.3) controls the boundary behaviour of the functions f that are included in the space. If $\psi(y_j)$ decays very quickly to 0 for large $|y_j|$ then the space can contain functions with very fast diverging mixed derivatives.
- The collection of *weight parameters* $\gamma_{\mathfrak{u}}$ associated with subsets $\mathfrak{u} \subset \mathbb{N}$ with finite cardinality $|\mathfrak{u}| < \infty$ controls the relative importance of various groups of variables, as discussed previously.

The reproducing kernel corresponding to the inner product (3.12) is given by

$$K_{\mathcal{F}}(\boldsymbol{x}, \boldsymbol{y}) = \sum_{\mathfrak{u} \subseteq \{1:s\}} \gamma_{s,\mathfrak{u}} \prod_{j \in \mathfrak{u}} \eta_j(x_j, y_j), \tag{3.13}$$

where for x, $y \in D$,

$$\eta_j(x,y) = \begin{cases} \displaystyle\int_c^{\min(x,y)} \frac{1}{\psi_j^2(t)} \,\mathrm{d}t & \text{if } x,\, y > c, \\ \displaystyle\int_{\max(x,y)}^{c} \frac{1}{\psi_j^2(t)} \,\mathrm{d}t & \text{if } x,\, y < c, \\ 0 & \text{otherwise.} \end{cases}$$

For this to be well-defined we must assume that for all j, ψ_j satisfies

$$\int_x^y \frac{1}{\psi_j^2(t)} \,\mathrm{d}t < \infty \quad \text{for all finite } x \text{ and } y, \tag{3.14}$$

which is satisfied if ψ_j is strictly positive and continuous on $(-\infty, \infty)$.

The kernel must satisfy the two conditions (3.9) and (3.10). Substituting (3.13) into (3.9), we find that

$$\int_{D^s} \int_{D^s} K_{\mathcal{F}}(\boldsymbol{x}, \boldsymbol{y}) \prod_{j=1}^{s} (\phi(x_j)\phi(y_j)) \,\mathrm{d}\boldsymbol{x} \,\mathrm{d}\boldsymbol{y} = \sum_{\mathfrak{u} \subseteq \{1:s\}} \gamma_{s,\mathfrak{u}} \prod_{j \in \mathfrak{u}} C_{0,j} < \infty, \tag{3.15}$$

where we define

$$C_{0,j} := \int_a^b \int_a^b \eta_j(x_j, y_j)\, \phi(x_j)\, \phi(y_j)\, \mathrm{d}x_j \mathrm{d}y_j \tag{3.16}$$
$$= \int_a^c \frac{\Phi^2(t)}{\psi_j^2(t)}\, \mathrm{d}t + \int_c^b \frac{(1-\Phi(t))^2}{\psi_j^2(t)}\, \mathrm{d}t,$$

with the last equality demonstrated in [39]. Similarly, we see that (3.10) reduces to

$$\int_{D^s} K_{\mathcal{F}}(\boldsymbol{y}, \boldsymbol{y}) \prod_{j=1}^{s} \phi(y_j)\, \mathrm{d}\boldsymbol{y} = \sum_{\mathfrak{u} \subseteq \{1:s\}} \gamma_{s,\mathfrak{u}} \prod_{j \in \mathfrak{u}} C_{1,j} \; < \; \infty \tag{3.17}$$

where we define

$$C_{1,j} := \int_a^b \eta_j(y_j, y_j)\, \phi(y_j)\, \mathrm{d}y_j \tag{3.18}$$
$$= \int_a^c \frac{\Phi(t)}{\psi_j^2(t)}\, \mathrm{d}t + \int_c^b \frac{1-\Phi(t)}{\psi_j^2(t)}\, \mathrm{d}t,$$

with the last equality again shown in [39]. Evidently, to satisfy (3.9) and (3.10) we require that $C_{0,j} < \infty$ and $C_{1,j} < \infty$ for all j.

Now we turn to the corresponding function space $\mathcal{G}$. The kernel $K_{\mathcal{G}}(\boldsymbol{u}, \boldsymbol{v})$ can be calculated as in (3.6), while the associated shift-invariant kernel is

$$K_{\mathcal{G}}^{\mathrm{sh}}(\{\boldsymbol{u} - \boldsymbol{v}\}) \; = \; \sum_{\mathfrak{u} \subseteq \{1:s\}} \gamma_{s,\mathfrak{u}} \prod_{j \in \mathfrak{u}} \theta_j(\{u_j - v_j\}), \tag{3.19}$$

where

$$\theta_j(u) \; := \int_0^1 \eta_j(\Phi^{-1}(\{u + \Delta\}), \Phi^{-1}(\Delta))\, \mathrm{d}\Delta \tag{3.20}$$
$$= \int_{\Phi^{-1}(u)}^{c} \frac{\Phi(t) - u}{\psi_j^2(t)}\, \mathrm{d}t \; + \int_{\Phi^{-1}(1-u)}^{c} \frac{\Phi(t) - 1 + u}{\psi_j^2(t)}\, \mathrm{d}t, \qquad u \in [0,1],$$

which was derived in [40]. It is important to note that

$$C_{0,j} \; = \; \int_0^1 \theta_j(u)\, \mathrm{d}u \qquad \text{and} \qquad C_{1,j} \; = \; \theta_j(0),$$

which applies regardless of the choice of kernel and η_j, hence this applies in the unanchored space, examined in the next section.

Now we are in a position to express the shift-averaged worst-case error for lattice rules in the anchored space. Substituting (3.15) and (3.19) into (3.11), we obtain the expression

$$[e_{s,n}^{\mathrm{sh}}(\boldsymbol{z})]^2 \; = \; \sum_{\mathfrak{u} \subseteq \{1:s\}} \gamma_{s,\mathfrak{u}} \left(-\prod_{j \in \mathfrak{u}} C_{0,j} + \frac{1}{n} \sum_{k=1}^{n} \prod_{j \in \mathfrak{u}} \theta_j \left(\left\{ \frac{k z_j}{n} \right\} \right) \right). \tag{3.21}$$

3.2.3 Unanchored spaces

Here we introduce the *weighted unanchored spaces*. We commence by deriving the reproducing kernel in one dimension, in a fashion inspired by the derivation in [67], and then proceed to higher dimensions where we also derive results for the shift-averaged worst-case errors for lattice rules.

Lemma 14 (Unanchored space – reproducing kernel) *We take $\mathcal{F}$ to be the space of functions from D to $\mathbb{R}$, where we define the inner product of f, $g \in \mathcal{F}$ as*

$$\langle f, g \rangle_{\mathcal{F}} := \left(\int_a^b f(y)\, \phi(y)\, \mathrm{d}y \right) \left(\int_a^b g(y)\, \phi(y)\, \mathrm{d}y \right) + \frac{1}{\gamma} \int_a^b f'(y)\, g'(y)\, \psi_j^2(y)\, \mathrm{d}y,$$

with $\gamma > 0$ and a weight function $\psi_j : D \to \mathbb{R}^+$ satisfying (3.14)*. Then the reproducing kernel in $\mathcal{F}$ is given by $K_{\mathcal{F}}(x, y) = 1 + \gamma \eta_j(x, y)$, where*

$$\eta_j(x, y) = \int_a^{\min(x,y)} \frac{\Phi(t)}{\psi_j^2(t)}\, \mathrm{d}t + \int_{\max(x,y)}^b \frac{1 - \Phi(t)}{\psi_j^2(t)}\, \mathrm{d}t - \int_a^b \frac{\Phi(t)(1 - \Phi(t))}{\psi_j^2(t)}\, \mathrm{d}t. \tag{3.22}$$

Proof. Since this reproducing kernel does not appear to exist in the literature, rather than simply verifying the reproducing property, we provide a derivation.

Suppose that $K_{\mathcal{F}}(x, y) = 1 + \gamma \eta_j(x, y)$, with

$$\eta_j(x, y) = \begin{cases} L_x(y) & \text{for } y \le x, \\ R_x(y) & \text{for } y \ge x. \end{cases}$$

The reproducing property in Definition 3 yields

$$\begin{aligned} f(x) = {} & \left(\int_a^b f(y)\, \phi(y)\, \mathrm{d}y \right) \left(1 + \gamma \left[\int_a^x L_x(y)\, \phi(y)\, \mathrm{d}y + \int_x^b R_x(y)\, \phi(y)\, \mathrm{d}y \right] \right) \\ & + \int_a^x f'(y)\, L_x'(y)\, \psi_j^2(y)\, \mathrm{d}y + \int_x^b f'(y)\, R_x'(y)\, \psi_j^2(y)\, \mathrm{d}y, \end{aligned}$$

which means that the following two properties must hold:

$$\begin{cases} \displaystyle\int_a^x L_x(y)\, \phi(y)\, \mathrm{d}y + \int_x^b R_x(y)\, \phi(y)\, \mathrm{d}y = 0, \\ \displaystyle\int_a^x f'(y)\, L_x'(y)\, \psi_j^2(y)\, \mathrm{d}y + \int_x^b f'(y)\, R_x'(y)\, \psi_j^2(y)\, \mathrm{d}y = f(x) - \int_a^b f(y)\, \phi(y)\, \mathrm{d}y. \end{cases} \tag{3.23}$$

As an initial guess, we assume that

$$L_x'(y) = \frac{\ell_x(y)}{\psi_j^2(y)} \qquad \text{and} \qquad R_x'(y) = \frac{r_x(y)}{\psi_j^2(y)}.$$

Since $L_x(x) = R_x(x) =: M(x)$, we can write

$$L_x(y) \;=\; M(x) - \int_y^x \frac{\ell_x(t)}{\psi_j^2(t)}\,\mathrm{d}t \qquad \text{and} \qquad R_x(y) \;=\; M(x) + \int_x^y \frac{r_x(t)}{\psi_j^2(t)}\,\mathrm{d}t. \tag{3.24}$$

Then the two required properties in (3.23) simplify to (assuming that all integrals are finite and therefore, by Fubini's theorem, interchanging the order of integration is allowed)

$$\begin{cases} \displaystyle\int_a^x \frac{\Phi(t)\,\ell_x(t)}{\psi_j^2(t)}\,\mathrm{d}t \;-\; \int_x^b \frac{(1-\Phi(t))\,r_x(t)}{\psi_j^2(t)}\,\mathrm{d}t \;=\; M(x) \\ \displaystyle\int_a^x f'(y)\,\ell_x(y)\,\mathrm{d}y \;+\; \int_x^b f'(y)\,r_x(y)\,\mathrm{d}y \;=\; f(x) - \int_a^b f(y)\,\phi(y)\,\mathrm{d}y. \end{cases} \tag{3.25}$$

If we take

$$\ell_x(y) \;=\; \Phi(y) \qquad \text{and} \qquad r_x(y) \;=\; \Phi(y) - 1, \tag{3.26}$$

and use integration by parts, we can verify that the second equation in (3.25) holds,

$$\begin{aligned} &\int_a^x f'(y)\,\ell_x(y)\,\mathrm{d}y + \int_x^b f'(y)\,r_x(y)\,\mathrm{d}y \\ &= \int_a^x f'(y)\,\Phi(y)\,\mathrm{d}y + \int_x^b f'(y)\,(\Phi(y)-1)\,\mathrm{d}y \\ &= \Big[f(y)\,\Phi(y)\Big]_a^x - \int_a^x f(y)\,\phi(y)\,\mathrm{d}y + \Big[f(y)\,(\Phi(y)-1)\Big]_x^b - \int_x^b f(y)\,\phi(y)\,\mathrm{d}y \\ &= f(x) - \int_a^b f(y)\,\phi(y)\,\mathrm{d}y, \end{aligned}$$

which is as required. Substituting (3.26) into the first equation in (3.25) determines $M(x)$.

Finally we show that $\eta_j(x,y)$ is as in (3.22). Consider first the case $y \le x$. Then from (3.24), (3.25), and (3.26) we have for $y \le x$ that

$$\begin{aligned} \eta_j(x,y) \;=\; L_x(y) \;&=\; \int_a^x \frac{\Phi^2(t)}{\psi_j^2(t)}\,\mathrm{d}t \;+\; \int_x^b \frac{(1-\Phi(t))^2}{\psi_j^2(t)}\,\mathrm{d}t \;-\; \int_y^x \frac{\Phi(t)}{\psi_j^2(t)}\,\mathrm{d}t \\ &=\; \int_a^y \frac{\Phi^2(t)}{\psi_j^2(t)}\,\mathrm{d}t \;+\; \int_x^b \frac{(1-\Phi(t))^2}{\psi_j^2(t)}\,\mathrm{d}t \;-\; \int_y^x \frac{\Phi(t)(1-\Phi(t))}{\psi_j^2(t)}\,\mathrm{d}t \\ &=\; \int_a^y \frac{\Phi^2(t)}{\psi_j^2(t)}\,\mathrm{d}t \;+\; \int_x^b \frac{(1-\Phi(t))^2}{\psi_j^2(t)}\,\mathrm{d}t \;-\; \int_a^b \frac{\Phi(t)(1-\Phi(t))}{\psi_j^2(t)}\,\mathrm{d}t \\ &\qquad +\; \int_a^y \frac{\Phi(t)(1-\Phi(t))}{\psi_j^2(t)}\,\mathrm{d}t \;+\; \int_x^b \frac{\Phi(t)(1-\Phi(t))}{\psi_j^2(t)}\,\mathrm{d}t \\ &=\; \int_a^y \frac{\Phi(t)}{\psi_j^2(t)}\,\mathrm{d}t \;+\; \int_x^b \frac{1-\Phi(t)}{\psi_j^2(t)}\,\mathrm{d}t \;-\; \int_a^b \frac{\Phi(t)(1-\Phi(t))}{\psi_j^2(t)}\,\mathrm{d}t. \end{aligned}$$

We easily obtain a similar expression for $x \le y$, hence obtaining (3.22). □

We can now generalize the unanchored setting to s dimensions. The unanchored inner product is

$$\langle f, g\rangle_{\mathcal{F}} := \sum_{\mathfrak{u}\subseteq\{1:s\}} \frac{1}{\gamma_{s,\mathfrak{u}}} \int_{D^{|\mathfrak{u}|}} \left(\int_{D^{s-|\mathfrak{u}|}} \frac{\partial^{|\mathfrak{u}|} f}{\partial \boldsymbol{y}_{\mathfrak{u}}}(\boldsymbol{y}_{\mathfrak{u}}; \boldsymbol{y}_{\text{-}\mathfrak{u}}) \prod_{j\notin\mathfrak{u}} \phi(y_j)\,\mathrm{d}\boldsymbol{y}_{\text{-}\mathfrak{u}} \right) \tag{3.27}$$
$$\times \left(\int_{D^{s-|\mathfrak{u}|}} \frac{\partial^{|\mathfrak{u}|} g}{\partial \boldsymbol{y}_{\mathfrak{u}}}(\boldsymbol{y}_{\mathfrak{u}}; \boldsymbol{y}_{\text{-}\mathfrak{u}}) \prod_{j\notin\mathfrak{u}} \phi(y_j)\,\mathrm{d}\boldsymbol{y}_{\text{-}\mathfrak{u}} \right) \prod_{j\in\mathfrak{u}} \psi_j^2(y_j)\,\mathrm{d}\boldsymbol{y}_{\mathfrak{u}},$$

where the notation is as in (2.31). The reproducing kernel takes the same form as (3.13), but now with the function η_j defined by (3.22). As before we require the two conditions (3.15) and (3.17), but now with different constants $C_{0,j}$ and $C_{1,j}$.

Lemma 15 (Unanchored space – constants) *For the function η_j given by* (3.22)*, the quantities $C_{0,j}$ and $C_{1,j}$ defined in* (3.16) *and* (3.18)*, respectively, are*

$$C_{0,j} = 0 \qquad \text{and} \qquad C_{1,j} = \int_a^b \frac{\Phi(t)(1-\Phi(t))}{\psi_j^2(t)}\,\mathrm{d}t.$$

Proof. Substituting (3.22) into (3.16), we have that

$$\begin{aligned}
C_{0,j} &= \int_a^b \int_a^b \int_a^{\min(x,y)} \frac{\Phi(t)}{\psi_j^2(t)}\,\mathrm{d}t\,\phi(x)\phi(y)\,\mathrm{d}x\,\mathrm{d}y \\
&\quad + \int_a^b \int_a^b \int_{\max(x,y)}^b \frac{1-\Phi(t)}{\psi_j^2(t)}\,\mathrm{d}t\,\phi(x)\phi(y)\,\mathrm{d}x\,\mathrm{d}y - \int_a^b \frac{\Phi(t)(1-\Phi(t))}{\psi_j^2(t)}\,\mathrm{d}t \\
&= \int_a^b \frac{\Phi(t)}{\psi_j^2(t)} \int_t^b \int_t^b \phi(x)\phi(y)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}t \\
&\quad + \int_a^b \frac{1-\Phi(t)}{\psi_j^2(t)} \int_a^t \int_a^t \phi(x)\phi(y)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}t - \int_a^b \frac{\Phi(t)(1-\Phi(t))}{\psi_j^2(t)}\,\mathrm{d}t \\
&= \int_a^b \frac{\Phi(t)(1-\Phi(t))^2}{\psi_j^2(t)}\,\mathrm{d}t + \int_a^b \frac{(1-\Phi(t))\Phi^2(t)}{\psi j^2(t)}\,\mathrm{d}t - \int_a^b \frac{\Phi(t)(1-\Phi(t))}{\psi_j^2(t)}\,\mathrm{d}t,
\end{aligned}$$

which equals 0. Similarly, substituting (3.22) into (3.18), we obtain

$$\begin{aligned}
&C_{1,j} \\
&= \int_a^b \int_a^x \frac{\Phi(t)}{\psi_j^2(t)}\,\mathrm{d}t\,\phi(x)\,\mathrm{d}x + \int_a^b \int_x^b \frac{1-\Phi(t)}{\psi_j^2(t)}\,\mathrm{d}t\,\phi(x)\,\mathrm{d}x - \int_a^b \frac{\Phi(t)(1-\Phi(t))}{\psi_j^2(t)}\,\mathrm{d}t \\
&= \int_a^b \frac{\Phi(t)}{\psi_j^2(t)} \int_t^b \phi(x)\,\mathrm{d}x\,\mathrm{d}t + \int_a^b \frac{1-\Phi(t)}{\psi_j^2(t)} \int_a^t \phi(x)\,\mathrm{d}x\,\mathrm{d}t - \int_a^b \frac{\Phi(t)(1-\Phi(t))}{\psi_j^2(t)}\,\mathrm{d}t, \\
&= \int_a^b \frac{\Phi(t)(1-\Phi(t))}{\psi_j^2(t)}\,\mathrm{d}t.
\end{aligned}$$

which is the required expression. □

Now we consider the space $\mathcal{G}$ of functions on the unit cube. The kernel $K_{\mathcal{G}}(\boldsymbol{u}, \boldsymbol{v})$ is given by (3.6), and the associated shift invariant kernel $K_{\mathcal{G}}^{\mathrm{sh}}(\{\boldsymbol{u}-\boldsymbol{v}\})$ is of the same form as (3.19), but the function θ_j takes a different form.

Lemma 16 (Unanchored space – shift-invariant kernel) *For the function η_j given by* (3.22)*, the function θ_j defined in* (3.20) *is*

$$\theta_j(u) = \int_{\Phi^{-1}(u)}^{b} \frac{\Phi(t)-u}{\psi_j^2(t)}\,\mathrm{d}t + \int_{\Phi^{-1}(1-u)}^{b} \frac{\Phi(t)-1+u}{\psi_j^2(t)}\,\mathrm{d}t - \int_a^b \frac{\Phi^2(t)}{\psi_j^2(t)}\,\mathrm{d}t. \tag{3.28}$$

Alternatively, given any arbitrary point $c \in D$, we can express θ_j as

$$\begin{aligned}\theta_j(u) = &\int_{\Phi^{-1}(u)}^{c} \frac{\Phi(t)-u}{\psi_j^2(t)}\,\mathrm{d}t + \int_{\Phi^{-1}(1-u)}^{c} \frac{\Phi(t)-1+u}{\psi_j^2(t)}\,\mathrm{d}t \\ &- \int_a^c \frac{\Phi^2(t)}{\psi_j^2(t)}\,\mathrm{d}t - \int_c^b \frac{(1-\Phi(t))^2}{\psi_j^2(t)}\,\mathrm{d}t.\end{aligned} \tag{3.29}$$

Note that the alternative expression (3.29) enables us to directly compare the unanchored space to the anchored space. If we use the superscripts 'anch' and 'unanch' to distinguish relevant quantities from the anchored and unanchored spaces, then

$$\theta_j^{\mathrm{unanch}} = \theta_j^{\mathrm{anch}} - C_{0,j}^{\mathrm{anch}}. \tag{3.30}$$

Proof. Substituting (3.22) into (3.20), we have

$$\begin{aligned}\theta_j(u) = &\int_0^1 \int_a^{\min(\Phi^{-1}(\{u+\Delta\}),\Phi^{-1}(\Delta))} \frac{\Phi(t)}{\psi_j^2(t)}\,\mathrm{d}t\,\mathrm{d}\Delta \\ &+ \int_0^1 \int_{\max(\Phi^{-1}(\{u+\Delta\}),\Phi^{-1}(\Delta))}^{b} \frac{1-\Phi(t)}{\psi_j^2(t)}\,\mathrm{d}t\,\mathrm{d}\Delta - \int_a^b \frac{\Phi(t)(1-\Phi(t))}{\psi_j^2(t)}\,\mathrm{d}t \\ = &\underbrace{\int_0^{1-u} \int_a^{\Phi^{-1}(\Delta)} \frac{\Phi(t)}{\psi_j^2(t)}\,\mathrm{d}t\,\mathrm{d}\Delta}_{A} + \int_{1-u}^1 \int_a^{\Phi^{-1}(u+\Delta-1)} \frac{\Phi(t)}{\psi_j^2(t)}\,\mathrm{d}t\,\mathrm{d}\Delta \\ &+ \int_0^{1-u} \int_{\Phi^{-1}(u+\Delta)}^{b} \frac{1-\Phi(t)}{\psi_j^2(t)}\,\mathrm{d}t\,\mathrm{d}\Delta + \int_{1-u}^1 \int_{\Phi^{-1}(\Delta)}^{b} \frac{1-\Phi(t)}{\psi_j^2(t)}\,\mathrm{d}t\,\mathrm{d}\Delta \\ &- \int_a^b \frac{\Phi(t)(1-\Phi(t))}{\psi_j^2(t)}\,\mathrm{d}t.\end{aligned} \tag{3.31}$$

For the expression labelled A, we substitute $w = \Phi^{-1}(\Delta)$ to obtain

$$\begin{aligned}A = \int_a^{\Phi^{-1}(1-u)} \int_a^w \frac{\Phi(t)}{\psi_j^2(t)}\,\mathrm{d}t\,\phi(w)\,\mathrm{d}w &= \int_a^{\Phi^{-1}(1-u)} \frac{\Phi(t)}{\psi_j^2(t)} \int_t^{\Phi^{-1}(1-u)} \phi(w)\,\mathrm{d}w\,\mathrm{d}t \\ &= \int_a^{\Phi^{-1}(1-u)} \frac{\Phi(t)}{\psi_j^2(t)}\,[(1-u)-\Phi(t)]\,\mathrm{d}t.\end{aligned}$$

Applying a similar procedure to the rest of (3.31), we obtain

$$\begin{aligned}
\theta_j(u) &= \int_a^{\Phi^{-1}(1-u)} \frac{\Phi(t)}{\psi_j^2(t)} \left[(1-u)-\Phi(t)\right] \mathrm{d}t + \int_a^{\Phi^{-1}(u)} \frac{\Phi(t)}{\psi_j^2(t)} \left[u-\Phi(t)\right] \mathrm{d}t \\
&\quad + \int_{\Phi^{-1}(u)}^b \frac{1-\Phi(t)}{\psi_j^2(t)} \left[\Phi(t)-u\right] \mathrm{d}t + \int_{\Phi^{-1}(1-u)}^b \frac{1-\Phi(t)}{\psi_j^2(t)} \left[\Phi(t)-(1-u)\right] \mathrm{d}t \\
&\quad - \int_a^b \frac{\Phi(t)(1-\Phi(t))}{\psi_j^2(t)} \, \mathrm{d}t, \\
&= \int_a^b \frac{\Phi(t)}{\psi_j^2(t)} \left[(1-u)-\Phi(t)\right] \mathrm{d}t + \int_a^b \frac{\Phi(t)}{\psi_j^2(t)} \left[u-\Phi(t)\right] \mathrm{d}t \\
&\quad + \int_{\Phi^{-1}(u)}^b \frac{\Phi(t)-u}{\psi_j^2(t)} \mathrm{d}t + \int_{\Phi^{-1}(1-u)}^b \frac{\Phi(t)-(1-u)}{\psi_j^2(t)} \mathrm{d}t - \int_a^b \frac{\Phi(t)(1-\Phi(t))}{\psi_j^2(t)} \, \mathrm{d}t \\
&= \int_{\Phi^{-1}(u)}^b \frac{\Phi(t)-u}{\psi_j^2(t)} \mathrm{d}t + \int_{\Phi^{-1}(1-u)}^b \frac{\Phi(t)-(1-u)}{\psi_j^2(t)} \mathrm{d}t - \int_a^b \frac{\Phi^2(t)}{\psi_j^2(t)} \, \mathrm{d}t
\end{aligned}$$

which gives us (3.28). Now, given some $c \in D$, we can rewrite (3.28) as

$$\begin{aligned}
\theta_j(u) &= \int_{\Phi^{-1}(u)}^c \frac{\Phi(t)-u}{\psi_j^2(t)} \mathrm{d}t + \int_{\Phi^{-1}(1-u)}^c \frac{\Phi(t)-1+u}{\psi_j^2(t)} \mathrm{d}t \\
&\quad + \int_c^b \frac{\Phi(t)-u}{\psi_j^2(t)} \mathrm{d}t + \int_c^b \frac{\Phi(t)-1+u}{\psi_j^2(t)} \mathrm{d}t - \int_a^c \frac{\Phi^2(t)}{\psi_j^2(t)} \, \mathrm{d}t - \int_c^b \frac{\Phi^2(t)}{\psi_j^2(t)} \, \mathrm{d}t, \\
&= \int_{\Phi^{-1}(u)}^c \frac{\Phi(t)-u}{\psi_j^2(t)} \mathrm{d}t + \int_{\Phi^{-1}(1-u)}^c \frac{\Phi(t)-(1-u)}{\psi_j^2(t)} \mathrm{d}t \\
&\quad - \int_a^c \frac{\Phi^2(t)}{\psi_j^2(t)} \, \mathrm{d}t - \int_c^b \frac{\Phi^2(t)-2\Phi(t)+1}{\psi_j^2(t)} \, \mathrm{d}t \\
&= \int_{\Phi^{-1}(u)}^c \frac{\Phi(t)-u}{\psi_j^2(t)} \mathrm{d}t + \int_{\Phi^{-1}(1-u)}^c \frac{\Phi(t)-(1-u)}{\psi_j^2(t)} \mathrm{d}t \\
&\quad - \int_a^c \frac{\Phi^2(t)}{\psi_j^2(t)} \, \mathrm{d}t - \int_c^b \frac{(1-\Phi(t))^2}{\psi_j^2(t)} \, \mathrm{d}t
\end{aligned}$$

which yields (3.29). □

Now we can express the shift-averaged worst-case error for lattice rules in the unanchored setting. It is also important to note that, as $C_{0,j} = 0$ in this space, we have $e_{s,0} = 1$. Substituting (3.15), (3.19), and $C_{0,j} = 0$ into (3.11), we obtain

$$[e_{s,n}^{\mathrm{sh}}(\boldsymbol{z})]^2 = \sum_{\emptyset \neq \mathfrak{u} \subseteq \{1:s\}} \frac{\gamma_{s,\mathfrak{u}}}{n} \sum_{k=1}^n \prod_{j \in \mathfrak{u}} \theta_j \left(\left\{ \frac{k z_j}{n} \right\} \right), \tag{3.32}$$

where θ_j is given by (3.28).

3.3 Main results

3.3.1 Reformulating the shift-averaged worst-case error for lattice rules

In Sections 3.2.2 and 3.2.3 we derived expressions for the shift-averaged worst-case error for lattice rules in the anchored and unanchored spaces. Here we reformulate the worst-case error in terms of the Fourier series coefficients of θ_j. As θ_j is continuous on the unit interval, the Fourier series converges uniformly. We denote the Fourier coefficients by $\widehat{\theta}_j(h)$, where $h \in \mathbb{Z}$. We also write $\widehat{\theta}_{\mathfrak{v}}(\boldsymbol{h}) = \prod_{j\in\mathfrak{v}} \widehat{\theta}_j(h_j)$ for $\boldsymbol{h} \in \mathbb{Z}^{|\mathfrak{v}|}$. Note that for both the anchored and unanchored spaces we have $C_{0,j} = \widehat{\theta}_j(0)$, while $C_{1,j} = \theta_j(0)$. In the following we use the notation $i \equiv_n j$ to mean $i \equiv j \pmod n$. First we consider the anchored case.

Lemma 17 *If we define a set of* auxiliary weights

$$\widetilde{\gamma}_{s,\mathfrak{v}} := \sum_{\mathfrak{v}\subseteq\mathfrak{u}\subseteq\{1:s\}} \gamma_{s,\mathfrak{u}} \prod_{j\in\mathfrak{u}\setminus\mathfrak{v}} C_{0,j}, \qquad \mathfrak{v} \subseteq \{1:s\}, \tag{3.33}$$

then we can rewrite the worst-case error (3.21) *for the anchored space as*

$$[e^{\mathrm{sh}}_{s,n}(\boldsymbol{z})]^2 = \sum_{\emptyset\neq\mathfrak{v}\subseteq\{1:s\}} \widetilde{\gamma}_{s,\mathfrak{v}} \sum_{\substack{\boldsymbol{h}\in(\mathbb{Z}\setminus\{0\})^{|\mathfrak{v}|} \\ \boldsymbol{h}\cdot\boldsymbol{z}_{\mathfrak{v}}\equiv_n 0}} \widehat{\theta}_{\mathfrak{v}}(\boldsymbol{h}),$$

where $\boldsymbol{z}_{\mathfrak{v}} \in \mathcal{Z}_n^{|\mathfrak{v}|}$ denotes the vector containing the components of the lattice generating vector $\boldsymbol{z} \in \mathcal{Z}_n^s$ whose indices are in $\mathfrak{v}$.

Proof. We rearrange the following sum over $\mathfrak{u}$ using the auxiliary weights such that the $h = 0$ term is removed from the Fourier representation:

$$\begin{aligned}
\sum_{\mathfrak{u}\subseteq\{1:s\}} \gamma_{s,\mathfrak{u}} \prod_{j\in\mathfrak{u}} \theta_j(x_j) &= \sum_{\mathfrak{u}\subseteq\{1:s\}} \gamma_{s,\mathfrak{u}} \prod_{j\in\mathfrak{u}} \Big((\theta_j(x_j) - \widehat{\theta}_j(0)) + \widehat{\theta}_j(0)\Big) \\
&= \sum_{\mathfrak{u}\subseteq\{1:s\}} \gamma_{s,\mathfrak{u}} \sum_{\mathfrak{v}\subseteq\mathfrak{u}} \Big(\prod_{j\in\mathfrak{u}\setminus\mathfrak{v}} \widehat{\theta}_j(0)\Big) \prod_{j\in\mathfrak{v}} \Big(\theta_j(x_j) - \widehat{\theta}_j(0)\Big) \\
&= \sum_{\mathfrak{v}\subseteq\{1:s\}} \sum_{\mathfrak{v}\subseteq\mathfrak{u}\subseteq\{1:s\}} \gamma_{s,\mathfrak{u}} \Big(\prod_{j\in\mathfrak{u}\setminus\mathfrak{v}} C_{0,j}\Big) \prod_{j\in\mathfrak{v}} \Big(\theta_j(x_j) - \widehat{\theta}_j(0)\Big) \\
&= \sum_{\mathfrak{v}\subseteq\{1:s\}} \widetilde{\gamma}_{s,\mathfrak{v}} \prod_{j\in\mathfrak{v}} \Big(\theta_j(x_j) - \widehat{\theta}_j(0)\Big),
\end{aligned}$$

where $x_j = \{kz_j/n\}$. Thus (3.21) becomes

$$
\begin{aligned}
[e^{\mathrm{sh}}_{s,n}(\boldsymbol{z})]^2 &= -\sum_{\mathfrak{u}\subseteq\{1:s\}} \gamma_{s,\mathfrak{u}} \prod_{j\in\mathfrak{u}} C_{0,j} + \frac{1}{n}\sum_{k=1}^{n} \sum_{\mathfrak{u}\subseteq\{1:s\}} \gamma_{s,\mathfrak{u}} \prod_{j\in\mathfrak{u}} \theta_j\left(\left\{\frac{kz_j}{n}\right\}\right) \\
&= -\widetilde{\gamma}_{s,\emptyset} + \frac{1}{n}\sum_{k=1}^{n} \sum_{\mathfrak{v}\subseteq\{1:s\}} \widetilde{\gamma}_{s,\mathfrak{v}} \prod_{j\in\mathfrak{v}} \left[\theta_j\left(\left\{\frac{kz_j}{n}\right\}\right) - \widehat{\theta}_j(0)\right] \qquad (3.34)\\
&= \sum_{\emptyset\neq\mathfrak{v}\subseteq\{1:s\}} \widetilde{\gamma}_{s,\mathfrak{v}} \frac{1}{n}\sum_{k=1}^{n} \prod_{j\in\mathfrak{v}} \left[\theta_j\left(\left\{\frac{kz_j}{n}\right\}\right) - \widehat{\theta}_j(0)\right] \qquad (3.35)\\
&= \sum_{\emptyset\neq\mathfrak{v}\subseteq\{1:s\}} \frac{\widetilde{\gamma}_{s,\mathfrak{v}}}{n}\sum_{k=1}^{n} \prod_{j\in\mathfrak{v}} \sum_{h\in\mathbb{Z}\setminus\{0\}} \widehat{\theta}_j(h)\, e^{2\pi \mathrm{i} k h z_j/n} \\
&= \sum_{\emptyset\neq\mathfrak{v}\subseteq\{1:s\}} \frac{\widetilde{\gamma}_{s,\mathfrak{v}}}{n}\sum_{k=1}^{n} \sum_{\boldsymbol{h}\in(\mathbb{Z}\setminus\{0\})^{|\mathfrak{v}|}} \widehat{\theta}_{\mathfrak{v}}(\boldsymbol{h})\, e^{2\pi \mathrm{i} k \boldsymbol{h}\cdot\boldsymbol{z}_{\mathfrak{v}}/n}. \\
&= \sum_{\emptyset\neq\mathfrak{v}\subseteq\{1:s\}} \widetilde{\gamma}_{s,\mathfrak{v}} \sum_{\substack{\boldsymbol{h}\in(\mathbb{Z}\setminus\{0\})^{|\mathfrak{v}|}\\ \boldsymbol{h}\cdot\boldsymbol{z}_{\mathfrak{v}}\equiv_n 0}} \widehat{\theta}(\boldsymbol{h}). \qquad (3.36)
\end{aligned}
$$

Here, in the last step, we applied the following special case of Identity 1,

$$
\frac{1}{n}\sum_{k=1}^{n} e^{2\pi \mathrm{i} k \boldsymbol{h}\cdot\boldsymbol{z}_{\mathfrak{v}}/n} = \begin{cases} 1 & \text{if } \boldsymbol{h}\cdot\boldsymbol{z}_{\mathfrak{v}} \equiv_n 0, \\ 0 & \text{otherwise.} \end{cases}
$$

We then obtain the formula (3.36). □

In the unanchored case, the use of the auxiliary weights $\widetilde{\gamma}_{s,\mathfrak{v}}$ is unnecessary due to $\widehat{\theta}_j(0) = C_{0,j} = 0$ and $\widetilde{\gamma}_{s,\mathfrak{v}} = \gamma_{s,\mathfrak{v}}$. Hence we have the following lemma.

Lemma 18 *The worst-case error* (3.32) *for the unanchored space can be written as*

$$
[e^{\mathrm{sh}}_{s,n}(\boldsymbol{z})]^2 = \sum_{\emptyset\neq\mathfrak{u}\subseteq\{1:s\}} \gamma_{s,\mathfrak{u}} \sum_{\substack{\boldsymbol{h}\in(\mathbb{Z}\setminus\{0\})^{|\mathfrak{u}|}\\ \boldsymbol{h}\cdot\boldsymbol{z}_{\mathfrak{u}}\equiv_n 0}} \widehat{\theta}_{\mathfrak{u}}(\boldsymbol{h}).
$$

3.3.2 Error bound for the CBC construction

In the previous subsection we showed that the worst-case errors for the anchored and unanchored spaces can be written in the same form in terms of the Fourier coefficients of θ_j. In this subsection we provide the error analysis for randomly-shifted lattice rules constructed using the CBC algorithm. For each $d = 1, 2, \ldots, s$ and $\boldsymbol{z} \in \mathcal{Z}_n^d$, we consider the quantity

$$
E^2_{d,s}(\boldsymbol{z}) := \sum_{\emptyset\neq\mathfrak{v}\subseteq\{1:d\}} \widetilde{\gamma}_{s,\mathfrak{v}} \sum_{\substack{\boldsymbol{h}\in(\mathbb{Z}\setminus\{0\})^{|\mathfrak{v}|}\\ \boldsymbol{h}\cdot\boldsymbol{z}_{\mathfrak{v}}\equiv_n 0}} \widehat{\theta}_{\mathfrak{v}}(\boldsymbol{h}), \qquad (3.37)
$$

noting that the weights $\widetilde{\gamma}_{s,\mathfrak{v}}$ depend on s and not on d. Evidently $E^2_{s,s}(\boldsymbol{z}) = [e^{\mathrm{sh}}_{s,n}(\boldsymbol{z})]^2$, but in general $E^2_{d,s}(\boldsymbol{z}) \neq [e^{\mathrm{sh}}_{d,n}(\boldsymbol{z})]^2$ for $d < s$. A notable exception occurs in the unanchored

space: if the weights are independent of s, then $\widetilde{\gamma}_{s,\mathfrak{v}} = \gamma_{s,\mathfrak{v}} = \gamma_{\mathfrak{v}}$ and hence $E^2_{d,s}(\boldsymbol{z}) = [e^{\mathrm{sh}}_{d,n}(\boldsymbol{z})]^2$ for all $d \le s$.

Algorithm 19 (CBC Algorithm for unbounded spaces) *For any $n \in \mathbb{N}$ and $s \in \mathbb{N}$*

1. *Set $z_1 = 1$.*
2. *For each $d = 2, 3, \ldots, s$ with $z_1, \ldots, z_{d-1}$ fixed, choose $z_d \in \mathcal{Z}_n$ such that $E^2_{d,s}(z_1, \ldots, z_{d-1}, z_d)$ is minimised.*

We note briefly that the CBC algorithm presented here differs from Algorithm 12 in the search criteria. Here we must minimise $E_{d,s}(\boldsymbol{z})$, which allows for unbounded spaces as well as general weights. We have also generalised for lattices with any $n \in \mathbb{N}$, that is, we no longer restrict ourselves to only prime n, as we did in Chapter 2.

Let $\varphi(n)$ denote Euler's totient function, the size of the set $\mathcal{Z}_n$. We have the following result.

Theorem 20 (CBC error bound) *Consider either the anchored or unanchored space. Let $r_2 > 1/2$ be such that for each $j \in \{1:s\}$ we have some $C_{2,j} > 0$ and $r_{2,j} \ge r_2$ such that*

$$\widehat{\theta}_j(h) \le \frac{C_{2,j}}{|h|^{2r_{2,j}}} \quad \text{for all} \quad h \ne 0. \tag{3.38}$$

Then a generating vector $\boldsymbol{z}^* \in \mathcal{Z}_n^s$ can be constructed by Algorithm 19 so that for any $\lambda \in (1/(2r_2), 1]$ and for every $d \in \{1:s\}$ we have

$$E^2_{d,s}(z_1^*, \ldots, z_d^*) \le \left(\frac{1}{\varphi(n)} \sum_{\emptyset \ne \mathfrak{v} \subseteq \{1:d\}} \widetilde{\gamma}^\lambda_{s,\mathfrak{v}} \prod_{j \in \mathfrak{v}} \Big(2C^\lambda_{2,j} \zeta(2r_{2,j}\lambda) \Big) \right)^{1/\lambda}. \tag{3.39}$$

Proof. First we demonstrate the bound in (3.39) for $d = 1$. We have

$$\begin{aligned} E^2_{1,s}(1) &= \widetilde{\gamma}_{s,\{1\}} \sum_{\substack{h \in \mathbb{Z}\setminus\{0\} \\ h \equiv_n 0}} \widehat{\theta}_1(h) \le \widetilde{\gamma}_{s,\{1\}} \left(\sum_{\substack{h \in \mathbb{Z}\setminus\{0\} \\ h \equiv_n 0}} \frac{C^\lambda_{2,1}}{|h|^{2r_{2,1}\lambda}} \right)^{1/\lambda} \\ &= \widetilde{\gamma}_{s,\{1\}} \left(\frac{2C^\lambda_{2,1}\zeta(2r_{2,1}\lambda)}{n^{2r_{2,1}\lambda}} \right)^{1/\lambda} \le \widetilde{\gamma}_{s,\{1\}} \left(\frac{2C^\lambda_{2,1}\zeta(2r_{2,1}\lambda)}{\varphi(n)} \right)^{1/\lambda}, \end{aligned}$$

where we used (3.38) and Jensen's inequality $\sum_k a_k \le (\sum_k a_k^\lambda)^{1/\lambda}$ for all nonnegative a_k and $\lambda \in (1/(2r_2), 1]$, as well as $2r_{2,1}\lambda \ge 2r_2\lambda > 1$ and $\varphi(n) < n$.

Suppose now that (3.39) holds for some $d < s$, and we proceed to prove that the choice of z^*_{d+1} obtained from Algorithm 19 satisfies the same error bound (3.39), but with

d replaced by $d+1$. We split the worst-case error in $d+1$ dimensions according to whether $d+1 \in \mathfrak{v}$ or not,

$$\begin{aligned} E^2_{d+1,s}(z_1,\dots,z_d,z_{d+1}) &= \sum_{\emptyset\neq\mathfrak{v}\subseteq\{1:d+1\}} \widetilde{\gamma}_{s,\mathfrak{v}} \sum_{\substack{\boldsymbol{h}\in(\mathbb{Z}\setminus\{0\})^{|\mathfrak{v}|}\\ \boldsymbol{h}\cdot\boldsymbol{z}_{\mathfrak{v}}\equiv_n 0}} \widehat{\theta}_{\mathfrak{v}}(\boldsymbol{h}) \\ &= \sum_{\mathfrak{v}\subseteq\{1:d\}} \widetilde{\gamma}_{s,\mathfrak{v}} \sum_{\substack{\boldsymbol{h}\in(\mathbb{Z}\setminus\{0\})^{|\mathfrak{v}|}\\ \boldsymbol{h}\cdot\boldsymbol{z}_{\mathfrak{v}}\equiv_n 0}} \widehat{\theta}_{\mathfrak{v}}(\boldsymbol{h}) + \sum_{d+1\in\mathfrak{v}\subseteq\{1:d+1\}} \widetilde{\gamma}_{s,\mathfrak{v}} \sum_{\substack{\boldsymbol{h}\in(\mathbb{Z}\setminus\{0\})^{|\mathfrak{v}|}\\ \boldsymbol{h}\cdot\boldsymbol{z}_{\mathfrak{v}}\equiv_n 0}} \widehat{\theta}(\boldsymbol{h}) \\ &= E^2_{d,s}(z_1,\dots,z_d) + T_{d+1,s}(z_{d+1}), \end{aligned} \tag{3.40}$$

where

$$T_{d+1,s}(z_{d+1}) := \sum_{d+1\in\mathfrak{v}\subseteq\{1:d+1\}} \widetilde{\gamma}_{s,\mathfrak{v}} \sum_{\substack{\boldsymbol{h}\in(\mathbb{Z}\setminus\{0\})^{|\mathfrak{v}|}\\ \boldsymbol{h}\cdot\boldsymbol{z}_{\mathfrak{v}}\equiv_n 0}} \widehat{\theta}_{\mathfrak{v}}(\boldsymbol{h}).$$

The choice of $z^*_{d+1} \in \mathcal{Z}_n$ that minimises $E^2_{d+1}(\boldsymbol{z}, z_{d+1})$ is also the choice that minimises the $T_{d+1,s}(z_{d+1})$ term, hence we have $T_{d+1,s}(z^*_{d+1}) \le T_{d+1,s}(z_{d+1})$ for all $z_{d+1} \in \mathcal{Z}_n$. It then also holds that $T^\lambda_{d+1,s}(z^*_{d+1}) \le T^\lambda_{d+1,s}(z_{d+1})$ for all $\lambda \in (1/(2r_2), 1]$, and thus as z^*_{d+1} minimises the $T^\lambda_{d+1,s}(z_{d+1})$, it also beats the average

$$\begin{aligned} T^\lambda_{d+1,s}(z^*_{d+1}) &\le \frac{1}{\varphi(n)} \sum_{z_{d+1}\in\mathcal{Z}_n} T^\lambda_{d+1,s}(z_{d+1}). \\ &= \frac{1}{\varphi(n)} \sum_{z_{d+1}\in\mathcal{Z}_n} \left(\sum_{d+1\in\mathfrak{v}\subseteq\{1:d+1\}} \widetilde{\gamma}_{s,\mathfrak{v}} \sum_{\substack{\boldsymbol{h}\in(\mathbb{Z}\setminus\{0\})^{|\mathfrak{v}|}\\ \boldsymbol{h}\cdot\boldsymbol{z}_{\mathfrak{v}}\equiv_n 0}} \widehat{\theta}(\boldsymbol{h}) \right)^\lambda. \end{aligned}$$

We now apply Jensen's inequality to obtain

$$T^\lambda_{d+1,s}(z^*_{d+1}) \le \frac{1}{\varphi(n)} \sum_{z_{d+1}\in\mathcal{Z}_n} \sum_{d+1\in\mathfrak{v}\subseteq\{1:d+1\}} \widetilde{\gamma}^\lambda_{s,\mathfrak{v}} \sum_{\substack{\boldsymbol{h}\in(\mathbb{Z}\setminus\{0\})^{|\mathfrak{v}|}\\ \boldsymbol{h}\cdot\boldsymbol{z}_{\mathfrak{v}}\equiv_n 0}} [\widehat{\theta}_{\mathfrak{v}}(\boldsymbol{h})]^\lambda.$$

Next we split the sum over $\boldsymbol{h}$ depending on whether or not h_{d+1} is a multiple of n, and use (3.38),

$$
\begin{aligned}
&T_{d+1,s}^{\lambda}(z_{d+1}^{*}) \\
&\le \frac{1}{\varphi(n)} \sum_{z_{d+1}\in\mathcal{Z}_n} \sum_{d+1\in\mathfrak{v}\subseteq\{1:d+1\}} \widetilde{\gamma}_{s,\mathfrak{v}}^{\lambda} \Bigg(\sum_{\substack{h_{d+1}\in\mathbb{Z}\setminus\{0\}\\ h_{d+1}\equiv_n 0}} [\widehat{\theta}_{d+1}(h_{d+1})]^{\lambda} \sum_{\substack{\boldsymbol{h}\in(\mathbb{Z}\setminus\{0\})^{|\mathfrak{v}|-1}\\ \boldsymbol{h}\cdot\boldsymbol{z}_{\mathfrak{v}\setminus\{d+1\}}\equiv_n 0}} [\widehat{\theta}_{\mathfrak{v}}(\boldsymbol{h})]^{\lambda} \\
&\qquad\qquad + \sum_{\substack{h_{d+1}\in\mathbb{Z}\setminus\{0\}\\ h_{d+1}\not\equiv_n 0}} [\widehat{\theta}_{d+1}(h_{d+1})]^{\lambda} \sum_{\substack{\boldsymbol{h}\in(\mathbb{Z}\setminus\{0\})^{|\mathfrak{v}|-1}\\ \boldsymbol{h}\cdot\boldsymbol{z}_{\mathfrak{v}\setminus\{d+1\}}\equiv_n -h_{d+1}z_{d+1}}} [\widehat{\theta}_{\mathfrak{v}}(\boldsymbol{h})]^{\lambda} \Bigg) \\
&\le \sum_{d+1\in\mathfrak{v}\subseteq\{1:d+1\}} \widetilde{\gamma}_{s,\mathfrak{v}}^{\lambda} \Bigg(\sum_{\substack{h_{d+1}\in\mathbb{Z}\setminus\{0\}\\ h_{d+1}\equiv_n 0}} \frac{C_{2,d+1}^{\lambda}}{|h_{d+1}|^{2r_{2,d+1}\lambda}} \sum_{\substack{\boldsymbol{h}\in(\mathbb{Z}\setminus\{0\})^{|\mathfrak{v}|-1}\\ \boldsymbol{h}\cdot\boldsymbol{z}_{\mathfrak{v}\setminus\{d+1\}}\equiv_n 0}} \prod_{j\in\mathfrak{v}\setminus\{d+1\}} \frac{C_{2,j}^{\lambda}}{|h_j|^{2r_{2,j}\lambda}} \\
&+ \frac{1}{\varphi(n)} \sum_{z_{d+1}\in\mathcal{Z}_n} \sum_{\substack{h_{d+1}\in\mathbb{Z}\setminus\{0\}\\ h_{d+1}\not\equiv_n 0}} \frac{C_{2,d+1}^{\lambda}}{|h_{d+1}|^{2r_{2,d+1}\lambda}} \sum_{\substack{\boldsymbol{h}\in(\mathbb{Z}\setminus\{0\})^{|\mathfrak{v}|-1}\\ \boldsymbol{h}\cdot\boldsymbol{z}_{\mathfrak{v}\setminus\{d+1\}}\equiv_n -h_{d+1}z_{d+1}}} \prod_{j\in\mathfrak{v}\setminus\{d+1\}} \frac{C_{2,j}^{\lambda}}{|h_j|^{2r_{2,j}\lambda}} \Bigg). \qquad (3.41)
\end{aligned}
$$

For the first term inside the brackets in (3.41), we have

$$
\sum_{\substack{h_{d+1}\in\mathbb{Z}\setminus\{0\}\\ h_{d+1}\equiv_n 0}} \frac{C_{2,d+1}^{\lambda}}{|h_{d+1}|^{2r_{2,d+1}\lambda}} = \frac{2C_{2,d+1}^{\lambda}\zeta(2r_{2,d+1}\lambda)}{n^{2r_{2,d+1}\lambda}},
$$

and we write

$$
B := \sum_{\substack{\boldsymbol{h}\in(\mathbb{Z}\setminus\{0\})^{|\mathfrak{v}|-1}\\ \boldsymbol{h}\cdot\boldsymbol{z}_{\mathfrak{v}\setminus\{d+1\}}\equiv_n 0}} \prod_{j\in\mathfrak{v}\setminus\{d+1\}} \frac{C_{2,j}^{\lambda}}{|h_j|^{2r_{2,j}\lambda}}.
$$

The second term inside the brackets in (3.41) can be rewritten as

$$
\frac{1}{\varphi(n)} \sum_{c=1}^{n-1} \sum_{z_{d+1}\in\mathcal{Z}_n} \sum_{\substack{h_{d+1}\in\mathbb{Z}\setminus\{0\}\\ h_{d+1}\equiv_n -cz_{d+1}^{-1}}} \frac{C_{2,d+1}^{\lambda}}{|h_{d+1}|^{2r_{2,d+1}\lambda}} \sum_{\substack{\boldsymbol{h}\in(\mathbb{Z}\setminus\{0\})^{|\mathfrak{v}|-1}\\ \boldsymbol{h}\cdot\boldsymbol{z}_{\mathfrak{v}\setminus\{d+1\}}\equiv_n c}} \prod_{j\in\mathfrak{v}\setminus\{d+1\}} \frac{C_{2,j}^{\lambda}}{|h_j|^{2r_{2,j}\lambda}}. \qquad (3.42)
$$

Now note that for any $c\in\{1,\dots,n-1\}$, we have equality of the two sets $\{cz_{d+1}^{-1} \pmod n : z_{d+1}\in\mathcal{Z}_n\} = \{cz \pmod n : z\in\mathcal{Z}_n\}$. Furthermore if we let $p=\gcd(c,n)$, then

$\gcd(c/p, n/p) = 1$ and we have

$$\begin{aligned}
&\sum_{z_{d+1}\in\mathcal{Z}_n} \sum_{\substack{h_{d+1}\in\mathbb{Z}\setminus\{0\}\\ h_{d+1}\equiv_n -cz_{d+1}^{-1}}} \frac{C_{2,d+1}^{\lambda}}{|h_{d+1}|^{2r_{2,d+1}\lambda}} = \sum_{z\in\mathcal{Z}_n}\sum_{m\in\mathbb{Z}} \frac{C_{2,d+1}^{\lambda}}{|mn-cz|^{2r_{2,d+1}\lambda}} \\
&= p^{-2r_{2,d+1}\lambda} \sum_{z\in\mathcal{Z}_n}\sum_{m\in\mathbb{Z}} \frac{C_{2,d+1}^{\lambda}}{|m(n/p)-(c/p)z|^{2r_{2,d+1}\lambda}} \\
&= p^{-2r_{2,d+1}\lambda} \sum_{z\in\mathcal{Z}_n} \sum_{\substack{h\in\mathbb{Z}\{0\}\\ h\equiv_{n/p}-(c/p)z}} \frac{C_{2,d+1}^{\lambda}}{|h|^{2r_{2,d+1}\lambda}} \\
&\leq p^{-2r_{2,d+1}\lambda}\, p \sum_{z=1}^{n/p-1} \sum_{\substack{h\in\mathbb{Z}\{0\}\\ h\equiv_{n/p} z}} \frac{C_{2,d+1}^{\lambda}}{|h|^{2r_{2,d+1}\lambda}} \\
&= C_{2,d+1}^{\lambda} p^{1-2r_{2,d+1}\lambda} \left(2\zeta(2r_{2,d+1}\lambda) - \frac{2\zeta(2r_{2,d+1}\lambda)}{(n/p)^{2r_{2,d+1}\lambda}}\right) \\
&\leq 2C_{2,d+1}^{\lambda}\zeta(2r_{2,d+1}\lambda)\left(1-\frac{1}{n^{2r_{2,d+1}\lambda}}\right). \qquad (3.43)
\end{aligned}$$

Note that as $\lambda \in (1/(2r_2), 1]$ and $r_{2,d+1} \geq r_2$, we know that $\zeta(2r_{2,d+1}\lambda) < \infty$ and $p^{1-2r_{2,d+1}\lambda} \leq 1$. Since the estimate (3.43) is independent of c, we can express the remaining factor in (3.42) as

$$\begin{aligned}
\sum_{c=1}^{n-1} \sum_{\substack{\boldsymbol{h}\in(\mathbb{Z}\setminus\{0\})^{|\mathfrak{v}|-1}\\ \boldsymbol{h}\cdot\boldsymbol{z}_{\mathfrak{v}\setminus\{d+1\}}\equiv_n c}} \prod_{j\in\mathfrak{v}\setminus\{d+1\}} \frac{C_{2,j}^{\lambda}}{|h_j|^{2r_{2,j}\lambda}} &= \sum_{\substack{\boldsymbol{h}\in(\mathbb{Z}\setminus\{0\})^{|\mathfrak{v}|-1}\\ \boldsymbol{h}\cdot\boldsymbol{z}_{\mathfrak{v}\setminus\{d+1\}}\not\equiv_n 0}} \prod_{j\in\mathfrak{v}\setminus\{d+1\}} \frac{C_{2,j}^{\lambda}}{|h_j|^{2r_{2,j}\lambda}} \\
&= \prod_{j\in\mathfrak{v}\setminus\{d+1\}} \Big(2C_{2,j}^{\lambda}\zeta(2r_{2,j}\lambda)\Big) - B.
\end{aligned}$$

Combining these elements back into (3.41), we obtain

$$\begin{aligned}
T_{d+1,s}^{\lambda}(z_{d+1}^*) &\leq \sum_{d+1\in\mathfrak{v}\subseteq\{1:d+1\}} \widetilde{\gamma}_{s,\mathfrak{v}}^{\lambda} \left(\frac{2C_{2,d+1}^{\lambda}\zeta(2r_{2,d+1}\lambda)}{n^{2r_{2,d+1}\lambda}} B \right. \\
&\quad \left. + \frac{2C_{2,d+1}^{\lambda}\zeta(2r_{2,d+1}\lambda)}{\varphi(n)}\left(1-\frac{1}{n^{2r_{2,d+1}\lambda}}\right)\left(\prod_{j\in\mathfrak{v}\setminus\{d+1\}}\Big(2C_{2,j}^{\lambda}\zeta(2r_{2,j}\lambda)\Big) - B\right)\right) \\
&\leq \frac{1}{\varphi(n)} \sum_{d+1\in\mathfrak{v}\subseteq\{1:d+1\}} \widetilde{\gamma}_{s,\mathfrak{v}}^{\lambda} \prod_{j\in\mathfrak{v}} \Big(2C_{2,j}^{\lambda}\zeta(2r_{2,j}\lambda)\Big). \qquad (3.44)
\end{aligned}$$

We assume by the inductive hypothesis that there is a particular $\boldsymbol{z}^* \in \mathbb{Z}^s$ for which (3.39) holds. We know that there is a particular z_{d+1}^* for which (3.44) holds, and combining

this with (3.39) and (3.40), we obtain

$$\begin{aligned}
E_{d+1}^2(\boldsymbol{z}^*, z_{d+1}^*) &= E_s^2(\boldsymbol{z}^*) + T_{d+1,s}(z_{d+1}^*) \\
&\le \frac{1}{[\varphi(n)]^{1/\lambda}} \left(\sum_{\emptyset \ne \mathfrak{v} \subseteq \{1:d\}} \widetilde{\gamma}_{s,\mathfrak{v}}^{\lambda} \prod_{j \in \mathfrak{v}} 2C_{2,j}^{\lambda} \zeta(2r_{2,j}\lambda) \right)^{1/\lambda} \\
&\quad + \frac{1}{[\varphi(n)]^{1/\lambda}} \left(\sum_{d+1 \in \mathfrak{v} \subseteq \{1:d+1\}} \widetilde{\gamma}_{s,\mathfrak{v}}^{\lambda} \prod_{j \in \mathfrak{v}} 2C_{2,j}^{\lambda} \zeta(2r_{2,j}\lambda) \right)^{1/\lambda} \\
&\le \frac{1}{[\varphi(n)]^{1/\lambda}} \left(\sum_{\emptyset \ne \mathfrak{v} \subseteq \{1:d+1\}} \widetilde{\gamma}_{s,\mathfrak{v}}^{\lambda} \prod_{j \in \mathfrak{v}} 2C_{2,j}^{\lambda} \zeta(2r_{2,j}\lambda) \right)^{1/\lambda},
\end{aligned}$$

for any $\lambda \in (1/(2r_2), 1]$. Again we have made use of Jensen's inequality. Thus by induction (3.39) holds for any $d \in \{1:s\}$. This completes the proof. □

Theorem 21 *Suppose that f belongs to the anchored or unanchored space for some weight parameters $\gamma_{s,\mathfrak{u}}$ and weight functions ψ_j, and suppose that (3.38) holds for constants $C_{2,j} > 0$ and $r_{2,j} \ge r_2 > 1/2$. Then a generating vector $\boldsymbol{z}^* \in \mathcal{Z}_n^s$ for a randomly-shifted lattice rule can be constructed by a CBC algorithm such that, for all $\lambda \in (1/(2r_2), 1]$,*

$$\begin{aligned}
&\sqrt{\mathbb{E}^{\boldsymbol{\Delta}} |I_{s,\phi}(f) - Q_{s,n}(f \circ \Phi^{-1})|^2} \\
&\le \left(\frac{1}{\varphi(n)} \sum_{\emptyset \ne \mathfrak{u} \subseteq \{1:s\}} \gamma_{s,\mathfrak{u}}^{\lambda} \prod_{j \in \mathfrak{u}} \Big(C_{0,j}^{\lambda} + 2C_{2,j}^{\lambda} \zeta(2r_{2,j}\lambda) \Big) \right)^{1/(2\lambda)} \|f\|_{\mathcal{F}}, \qquad (3.45)
\end{aligned}$$

where the expectation is taken with respect to the random shift which is uniformly distributed on $[0,1]^s$, and $C_{0,j}$ is given by (3.16) for the anchored variant and $C_{0,j} = 0$ for the unanchored variant.

Proof. From (3.8) and (3.11) we see that

$$\begin{aligned}
\mathbb{E}^{\boldsymbol{\Delta}} |I_{s,\phi}(f) - Q_{s,n}(f \circ \Phi^{-1})|^2 &\le \int_{[0,1]^s} [e_{s,n}(Q_{s,n}; \mathcal{F})]^2 \, \|f\|_{\mathcal{F}}^2 \, \mathrm{d}\boldsymbol{\Delta} \\
&= [e_{s,n}^{\mathrm{sh}}(\boldsymbol{z})]^2 \, \|f\|_{\mathcal{F}}^2.
\end{aligned}$$

Substituting (3.33) into (3.39) and applying Jensen's inequality, we obtain

$$\begin{aligned}
[e_{s,n}^{\mathrm{sh}}(\boldsymbol{z}^*)]^2 &= E_{s,s}^2(\boldsymbol{z}^*) \\
&\le \left(\frac{1}{\varphi(n)} \sum_{\emptyset\ne\mathfrak{v}\subseteq\{1:s\}} \sum_{\mathfrak{v}\subseteq\mathfrak{u}\subseteq\{1:s\}} \gamma_{s,\mathfrak{u}}^\lambda \Big(\prod_{j\in\mathfrak{v}\setminus\mathfrak{u}} C_{0,j}^\lambda\Big) \prod_{j\in\mathfrak{v}} \Big(2C_{2,j}^\lambda \zeta(2r_{2,j}\lambda)\Big)\right)^{1/\lambda} \\
&= \left(\frac{1}{\varphi(n)} \sum_{\emptyset\ne\mathfrak{u}\subseteq\{1:s\}} \gamma_{s,\mathfrak{u}}^\lambda \sum_{\mathfrak{v}\subseteq\mathfrak{u}} \Big(\prod_{j\in\mathfrak{v}\setminus\mathfrak{u}} C_{0,j}^\lambda\Big) \prod_{j\in\mathfrak{v}} \Big(2C_{2,j}^\lambda \zeta(2r_{2,j}\lambda)\Big)\right)^{1/\lambda}, \\
&= \left(\frac{1}{\varphi(n)} \sum_{\emptyset\ne\mathfrak{u}\subseteq\{1:s\}} \gamma_{s,\mathfrak{u}}^\lambda \prod_{j\in\mathfrak{u}} \Big(C_{0,j}^\lambda + 2C_{2,j}^\lambda \zeta(2r_{2,j}\lambda)\Big)\right)^{1/\lambda},
\end{aligned}$$

which equals the square of the first factor in (3.45). □

The CBC construction and error bound we have presented thus far depend on the final dimension s. If the weights are independent of s, i.e., $\gamma_{s,\mathfrak{u}} = \gamma_{\mathfrak{u}}$, and if

$$\sum_{|\mathfrak{u}|<\infty} \gamma_{\mathfrak{u}}^\lambda \prod_{j\in\mathfrak{u}} \Big(C_{0,j}^\lambda + 2C_{2,j}^\lambda \zeta(2r_{2,j}\lambda)\Big) < \infty, \tag{3.46}$$

then for every s the CBC algorithm yields a generating vector $\boldsymbol{z}^*$ for which $e_{s,n}^{\mathrm{sh}}(\boldsymbol{z}^*) = \mathcal{O}(n^{-1/(2\lambda)})$, with the implied constant independent of s. If the weights are of a product form, i.e., $\gamma_{\mathfrak{u}} = \prod_{j\in\mathfrak{u}} \gamma_j$, then the condition (3.46) simplifies to $\sum_{j=1}^\infty \gamma_j^\lambda < \infty$, as seen in earlier papers.

In the unanchored space with weights $\gamma_{\mathfrak{u}}$ independent of s, since there is no need for auxiliary weights (we have $C_{0,j} = 0$ and $\widetilde{\gamma}_{s,\mathfrak{v}} = \gamma_{\mathfrak{v}}$), the CBC algorithm actually works directly with $[e_{d,n}^{\mathrm{sh}}(\boldsymbol{z})]^2$, and the resulting generating vector is *extensible* in dimension. That is, if a fixed $\boldsymbol{z} \in \mathcal{Z}_n^d$ minimizes $[e_{d,n}^{\mathrm{sh}}(\boldsymbol{z})]^2$, then there is some $z_{d+1} \in \mathcal{Z}_n$ such that $(\boldsymbol{z}, z_{d+1})$ will minimize $[e_{d+1,n}^{\mathrm{sh}}(\boldsymbol{z}, z_{d+1})]^2$. For general non-product weights $\gamma_{\mathfrak{u}}$, the cost of the CBC algorithm can be prohibitively expensive. In §5.1 we will discuss the fast CBC implementation for POD weights (2.23).

In the anchored space, however, for Theorem 20 to be valid the CBC algorithm must work with the auxiliary quantity $E_{d,s}^2(\boldsymbol{z})$ which involves the auxiliary weights $\widetilde{\gamma}_{s,\mathfrak{v}}$. Even if the original weights $\gamma_{\mathfrak{u}}$ are independent of s, the auxiliary weights $\widetilde{\gamma}_{s,\mathfrak{v}}$ needed in the CBC construction still depend on s by definition, see (3.33). Thus the resulting generating vector is *not extensible* in dimension, even though the error bound can be independent of s when (3.46) holds. This means that unique generating vectors $\boldsymbol{z} \in \mathcal{Z}_n^s$ must be built from the bottom up for each s, using the CBC construction. We stress that an implementation based on minimizing $[e_{d,n}^{\mathrm{sh}}(\boldsymbol{z})]^2$ in each step, although intuitively sound, cannot be justified by Theorem 20. Unfortunately, even if the original weights have some nice structure such as POD weights, this structure is not preserved by the auxiliary weights. A method of tackling this issue for the anchored space with POD weights will be discussed §5.1.2.

3.3.3 *Examples of ψ_j and ϕ*

In [39] a study is undertaken for various combinations of the weight function ψ_j and probability density ϕ. In particular it is examined whether conditions (3.9) and (3.10) are satisfied, and then rates of decay of $\widehat{\theta}_j(h)$ are calculated. For the anchored space it was shown in [39] that

$$\widehat{\theta}_j(h) = \frac{1}{\pi^2 h^2} \int_a^b \frac{1}{\psi_j^2(t)} \sin^2(\pi h \Phi(t)) \, \mathrm{d}t \qquad \text{for} \quad h \neq 0. \tag{3.47}$$

We see from (3.30) that the function θ_j for the unanchored space only differs from the anchored case by a constant. Hence the formula (3.47) also applies in the unanchored space.

We present in Table 3.1 summary of a number useful probability densities ϕ. Some of these choices have been used in applications of the QMC theory. We leave a free parameter ν to be able to adjust the distributions.

Table 3.1: Various possible choices for $\phi(y)$

Distribution	Formula
Normal	$\phi_{\mathrm{nor},\nu}(y) = \frac{e^{-y^2/2\nu}}{\sqrt{2\pi\nu}}$
Logistic	$\phi_{\mathrm{logit},\nu}(y) = \frac{e^{y/\nu}}{\nu(1+e^{y/\nu})^2}$
Exponential	$\phi_{\mathrm{exp},\nu}(y) = \frac{e^{-\lvert y\rvert/\nu}}{2\nu}$
Student	$\phi_{\mathrm{stu},\nu}(y) = \frac{1}{\sqrt{\nu\pi}} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \left(1 + \frac{y^2}{\nu}\right)^{-(\nu+1)/2}$
Rational	$\phi_{\mathrm{rat},\nu}(y) = \frac{\nu}{2}(1 + \lvert y\rvert)^{-(\lvert\nu\rvert+1)}$

In Table 3.2 present a summary of the conditions (3.9) and (3.10) as well as estimates of $r_{2,j}$ for selected combinations of ϕ from Table 3.1 and ψ_j. Full details of these calculations can be found in [39], with the exception that the cases with $\psi_j(y) = 1$ are given in [57]. The asterisks in $r^*_{2,j}$ in Table 3.2 mark those cases where matching lower bounds on $\widehat{\theta}_j(h)$ have been obtained (up to $\delta > 0$), indicating that those estimates on $r_{2,j}$ are sharp.

Full details of the calculations of bounds of $\hat{\theta}_j(h)$ and hence the constants $C_{2,j}$, $C_{3,j}$, $r_{2,j}$, and $r_{3,j}$ for most of the choices in Table 3.2 can be found in [39]. These calculations were not, however, performed for the cases where $\psi_j(y) = 1$. We present these calculations here in the Examples 23-25.

First, we can see that due to symmetry of all our choices of ϕ, we can write

$$\hat{\theta}_j(h) = \frac{2}{\pi^2 h^2} \int_0^{1/2} \frac{\sin^2(\pi h u)}{\psi_j(\Phi^{-1}(u))\phi(\Phi^{-1}(u))} \, \mathrm{d}u.$$

In the examples below we make use of the following lemma

Table 3.2: Selected combinations of the probability density ϕ and weight function ψ_j. Do conditions (3.9) and (3.10) hold? What are the estimates for $r_{2,j}$? The asterisk in $r^*_{2,j}$ below indicates a sharp estimate of $r_{2,j}$.

	$\phi_{\mathrm{nor},\nu}(y)$	$\phi_{\mathrm{logit},\nu}(y)$ or $\phi_{\mathrm{exp},\nu}(y)$	$\phi_{\mathrm{stu},\nu}(y)$ or $\phi_{\mathrm{rat},\nu}(y)$
$\psi_j(y) = e^{-y^2/(2\alpha)}$	Yes if $\alpha > 2\nu$ $r_{2,j} = 1 - \frac{\nu}{\alpha}$	-	-
$\psi_j(y) = e^{-\lvert y\rvert/\alpha}$	Yes $r_{2,j} = 1-\delta,\ \forall \delta \in (0, \frac{1}{2})$	Yes if $\alpha > 2\nu$ $r^*_{2,j} = 1 - \frac{\nu}{\alpha}$	-
$\psi_j(y) = (1+\lvert y\rvert)^{-\alpha}$	Yes $r_{2,j} = 1-\delta,$ $\forall \delta \in (0, \min(\frac{1}{2}, \frac{9}{8}\alpha\nu))$	Yes $r^*_{2,j} = 1-\delta,$ $\forall \delta \in (0, \min(\frac{1}{2}, \alpha\nu))$	Yes if $2\alpha + 1 < \nu$ $r^*_{2,j} = 1 - \frac{2\alpha+1}{2\nu}$
$\psi_j(y) = 1$	Yes $r_{2,j} = 1-\delta,\ \forall \delta \in (0, \frac{1}{2})$	Yes $r^*_{2,j} = 1-\delta,\ \forall \delta \in (0, \frac{1}{2})$	Yes if $\nu > 1$ $r^*_{2,j} = 1 - \frac{1}{2\nu}$

Lemma 22 *For any $h \geq 1$, we have*

$$\int_0^{1/2} \frac{\sin^2(\pi h u)}{u}\,\mathrm{d}u \;\leq\; \frac{1}{2} + \ln\left(\frac{\pi h}{2}\right)$$

and for any $0 < b \leq 1$,

$$\int_0^{b/2} \frac{\sin^2(\pi h u)}{u}\,\mathrm{d}u \;\geq\; \frac{b^2}{2}$$

Proof. The first inequality we use $\sin^2(\pi h u) \leq (\pi h u)^2$ for $u \in [0, 1/(\pi h)]$ and $\sin^2(\pi h u) \leq 1$ for $u \in [1/(\pi h), 1]$,

$$\begin{aligned}\int_0^{1/2} \frac{\sin^2(\pi h u)}{u}\,\mathrm{d}u \;&\leq\; \int_0^{1/(\pi h)} \pi^2 h^2 u\,\mathrm{d}u \;+\; \int_{1/(\pi h)}^{1/2} u^{-1}\,\mathrm{d}u \\ &=\; \frac{1}{2} + \ln\left(\frac{\pi h}{2}\right)\end{aligned}$$

For the lower bound, we use Lemma 3 in [39] and take $a = 0$. □

Example 23 We take the combination

$$\phi(y) = \phi_{\mathrm{nor},\nu}(y), \qquad \psi_j = 1.$$

We see that

$$\hat{\theta}_j(h) = \frac{2\sqrt{2\pi\nu}}{\pi^2 h^2} \int_0^{1/2} \exp\left(\frac{(\Phi^{-1}_{\mathrm{nor},\nu}(u))^2}{2\nu}\right) \sin^2(\pi h u)\,\mathrm{d}u$$
$$\leq \frac{2\sqrt{2\pi\nu}}{\pi^2 h^2} \int_0^{1/2} \frac{\sin^2(\pi h u)}{u}\,\mathrm{d}u \leq \frac{2\sqrt{2\pi\nu}}{\pi^2 h^2}\left(\frac{1}{2} + \ln\left(\frac{\pi h}{2}\right)\right),$$

where we used $\exp\left(\frac{(\Phi^{-1}_{\mathrm{nor},\nu}(u))^2}{2\nu}\right) < \frac{1}{u}$ for $u \in (0, 1/2)$, as demonstrated in Example 4 of [39]. Thus for any $\delta \in (0, 1/2)$, using $\ln x \leq x^\delta/\delta$, we can take

$$C_{2,j} = \frac{2^{3/2-\delta}\,\nu^{1/2}}{\delta\,\pi^{3/2-\delta}\,h^{2-\delta}\,e^{1-2\delta}} \qquad r_{2,j} = 1 - \delta.$$

A useful lower bound proved elusive.

Example 24 Now consider the pair

$$\phi(y) = \phi_{\mathrm{logit},\nu}(y), \qquad \psi_j = 1.$$

From the definition in Table 3.1 we see that for $u \in (0, 1/2]$

$$\phi_{\mathrm{logit},\nu}(\Phi^{-1}_{\mathrm{logit},\nu}(u)) = \frac{u(1-u)}{\nu}.$$

Thus, using $1/2 \leq 1 - u \leq 1$ for $u \in (0, 1/2)$,

$$\frac{2}{\pi^2 h^2} \int_0^{1/2} u^{-1} \sin^2(\pi h u)\,\mathrm{d}u \leq \hat{\theta}_j(h) \leq \frac{4}{\pi^2 h^2} \int_0^{1/2} u^{-1} \sin^2(\pi h u)\,\mathrm{d}u,$$

and using $b = 1$ in Lemma 22 we obtain, for any $\delta > 0$

$$C_{3,j} = \frac{1}{2\pi^2} \qquad \text{and} \qquad r_{3,j} = 1.$$

and we see that $C_{2,j}$ and $r_{2,j}$ can be derived in exactly the same way as Example 23. The same results can be obtained for $r_{2,j}$ and $r_{3,j}$ if we consider $\phi(y) = \phi_{\exp,\nu}(y)$.

Example 25 Finally, consider the pair

$$\phi(y) = \phi_{\mathrm{rat},\nu}(y), \qquad \psi_j = 1.$$

Again, using explicit formulas in Table 3.1, we find, assuming $\nu > 1$

$$\phi_{\mathrm{rat},\nu}(\Phi^{-1}_{\mathrm{rat},\nu}(u)) = \frac{2}{\nu}(2u)^{1+1/\nu},$$

hence we have that

$$\hat{\theta}_j(h) = \frac{2^{1-1/\nu}}{\pi^2 h^2 \nu} \int_0^{1/2} u^{-(1+1/\nu)} \sin^2(\pi h u)\,\mathrm{d}u,$$

Using Lemma 3 from [39], we can derive

$$C_{2,j} = \frac{1}{2 - 1/\nu} \left(\frac{2}{\pi} \right)^{2-1/\nu}, \quad C_{3,j} = \frac{2}{(2/\nu - 1)\pi^2}, \quad \text{and} \quad r_{2,j} = r_{3,j} = 1 - \frac{1}{2\nu}$$

Relationships between $\phi_{\text{stu},\nu}(y)$ and $\phi_{\text{rat},\nu}(y)$, derived in [39], shows that we obtain the same values for $r_{2,j}$ and $r_{3,j}$, but different $C_{2,j}$ and $C_{3,j}$, when we consider $\phi(y) = \phi_{\text{stu},\nu}(y)$.

Chapter 4

The porous flow problem

In this chapter we analyse the porous-flow problem first presented in the introduction. We are concerned with modelling the behaviour of single-phase fluid flow in saturated media, where the permeability underlying medium is allowed to be a random field.

This model is primarily intended to simulate the flow of fluids through porous rock or soils in the earth's crust. We are motivated to understand the dynamics of fluids in these media for various endeavours, particularly for understanding the spread of pollutants within groundwater deposits, or for management of water resources.

The model is based on Darcy's law

$$q(\boldsymbol{x}) \;=\; -a(\boldsymbol{x})(\nabla u(\boldsymbol{x}))$$

coupled with the incompressibility constraint,

$$\nabla \cdot q(\boldsymbol{x}) \;=\; 0$$

Here q represents the Darcy flux, u is the residual pressure field, and a the permeability of the underlying medium. In our model we are interested in letting the field a be a homogeneous random field, reflecting the complex nature of rock or soil formations. Rather than having constant permeability, these formations are known to have strong spatial variation in permeability, a property we call *heterogeneity*. This property is also known to apply on a variety of scales of measurement. In practice it is impossible to measure exactly the permeability of a given medium with heterogeneity. Further, it is known that taking sparsely spaced measurements and performing some sort of interpolation yields poor results that do not reflect the reality of the medium being simulated. Thus we use random fields, and assume that a single realisation of the field bears some similarity to the permeability field of our rock formation. We then ensure that statistically our random field model bears a resemblance to the permeability field, and our simulation consists of studying statistical properties of the solutions to the Darcy flow equations.

This leads us to the following PDE, which is the focus of interest in this chapter,

$$-\nabla \cdot (a(\boldsymbol{x}, \boldsymbol{y})\nabla u(\boldsymbol{x}, \boldsymbol{y})) \;=\; f(\boldsymbol{x})\,, \qquad \text{for } \boldsymbol{y} \in \mathbb{R}^{\mathbb{N}} \text{ and } \boldsymbol{x} \in D \tag{4.1}$$

where D is a bounded spatial domain in $\mathbb{R}^d$, $d = 1, 2,$ or 3, and $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}), \rho_G)$ is our parameterised probability space of sequences $\boldsymbol{y}$, to be discussed shortly. Here we take the

boundary conditions to be homogeneous Dirichlet conditions. The field here is taken to be of the form

$$a(\boldsymbol{x}, \boldsymbol{y}) := a_*(\boldsymbol{x}) + a_0(\boldsymbol{x}) \exp(Z(\boldsymbol{x}, \boldsymbol{y})), \tag{4.2}$$

where and a_*, a_0 are given functions that are continuous on $\overline{D}$ with a_* non-negative and a_0 strictly positive on $\overline{D}$, and

$$Z(\boldsymbol{x}, \boldsymbol{y}) := \sum_{j=1}^{\infty} \sqrt{\mu_j}\, \xi_j(\boldsymbol{x})\, y_j, \tag{4.3}$$

where μ_j are real, positive and non-increasing in j, and we assume the ξ_j are orthonormal in $L^2(D)$. We discuss the conditions for point-wise convergence of this expression later in Lemma 26. The ξ_j, in some sense, can be considered to be "basis" functions for construction of the random field. The y_j are i.i.d. zero-mean Gaussian normal (i.e. $\mathcal{N}(0,1)$) variables, and our probability space is $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}), \rho_G)$, where $\mathcal{B}(\mathbb{R}^{\mathbb{N}})$ is the Borel sigma-algebra on $\mathbb{R}^{\mathbb{N}}$, taken here to be the sigma-algebra generated by countable products of intervals $I \in \mathcal{B}(\mathbb{R})$. The measure ρ_G is the product Gaussian measure

$$\rho_G = \bigotimes_{j=1}^{\infty} \mathcal{N}(0,1). \tag{4.4}$$

We see that ρ_G is a Gaussian measure on an infinite dimensional space, in the sense that for any bounded linear functional $\ell : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$, the push-forward measure defined on $\mathcal{B}(\mathbb{R})$ by $(\ell_* \rho_G)(A) = \rho_G(\ell^{-1}(A))$, where $A \in \mathcal{B}(\mathbb{R})$, is a Gaussian measure on $\mathbb{R}$ with zero mean. We note also that $Z(\boldsymbol{x}, \cdot)$ is a Gaussian random variable for any $\boldsymbol{x} \in D$ with zero mean, and indeed $(Z(\boldsymbol{x}_1, \cdot), \ldots, Z(\boldsymbol{x}_k, \cdot))$ is a Gaussian random vector for any collection of points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k \in D$. Thus, in accordance with the standard definition, we call $Z : D \times \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ a *Gaussian field*, see e.g. [2, §1.6]. Thus the field a is said to be a *lognormal random field*.

We note briefly that, in contrast to our "parametrically defined" field, as presented in (4.2) and (4.3), most of the literature presents random fields where Z is specified as a Gaussian field and given covariance structure. The covariance of a Gaussian field is represented by a function, $c(\boldsymbol{x}_0, \boldsymbol{x}_1) = \mathbb{E}[Z(\boldsymbol{x}_0, \cdot)\, Z(\boldsymbol{x}_1, \cdot)]$. We assume c to be continuous on $\overline{D} \times \overline{D}$ and note that $c(\boldsymbol{x}_0, \boldsymbol{x}_1) = c(\boldsymbol{x}_1, \boldsymbol{x}_0)$. The Karhunen–Loève expansion then allows us to find the parameters μ_j and ξ_j which can be said to be the eigenvalues and eigenvectors, respectively, of the *covariance integral operator* for which $c(\boldsymbol{x}_0, \boldsymbol{x}_1)$ is the kernel, in fact we have

$$\mu_j \xi_j(\boldsymbol{x}_0) = \int_D c(\boldsymbol{x}_0, \boldsymbol{x}_1)\, \xi_j(\boldsymbol{x}_1)\, \mathrm{d}\boldsymbol{x}_1$$

We assume that the covariance functions are stationary and isotropic, meaning that we can write $c(\boldsymbol{x}_0, \boldsymbol{x}_1) = \rho(|\boldsymbol{x}_0 - \boldsymbol{x}_1|)$. For further information on this subject we refer the reader to [2].

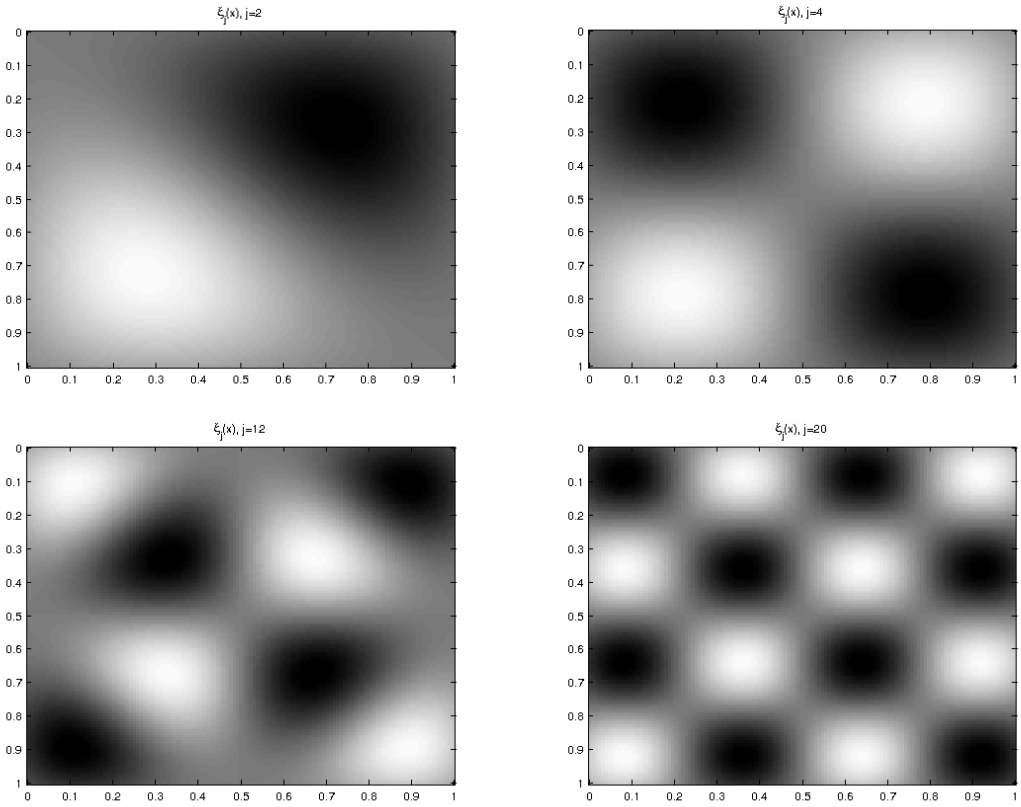

Figure 4.1: Grey-scale plots of example ξ_j in 2-dimensions

In Figure 4.1 we present some examples of ξ_j for a given set, where the value of the function is given in grey-scale. In fact these examples here were derived using the Karhunen–Loève expansion for an exponential covariance. We see that they are periodic functions increasing in frequency for higher j, the sort of behaviour we might expect. Let us define the partial sums

$$Z^s(\boldsymbol{x}, \boldsymbol{y}) := \sum_{j=1}^{s} \sqrt{\mu_j}\, \xi_j(\boldsymbol{x})\, y_j \qquad \text{for } \boldsymbol{y} \in \mathbb{R}^{\mathbb{N}} \tag{4.5}$$

and define $a^s(\boldsymbol{x}, \boldsymbol{y})$ as in (4.2) but with $Z^s(\boldsymbol{x}, \boldsymbol{y})$. In Figure 4.2 we have a sequence Z^s for increasing s for a fixed $\boldsymbol{y} \in \mathbb{R}^{\mathbb{N}}$. As s increases we clearly see more "resolution" towards some final random field, although even for small s we see some key features of the field emerge already, such as impermeable clusters.

We study problem (4.1) in its weak form. Thus we seek $u(\cdot, \boldsymbol{y}) \in H_0^1(D)$ (these spaces will be defined shortly, in §4.1) such that

$$\mathscr{A}(\boldsymbol{y}; u, v) \;=\; \langle f, v \rangle \;, \quad \text{for all } v \in H_0^1(D) \text{ and for almost all } \boldsymbol{y} \in \mathbb{R}^{\mathbb{N}}, \tag{4.6}$$

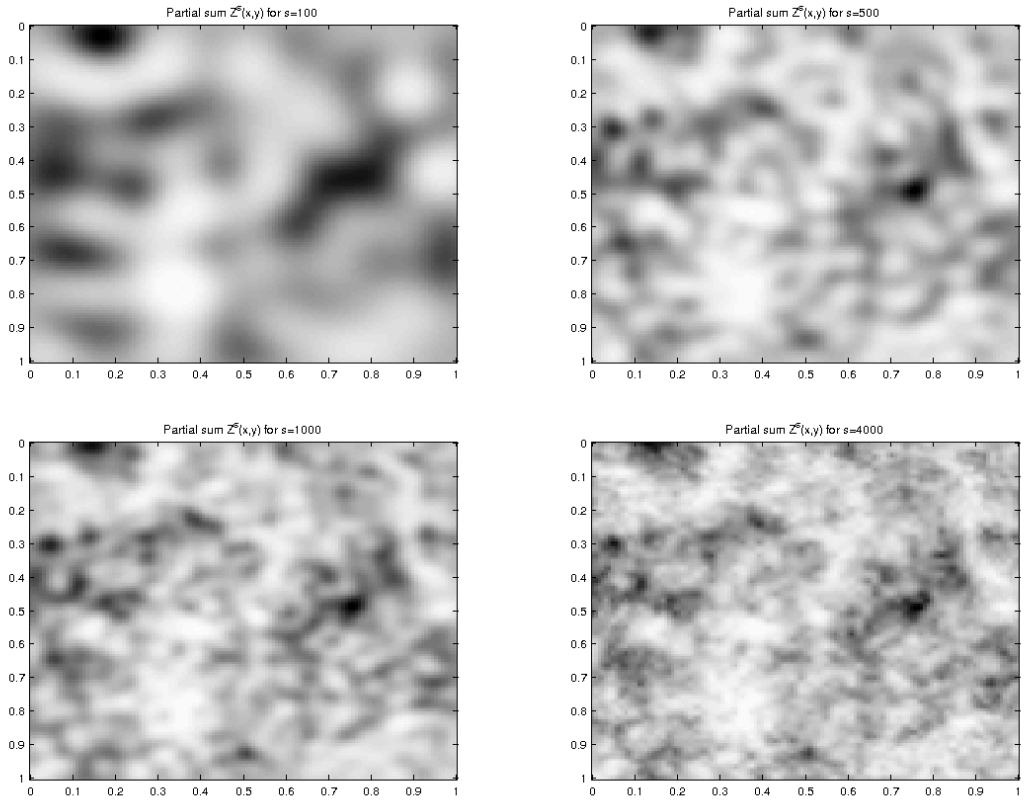

Figure 4.2: Grey-scale plots of example partial sums $Z^s(\boldsymbol{x},\boldsymbol{y}) = \sum_{j=1}^{s} \sqrt{\mu_j}\xi_j(\boldsymbol{x})y_j$ for various s

where $\langle\cdot,\cdot\rangle$ denotes the duality pairing between $H^s(D)$ and $(H^s(D))'$, which can simply be taken to be the $L^2(D)$ inner product, and we have the bilinear-form

$$\mathscr{A}(\boldsymbol{y};w,v) := \int_D a(\boldsymbol{x},\boldsymbol{y})\nabla w(\boldsymbol{x})\cdot\nabla v(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\,, \quad w,v\in H^1(D)\,,$$

and we assume that $f \in (H_0^1(D))'$.

For any $\varepsilon > 0$ we have $\rho_G(\{\boldsymbol{y} : a(\boldsymbol{x},\cdot) > \varepsilon^{-1}\}) > 0$ so that problem (4.6) is not uniformly bounded over all possible realisations of a. If $a_*(\boldsymbol{x}) = 0$, then we also have $\rho_G(\{\boldsymbol{y} : a(\boldsymbol{x},\cdot) < \varepsilon\}) > 0$ so that (4.6) is not uniformly elliptic either. This loss of ellipticity and boundedness is one of the main difficulties in the (numerical) analysis of (4.6).

We are interested in expected values of functionals of the solution of (4.6). That is, if $\mathcal{G} \in (H_0^1(D))'$, we will be interested in the expected value $\mathbb{E}[\mathcal{G}(u)]$ of the random variable $\mathcal{G}(u(\cdot,\boldsymbol{y}))$. The choices of functionals are generally inspired by the literature in groundwater flow modelling, and include but are not limited to effective permeability of a medium, the pressure at a point (point evaluation), and positions or travel times of a suspended particle in the fluid. We shall use sampling methods for the computation of $\mathbb{E}[\mathcal{G}(u)]$. That is, we will compute realisations of $a(\boldsymbol{x},\boldsymbol{y})$, which yield realisations of $u(\boldsymbol{x},\boldsymbol{y})$, via the solution of the elliptic problem (4.6), and from these we shall compute

an approximation of $\mathbb{E}[\mathcal{G}(u)]$ by an appropriate averaging over $\boldsymbol{y}$. However, in contrast to standard Monte Carlo (MC) methods, we will sample $a(\boldsymbol{x}, \boldsymbol{y})$ using quasi-Monte Carlo (QMC) methods. We demonstrate here that under suitable assumptions, QMC methods are faster than MC methods for this class of problems. For simplicity in the theory, we must assume the functionals $\mathcal{G}$ to be linear.

We summarise here the approach to approximating the problem. First we approximate (4.6) for a fixed $\boldsymbol{y}$ using the finite element method. Following conventional notation, a solution is found in a finite-dimensional subspace $V_h \subset H_0^1(D)$ of piecewise-linear functions on a triangulation of D, where h is a parameter for the the maximum diameter of all the triangles, and the finite dimensional solution is labelled u_h. Next we we sample the random field by truncating the sum in (4.2) to s terms. The resulting approximation of the field, which we denote a^s, is substituted in to (4.6), and the resulting finite element solution is then denoted $u_h^s(\boldsymbol{x}, \boldsymbol{y})$. The corresponding approximation of $\mathbb{E}[\mathcal{G}(u)]$ is then taken to be the expected value of the random variable $\mathcal{G}(u_h^s(\cdot, \boldsymbol{y}))$, written $\mathbb{E}[\mathcal{G}(u_h^s)]$. In fact, since u_h^s is a random field derived from the s i.i.d. $\mathcal{N}(0,1)$ random variables $\{Y_j\}_{j\geq 1}$, we have the formula

$$\mathbb{E}[\mathcal{G}(u_h^s)] \;=\; \int_{\mathbb{R}^s} \mathcal{G}(u_h^s(\cdot, \boldsymbol{y})) \prod_{j=1}^{s} \phi(y_j)\, \mathrm{d}\boldsymbol{y} \;, \tag{4.7}$$

where $\phi(y) = \exp(-y^2/2)/\sqrt{2\pi}$ is the Gaussian normal probability density.

The computation of this (possibly high dimensional) integral, leads us to the use of the QMC methods outlined through this thesis, in particular randomly shifted lattice rules. Our aim in this chapter is to bound the root mean square error of this QMC evaluation,

$$\sqrt{\mathbb{E}^{\Delta}\Big[\big(\mathbb{E}[\mathcal{G}(u)] - Q_{s,n}(\boldsymbol{z}, \Delta;\, \mathcal{G}(u_h^s))\big)^2\Big]}, \tag{4.8}$$

where $\mathbb{E}^{\Delta}$ denotes expectation with respect to the random shift Δ, and where $Q_{s,n}(\boldsymbol{z}, \Delta;\, \cdot)$ a lattice rule, as defined in (2.10). Thus $Q_{s,n}(\boldsymbol{z}, \Delta;\, \mathcal{G}(u_h^s))$ represents our proposed QMC-FE method in full.

The starting point for our analysis is the observation that we can break up our analysis in to error terms due to the various approximations. We have that

$$\mathbb{E}[\mathcal{G}(u)] - Q_{s,n}(\boldsymbol{z}, \Delta;\, \mathcal{G}(u_h^s)) \;=\; \mathbb{E}[\mathcal{G}(u) - \mathcal{G}(u_h^s)] \;+\; (\mathbb{E}[\mathcal{G}(u_h^s)] \;-\; Q_{s,n}(\boldsymbol{z}, \Delta;\, \mathcal{G}(u_h^s)))$$

Since the random diffusion coefficient $a(\boldsymbol{x}, \boldsymbol{y})$ in (4.1) and the random shift Δ in the QMC rule are statistically independent, we can write

$$\mathbb{E}[\mathcal{G}(u)] - Q_{s,n}(\boldsymbol{z}, \Delta;\, \mathcal{G}(u_h^s)) \;=\; \big(\mathbb{E}[\mathcal{G}(u) - \mathcal{G}(u_h^s)]\big)^2 \;+\; \mathbb{E}^{\Delta}\,|\mathbb{E}[\mathcal{G}(u_h^s)] - Q_{s,n}(\boldsymbol{z}, \cdot;\, \mathcal{G}(u_h^s))|^2 \;. \tag{4.9}$$

Related publications

This problem has attracted great interest recently. The results in this thesis draw directly from a collaboration that resulted in [24]. In this thesis we undertake the analysis of the problem, in particular analytic results for error analysis of QMC quadrature applied to the PDE problem in (4.1) with lognormal random fields. Many results in [24] are more general than the somewhat more simplified results presented in this chapter.

Early papers [44, 45, 13] provide the foundations of the problem, and explore the use of Monte Carlo methods when applied to the problem. Extensive analysis of the FEM problem that arises from the solution of the PDE for a single instance of a permeability field, that is, without the statistical analysis involving the random field, can be found in [10] and [55]. In [25] the same approach to the spatial problem is taken, but then using QMC methods for the statistical simulation of the random field, but did not include error analysis. A full analysis of QMC methods when applied to this problem, but without the allowance of lognormal fields (that is, uniformly bounded fields are considered) is considered in [36]. Further analysis of the problem, including extensive results on the regularity of the problem in its spatial domain, with applications towards using multilevel Monte Carlo methods for the quadrature on the probability space can be found in [66] and [9].

4.1 Preliminaries

Here we introduce some standard notations and concepts, particularly the relevant function spaces and norms to be used throughout this chapter.

First we introduce multi-index notation. Let $\boldsymbol{\nu} = (\nu_j)_{j\in\mathbb{N}}$ denote a multi-index of non-negative integers, with finitely many nonzero elements, i.e. $|\boldsymbol{\nu}| := \sum_{j\ge 1} \nu_j < \infty$. As usual, the value of ν_j will determine the number of derivatives to be taken with respect to y_j, for any function $u : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$, we use the shorthand notation

$$\partial^{\boldsymbol{\nu}} u := (\partial^{\nu_1}\partial^{\nu_2}\dots)\, u \qquad \text{where} \qquad \partial^{\nu_i} u := \frac{\partial^{\nu_i} u}{\partial y_i^{\nu_i}}$$

If u is a function on a finite dimensional space $\mathbb{R}^s$, we evidently consider $\boldsymbol{\nu} = (\nu_j)_{j=1}^s$. We write $\boldsymbol{m} \preceq \boldsymbol{\nu}$ to mean that the multi-index $\boldsymbol{m}$ satisfies $m_j \le \nu_j$ for all j. Let $\boldsymbol{\nu} - \boldsymbol{m}$ denote a multi-index with the elements $\nu_j - m_j$, and $\binom{\boldsymbol{\nu}}{\boldsymbol{m}} := \prod_{j\ge 1} \binom{\nu_j}{m_j}$. Now, given any multi-index $\boldsymbol{\nu}$ with $|\boldsymbol{\nu}| = n$, we have the Leibniz product rule

$$\partial^{\boldsymbol{\nu}}(AB) = \sum_{\boldsymbol{m}\preceq\boldsymbol{\nu}} \binom{\boldsymbol{\nu}}{\boldsymbol{m}} (\partial^{\boldsymbol{\nu}-\boldsymbol{m}} A)(\partial^{\boldsymbol{m}} B)\,.$$

Note that this multi-index notation is quite different to the set notation $\mathfrak{u}$ used in earlier chapters.

Let $\boldsymbol{\mathcal{C}}^0(D)$ denote the space of continuous functions on D, and let $\boldsymbol{\mathcal{C}}^k(D)$ represent the space of k-times continuously differentiable functions on D. We say that a function

$v : D \to \mathbb{R}$ is Hölder continuous with exponent t in D if we have

$$|v|_{\boldsymbol{\mathcal{C}}^{0,t}(D)} := \sup_{\boldsymbol{x}_0,\boldsymbol{x}_1\in\overline{D}:\,\boldsymbol{x}_0\neq\boldsymbol{x}_1} \frac{|v(\boldsymbol{x}_0)-v(\boldsymbol{x}_1)|}{|\boldsymbol{x}_0-\boldsymbol{x}_1|^t} < \infty, \tag{4.10}$$

where $|\boldsymbol{x}|$ is the usual Euclidean distance in $\mathbb{R}^d$. The quantity $|v|_{\boldsymbol{\mathcal{C}}^{0,t}(D)}$ serves as a seminorm on $\boldsymbol{\mathcal{C}}^0(D)$. We can define the related norm as

$$\|v\|_{\boldsymbol{\mathcal{C}}^{k,t}(D)} := \|v\|_{\boldsymbol{\mathcal{C}}^k(D)} + \max_{|\boldsymbol{\nu}|=k} |\partial^{\boldsymbol{\nu}} v|_{\boldsymbol{\mathcal{C}}^{0,t}(D)},$$

where we take

$$\|v\|_{\boldsymbol{\mathcal{C}}^k(D)} := \max_{|\boldsymbol{\nu}|\le k} \sup_{\boldsymbol{x}\in D} |\partial^{\boldsymbol{\nu}} v(\boldsymbol{x})|.$$

We write $H^s(D)$, where $s \ge 0$, for the Sobolev space of functions where, for every $\boldsymbol{\nu}$ such that $|\boldsymbol{\nu}| \le s$, we have that $\partial^{\boldsymbol{\nu}} v \in L^2(D)$ (note that we can allow $\partial^{\boldsymbol{\nu}}$ to include differentiation in the weak sense here). We take the usual norm as follows,

$$\|v\|_{H^s(D)} := \sum_{|\boldsymbol{\nu}|\le s} \|\partial^{\boldsymbol{\nu}} v\|_{L^2(D)}. \tag{4.11}$$

Let us also write $H_0^1(D)$ for the subspace of functions in $H^1(D)$ with vanishing trace on the boundary ∂D. We define the norm on $H_0^1(D)$ as

$$\|v\|_{H_0^1(D)} := \left(\int_D |\nabla v|^2 \,\mathrm{d}\boldsymbol{x}\right)^{1/2}, \tag{4.12}$$

and remark that it is well known (see e.g. [22, Theorem 7.10]) that in $H_0^1(D)$ this norm is equivalent to the norm defined in (4.11). As $H_0^1(D)$ is the natural setting for much of the theory to come, for brevity we adopt the notation $V = H_0^1(D)$ and let V' indicate its dual space from now on.

Finally, for some regularity results we will also require spaces of Bochner integrable functions. For any Banach space X with norm $\|\cdot\|_X$ and for $1 \le q < \infty$, we denote by $L^q(\Omega, \mathbb{P}; X)$ the space of all strongly $\mathbb{P}$–measurable mappings v from $(\Omega, \mathcal{A})$ to $(X, \mathcal{B}(X))$ (where $\mathcal{B}(X)$ denotes the Borel sigma algebra over X), for which the Bochner integral

$$\|v\|_{L^q(\Omega,\mathbb{P};X)} = \begin{cases} \left(\int_\Omega \|v\|_X^q \mathrm{d}\mathbb{P}\right)^{1/q}, & \text{for} \quad 1 \le q < \infty, \\ \operatorname{esssup}_{\omega\in\Omega} \|v\|_X, & \text{for} \quad q = \infty, \end{cases}$$

is finite. When there is no ambiguity about the measure, we shall denote this space by $L^q(\Omega; X)$. In the particular case $X = \mathbb{R}$, we shall simply write $L^q(\Omega)$ in place of $L^q(\Omega; \mathbb{R})$.

4.2 Discretisation and truncation

In this section we are concerned with finding an upper bound for the first term in (4.9). We do so by separating the contribution towards the error of the finite element approximation,

followed by the truncation, that is we write

$$\mathcal{G}(u) - \mathcal{G}(u_h^s) \;=\; (\mathcal{G}(u) - \mathcal{G}(u_h)) \;+\; (\mathcal{G}(u_h) - \mathcal{G}(u_h^s)) \tag{4.13}$$

and estimate the expectation for each of these two terms separately. Other publications, for example [36, 7], take the approach of estimating $\mathcal{G}(u) - \mathcal{G}(u^s)$ and $\mathcal{G}(u) - \mathcal{G}(u_h)$ separately.

For ease of notation we define

$$\beta_j(\boldsymbol{x}) \;:=\; \sqrt{\mu_j}\,\xi_j(\boldsymbol{x}). \tag{4.14}$$

We now make some suitable assumptions on the regularity of the field parameters μ_j and ξ_j. These assumptions are inspired by what is commonly seen in the literature, but are simplified here for the sake of exposition in this thesis. These assumptions are somewhat less general than, for example, those found in [7] or [24].

Assumption A1 *(a) The functions ξ_j are continuously differentiable, i.e. $\{\xi_j\}_{j\geq 1} \subset \boldsymbol{\mathcal{C}}^1(\overline{D})$.*

(b) $\sum_{j=1}^{\infty} \|\beta_j\|^p_{\boldsymbol{\mathcal{C}}^0(\overline{D})} < \infty$ for some $p \in (0,1]$, and $\sum_{j=1}^{\infty} \|\nabla\beta_j\|_{\boldsymbol{\mathcal{C}}^0(\overline{D})} < \infty$.

We can now define the *admissible* parameter set. This will be a set of "good" realisations of the field a, in the sense that $a(\cdot, \boldsymbol{y})$ is bounded away from 0 and ∞, and has appropriate smoothness properties. We define

$$U_{\boldsymbol{\beta}} \;:=\; \left\{ \boldsymbol{y} \in \mathbb{R}^{\mathbb{N}} : \sum_{j=1}^{\infty} \left(\|\beta_j\|_{\boldsymbol{\mathcal{C}}^0(\overline{D})} + \|\nabla\beta\|_{\boldsymbol{\mathcal{C}}^0(\overline{D})} \right) |y_j| < \infty \right\} \subset \mathbb{R}^{\mathbb{N}}\,. \tag{4.15}$$

This set $U_{\boldsymbol{\beta}} \subset \mathbb{R}^{\mathbb{N}}$ is not a simple set, in the sense of being able to be constructed through a countable product of intervals in $\mathbb{R}$, but as is shown in the following lemma, $U_{\boldsymbol{\beta}}$ is measurable and contains ρ_G-almost all realisations $\boldsymbol{y}$. This enables to consider all results going forward to be for realisations in $U_{\boldsymbol{\beta}}$, and can employ our regularity results "pointwise" for $\boldsymbol{y} \in U_{\boldsymbol{\beta}}$.

Lemma 26 *If Assumption A1 holds, then $U_{\boldsymbol{\beta}} \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})$ and $\rho_G(U_{\boldsymbol{\beta}}) = 1$.*

Proof. This proof is based on the proof of [56, Lemma 2.28]. We define

$$U_{\boldsymbol{\beta},M,N} \;:=\; \left\{ \boldsymbol{y} \in \mathbb{R}^{\mathbb{N}} : \sum_{j=1}^{M} \left(\|\beta_j\|_{\boldsymbol{\mathcal{C}}^0(\overline{D})} + \|\nabla\beta\|_{\boldsymbol{\mathcal{C}}^0(\overline{D})} \right) |y_j| \leq N \right\},$$

which is clearly measurable, and measurability of $U_{\boldsymbol{\beta}}$ then follows from the fact that we can express it as a countable intersection and union,

$$U_{\boldsymbol{\beta}} = \bigcup_{N=1}^{\infty} \bigcap_{M=1}^{\infty} U_{\boldsymbol{\beta},M,N}.$$

We have that

$$\mathbb{E}_{\rho_G}|y_j| \;=\; \int_{\mathbb{R}^{\mathbb{N}}} |y_j|\,\mathrm{d}\rho_G(\boldsymbol{y}) \;=\; \frac{2}{\sqrt{2\pi}}\int_0^\infty y\exp(-y^2/2)\mathrm{d}y \;=\; \sqrt{\frac{2}{\pi}}.$$

Now, considering the partial sums we can apply the monotone convergence theorem to swap the sum and the integral,

$$\begin{aligned}
&\int_{\mathbb{R}^{\mathbb{N}}} \sum_{j=1}^{\infty} \left(\|\beta_j\|_{\boldsymbol{\mathcal{C}}^0(\overline{D})} + \|\nabla\beta\|_{\boldsymbol{\mathcal{C}}^0(\overline{D})}\right)|y_j|\,\mathrm{d}\rho_G(\boldsymbol{y}) \\
&\qquad = \sum_{j=1}^{\infty} \left(\|\beta_j\|_{\boldsymbol{\mathcal{C}}^0(\overline{D})} + \|\nabla\beta\|_{\boldsymbol{\mathcal{C}}^0(\overline{D})}\right) \int_{\mathbb{R}^{\mathbb{N}}} |y_j|\,\mathrm{d}\rho_G(\boldsymbol{y}) \\
&\qquad = \sqrt{\frac{2}{\pi}} \sum_{j=1}^{\infty} \left(\|\beta_j\|_{\boldsymbol{\mathcal{C}}^0(\overline{D})} + \|\nabla\beta\|_{\boldsymbol{\mathcal{C}}^0(\overline{D})}\right) < \infty.
\end{aligned}$$

This implies that the sum must be finite for ρ_G-almost all $\boldsymbol{y}$, thus $\rho_G(U_{\boldsymbol{\beta}}) = 1$. □

We have the following important consequence,

Lemma 27 *Let Assumption 1 hold, then the partial sums Z^s converge uniformly (in $\boldsymbol{x}$) to Z, for ρ_G-almost all $\boldsymbol{y}$.*

Proof. We have that $|Z(\boldsymbol{x},\boldsymbol{y}) - Z^s(\boldsymbol{x},\boldsymbol{y})| \;\leq\; \sum_{j=s+1}^{\infty} \|\beta_j\|_{\boldsymbol{\mathcal{C}}^0(\overline{D})}\,|y_j|$ which necessarily must converge to 0 by Assumption A1, irrespective of $\boldsymbol{x}$, for any $\boldsymbol{y} \in U_{\boldsymbol{\beta}}$. □

4.2.1 Spatial regularity

The regularity of the random field $a(\boldsymbol{x},\boldsymbol{y})$, in the spatial variable $\boldsymbol{x}$, depends on the properties of the parameters μ_j and ξ_j. As we shall see, Assumption A1 enables us to demonstrate a degree of spatial regularity. Similar results with more generality can be found in [24, 9]. Our approach here is more akin to [36] in terms of our assumptions of spatial regularity. It should be noted that Proposition 28 is demonstrated in [2, Theorem 8.3.2], however this result assumes the statistics of the random field are given by a covariance kernel with a certain smoothness, rather than our approach of building our random field directly from the parametric representation (4.3).

Proposition 28 *Under Assumption A1 the realisations of $Z(\cdot,\boldsymbol{y})$, as defined in* (4.3)*, are in $\boldsymbol{\mathcal{C}}^1(\overline{D})$, ρ_G-almost surely. If, in addition, a_*, $a_0 \in \boldsymbol{\mathcal{C}}^1$, then $a(\cdot,\boldsymbol{y})$, as defined in* (4.2)*, are also in $\boldsymbol{\mathcal{C}}^1(\overline{D})$, ρ_G-almost surely.*

Proof. To show that $Z(\cdot,\boldsymbol{y}) \in \boldsymbol{\mathcal{C}}^1(\overline{D})$, it suffices to show that $\|Z(\cdot,\boldsymbol{y})\|_{\boldsymbol{\mathcal{C}}^0(\overline{D})} < \infty$ and that $\|\nabla_{\boldsymbol{x}} Z(\cdot,\boldsymbol{y})\|_{\boldsymbol{\mathcal{C}}^0(\overline{D})} < \infty$, for ρ_G-almost all $\boldsymbol{y}$. Using (4.3) we have that

$$\|Z(\cdot,\boldsymbol{y})\|_{\boldsymbol{\mathcal{C}}^0(\overline{D})} \;\leq\; \sum_{j=1}^{\infty} \|\beta_j\|_{\boldsymbol{\mathcal{C}}^0(\overline{D})}|y_j| \;<\; \infty$$

for all $\boldsymbol{y} \in U_{\boldsymbol{\beta}}$. By Lemma 27 we have uniform convergence of the sum for $\boldsymbol{y} \in U_{\boldsymbol{\beta}}$, hence we can exchange differentiation with summation, thus we have

$$\|\nabla_{\boldsymbol{x}} Z(\cdot,\boldsymbol{y})\|_{\boldsymbol{\mathcal{C}}^0(\overline{D})} = \left\| \sum_{j=1}^{\infty} \nabla_{\boldsymbol{x}} \beta_j(\boldsymbol{x})\, y_j \right\|_{\boldsymbol{\mathcal{C}}^0(\overline{D})} \leq \sum_{j=1}^{\infty} \|\nabla \beta_j\|_{\boldsymbol{\mathcal{C}}^0(\overline{D})}\, |y_j| < \infty$$

for all $\boldsymbol{y} \in U_{\boldsymbol{\beta}}$, i.e. ρ_G-almost all $\boldsymbol{y}$. The regularity of $a(\cdot,\boldsymbol{y})$ follows from the regularity of a_*, a_0 and $\exp(\cdot)$. □

Note that as $\overline{D}$ is compact we also have that $a(\cdot,\boldsymbol{y}) \in \boldsymbol{\mathcal{C}}^{0,t}(\overline{D})$ for any $0 \leq t \leq 1$.

Consider the weak form of (4.1) as defined in (4.6). To prove well-posedness of this variational problem, we define, for ρ_G-almost every $\boldsymbol{y}$,

$$\check{a}(\boldsymbol{y}) := \min_{\boldsymbol{x}\in\overline{D}} a(\boldsymbol{x},\boldsymbol{y}) \qquad \text{and} \qquad \hat{a}(\boldsymbol{y}) := \max_{\boldsymbol{x}\in\overline{D}} a(\boldsymbol{x},\boldsymbol{y}). \tag{4.16}$$

Under the assumptions of Proposition 28, for almost all $\boldsymbol{y} \in \mathbb{R}^{\mathbb{N}}$, $Z(\cdot,\boldsymbol{y})$ is a continuous function on $\overline{D}$ and hence attains its finite maximum and minimum on $\overline{D}$. Thus the quantities $\check{a}$ and $\hat{a}$ defined in (4.16) are ρ_G-measurable and, hence, random variables which satisfy $\check{a}(\boldsymbol{y}) > 0$ and $\hat{a}(\boldsymbol{y}) < \infty$ ρ_G-almost surely. Furthermore for all $x \in D$ and ρ_G-almost all $\boldsymbol{y}$ we have

$$0 < \check{a}(\boldsymbol{y}) \leq a(\boldsymbol{x},\boldsymbol{y}) \leq \hat{a}(\boldsymbol{y}) < \infty, \quad \text{for all } \boldsymbol{x} \in D \text{ and } \boldsymbol{y} \in U_{\boldsymbol{\beta}}\,. \tag{4.17}$$

Therefore for each $\boldsymbol{y}$ the Lax-Milgram Lemma holds, and we can infer the existence of a unique solution $u(\cdot,\boldsymbol{y})$ of (4.1).

Theorem 29 *Let Assumption 1 hold, then for every $\boldsymbol{y} \in U_{\boldsymbol{b}}$, $s \in \mathbb{N}$ and $h > 0$, the weak-form problem* (4.6) *admits unique solutions $u(\cdot,\boldsymbol{y}) \in V$. Moreover,*

$$\|u(\cdot,\boldsymbol{y})\|_V \leq \frac{1}{\check{a}(\boldsymbol{y})}\, \|f\|_{V'}\,, \qquad \forall \boldsymbol{y} \in U_{\boldsymbol{\beta}}\,, \tag{4.18}$$

We have, however, no uniform bound of $\check{a}$ and $\hat{a}$ for all $\boldsymbol{y}$, as discussed earlier in this chapter. Hence we can not provide uniform bounds for the solution u of (4.1). However, we can see that the mapping given by (4.1), between $a \in \boldsymbol{\mathcal{C}}^{0,t}(\overline{D})$ and $u \in V$, is Lipschitz continuous. This guarantees ρ_G-measurability and hence u is a random field on the probability space $(U_{\boldsymbol{\beta}}, \mathcal{B}(U_{\boldsymbol{\beta}}), \rho_G)$ which takes values in the space V.

In the next theorem we take this further and use an application of Fernique's Theorem, see [12, Theorem 2.6], which allows us to extend ρ_G-almost sure bounds on $u(\cdot,\boldsymbol{y})$ to infer boundedness of $\|u\|_{L^q(U_{\boldsymbol{\beta}},V)}$ for any $0 < q < \infty$. We refer to [7, §2] for details, from which the following theorem originated.

Proposition 30 *Let Assumption 1 hold and assume that* $a_*, a_0 \in \boldsymbol{\mathcal{C}}^0(\overline{D})$ *in* (4.2)*. Then, for all* q *in the range* $1 \leq q < \infty$*,* $1/\check{a} \in L^q(U_{\boldsymbol{\beta}})$ *and* $\hat{a} \in L^q(U_{\boldsymbol{\beta}})$*, and for every* $f \in V'$ *the problem* (4.6) *admits a unique solution* $u \in L^q(U_{\boldsymbol{\beta}}, V)$ *that satisfies*

$$\|u\|_{L^q(U_{\boldsymbol{\beta}},V)} \;\leq\; \|1/\check{a}\|_{L^q(U_{\boldsymbol{\beta}})} \|f\|_{V'}\,.$$

Proof. Recall that Z is a Gaussian field, and since D is a bounded domain, by Proposition 28 $Z(\cdot, \boldsymbol{y}) \in C^0(\overline{D})$, ρ_G-almost surely. Hence we can apply [7, Prop. 2.3 & 2.4] to obtain the result. □

To quantify the rate of convergence of finite element solutions of (4.6) with respect to the triangulation parameter h, additional regularity of the solution u is required. For this we use a simplified form of a result from [66] and [9] that extends the ρ_G-almost sure Hölder regularity of a. First we require the following lemma, based on [9, Lemma 2.3], which we then use to demonstrate the following result.

Lemma 31 *If Assumption 1 holds then for any* $t \in (0, 1]$

$$\|a(\cdot, \boldsymbol{y})\|_{\boldsymbol{\mathcal{C}}^{0,t}} \;\leq\; (1 + 2\|Z(\cdot, \boldsymbol{y})\|_{\boldsymbol{\mathcal{C}}^{0,t}})\, a_{\max}(\boldsymbol{y})$$

Proof. The proof follows as in the proof of [9, Lemma 2.3], noting that we have $a(\cdot, \boldsymbol{y}) \in \boldsymbol{\mathcal{C}}^{0,t}(\overline{D})$ for any $t \leq 1$ from Proposition 28. Thus we have, for $\boldsymbol{y} \in U_{\boldsymbol{\beta}}$ and $\boldsymbol{x}_0, \boldsymbol{x}_1 \in D$

$$\begin{aligned}|\exp(Z(\boldsymbol{x}_0, \boldsymbol{y})) - \exp(Z(\boldsymbol{x}_1, \boldsymbol{y}))| \;&\leq\; |Z(\boldsymbol{x}_0, \boldsymbol{y}) - Z(\boldsymbol{x}_1, \boldsymbol{y})|\, (\exp(Z(\boldsymbol{x}_0, \boldsymbol{y})) - \exp(Z(\boldsymbol{x}_1, \boldsymbol{y})))\\ &\leq\; 2\,\hat{a}(\boldsymbol{y})\, \|Z(\cdot, \boldsymbol{y})\|_{\boldsymbol{\mathcal{C}}^{0,t}(\overline{D})}\, |\boldsymbol{x}_0 - \boldsymbol{x}_1|\end{aligned}$$

Since we know that $\hat{a}\|Z(\cdot, \boldsymbol{y})\|_{\boldsymbol{\mathcal{C}}^{0,t}(\overline{D})} < \infty$ for all $\boldsymbol{y} \in U_{\boldsymbol{\beta}}$, we can retrieve the result when we take the supremum over $\boldsymbol{x}_0$, $\boldsymbol{x}_1 \in D$. □

The next two results have certain requirements of the spatial domain D, which we state in the following assumption

Assumption A2 $D \subset \mathbb{R}^d$ *for* $d = 1, 2$ *or* 3*, is a bounded convex polyhedral domain with plane faces.*

Note that the results also follow if the boundary of D is continuously twice-differentiable everywhere. This assumption on the domain is the primary focus of [9], while [66] extends many of the same results to the polyhedral domains.

Proposition 32 *Let Assumptions A1 and A2 hold, then* $a \in L^p(U_{\boldsymbol{\beta}}, \boldsymbol{\mathcal{C}}^{0,t}(\overline{D}))$ *and* $u \in L^p(U_{\boldsymbol{\beta}}, H^2(D))$*.*

Proof. We see that [66, Assumption A1] (that $1/\check{a}(\boldsymbol{y}) \in L^p(U_{\boldsymbol{\beta}})$) is satisfied by (4.17) and Proposition 30. As we assume here that f is smooth over $\boldsymbol{x}$ and has no dependence on $\boldsymbol{y}$, we see also that [66, Assumption A3] is also satisfied (that $f \in L^p(U_{\boldsymbol{\beta}}, V')$).

We have that $Z \in L^p(U_{\boldsymbol{\beta}}, \boldsymbol{\mathcal{C}}^{0,t}(\overline{D}))$ from [7, Proposition 3.8], and that $a_{\max} \in L^p(U_{\boldsymbol{\beta}})$ from Proposition 30, thus using Lemma 31 and an application of Hölder's inequality, we have that $\|a\|_{L^p(U_{\boldsymbol{\beta}}, \boldsymbol{\mathcal{C}}^{0,t}(\overline{D}))} < \infty$. Thus [66, Assumption A2] is also satisfied. Thus the assumptions for a in [66, Theorem 2.2] are satisfied, and further noting that the convexity of D means we have $\lambda_\Delta(D) = 1$ (as defined in [66, Definition 2.1]), which gives us the final result. □

4.2.2 Discretisation error

To discretise (4.6) in the physical domain D we consider now finite element approximations with standard, continuous, piecewise linear finite elements. We denote by $\{\mathcal{T}_h\}_{h>0}$ a shape-regular family of simplicial triangulations of the domain D, parametrised by the mesh width $h := \max_{T\in\mathcal{T}_h} \operatorname{diam}(T)$. Associated with each triangulation $\mathcal{T}_h$ we define the space $V_h \subset V$ of piecewise linear, continuous functions on this mesh, which vanish on ∂D. For any $\boldsymbol{y} \in U_{\boldsymbol{\beta}}$, we denote by $u_h(\boldsymbol{y}, \cdot) \in V_h$ the solution of

$$\mathscr{A}(\boldsymbol{y}; u_h(\cdot, \boldsymbol{y}), v_h) = \langle f, v_h \rangle, \qquad \text{for all} \quad v_h \in V_h\,. \tag{4.19}$$

As in Proposition 30, for every h and for ρ_G-almost every realization $a(\cdot, \boldsymbol{y})$, the FE solution $u_h(\boldsymbol{y}, \cdot) \in V_h$ exists, is unique and (like the exact solution $u(\boldsymbol{y}, \cdot)$) satisfies the a priori bound

$$\|u_h\|_{L^q(U_{\boldsymbol{\beta}};V)} \;\le\; \|1/\check{a}\|_{L^q(U_{\boldsymbol{\beta}})} \|f\|_{V'}\,, \quad \text{for all} \quad 1 \le q < \infty\,, \tag{4.20}$$

First we have the following result, a simplification of [66, Theorem 2.3].

Proposition 33 *Let Assumptions A1 and A2 hold, then for any $q > 0$ there is a constant C such that*

$$\|u - u_h\|_{L^q(U_{\boldsymbol{\beta}},V)} \;\le\; Ch.$$

We are now in a position to bound the first term in the overall error bound (4.13) for our method. A proof can be found in [66], and uses an Aubin-Nitsche duality argument to boost the order of convergence.

Theorem 34 *Let Assumption A1 hold. Suppose $\mathcal{G}(\cdot)$ is a continuous linear functional on $H^{1-\tau}(D)$ for some $\tau \le 1$, i.e. there exists a constant $C_{\mathcal{G}}$ such that $|\mathcal{G}(v)| \le C_{\mathcal{G}} \|v\|_{H^{1-\tau}(D)}$ for all $v \in H^{1-\tau}(D)$. Then*

$$\left|\mathbb{E}[\mathcal{G}(u) - \mathcal{G}(u_h)]\right| \;\le\; Ch^2\,. \tag{4.21}$$

4.2.3 Dimension Truncation Error

To perform any numerical experiments it is of course necessary to truncate the infinite series expansion (4.3), creating a finite dimensional expression upon which we can perform our QMC quadrature. Here we examine the resulting truncation error.

Recalling (4.2) and (4.3), the approximation of a obtained by the dimensionally truncated expansion of Z is

$$a^s(\boldsymbol{x},\boldsymbol{y}) \;:=\; a_*(\boldsymbol{x}) \;+\; a_0(\boldsymbol{x})\exp\left(Z^s(\boldsymbol{x},\boldsymbol{y})\right), \qquad \text{for some } s\in\mathbb{N}\,. \tag{4.22}$$

The number of terms s is the dimension of the parameter domain for QMC integration in (4.7).

Note that $a^s(\boldsymbol{x},\boldsymbol{y})$ can be considered as the exact coefficient $a(\boldsymbol{x},\boldsymbol{y})$ evaluated at the particular vector $\boldsymbol{y}=(y_1,...,y_s,0,0,...)$. More generally, denoting by $\mathfrak{u}\subset\mathbb{N}$ any set of "active" coordinates, as in Chapter 2, we denote by $(\boldsymbol{y}_\mathfrak{u};\mathbf{0})$ the vectors $\boldsymbol{y}\in U_{\boldsymbol{b}}$ with the constraint that $y_j=0$ if $j\notin\mathfrak{u}$. This is in line with notation used in the previous chapters.

For any $\boldsymbol{y}\in U_{\boldsymbol{\beta}}$, we can now define $u_h^s(\cdot,\boldsymbol{y})\in V_h$ to be the solution of the dimensionally truncated, discretised boundary value problem

$$\mathscr{A}^s(\boldsymbol{y};u_h^s,v_h)=\langle f,v_h\rangle, \qquad \text{for all} \quad v_h\in V_h\,, \tag{4.23}$$

where

$$\mathscr{A}^s(\boldsymbol{y};w,v) \;:=\; \int_D a^s(\boldsymbol{x},\boldsymbol{y})\nabla w(\boldsymbol{x})\cdot\nabla v(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}, \quad \text{for any} \;\; v,w\in V.$$

For simplicity, we work here under the assumption that, for any $v_h,w_h\in V_h$, we evaluate the integrals in $\mathscr{A}^s(\boldsymbol{y};w_h,v_h)$ exactly. It is possible to also include quadrature errors in the analysis (see [9, §3.3] for details). Existence and uniqueness for $u_h^s(\cdot,\boldsymbol{y})$ ρ_G-almost everywhere follows again by the Lax-Milgram Lemma, in fact, Theorem 29 applies to the truncated problem, with the bound (4.18) also holding for $\|u_h^s(\cdot,\boldsymbol{y})\|_V$.

To obtain a bound on $|\mathbb{E}[\mathcal{G}(u_h)-\mathcal{G}(u_h^s)]|$ we apply the truncation error analysis in [7, 9].

Theorem 35 *Let Assumption A1 hold. Then $\|1/\breve{a}^s\|_{L^q(U_{\boldsymbol{\beta}})}$ is bounded independently of s, for all $1\le q<\infty$. Suppose further that $\mathcal{G}\in V'$. Then*

$$|\mathbb{E}[\mathcal{G}(u_h)-\mathcal{G}(u_h^s)]| \;\le\; C_\chi\, s^{-\chi}, \quad \textit{for all} \;\; 0<\chi<\frac{1}{p}-\frac{1}{2}. \tag{4.24}$$

Proof. We define

$$R_{s,\alpha} \;:=\; \sum_{j>s} \|\beta_j\|^{2(1-\alpha)}_{\boldsymbol{\mathcal{C}}^0(\overline{D})}\,\|\nabla\beta_j\|^{2\alpha}_{\boldsymbol{\mathcal{C}}^0(\overline{D})}$$

Note that Assumption A1 implies that

$$R_{s,0} \;=\; \sum_{j\ge1}\|\beta_j\|^2_{\boldsymbol{\mathcal{C}}^0(\overline{D})} \;<\; \infty,$$

and further, since clearly $\|\nabla\beta_j\|_{\boldsymbol{\mathcal{C}}^0(\overline{D})}$ is bounded,

$$R_{s,\alpha} \;=\; \sum_{j>s} \|\beta_j\|^{2(1-\alpha)}_{\boldsymbol{\mathcal{C}}^0(\overline{D})}\,\|\nabla\beta_j\|^{2\alpha}_{\boldsymbol{\mathcal{C}}^0(\overline{D})} \;\le\; C\sum_{j\ge1}\|\beta_j\|^{2(1-\alpha)}_{\boldsymbol{\mathcal{C}}^0(\overline{D})} \;<\; \infty, \tag{4.25}$$

for any $\alpha \in (0, 1 - \frac{p}{2}]$. Thus Assumption 3.1 of [7] holds. The required result that $\|1/\breve{a}^s\|_{L^q(U_{\boldsymbol{\beta}})}$ is bounded for $0 < q < \infty$, independently of s, then follows from [7, Proposition 3.10]. Moreover (4.25) implies that for s large enough

$$\begin{aligned} R_{s,\alpha} &\le \sum_{j>s} \|\beta_j\|^{2(1-\alpha)}_{\boldsymbol{\mathcal{C}}^0(\overline{D})} \\ &\le \|\beta_s\|^{2(1-\alpha)-p}_{\boldsymbol{\mathcal{C}}^0(\overline{D})} \sum_{j>s} \|\beta_j\|^p_{\boldsymbol{\mathcal{C}}^0(\overline{D})} \\ &\le s^{-2(1-\alpha)/p+1} \left(\sum_{j>s} \|\beta_j\|^p_{\boldsymbol{\mathcal{C}}^0(\overline{D})} \right)^{2(1-\alpha)/p-1} \sum_{j>s} \|\beta_j\|^p_{\boldsymbol{\mathcal{C}}^0(\overline{D})} \\ &\le C s^{-2(1-\alpha)/p+1}, \end{aligned}$$

where we used the fact that $\|\beta_s\|^p_{\boldsymbol{\mathcal{C}}^0(\overline{D})} \le \frac{1}{s} \sum_{j=1}^s \|\beta_j\|^p_{\boldsymbol{\mathcal{C}}^0(\overline{D})}$, which we infer from the summability of $\|\beta_j\|^p_{\boldsymbol{\mathcal{C}}^0(\overline{D})}$. Hence we have that

$$\sum_{s>0} (\max(R_{s,0}, R_{s,\alpha}))^{p_0} < \infty$$

for arbitrary $p_0 > (2(1-\alpha)/p - 1)^{-1}$, which is precisely Assumption 3.5 of [7]. This allows us to use [7, Theorem 4.2] to obtain

$$\|u - u^s\|_{L^q(U_{\boldsymbol{\beta}};V)} \le C_{q,\chi} s^{-\chi}, \tag{4.26}$$

where $\chi = (1-\alpha)/p - 1/2$, and u^s is the solution of the dimensionally truncated problem

$$\mathscr{A}^s(\boldsymbol{y}; u^s(\cdot, \boldsymbol{y}), v) = \langle f, v \rangle, \quad \text{for all } \ v \in V.$$

Finally, since the finite element solution u_h satisfies the same a priori bound (4.20) as the exact solution u (in Proposition 30) and since the right hand sides in (4.19) and in (4.23) are identical, the bound (4.26) holds also for the finite dimensional solution, that is $\|u_h - u^s_h\|_{L^q(U_{\boldsymbol{\beta}};V)}$ with constant $C_{q,\chi} > 0$ being independent of s and h. This follows immediately from the proofs of [7, Theorems 4.1 and 4.2], since all the identities and bounds involving $u - u^s$ there, hold equally for $u_h - u^s_h$. The final result (4.24) then follows upon taking $q = 1$ (note that $\mathbb{E}[\cdot] = \|\cdot\|_{L^1(U_{\boldsymbol{\beta}})}$) and from the fact that $\mathcal{G} \in V'$. □

Combining Theorems 34 and 35, we obtain the following estimate of the first term in (4.9).

Corollary 36 *Under Assumption A1 and with $\chi > 0$ as defined in Theorems 34 and 35 we have*

$$|\mathbb{E}[\mathcal{G}(u) - \mathcal{G}(u^s_h)]| \le C \left(h^2 + s^{-\chi} \right) .$$

4.3 Quadrature error

In this section, we perform the analysis of the second term in (4.9), which is the error in approximating the expectation $\mathbb{E}[\mathcal{G}(u_h^s)]$ by a suitable randomly shifted QMC quadrature approximation. We make extensive use of the analysis introduced in Chapters 2 and 3. First we must prepare for this analysis by proving certain regularity results, which follow in the next section.

4.3.1 Regularity with respect to the parametric variables

To estimate the second term in (4.9), it is crucial to bound the mixed first derivatives of $u_h^s(\cdot, \boldsymbol{y})$ with respect to $\boldsymbol{y}$. Here we state and prove a more general result which gives bounds also for higher order mixed derivatives. We prove the result for $u(\cdot, \boldsymbol{y})$ and explain subsequently why the argument also applies (with constants that are independent of h and of s) to $u_h^s(\cdot, \boldsymbol{y})$.

Recalling the multi-index notation, defined in §4.1, We will make use of the following identity.

Identity 37

$$\sum_{\substack{\boldsymbol{m} \preceq \boldsymbol{\nu} \\ |\boldsymbol{m}|=i}} \binom{\boldsymbol{\nu}}{\boldsymbol{m}} = \binom{|\boldsymbol{\nu}|}{i}.$$

Proof. This result can be demonstrated using a counting argument. For each $\boldsymbol{m}$ in the sum on the left hand, $\binom{\boldsymbol{\nu}}{\boldsymbol{m}}$ side counts all possible combinations of m_j elements that can be drawn from ν_j distinct elements, totaling to i elements from a collection of size $|\boldsymbol{\nu}|$. The sum over all possible $\boldsymbol{m}$ gives us every possible combination of i elements that can be drawn from a set of size $|\boldsymbol{\nu}|$, which is the same as the right hand side. □

From (4.2) and (4.3) we see that

$$\begin{aligned}(\partial^{\boldsymbol{\nu}} a)(\boldsymbol{x}, \boldsymbol{y}) &= a_0(\boldsymbol{x}) \left(\prod_{j \geq 1} (\sqrt{\mu_j} \xi_j(\boldsymbol{x}))^{\nu_j} \right) \exp(Z(\boldsymbol{x}, \boldsymbol{y})) \\ &= (a(\boldsymbol{x}, \boldsymbol{y}) - a_*(\boldsymbol{x})) \prod_{j \geq 1} (\sqrt{\mu_j}\, \xi_j(\boldsymbol{x}))^{\nu_j}\,.\end{aligned}$$

Since from (4.2) we have $0 \leq a_*(\boldsymbol{x}) \leq a(\boldsymbol{x}, \boldsymbol{y})$ for all $\boldsymbol{x} \in D$ and $\boldsymbol{y} \in U_{\boldsymbol{b}}$, it follows that

$$\left\| \frac{\partial^{\boldsymbol{\nu}} a(\cdot, \boldsymbol{y})}{a(\cdot, \boldsymbol{y})} \right\|_{L^\infty(D)} \leq \prod_{j \geq 1} b_j^{\nu_j} \qquad \forall \boldsymbol{y} \in U_{\boldsymbol{b}}\,, \tag{4.27}$$

where, for our convenience, we have written

$$b_j := \|\beta_j\|_{\boldsymbol{\mathcal{C}}^0(\overline{D})} \tag{4.28}$$

The bound in (4.27) also holds with $a(\cdot, \boldsymbol{y})$ replaced by the truncated parametric coefficient $a^s(\cdot, \boldsymbol{y})$, uniformly with respect to $s \in \mathbb{N}$. In the case of $a^s(\cdot, \boldsymbol{y})$, clearly there is no

dependence on coordinates $j > s$, hence if $\nu_j > 0$ for any $j > s$ then the left-hand side of (4.27) vanishes and the bound holds trivially. This leads to the following regularity result with respect to the parameters.

Theorem 38 *For any $\boldsymbol{y} \in U_{\boldsymbol{b}}$, any $f \in V'$, and for any multi-index $\boldsymbol{\nu}$ with $|\boldsymbol{\nu}| := \sum_{j\geq 1} \nu_j < \infty$, the solution $u(\cdot, \boldsymbol{y})$ of the parametric weak problem (4.6) satisfies the a-priori estimate*

$$\|\partial^{\boldsymbol{\nu}} u(\cdot, \boldsymbol{y})\|_V \;\leq\; \frac{|\boldsymbol{\nu}|!}{(\ln 2)^{|\boldsymbol{\nu}|}} \left(\prod_{j\geq 1} b_j^{\nu_j} \right) \frac{\|f\|_{V'}}{\breve{a}(\boldsymbol{y})} \,. \tag{4.29}$$

Moreover, the estimate (4.29) also holds with u replaced by u_h^s.

Proof. We only establish in detail the result for u as an identical argument will apply to u_h^s with all constants appearing in the bounds being independent of s and of h. Note that below the gradient operator ∇ will relate exclusively to the spatial variable $\boldsymbol{x}$, while $\partial^{\boldsymbol{\nu}}$ will relate only to the probability parameters $\boldsymbol{y}$. We first prove by induction on $|\boldsymbol{\nu}|$ that, for any fixed $\boldsymbol{y} \in U_{\boldsymbol{b}}$,

$$\left(\int_D a(\boldsymbol{x}, \boldsymbol{y}) |\nabla(\partial^{\boldsymbol{\nu}} u)(\boldsymbol{x}, \boldsymbol{y})|^2 \,\mathrm{d}\boldsymbol{x} \right)^{1/2} \;\leq\; \Lambda_{|\boldsymbol{\nu}|} \left(\prod_{j\geq 1} b_j^{\nu_j} \right) \frac{\|f\|_{V'}}{\sqrt{\breve{a}(\boldsymbol{y})}} \,, \tag{4.30}$$

where the sequence $(\Lambda_n)_{n\geq 0}$ is defined recursively by

$$\Lambda_0 \;:=\; 1 \qquad \text{and} \qquad \Lambda_n \;:=\; \sum_{i=0}^{n-1} \binom{n}{i} \Lambda_i \quad \text{for all} \quad n \geq 1 \,. \tag{4.31}$$

To obtain (4.30) for the base case $|\boldsymbol{\nu}| = 0$, we set $v = u(\cdot, \boldsymbol{y})$ in the variational form (4.6):

$$\int_D a(\boldsymbol{x}, \boldsymbol{y}) \, |\nabla u(\boldsymbol{x}, \boldsymbol{y})|^2 \,\mathrm{d}\boldsymbol{x} \;=\; \langle f, u(\cdot, \boldsymbol{y}) \rangle \;\leq\; \|f\|_{V'} \, \|u(\cdot, \boldsymbol{y})\|_V,$$

Remembering that $\langle \cdot, \cdot \rangle$ denotes the duality-pairing between V and V'. Noting that $a(\boldsymbol{x}, \boldsymbol{y})/\breve{a}(\boldsymbol{y}) \geq 1$, we now see from (4.12) that

$$\|u\|_V \;=\; \left(\int_D |\nabla u(\boldsymbol{x}, \boldsymbol{y})|^2 \,\mathrm{d}\boldsymbol{x} \right)^2 \;\leq\; \left(\int_D \frac{a(\boldsymbol{x}, \boldsymbol{y})}{\breve{a}(\boldsymbol{y})} |\nabla u(\boldsymbol{x}, \boldsymbol{y})|^2 \,\mathrm{d}\boldsymbol{x} \right)^{1/2},$$

hence we obtain

$$\int_D a(\boldsymbol{x}, \boldsymbol{y}) \, |\nabla u(\boldsymbol{x}, \boldsymbol{y})|^2 \,\mathrm{d}\boldsymbol{x} \;\leq\; \frac{\|f\|_{V'}}{\sqrt{\breve{a}(\boldsymbol{y})}} \left(\int_D a \, |\nabla u(\boldsymbol{x}, \boldsymbol{y})|^2 \,\mathrm{d}\boldsymbol{x} \right)^{1/2}.$$

Cancelling the common factor from both sides yields (4.30) for $|\boldsymbol{\nu}| = 0$.

For the following results we do not explicitly write the dependence on the variables, that is we write u for $u(\boldsymbol{x}, \boldsymbol{y})$, unless the dependence is not understood from the context.

Now, applying $\partial^{\boldsymbol{\nu}}$ to the variational formulation (4.6), and recalling that f is independent of $\boldsymbol{y}$, we obtain the identity

$$\int_D \left(\sum_{\boldsymbol{m} \preceq \boldsymbol{\nu}} \binom{\boldsymbol{\nu}}{\boldsymbol{m}} (\partial^{\boldsymbol{\nu}-\boldsymbol{m}} a)\, \nabla(\partial^{\boldsymbol{m}} u) \cdot \nabla v(\boldsymbol{x}) \right) \mathrm{d}\boldsymbol{x} \;=\; 0 \qquad \text{for all} \quad v \in V\,.$$

Taking $v = \partial^{\boldsymbol{\nu}} u(\cdot, \boldsymbol{y})$, separating out the $\boldsymbol{m} = \boldsymbol{\nu}$ term, dividing and multiplying by a, and using the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}\int_D a\, |\nabla(\partial^{\boldsymbol{\nu}} u)|^2 \,\mathrm{d}\boldsymbol{x} \;&=\; -\sum_{\substack{\boldsymbol{m} \preceq \boldsymbol{\nu} \\ \boldsymbol{m} \neq \boldsymbol{\nu}}} \binom{\boldsymbol{\nu}}{\boldsymbol{m}} \int_D (\partial^{\boldsymbol{\nu}-\boldsymbol{m}} a)\, \nabla(\partial^{\boldsymbol{m}} u) \cdot \nabla(\partial^{\boldsymbol{\nu}} u)\, \mathrm{d}\boldsymbol{x} \\ &\le \sum_{\substack{\boldsymbol{m} \preceq \boldsymbol{\nu} \\ \boldsymbol{m} \neq \boldsymbol{\nu}}} \binom{\boldsymbol{\nu}}{\boldsymbol{m}} \left\| \frac{(\partial^{\boldsymbol{\nu}-\boldsymbol{m}} a)(\cdot, \boldsymbol{y})}{a(\cdot, \boldsymbol{y})} \right\|_{L^\infty(D)} \left(\int_D a |\nabla(\partial^{\boldsymbol{m}} u)|^2 \,\mathrm{d}\boldsymbol{x} \right)^{1/2} \left(\int_D a |\nabla(\partial^{\boldsymbol{\nu}} u)|^2 \,\mathrm{d}\boldsymbol{x} \right)^{1/2}.\end{aligned}$$

Canceling the common factor on both sides and using (4.27), we arrive at

$$\left(\int_D a |\nabla(\partial^{\boldsymbol{\nu}} u)|^2 \,\mathrm{d}\boldsymbol{x} \right)^{1/2} \le \sum_{\substack{\boldsymbol{m} \preceq \boldsymbol{\nu} \\ \boldsymbol{m} \neq \boldsymbol{\nu}}} \binom{\boldsymbol{\nu}}{\boldsymbol{m}} \left(\prod_{j \ge 1} b_j^{\nu_j - m_j} \right) \left(\int_D a |\nabla(\partial^{\boldsymbol{m}} u)|^2 \,\mathrm{d}\boldsymbol{x} \right)^{1/2}.$$

We now use the inductive hypothesis (that (4.30) holds when $|\boldsymbol{\nu}| \le n-1$) in each of the terms on the right-hand side and use Identity 37 to obtain

$$\begin{aligned}\left(\int_D a |\nabla(\partial^{\boldsymbol{\nu}} u)|^2 \,\mathrm{d}\boldsymbol{x} \right)^{1/2} &\le \sum_{i=0}^{n-1} \sum_{\substack{\boldsymbol{m} \preceq \boldsymbol{\nu} \\ |\boldsymbol{m}| = i}} \binom{\boldsymbol{\nu}}{\boldsymbol{m}} \left(\prod_{j \ge 1} b_j^{\nu_j - m_j} \right) \Lambda_i \left(\prod_{j \ge 1} b_j^{m_j} \right) \frac{\|f\|_{V'}}{\sqrt{\check{a}(\boldsymbol{y})}} \\ &= \sum_{i=0}^{n-1} \binom{n}{i} \Lambda_i \left(\prod_{j \ge 1} b_j^{\nu_j} \right) \frac{\|f\|_{V'}}{\sqrt{\check{a}(\boldsymbol{y})}} \;=\; \Lambda_n \left(\prod_{j \ge 1} b_j^{\nu_j} \right) \frac{\|f\|_{V'}}{\sqrt{\check{a}(\boldsymbol{y})}}\,.\end{aligned}$$

This completes the proof of (4.30).

Next we prove by induction that

$$\Lambda_n \;\le\; \frac{n!}{(\ln 2)^n} \qquad \text{for all} \quad n \ge 0\,. \tag{4.32}$$

Clearly the result holds for Λ_0. Suppose the result holds for all Λ_i with $i \le n-1$. Then we have

$$\Lambda_n \;\le\; \sum_{i=0}^{n-1} \binom{n}{i} \frac{i!}{(\ln 2)^i} \;=\; \frac{n!}{(\ln 2)^n} \sum_{i=0}^{n-1} \frac{(\ln 2)^{n-i}}{(n-i)!} \;=\; \frac{n!}{(\ln 2)^n} \sum_{k=1}^{n} \frac{(\ln 2)^k}{k!} \;\le\; \frac{n!}{(\ln 2)^n} \left(e^{\ln 2} - 1 \right),$$

and so (4.32) holds for all n.

The final result (4.29) can be obtained by noting that

$$\|\partial^{\boldsymbol{\nu}} u(\cdot,\boldsymbol{y})\|_V \;\le\; \frac{1}{\sqrt{\breve{a}(\boldsymbol{y})}} \left(\int_D a(\boldsymbol{x},\boldsymbol{y})|\partial^{\boldsymbol{\nu}} u(\boldsymbol{x},\boldsymbol{y})|^2 \,\mathrm{d}\boldsymbol{x}\right)^{1/2}$$

then inserting (4.32) into the right-hand side of (4.30).

Our proof argument is based entirely on the weak form (4.6) which is satisfied also by $u_h^s(\cdot,\boldsymbol{y})$ if V is replaced by V_h, $\boldsymbol{y} \in U_{\boldsymbol{b}}$ is such that $y_j = 0$ for $j > s$, and a is replaced by a^s. Thus the result holds also for the finite element solution $u_h^s(\cdot,\boldsymbol{y})$ of the dimensionally truncated problem, with all constants independent of s and of h. □

4.3.2 Analysis of the QMC integration error for $\mathcal{G}(u_h^s)$

In this section we use the regularity results of §4.3.1 to bound the QMC integration error, which is the second term on the right-hand side of (4.9). Recalling (4.7), we address the efficient numerical evaluation, for large s, of integrals

$$I_{s,\phi}(F) \;:=\; \int_{\boldsymbol{y}\in\mathbb{R}^s} F(\boldsymbol{y}) \prod_{j=1}^{s} \phi(y_j)\,\mathrm{d}\boldsymbol{y}\,, \quad \text{with} \quad F(\boldsymbol{y}) \;:=\; \mathcal{G}(u_h^s(\cdot,\boldsymbol{y}))\,, \tag{4.33}$$

where $\phi(y) = e^{-y^2/2}/\sqrt{2\pi}$ is the standard normal probability density function,

$$\Phi(y) \;=\; \int_{-\infty}^{y} \frac{e^{-t^2}}{\sqrt{2\pi}}\,\mathrm{d}t$$

denotes the cumulative normal distribution function and let Φ^{-1} denote its inverse. The integral $I_s(F)$ is transformed to the unit cube by applying Φ_s^{-1} component-wise. This is precisely the problem presented in Chapter 3, accordingly we approximate this truncated integration problem using randomly shifted lattice rules $Q_{s,n}(\boldsymbol{z},\boldsymbol{\Delta};F)$.

As explored throughout this thesis, analysis of QMC quadrature relies on a suitable function space for which worst-case error analysis can be performed. We see quite clearly here that the "standard" weighted Sobolev spaces, as introduced in Chapter 2, are unlikely to be suitable for this problem. In an integral of the form (4.33) over the unbounded domain $\mathbb{R}^s$, the transformation to the unit cube yields the transformed integrand $F(\Phi_s^{-1}(\cdot))$ that may be unbounded near the boundary of the unit cube, and thus does *not* belong to the weighted Sobolev space. We also see that the spaces presented in [39] turn out to be sub-optimal, as they only allow for product weights. Here we will see, particularly in Theorem 43, that to be able to best minimise (4.8), we require the use of more general non-product weights.

A suitable function space setting for the integral (4.33) is the triple weighted unanchored space of functions on $\mathbb{R}^s$, introduced in Chapter 3. We recall the norm for this

space from (3.27),

$$\|F\|_{\mathcal{W}_s}^2 := \sum_{\mathfrak{u}\subseteq\{1:s\}} \frac{1}{\gamma_{\mathfrak{u}}} \int_{\mathbb{R}^{|\mathfrak{u}|}} \left(\int_{\mathbb{R}^{s-|\mathfrak{u}|}} \frac{\partial^{|\mathfrak{u}|} F}{\partial \boldsymbol{y}_{\mathfrak{u}}}(\boldsymbol{y}_{\mathfrak{u}}; \boldsymbol{y}_{-\mathfrak{u}}) \prod_{j\in -\mathfrak{u}} \phi(y_j)\, \mathrm{d}\boldsymbol{y}_{-\mathfrak{u}} \right)^2 \prod_{j\in\mathfrak{u}} \psi_j^2(y_j)\, \mathrm{d}\boldsymbol{y}_{\mathfrak{u}}\,, \tag{4.34}$$

where for each $j \geq 1$, the function $\psi_j : \mathbb{R} \to \mathbb{R}^+$ is the positive and continuous weight function, and for each $\mathfrak{u} \subseteq \{1:s\}$, $\gamma_{\mathfrak{u}} > 0$ is our set of weight parameters. We shall go on to specify how to best set these parameters throughout this section.

Now recalling the shift averaged worst-case error (2.2) and Theorem 21, we have the following for the second term of (4.9):

$$\mathbb{E}^{\boldsymbol{\Delta}} |\mathbb{E}[\mathcal{G}(u_h^s)] - Q_{s,n}(\boldsymbol{z}, \Delta; \mathcal{G}(u_h^s))|^2 \;=\; \mathbb{E}^{\boldsymbol{\Delta}} |I_s(F) - Q_{s,n}(\boldsymbol{z}, \cdot; F)|^2 \;\leq\; [e_{s,n}^{\mathrm{sh}}(\boldsymbol{z})]^2 \, \|F\|_{\mathcal{W}_s}^2\,,$$

where the expectation is taken over the random shift $\boldsymbol{\Delta}$ which is uniformly distributed over $[0,1]^s$,

First we show that, under a suitable assumption on the weight functions ψ_j, we have $\|F\|_{\mathcal{W}_s} < \infty$, regardless of our choice of weight parameters $\gamma_{\mathfrak{u}}$. This result is stated in Theorem 39 and makes use of the regularity results in §4.3.1, particularly Theorem 38. We can then use Theorem 39 with Thorem 21 of Chapter 3 to lead to Theorem 40, which gives an estimate for the root-mean-square error and shows that this attains a rate of convergence arbitrarily close to $\mathcal{O}(n^{-1})$, but with a possibly s-dependent asymptotic constant. Then in the following subsection we show that a careful choice of the weight parameters $\gamma_{\mathfrak{u}}$ can be made so that the asymptotic constant in the convergence estimate is bounded uniformly with respect to s, leading to the main result, Theorem 43. Throughout this work we allow for arbitrary ψ_j, as long as some conditions are maintained. Furthermore, we have some parameters, defined shortly, that depend entirely on the choice of ψ_j. We will assert some conditions on these parameters, and allow them to be unspecified through the following results.

We assume throughout the remainder of this chapter that Assumption A1 holds for some $p \leq 1$. We also require exponential decay of the ψ_j, hence we define

$$\Psi_j \;:=\; \int_{-\infty}^{\infty} \exp(2b_j\,|y|)\,\psi_j^2(y)\,\mathrm{d}y \;<\; \infty, \tag{4.35}$$

and further, let $\Psi_{\max} = \sup_{j\geq 1} \Psi_j$ and $\Psi_{\min} = \inf_{j\geq 1} \Psi_j$. We require the following,

Assumption A3 *For all $j \geq 1$ we have $0 < \Psi_{\min} \leq \Psi_j \leq \Psi_{\max} < \infty$.*

Later in this section we shall make specific choices for ψ_j, with a corresponding family of parameters α_j, for which we shall prove that this assumption holds.

In Theorem 21, the bound (3.45) contains the values $C_{2,j}$ and $r_{2,j}$, which are parameters that depend on the specific choices of ψ_j and ϕ. For simplicity of notation, from here on we write

$$\varrho_j(\lambda) = 2C_{2,j}^{\lambda}\zeta(2r_{2,j}\lambda), \tag{4.36}$$

where once again ζ denotes the Riemann Zeta function. The parameter $\varrho_j(\lambda)$ is not necessarily finite for all $\lambda \in (1/2, 1]$, see [40, 39], as well as examples 23–25 in this thesis. In fact we always require, at the very least, that $\lambda \in (1/(2r_2), 1]$, where r_2 is defined in Theorem 20

With this notation we see that (3.45) of Theorem 21 becomes

$$\sqrt{\mathbb{E}^{\boldsymbol{\Delta}}|I_s(F) - Q_{s,n}(\boldsymbol{z}, \cdot; F)|^2} \;\le\; \left(\sum_{\emptyset\neq\mathfrak{u}\subseteq\{1:s\}} \gamma_{\mathfrak{u}}^{\lambda} \prod_{j\in\mathfrak{u}} \varrho_j(\lambda)\right)^{1/(2\lambda)} [\varphi_{\mathrm{tot}}(n)]^{-1/(2\lambda)}\, \|F\|_{\mathcal{W}_s}\,, \tag{4.37}$$

where again $\varphi(n)$ denotes the Euler totient function. Evidently we see that the smaller λ, the faster our best convergence rate of the quadrature error becomes. Ideally we want $\lambda = 1/(2-2\delta)$ for some small $\delta > 0$, as this allows for optimal $\mathcal{O}(n^{-1+\delta})$ convergence. In §4.3.4 we will examine the values of r_2 and hence possible values of λ for various choices of ψ_j.

Now we show that under the appropriate assumptions, the norm of F is finite, independently of the choice of the weights $\gamma_{\mathfrak{u}}$.

Theorem 39 *For each $j \ge 1$, let b_j be defined by* (4.14) *and let Assumption 3 hold. Then the norm* (4.34) *of the integrand F in* (4.33) *satisfies the bound*

$$\|F\|_{\mathcal{W}_s}^2 \;\le\; (C^*)^2 \sum_{\mathfrak{u}\subseteq\{1:s\}} \frac{(|\mathfrak{u}|!)^2}{\gamma_{\mathfrak{u}}\,(\ln 2)^{2|\mathfrak{u}|}} \prod_{j\in\mathfrak{u}} \tilde{b}_j^2\, \Psi_j \tag{4.38}$$

where

$$\tilde{b}_j^2 \;:=\; \frac{b_j^2}{2\exp(b_j^2/2)\Phi(b_j)}, \tag{4.39}$$

with $\Phi(\cdot)$ denoting the cumulative standard normal distribution function, and with

$$C^* \;:=\; \frac{\|f\|_{V'}\,\|\mathcal{G}(\cdot)\|_{V'}}{\min_{\boldsymbol{x}\in\overline{D}} a_0(\boldsymbol{x})} \exp\left(\frac{1}{2}\sum_{j\ge1} b_j^2 + \frac{2}{\sqrt{2\pi}}\sum_{j\ge1} b_j\right)\,. \tag{4.40}$$

Proof. Now, for the integrand F from (4.33) and for any $\boldsymbol{y} \in \mathbb{R}^s$ (which we identify throughout this proof, with slight abuse of notation, with the sequence $\boldsymbol{y} \in \mathbb{R}^{\mathbb{N}}$ with $y_j = 0$ for $j > s$), we have from Theorem 38 with $\nu_j \in \{0, 1\}$ and the linearity of $\mathcal{G}$, that

$$\left|\frac{\partial^{|\mathfrak{u}|}F}{\partial \boldsymbol{y}_{\mathfrak{u}}}(\boldsymbol{y})\right| \;\le\; \|\mathcal{G}\|_{V'}\left\|\frac{\partial^{|\mathfrak{u}|}u_h^s}{\partial \boldsymbol{y}_{\mathfrak{u}}}(\cdot,\boldsymbol{y})\right\|_V \;\le\; \|f\|_{V'}\,\|\mathcal{G}\|_{V'}\frac{|\mathfrak{u}|!}{(\ln 2)^{|\mathfrak{u}|}}\left(\prod_{j\in\mathfrak{u}} b_j\right)\frac{1}{\breve{a}(\boldsymbol{y})}.$$

Since a_* in (4.2) was assumed to be non-negative and since $\min_{\boldsymbol{x}\in\overline{D}} \sqrt{\mu_j}\,\xi_j(\boldsymbol{x}) \geq -b_j$, we have

$$\check{a}(\boldsymbol{y}) \geq \min_{\boldsymbol{x}\in\overline{D}} a_0(\boldsymbol{x}) \prod_{j\geq 1} \exp(-b_j y_j).$$

Using this, introducing the notation $K^* := \|f\|_{V'}\,\|\mathcal{G}(\cdot)\|_{V'}/\min_{\boldsymbol{x}\in\overline{D}} a_0(\boldsymbol{x})$ for the sake of readability, we now have

$$\left|\frac{\partial^{|\mathfrak{u}|}F}{\partial \boldsymbol{y}_{\mathfrak{u}}}(\boldsymbol{y})\right| \;\leq\; K^* \frac{|\mathfrak{u}|!}{(\ln 2)^{|\mathfrak{u}|}} \left(\prod_{j\in\mathfrak{u}} b_j\right) \left(\prod_{j\in\{1:s\}} \exp(b_j\,|y_j|)\right). \tag{4.41}$$

Since the final term on the right-hand side of (4.41) is separable, we can group the factors corresponding to $j \in \mathfrak{u}$ and $j \in -\mathfrak{u}$ separately, allowing us to estimate the norm (4.34) as

$$\begin{aligned}\|F\|^2_{\mathcal{W}_s} \;\leq\; (K^*)^2 \sum_{\mathfrak{u}\subseteq\{1:s\}} \frac{1}{\gamma_{\mathfrak{u}}} \frac{|\mathfrak{u}|!^2}{(\ln 2)^{2|\mathfrak{u}|}} \left(\prod_{j\in\mathfrak{u}} b_j\right)^2 \left(\int_{\mathbb{R}^{s-|\mathfrak{u}|}} \prod_{j\in-\mathfrak{u}} \exp(b_j\,|y_j|)\phi(y_j)\,\mathrm{d}\boldsymbol{y}_{-\mathfrak{u}}\right)^2 \\ \times \int_{\mathbb{R}^{|\mathfrak{u}|}} \prod_{j\in\mathfrak{u}} \exp(2b_j\,|y_j|)\psi_j^2(y_j)\,\mathrm{d}\boldsymbol{y}_{\mathfrak{u}}. \end{aligned} \tag{4.42}$$

The integrals on the right hand side of (4.42) can be readily estimated. Firstly,

$$\begin{aligned}\int_{\mathbb{R}^{s-|\mathfrak{u}|}} \prod_{j\in-\mathfrak{u}} \exp(b_j\,|y_j|)\phi(y_j)\,\mathrm{d}\boldsymbol{y}_{-\mathfrak{u}} &= \prod_{j\in-\mathfrak{u}} \left(\int_{-\infty}^{\infty} \exp(b_j\,|y|)\frac{\exp(-y^2/2)}{\sqrt{2\pi}}\,\mathrm{d}y\right) \\ &= \prod_{j\in-\mathfrak{u}} \left(2\,\exp(b_j^2/2)\int_0^{\infty} \frac{\exp(-(y-b_j)^2/2)}{\sqrt{2\pi}}\,\mathrm{d}y\right) \\ &= \prod_{j\in-\mathfrak{u}} \left(2\,\exp(b_j^2/2)\,\Phi(b_j)\right). \end{aligned} \tag{4.43}$$

Secondly,

$$\int_{\mathbb{R}^{|\mathfrak{u}|}} \prod_{j\in\mathfrak{u}} \exp(2b_j\,|y_j|)\psi_j^2(y_j)\,\mathrm{d}\boldsymbol{y}_{\mathfrak{u}} \;=\; \prod_{j\in\mathfrak{u}} \left(\int_{-\infty}^{\infty} \exp(2b_j\,|y|)\,\psi_j^2(y)\,\mathrm{d}y\right) \;=\; \prod_{j\in\mathfrak{u}} \Psi_j\;, \tag{4.44}$$

Combining (4.42) with (4.43) and (4.44), we obtain

$$\|F\|^2_{\mathcal{W}_s} \;\leq\; (K^*)^2 \prod_{j\in\{1:s\}} \left(2\exp(b_j^2/2)\Phi(b_j)\right) \sum_{\mathfrak{u}\subseteq\{1:s\}} \left(\frac{1}{\gamma_{\mathfrak{u}}} \frac{|\mathfrak{u}|!^2}{(\ln 2)^{2|\mathfrak{u}|}} \prod_{j\in\mathfrak{u}} \tilde{b}_j^2\,\Psi_j\right). \tag{4.45}$$

Now, to obtain the bound (4.38), it remains to bound the product in (4.45) independently of s. To do this we note that $2\exp(b_j^2/2)\Phi(b_j) \geq 1$ and

$$\Phi(b_j) \;\leq\; \frac{1}{2}\left(1 + \frac{2\,b_j}{\sqrt{2\pi}}\right) \;\leq\; \frac{1}{2}\exp\left(\frac{2\,b_j}{\sqrt{2\pi}}\right) \quad \text{since} \quad b_j \geq 0\;.$$

Thus we have $\prod_{j\in\{1:s\}}(2\exp(b_j^2/2)\Phi(b_j)) \le \prod_{j\ge1}\exp(b_j^2/2+2\,b_j/\sqrt{2\pi})$ and the bound (4.38) then follows. □

The root-mean-square error can now be estimated by combining our results thus far.

Theorem 40 *Let F be the integrand in* (4.33) *and let Assumption 3 hold. Given $s,n\in\mathbb{N}$ with $n\le 10^{30}$, weights $\boldsymbol{\gamma}=(\gamma_{\mathfrak{u}})_{\mathfrak{u}\subset\mathbb{N}}$, and standard normal density function ϕ, a randomly shifted lattice rule with n points in s dimensions can be constructed by the CBC algorithm such that, for all $\lambda\in(1/(2r_2),1]$,*

$$\sqrt{\mathbb{E}^{\boldsymbol{\Delta}}|I_s(F)-Q_{s,n}(\cdot;F)|^2}\;\le\;9\,C^*\,C_{\boldsymbol{\gamma},s}(\lambda)\,n^{-1/(2\lambda)}, \tag{4.46}$$

with

$$C_{\boldsymbol{\gamma},s}(\lambda)\;:=\;\left(\sum_{\emptyset\ne\mathfrak{u}\subseteq\{1:s\}}\gamma_{\mathfrak{u}}^{\lambda}\prod_{j\in\mathfrak{u}}\varrho_j(\lambda)\right)^{1/(2\lambda)}\left(\sum_{\mathfrak{u}\subseteq\{1:s\}}\frac{(|\mathfrak{u}|!)^2}{\gamma_{\mathfrak{u}}\,(\ln 2)^{2|\mathfrak{u}|}}\prod_{j\in\mathfrak{u}}\tilde{b}_j^2\,\Psi_j\right)^{1/2},$$

where b_j is defined in (4.14), *$\tilde{b}_j$ is defined in* (4.39), *C^* is defined in* (4.40), *and $\varrho_j(\lambda)$ is defined in* (4.36).

Proof. The result follows immediately from Theorem 39 and (4.37) with the main result of Chapter 3, Theorem 21, also noting that for all $n\le 10^{30}$ we have that $1/\varphi(n)\le 9/n$. □

Without a careful choice of the weight parameters $\gamma_{\mathfrak{u}}$, the quantity $C_{\boldsymbol{\gamma},s}(\lambda)$ might grow (even exponentially) with increasing s. To ensure that $C_{\boldsymbol{\gamma},s}(\lambda)$ is bounded independently of s, we choose the weight parameters to ensure that

$$C_{\boldsymbol{\gamma}}(\lambda)\;:=\;\left(\sum_{|\mathfrak{u}|<\infty}\gamma_{\mathfrak{u}}^{\lambda}\prod_{j\in\mathfrak{u}}\varrho_j(\lambda)\right)^{1/(2\lambda)}\left(\sum_{|\mathfrak{u}|<\infty}\frac{(|\mathfrak{u}|!)^2}{\gamma_{\mathfrak{u}}\,(\ln 2)^{2|\mathfrak{u}|}}\prod_{j\in\mathfrak{u}}\left(\tilde{b}_j^2\,\Psi_j\right)\right)^{1/2}<\infty. \tag{4.47}$$

(Note that $\tilde{b}_j\le b_j$ and that it tends to b_j rapidly as $j\to\infty$.) Provided (4.47) holds then it follows immediately that $C_{\boldsymbol{\gamma},s}(\lambda)\le C_{\boldsymbol{\gamma}}(\lambda)<\infty$ for all s, and so the asymptotic constant in the convergence estimate (4.46) is independent of the truncation dimension s.

4.3.3 Choosing the weight parameters $\gamma_{\mathfrak{u}}$

For any given $\lambda\in(1/(2r_2),1]$, we now follow the strategy in [36] and choose the weight parameters $\gamma_{\mathfrak{u}}$ to minimise the constant $C_{\boldsymbol{\gamma}}(\lambda)$ given in (4.47). We shall see that the resulting minimal value of C_λ is finite. To do this we will use the following two lemmas, both of which can be found in [24, 36].

Lemma 41 *Let $m \in \mathbb{N}$, $\lambda > 0$, and $A_i, B_i > 0$ for all i. Then the function*

$$p(x_1, \ldots, x_m) \;=\; \left(\sum_{i=1}^{m} {x_i}^{\lambda} A_i\right)^{1/\lambda} \left(\sum_{i=1}^{m} \frac{B_i}{x_i}\right) \tag{4.48}$$

is minimised over all sequences $(x_i)_{1\le i\le m}$ when

$$x_i \;=\; c\left(\frac{B_i}{A_i}\right)^{1/(1+\lambda)} \qquad \textit{for any} \quad c > 0\,. \tag{4.49}$$

The function obtained by letting $m \to \infty$ in (4.48) is minimised when x_i is given by (4.49) for all i and has a finite value if and only if the series $\sum_{i=1}^{\infty}(A_i B_i^{\lambda})^{1/(1+\lambda)}$ converges.

Proof. We show that the choice in (4.49) is by first differentiating (4.48) by x_j for $1 \le j \le m$ to obtain

$$\frac{\partial}{\partial x_j} p(x_1, \ldots, x_m) = \lambda x_j^{\lambda-1} A_j \frac{1}{\lambda}\left(\sum_{i=1}^{m} x_i^{\lambda} A_i\right)^{1/\lambda-1} - \frac{B_j}{x_j^2}\left(\sum_{i=1}^{m} x_i^{\lambda} A_i\right)^{1/\lambda},$$

and observing that we have minima when this is set to zero, which we find happens uniquely at

$$x_j^{1+\lambda} \;=\; \frac{B_j \sum_{i=1}^{m} x_i^{\lambda} A_i}{A_j \sum_{i=1}^{m} B_i/x_i}.$$

However, we note that for any constant $c > 0$, we have that $p(cx_1, \ldots, cx_m) = p(x_1, \ldots, x_m)$, thus p is minimised regardless of the scaling of the x_i, thus we can simply chose $x_i = c(B_i/A_i)^{1/(1+\lambda)}$ for any $c > 0$. □

Lemma 42 *For all $A_j > 0$ with $\sum_{j\ge1} A_j < 1$ we have*

$$\sum_{|\mathfrak{u}|<\infty} |\mathfrak{u}|! \prod_{j\in\mathfrak{u}} A_j \;\le\; \sum_{k=0}^{\infty}\left(\sum_{j\ge1} A_j\right)^k \;=\; \frac{1}{1-\sum_{j\ge1} A_j}\,,$$

and for all $B_j > 0$ with $\sum_{j\ge1} B_j < \infty$ we have

$$\sum_{|\mathfrak{u}|<\infty} \prod_{j\in\mathfrak{u}} B_j \;=\; \prod_{j\ge1}(1+B_j) \;=\; \exp\left(\sum_{j\ge1}\log(1+B_j)\right) \;\le\; \exp\left(\sum_{j\ge1} B_j\right)$$

Proof. Note that for every set $\mathfrak{u} \subset \mathbb{N}$, there are $|\mathfrak{u}|!$ permuted equivalents in $\mathbb{N}^k$. In addition to this, there are $\boldsymbol{u} \in \mathbb{N}^k$ that have repeated elements, thus we see that

$$\left(\sum_{j\ge0} A_j\right)^k \;=\; \sum_{\boldsymbol{u}\in\mathbb{N}^k}\prod_{j\in\boldsymbol{u}} A_j \ge k! \sum_{\substack{\mathfrak{u}\subset\mathbb{N}\\ |\mathfrak{u}|=k}}\prod_{j\in\boldsymbol{u}} A_j.$$

Hence, we can conclude that

$$\sum_{|\mathfrak{u}|<\infty} |\mathfrak{u}|! \prod_{j\in\mathfrak{u}} A_j \;\le\; \sum_{k=0}^{\infty} k! \Bigg(\sum_{\substack{\mathfrak{u}\subset\mathbb{N}\\ |\mathfrak{u}|=k}} \prod_{j\in\boldsymbol{u}} A_j \Bigg) \;\le\; \sum_{k=0}^{\infty} \Bigg(\sum_{j\ge 1} A_j \Bigg)^k .$$

The final equality comes from the standard result for geometric series, provided that $\sum_{j\ge1} A_j < 1$. The second estimate follows from the observation that $\log(1+x) \le x$ for $x \ge 1$. □

Since $C_{\boldsymbol{\gamma}}(\lambda)$ in Theorem 40 is of the same general form as the function in Lemma 41, we obtain the weights (4.51) below.

Theorem 43 *Suppose that Assumption A1 holds for some $p \le 1$ and that Assumption A3 holds. If $p = 1$ assume additionally that*

$$\sum_{j\ge1} b_j \;<\; \ln 2 \sqrt{\frac{1}{\Psi_{\max}\, \varrho_{\max}(1)}}\,, \tag{4.50}$$

where $\Psi_{\max} = \sup_{j\ge1} \Psi_j$ and $\varrho_{\max}(\lambda) = \sup_{j\ge1} \varrho_j(\lambda)$. Then, for any given $\lambda \in (1/(2r_2), 1]$, the choice of weights

$$\gamma_{\mathfrak{u}} \;=\; \gamma^*_{\mathfrak{u}}(\lambda) \;:=\; \left(\frac{(|\mathfrak{u}|!)^2}{(\ln 2)^{2|\mathfrak{u}|}} \prod_{j\in\mathfrak{u}} \frac{\tilde{b}_j^2\, \Psi_j}{\varrho_j(\lambda)} \right)^{1/(1+\lambda)} \tag{4.51}$$

minimises $C_{\boldsymbol{\gamma}}(\lambda)$ given in (4.47), if a finite minimum exists. If we furthermore choose

$$\lambda_* \;:=\; \begin{cases} \frac{1}{2-2\delta} \quad \textit{for arbitrary} \quad \delta \in (0, 1/2] & \textit{when } p \in (0, 2/3]\,, \\ \frac{p}{2-p} & \textit{when } p \in (2/3, 1)\,, \\ 1 & \textit{when } p = 1\,, \end{cases} \tag{4.52}$$

and set $\lambda = \max\{\lambda_*, 1/(2r_2)\}$ and $\gamma_{\mathfrak{u}} = \gamma^*_{\mathfrak{u}}(\lambda)$, then $C_{\boldsymbol{\gamma}}(\lambda) < \infty$.

Proof. The fact that the choice of weights (4.51) minimises $C_{\boldsymbol{\gamma}}(\lambda)$ follows from Lemma 41, as in [36, Theorem 6.4], on observing that the finite subsets of $\mathbb{N}$ in (4.47) can be ordered (i.e. are countable), and that the particular ordering is immaterial, as the convergence is absolute and hence unconditional.

Let us define

$$S_\lambda \;:=\; \sum_{|\mathfrak{u}|<\infty} (\gamma^*_{\mathfrak{u}})^\lambda \prod_{j\in\mathfrak{u}} \varrho_j(\lambda) \;=\; \sum_{|\mathfrak{u}|<\infty} \left(\frac{(|\mathfrak{u}|!)^2}{(\ln 2)^{2|\mathfrak{u}|}} \prod_{j\in\mathfrak{u}} \Big([\varrho_j(\lambda)]^{1/\lambda}\, \tilde{b}_j^2\, \Psi_j \Big) \right)^{\lambda/(1+\lambda)} . \tag{4.53}$$

Then $S_\lambda^{1/(2\lambda)}$ is the first factor of $C_{\boldsymbol{\gamma}}(\lambda)$ in (4.47) with the choice of weight parameters (4.51). Moreover, the second factor in $C_{\boldsymbol{\gamma}}(\lambda)$ can also be shown to reduce to $S_\lambda^{1/2}$. Thus

we have $C_{\boldsymbol{\gamma}}(\lambda) = S_\lambda^{1/(2\lambda)+1/2}$. So, to prove $C_\gamma(\lambda)$ is finite it suffices to prove that S_λ is finite.

By definition we have $\Psi_j \le \Psi_{\max}$ and $\varrho_j(\lambda) \le \varrho_{\max}(\lambda)$ for all $1 \le j \le s$, and further we see that $\tilde{b}_j \le b_j$ for all $j \le s$. Applying these estimates to S_λ in (4.53) yields

$$S_\lambda \;\le\; \sum_{|\mathfrak{u}|<\infty} (|\mathfrak{u}|!)^{2\lambda/(1+\lambda)} \prod_{j\in\mathfrak{u}} \left(\frac{\Psi_{\max}\,[\varrho_{\max}(\lambda)]^{1/\lambda}}{(\ln 2)^2}\, b_j^2\right)^{\lambda/(1+\lambda)} . \tag{4.54}$$

In the following we consider the cases $\lambda \ne 1$ and $\lambda = 1$ separately.

For $\lambda \in (1/(2r_2), 1)$, we have $2\lambda/(1+\lambda) < 1$ and we further estimate S_λ as follows: we multiply and divide the terms on the right-hand side of (4.54) by $\prod_{j\in\mathfrak{u}} A_j^{2\lambda/(1+\lambda)}$, where $A_j > 0$ will be specified later, and then apply Hölder's inequality with conjugate exponents $(1+\lambda)/(2\lambda)$ and $(1+\lambda)/(1-\lambda)$, to obtain

$$\begin{aligned}
S_\lambda &\le \sum_{|\mathfrak{u}|<\infty} (|\mathfrak{u}|!)^{2\lambda/(1+\lambda)} \prod_{j\in\mathfrak{u}} A_j^{2\lambda/(1+\lambda)} \prod_{j\in\mathfrak{u}} \left(\frac{\Psi_{\max}\,[\varrho_{\max}(\lambda)]^{1/\lambda}}{(\ln 2)^2}\, \frac{b_j^2}{A_j^2}\right)^{\lambda/(1+\lambda)} \\
&\le \left(\sum_{|\mathfrak{u}|<\infty} |\mathfrak{u}|! \prod_{j\in\mathfrak{u}} A_j\right)^{2\lambda/(1+\lambda)} \left(\sum_{|\mathfrak{u}|<\infty} \prod_{j\in\mathfrak{u}} \left(\frac{\Psi_{\max}\,[\varrho_{\max}(\lambda)]^{1/\lambda}}{(\ln 2)^2}\, \frac{b_j^2}{A_j^2}\right)^{\lambda/(1-\lambda)}\right)^{(1-\lambda)/(1+\lambda)} \\
&\le \left(\frac{1}{1-\sum_{j\ge1} A_j}\right)^{2\lambda/(1+\lambda)} \exp\left(\frac{1-\lambda}{1+\lambda} \left(\frac{\Psi_{\max}\,[\varrho_{\max}(\lambda)]^{1/\lambda}}{(\ln 2)^2}\right)^{\lambda/(1-\lambda)} \sum_{j\ge1} \left(\frac{b_j}{A_j}\right)^{2\lambda/(1-\lambda)}\right).
\end{aligned}$$

In the last step we applied Lemma 42 which holds and guarantees that S_λ is finite, provided that

$$\sum_{j\ge1} A_j \;<\; 1 \qquad \text{and} \qquad \sum_{j\ge1} \left(\frac{b_j}{A_j}\right)^{2\lambda/(1-\lambda)} \;<\; \infty\,. \tag{4.55}$$

We now choose

$$A_j \;:=\; \frac{b_j^p}{\varpi} \qquad \text{for some} \quad \varpi > \sum_{j\ge1} b_j^p\,. \tag{4.56}$$

Then we have $\sum_{j\ge1} A_j < 1$ due to Assumption A1. Noting that Assumption A1 also implies that $\sum_{j\ge1} b_j^{p'} < \infty$ for all $p' \ge p$, we conclude that the second sum in (4.55) converges for

$$\frac{2\lambda}{1-\lambda}(1-p) \;\ge\; p \qquad \Longleftrightarrow \qquad p \;\le\; \frac{2\lambda}{1+\lambda} \qquad \Longleftrightarrow \qquad \lambda \;\ge\; \frac{p}{2-p}\,. \tag{4.57}$$

Recall that λ must be strictly between $1/(2r_2)$ and 1 for the argument above. Thus when $p \in (0, 2/3]$, we choose $\lambda = \max\{1/(2-2\delta), 1/(2r_2)\}$ for some $\delta \in (0, 1/2)$. When $p \in (2/3, 1)$, we set $\lambda = \max\{p/(2-p), 1/(2r_2)\}$.

In the case $p = 1$ we take $\lambda = 1$. Then, using Lemma 42, we obtain from (4.54) that

$$S_1 \leq \sum_{|\mathfrak{u}|<\infty} |\mathfrak{u}|! \prod_{j\in\mathfrak{u}} \left(\frac{\Psi_{\max}\, \varrho_{\max}(1)}{(\ln 2)^2} b_j^2 \right)^{1/2} \leq \left(1 - \sum_{j\geq 1} \sqrt{\Psi_{\max}\, \varrho_{\max}(1)}\, \frac{b_j}{\ln 2} \right)^{-1},$$

which is finite due to the assumption (4.50). This completes the proof. □

Following the argument in the proof of [36, Theorem 6.5], we can prove that the alternative choice of weights

$$\gamma_{\mathfrak{u}} = \gamma_{\mathfrak{u}}^{**} := \left(|\mathfrak{u}|! \prod_{j\in\mathfrak{u}} (\kappa\, b_j) \right)^{2-p} \qquad \text{for arbitrary} \quad \kappa > 0\,,$$

while not minimizing $C_{\boldsymbol{\gamma}}(\lambda)$, still ensures that $C_{\boldsymbol{\gamma}}(\lambda) < \infty$ and yields the same convergence rates under the same conditions on b_j. The form of these weights makes them much more simple to implement than the optimised weights of (4.51). This result might also seem to indicate that the approximation is somewhat robust with respect to the scaling parameters κ. However, numerical experiments indicate that arbitrary choices of κ can lead to very poor lattice rules, due to numerical instability of the worst-case error. Therefore, we recommend the choice of weight parameters (4.51) that minimises the bound. This will be discussed further in Chapter 5

4.3.4 Choosing the weight functions ψ_j

In this section we make two specific choices for the weight functions ψ_j and analyse the quantity Ψ_j and examine when Assumption A3 is satisfied and hence when $\|F\|_{\mathcal{W}_s} < \infty$ from Theorem 39. We also derive the values for $\varrho_j(\lambda)$ and the associated parameter r_2 that arise from from these choices. We see from (4.51) that our weights $\gamma_{\mathfrak{u}}$ depend directly on Ψ_j and $\varrho_j(\lambda)$, and the parameter r_2 can affect the convergence rate we attain for the error estimate.

Exponential ψ_j

Informed by the condition in Assumption A3, we chose

$$\psi_j^2(y) = \exp(-2\alpha_j|y|) \qquad \text{for some} \quad \alpha_j > 0\,. \tag{4.58}$$

It is a simple calculation to verify that

$$\Psi_j = \frac{1}{\alpha_j - b_j},$$

and evidently to satisfy Assumption A3 we require that α_j satisfy, for some constants $0 < \alpha_{\min} < \alpha_{\max} < \infty$,

$$\max(b_j, \alpha_{\min}) < \alpha_j \leq \alpha_{\max}\,, \quad j \in \mathbb{N}\,. \tag{4.59}$$

In this case have from [39, Example 5] that Theorem 21 holds for any $\lambda \in (1/2, 1]$

$$C_{2,j} = \frac{\sqrt{2\pi}\exp(\alpha_j^2/\eta)}{\pi^{2-2\eta}(1-\eta)\eta} \quad \text{and} \quad r_2 = 1-\eta \quad \text{for any} \quad \eta \in (0, 1-1/(2\lambda)) \,,$$

For simplicity we choose η to be at the mid-point of its allowable range, that is $\eta = \eta^* = 1/2 - 1/(4\lambda)$. Thus, we have

$$\varrho_j(\lambda) = 2\left(\frac{\sqrt{2\pi}\exp(\alpha_j^2/\eta^*)}{\pi^{2-2\eta^*}(1-\eta^*)\eta^*}\right)^{\lambda} \zeta(2\,(1-\eta^*)\lambda). \tag{4.60}$$

There is a trade-off implicit in these constants - the closer we choose λ to $1/2$, the smaller η, and hence the larger $\varrho_j(\lambda)$ grows.

The biggest difficulty, however, is in specifying α_j. Here we specify the choice of parameters α_j which minimises the constant $C_{\boldsymbol{\gamma}}(\lambda)$.

Corollary 44 *Following Theorem 43, the constant $C_{\boldsymbol{\gamma}}(\lambda)$, with $\lambda = \lambda_*$ given by* (4.52) *and $\gamma_{\mathrm{u}} = \gamma_{\mathrm{u}}^*(\lambda_*)$ given by* (4.51)*, is minimised by choosing*

$$\alpha_j \;=\; \frac{1}{2}\left(b_j + \sqrt{b_j^2 + 1 - \frac{1}{2\lambda_*}}\right). \tag{4.61}$$

Proof. We have $C_{\boldsymbol{\gamma}}(\lambda) = S_\lambda^{1/(2\lambda)+1/2}$ with S_λ given by (4.53), which is minimised if each factor $[\varrho_j(\lambda)]^{1/\lambda}\Psi_j = [\varrho_j(\lambda)]^{1/\lambda}/(\alpha_j - b_j)$ is minimised with respect to α_j. Thus from (4.36) we see that α_j should be chosen to minimise $e^{\alpha_j^2/\eta_*}/(\alpha_j - b_j)$. This yields the choice (4.61). □

Before discovering the scheme outlined in Corollary 44 many attempts were made to find a scheme for setting α_j that would produce decent results. Although Theorem 43 holds for any choice of α_j that satisfy (4.59), in practice and for concrete values of n, we should of course strive to minimise the constants to obtain the best performance. Furthermore, the strong sensitivity of $\varrho_j(\lambda)$, and hence $C_{\boldsymbol{\gamma}}(\lambda)$, to α_j meant that the CBC algorithm could succumb to floating-point overflow errors under an arbitrary or poor scheme for α_j. Choices in past experiments, such as setting $\alpha_j = 2b_j$, have lead to suboptimal results in our numerics. These issues are discussed in depth in Chapter 5.

For convenience we can summarise the following final result.

Theorem 45 *Under the assumptions of Theorem 43, with the exponential ψ_j* (4.58)*, a randomly shifted lattice rule can be constructed for the approximation of the integral* (4.33) *such that*

$$\sqrt{\mathbb{E}^{\boldsymbol{\Delta}}|I_s(F) - Q_{s,n}(\cdot;F)|^2} \;=\; \begin{cases} \mathcal{O}(n^{-(1-\delta)}) & \text{when } p \in (0, 2/3] \,, \\ \mathcal{O}(n^{-(1/p-1/2)}) & \text{when } p \in (2/3, 1) \,, \\ \mathcal{O}(n^{-1/2}) & \text{when } p = 1 \,, \end{cases}$$

with the implied constant independent of s, but depending on p and, when relevant, δ.

Proof. Using the value given for λ and the weights $\gamma_{\mathfrak{u}}$ defined in Theorem 43, we have that $C_{\boldsymbol{\gamma}}(\lambda)$ is finite and minimised, thus the result clearly follows from Theorem 40, noting in particular the relationship between λ and p in (4.57). □

Gaussian ψ_j

Here we explore the alternative of setting

$$\psi_j^2(y) \;=\; \exp(-\alpha_j y^2) \qquad \text{for some} \quad 0 < \alpha_j < 1/2\,. \tag{4.62}$$

It is a somewhat more involved calculation, but we see that

$$\begin{aligned}\Psi_j \;=\; \int_{-\infty}^{\infty} \exp(-(\alpha_j y^2 - 2b_j|y|)\,\mathrm{d}y \;&=\; 2\exp\left(\frac{b_j^2}{\alpha_j}\right)\int_0^{\infty} \exp(-\alpha_j(y - b_j/\alpha_j)^2)\,\mathrm{d}y \\ &=\; 2\exp\left(\frac{b_j^2}{\alpha_j}\right)\sqrt{\frac{\pi}{\alpha_j}}\int_{-b_j\sqrt{2/\alpha_j}}^{\infty} \frac{\exp(-u^2/2)}{\sqrt{2\pi}}\,\mathrm{d}y \\ &=\; 2\exp\left(\frac{b_j^2}{\alpha_j}\right)\sqrt{\frac{\pi}{\alpha_j}}\,\Phi\left(b_j\sqrt{\frac{2}{\alpha_j}}\right).\end{aligned}$$

To satisfy Assumption A3, we merely require $0 < \alpha_{\min} \le \alpha_j \le \alpha_{\max} < \infty$ for all $n \in \mathbb{N}$. However we have from [39, Example 4] that with this choice of ψ_j

$$C_{2,j} = \frac{\sqrt{2\pi}}{\pi^{2-2\alpha_j}(1-\alpha_j)\alpha_j} \quad \text{and} \quad r_{2,j} = 1 - \alpha_j\,,$$

and further, we have the following

$$\varrho_j(\lambda) = 2\left(\frac{\sqrt{2\pi}}{\pi^{2-2\alpha_j}(1-\alpha_j)\alpha_j}\right)^{\lambda} \zeta(2\,(1-\alpha_j)\lambda). \tag{4.63}$$

For our chosen λ, we require that $\varrho_j(\lambda)$ be bounded, from which we see that we require that $2(1-\alpha_j)\lambda > 1$. Thus from this relationship we see that we have $\alpha_{\max} < 1 - 1/(2\lambda)$, thus we require

$$0 < \alpha_{\min} \le \alpha_j \le \alpha_{\max} < 1 - 1/(2\lambda)\,, \quad j \in \mathbb{N}\,. \tag{4.64}$$

That is, we take λ to be the "free" parameter and accordingly restrict α_j as above. Under these conditions, $C_{2,j}$ will also remain positive as they ensure $\alpha_{\max} < 1/2$.

To produce a result as in Corollary 44, where we find a choice of α_j that minimises S_λ, we must minimise the expression $[\varrho_j(\lambda)]^{1/\lambda}\Psi_j$ for each α_j. We quickly see that this minimising choice can not be calculated analytically with Gaussian weight functions ψ_j, we have no closed form for the choice of α_j based on the b_j and λ. It is not hard, however, to implement a numerical minimisation routine to find an appropriate α_j for each j. This step is discussed further in the next chapter, and results of this minimisation are discussed and compared with the results from the exponential setting.

Regardless, we summarise our final result in the following theorem.

Theorem 46 *Under the assumptions of Theorem 43, with the Gaussian ψ_j (4.62), a randomly shifted lattice rule can be constructed for the approximation of the integral (4.33) such that*

$$\sqrt{\mathbb{E}^{\boldsymbol{\Delta}}|I_s(F) - Q_{s,n}(\cdot;F)|^2} = \begin{cases} \mathcal{O}(n^{-1-\delta}) & \text{when } p \in (0, 2/3] \,, \\ \mathcal{O}(n^{-(1/p-1/2)}) & \text{when } p \in (2/3, 1) \,, \\ \mathcal{O}(n^{-1/2}) & \text{when } p = 1 \,, \end{cases}$$

where $\alpha_{\max} < 1 - 1/(2\lambda)$, and again the implied constant independent of s, but depending on p and, when relevant, δ.

Proof. The result follows in the same fashion as Theorem 45. □

If $p \in (0, 2/3]$ then we can choose any $\lambda \in (1/2, 1]$, so we set $\lambda = 1/2 + q$, with $q > 0$ and not too small, as we can see from (4.64) that $\alpha_{\max} < 1 - 1/(1 + 2q)$, hence small q means $\alpha_{\max}$ approaches 0. We see however that Ψ_j blows up exponentially with small α_j, hence for the numerical implementation we will want q to be of some "reasonable" size. This will be discussed further in Chapter 5.

4.4 Final result

We now summarise our theoretical results and state our combined bound for the root-mean-square error, which includes the finite element error, the dimension truncation error and the QMC quadrature error, estimated in Theorems 34, 35, 43, 45 and 46, respectively.

Theorem 47 *We consider approximations of the expected value of $\mathcal{G}(u)$ via quasi-Monte Carlo finite element methods. In particular, we apply a randomly shifted lattice rule $Q_{s,n}$ to $\mathcal{G}(u_h^s)$. Then, under the same assumptions and definitions as in Theorems 34, 35, 43, 45 and 46, the root-mean-square error with respect to the uniformly distributed shift $\boldsymbol{\Delta} \in [0,1]^s$ can be bounded by*

$$\sqrt{\mathbb{E}^{\boldsymbol{\Delta}}\left[\left(\mathbb{E}[\mathcal{G}(u)] - Q_{s,n}(\cdot;\mathcal{G}(u_h^s))\right)^2\right]} \le C\left(h^2 + s^{-\chi} + n^{-r}\right),$$

for some $0 < \chi < 1/p - 1/2$, and with $r = 1/p - 1/2$ for $p \in (2/3, 1]$ and $r = 1 - \delta$ for $p < 2/3$, with δ arbitrarily small. The rates χ and r depend on the parameter p which in turn depend on the asymptotics of the parameters (μ_j, ξ_j) in Assumption A1. The constant C is independent of h, s, and n.

Note that the rate r is capped at 1 even for $p < 2/3$ because we are using only QMC methods of order one. With higher order QMC methods we might expect to have $r = 1/p - 1/2$ also for $p < 2/3$. Finally, a recent result in [8] shows that under slightly stronger conditions on the data, the rate χ in the truncation error can also be increased to 2χ.

Chapter 5

Implementation and numerical results

In this section we present details of implementations of various algorithms presented through this thesis, including the CBC algorithm for the unbounded spaces with POD weights and the application of lattice rules to the porous flow problem. We shall also discuss numerous hurdles and issues with numerical stability encountered throughout the development and testing of these algorithms.

The bulk of this chapter presents original and novel work. The implementation of these algorithms, particularly the CBC algorithm as presented in Chapter 3, is original work by the author, and the numerical stability issues involved in this work presented new and exciting challenges for the field. In addition to this, the premise tuning process for the weights, as outlined in Chapter 4, is relatively new territory in the field of QMC quadrature.

We note that all the code for these numerical experiments was written in Python 2.7, using the Scipy and Numpy packages, and the MatPlotLib library for plotting our results. Other than standard library function calls in the Scipy and Numpy libraries, all code is original and written by the author.

The outline of the chapter is as follows. In §5.1 we discuss the implementation of the CBC algorithm for the settings introduced in Chapter 3. We present well known optimisations of the algorithm, including the use of FFT techniques. In §5.1.1 we present details of the CBC when using POD weights in unanchored space, while in §5.1.2 we discuss the difficulties inherent in the anchored setting, but present possible techniques to proceed nonetheless. We present the results of some of the numerical experiments in §5.2, examining the impact of worst-case errors for different choices of ϕ, ψ_j, and weights $\gamma_{\mathfrak{u}}$. In §5.3 we specify a model problem to highlight the theory studied in Chapter 4. In §5.3.1 we present the results for the choice of exponential weight functions, both the worst-case error results from the CBC algorithm, and the standard error results from the QMC algorithm. In §5.3.2 we discuss the regimes for setting the parameters of the weights functions. Next we present similar results for the Gaussian weight functions, in §5.3.3. Finally we summarise our discoveries in §5.4.

5.1 Implementing the CBC algorithm

Here we specify further technical details relevant to the implementation of the CBC algorithm. This includes a full description of the fast CBC algorithm for the unanchored space, including the matrix permutations necessary to use FFT methods to speed up

the matrix-vector multiplications implicit in Algorithm 19. We discuss some numerical challenges that we had and propose their remedies, including overflow problem in the weights. We then discuss a regime for the numerical quadrature to calculate the shift-invariant kernel, a non-trivial matter in the unbounded spaces. Finally we present worst-case errors for a set of model problems.

5.1.1 Fast CBC construction for POD weights in the unanchored space

Implementation of the CBC construction as described in Algorithm 19 is infeasible unless some structure is assumed for the weights $\gamma_{\mathfrak{u}}$. For product weights $\gamma_{\mathfrak{u}} = \prod_{j\in\mathfrak{u}} \gamma_j$, a fast CBC implementation based on FFT is known from [53, 54], which requires only $\mathcal{O}(s\, n \log n)$ operations and $\mathcal{O}(n)$ memory. For order-dependent weights $\gamma_{\mathfrak{u}} = \Gamma_{|\mathfrak{u}|}$, a similar fast CBC implementation is discussed in [11], which requires $\mathcal{O}(s\, n \log n + s^2\, n)$ operations and $\mathcal{O}(s\, n)$ memory. The CBC implementation for POD weights (2.23) is presented in [37] and has the same cost as order-dependent weights.

Here we present a refinement of the strategy from [37] to avoid numerical overflow when the order-dependent parts of POD weights grow very quickly. This consideration is motivated by the form of POD weights which arise from the application to PDEs with random coefficients, see [36],

$$\gamma_{\mathfrak{u}} \;=\; (|\mathfrak{u}|!)^a \prod_{j\in\mathfrak{u}} \gamma_j, \qquad a > 0. \tag{5.1}$$

That is, we have POD weights (2.23) where $\Gamma_{|\mathfrak{u}|} = (|\mathfrak{u}|!)^a$. In general, rather than working directly with the sequence $\{\Gamma_\ell\}_{\ell\geq 0}$ in the CBC construction, we shall work with their ratios

$$\tau_\ell \;:=\; \frac{\Gamma_\ell}{\Gamma_{\ell-1}}, \quad \text{so that} \quad \Gamma_\ell = \prod_{i=1}^{\ell} \tau_i. \tag{5.2}$$

Thus in the particular case (5.1) we have $\tau_\ell = \ell^a$.

For the unanchored space there are no auxiliary weights, and we have from (3.37) $E^2_{d,s}(\boldsymbol{z}) = [e^{\mathrm{sh}}_{d,n}(\boldsymbol{z})]^2$. At each step of the CBC algorithm we consider the squared worst-case error $e^{\mathrm{sh}}_{d+1,n}(\boldsymbol{z}, z_{d+1})$ to be a function of z_{d+1}, with $\boldsymbol{z}$ fixed, and for simplicity we write $e^2_{d+1}(z_{d+1}) = [e^{\mathrm{sh}}_{d+1,n}(\boldsymbol{z}, z_{d+1})]^2$. Substituting POD weights (2.23) into (3.32) and

using (5.2), we have

$$
\begin{aligned}
e^2_{d+1}(z_{d+1}) \;&=\; \frac{1}{n}\sum_{k=1}^{n}\sum_{\ell=1}^{d+1}\sum_{\substack{\mathfrak{u}\subseteq\{1:d+1\}\\|\mathfrak{u}|=\ell}}\Big(\prod_{i=1}^{\ell}\tau_i\Big)\Big(\prod_{j\in\mathfrak{u}}\gamma_j\,\theta_j\Big(\Big\{\frac{kz_j}{n}\Big\}\Big)\Big)\\
&=\; \frac{1}{n}\sum_{k=1}^{n}\sum_{\ell=1}^{d+1}\Bigg(\underbrace{\sum_{\substack{\mathfrak{u}\subseteq\{1:d\}\\|\mathfrak{u}|=\ell}}\Big(\prod_{i=1}^{\ell}\tau_i\Big)\Big(\prod_{j\in\mathfrak{u}}\gamma_j\,\theta_j\Big(\Big\{\frac{kz_j}{n}\Big\}\Big)\Big)}_{q_{d,\ell}(k)}\\
&\qquad\qquad +\,\gamma_{d+1}\tau_\ell\,\theta_{d+1}\Big(\Big\{\frac{kz_{d+1}}{n}\Big\}\Big)\underbrace{\sum_{\substack{\mathfrak{u}\subseteq\{1:d\}\\|\mathfrak{u}|=\ell-1}}\Big(\prod_{i=1}^{\ell-1}\tau_i\Big)\Big(\prod_{j\in\mathfrak{u}}\gamma_j\,\theta_j\Big(\Big\{\frac{kz_j}{n}\Big\}\Big)\Big)}_{q_{d,\ell-1}(k)}\Bigg)\\
&=\; [e^{\mathrm{sh}}_{d,n}(\boldsymbol{z})]^2 \;+\; \frac{\gamma_{d+1}}{n}\sum_{k=1}^{n}\theta_{d+1}\Big(\Big\{\frac{kz_{d+1}}{n}\Big\}\Big)\sum_{\ell=1}^{d+1}\tau_\ell\,q_{d,\ell-1}(k).
\end{aligned}
$$

We deduce the following recursions to compute $q_{d,\ell}(k)$

$$
\begin{aligned}
q_{d,0}(k) \;&:=\; 1,\\
q_{d+1,\ell}(k) \;&:=\; q_{d,\ell}(k) + \gamma_{d+1}\,\tau_\ell\,\theta_{d+1}\Big(\Big\{\frac{kz_{d+1}}{n}\Big\}\Big)\,q_{d,\ell-1}(k),
\end{aligned}
\tag{5.3}
$$

with $q_{d,\ell}(k) := 0$ if $\ell > d$ or $\ell < 0$. We need to evaluate $e^2_{d+1}(z_{d+1})$ for all $z_{d+1} \in \mathcal{Z}_n$. This suggests a matrix-vector operation, where we have the vectors

$$
\boldsymbol{e}^2_{d+1} \;:=\; [e^2_{d+1}(z)]_{z\in\mathcal{Z}_n}, \quad \boldsymbol{q}_{d,\ell} \;:=\; [q_{d,\ell}(k)]_{1\le k\le n},
$$

and the matrix

$$
\boldsymbol{\Omega}_{n,d+1} \;:=\; \Big[\theta_{d+1}\Big(\Big\{\frac{kz}{n}\Big\}\Big)\Big]_{\substack{z\in\mathcal{Z}_n\\1\le k\le n}} \;=\; \Big[\theta_{d+1}\Big(\frac{kz \bmod n}{n}\Big)\Big]_{\substack{z\in\mathcal{Z}_n\\1\le k\le n}}.
$$

We can now write the calculation of $\boldsymbol{e}^2_{d+1}$ as follows

$$
\boldsymbol{e}^2_{d+1} \;=\; [e^{\mathrm{sh}}_{d,n}(\boldsymbol{z})]^2\,\mathbf{1}_{\varphi(n)} \;+\; \frac{\gamma_{d+1}}{n}\,\boldsymbol{\Omega}_{n,d+1}\Big(\sum_{\ell=1}^{d+1}\tau_\ell\,\boldsymbol{q}_{d,\ell-1}\Big),
\tag{5.4}
$$

where $\mathbf{1}_{\varphi(n)}$ is a vector of ones of length $\varphi(n)$ (i.e., with as many elements as there are rows in $\boldsymbol{\Omega}_{n,d+1}$). Now we choose the value of $z_{d+1} \in \mathcal{Z}_n$ that corresponds to the the smallest entry in $\boldsymbol{e}^2_{d+1}$. For this crucial step, we can reduce the cost of the matrix-vector multiplication by using FFT methods. This reduces the cost of the multiplication from $\mathcal{O}(n^2)$ to $\mathcal{O}(n\log n)$. The key is to permute the rows and columns of the matrix $\boldsymbol{\Omega}_{n,d+1}$

so that it becomes a circulant matrix (except for the column of zeros corresponding to $k = n$); this method was discovered and outlined in [53, 54, 11]. Here we summarise the permutation procedure. For the sake of simplicity, we restrict ourselves again to the case where n is prime, again noting then that we have $\mathcal{Z}_n = \{1, \ldots, n-1\}$. We also ignore, for the time being, the dimensional subscript d on $\mathbf{\Omega}$ and θ, that is we simply write $\mathbf{\Omega}_n$ and $\theta(u)$.

We note that we can write $\mathbf{\Omega}_n$ in the form

$$\mathbf{\Omega}_n = \begin{bmatrix} \omega_{1,1} & \omega_{1,2} & \cdots & \omega_{1,n} \\ \omega_{2,1} & \omega_{2,2} & & \omega_{2,n} \\ \vdots & & \ddots & \vdots \\ \omega_{n-1,1} & \omega_{n-1,2} & \cdots & \omega_{n-1,n} \end{bmatrix} = \begin{bmatrix} \omega_1 & \omega_2 & \cdots & \omega_{n-1} & \omega_n \\ \omega_2 & \omega_4 & \cdots & \omega_{n-2} & \omega_n \\ \omega_3 & \omega_6 & \cdots & \omega_{n-3} & \omega_n \\ \vdots & & \ddots & \vdots & \vdots \\ \omega_{n-1} & \omega_{n-2} & \cdots & \omega_1 & \omega_0 \end{bmatrix} = \left[\begin{array}{c:c} \mathbf{\Omega}'_n & \omega_n \mathbf{1}_{n-1} \end{array} \right], \tag{5.5}$$

where we have used the notation

$$\omega_{i,j} = \theta\left(\frac{ij \bmod n}{n}\right) \qquad \text{and} \qquad \omega_j = \theta\left(\frac{j}{n}\right),$$

and further we have written $\mathbf{\Omega}'_n$ to indicate $\mathbf{\Omega}_n$ with the last column removed, and $\mathbf{1}_{n-1}$ denotes the vector of ones of length $n-1$. Also note from the definition that $\omega_n = \omega_0$. Later we also use the notation $\boldsymbol{\omega}_n := [\omega_1, \omega_2, \ldots, \omega_n]^T$.

For prime n the set $\mathcal{Z}_n$ has a *generator* g, that is, an integer $g \in \mathcal{Z}_n$ such that $g^k \bmod n$ covers all of $\mathcal{Z}_n$, or $\{g^k \bmod n : 0 \leq k < n-1\} = \mathcal{Z}_n$. Using this generator we can create a permutation matrix that performs the *Rader factorisation*. This is best illustrated with an example. If we take $n = 7$, then we have the set of possible components of the generating vector $\mathcal{Z}_7 = \{1, 2, 3, 4, 5, 6\}$. We take the generator to be $g = 5$, and we have the ordered vector

$$[g^k \bmod 7 : 0 \leq k < 7] = [1, 5, 4, 6, 2, 3] := \boldsymbol{g}\,,$$

and we see that taking the negative powers simply reverses part of the vector,

$$[g^{-k} \bmod 7 : 0 \leq k < 7] = [1, 3, 2, 6, 4, 5] := \boldsymbol{g}_{\text{inv}}\,.$$

We write $\mathbf{\Pi}'_g$ and $\mathbf{\Pi}'_{g^{-1}}$ for the matrix that permutes the ordered vector $[1, 2, 3, 4, 5, 6]$ to $\boldsymbol{g}$ and $\boldsymbol{g}_{\text{inv}}$ respectively, that is,

$$\mathbf{\Pi}'_{7,g} \cdot [1, 2, 3, 4, 5, 6]^T = \boldsymbol{g} \qquad \text{and} \qquad \mathbf{\Pi}'_{7,g^{-1}} \cdot [1, 2, 3, 4, 5, 6]^T = \boldsymbol{g}_{\text{inv}}\,.$$

It can be checked that

$$\boldsymbol{\Pi}'_{7,g^{-1}}\,\boldsymbol{\Omega}'_7\,\boldsymbol{\Pi}'^{T}_{7,g} = \begin{bmatrix} 1&0&0&0&0&0\\ 0&0&1&0&0&0\\ 0&1&0&0&0&0\\ 0&0&0&0&0&1\\ 0&0&0&1&0&0\\ 0&0&0&0&1&0 \end{bmatrix} \begin{bmatrix} \omega_1&\omega_2&\omega_3&\omega_4&\omega_5&\omega_6\\ \omega_2&\omega_4&\omega_6&\omega_1&\omega_3&\omega_5\\ \omega_3&\omega_6&\omega_2&\omega_5&\omega_1&\omega_4\\ \omega_4&\omega_1&\omega_5&\omega_2&\omega_6&\omega_3\\ \omega_5&\omega_3&\omega_1&\omega_6&\omega_4&\omega_2\\ \omega_6&\omega_5&\omega_4&\omega_3&\omega_2&\omega_1 \end{bmatrix} \begin{bmatrix} 1&0&0&0&0&0\\ 0&0&0&0&1&0\\ 0&0&0&0&0&1\\ 0&0&1&0&0&0\\ 0&1&0&0&0&0\\ 0&0&0&1&0&0 \end{bmatrix}$$

$$= \begin{bmatrix} \omega_1&\omega_5&\omega_4&\omega_6&\omega_2&\omega_3\\ \omega_3&\omega_1&\omega_5&\omega_4&\omega_6&\omega_2\\ \omega_2&\omega_3&\omega_1&\omega_5&\omega_4&\omega_6\\ \omega_6&\omega_2&\omega_3&\omega_1&\omega_5&\omega_4\\ \omega_4&\omega_6&\omega_2&\omega_3&\omega_1&\omega_5\\ \omega_5&\omega_4&\omega_6&\omega_2&\omega_3&\omega_1 \end{bmatrix} =: \mathbf{C}_7\,.$$

Note that $\mathbf{C}_7$ is a matrix of *circulant* form, it is uniquely determined by its top row, $\boldsymbol{c}_7 = [\omega_1, \omega_5, \omega_4, \omega_6, \omega_2, \omega_3]$. The remaining rows of their matrix are made up of successive right-translates of $\boldsymbol{c}_7$.

The ability to permute the matrix $\boldsymbol{\Omega}_n$ into circulant form applies in this form for any prime n. Given prime n, and its generator g, we write $\boldsymbol{\Pi}_{n,g}$ for the permutation such that

$$\text{if} \quad \boldsymbol{z} = \boldsymbol{\Pi}_{n,g}\boldsymbol{x} \quad \text{then} \quad z_i = x_{k(i)} \quad \text{where} \quad k(i) = \begin{cases} g^{i-1} \bmod n & \text{if } 1 \le i \le n-1, \\ n & \text{if } i = n. \end{cases}$$

Note the use of the prime notation, that is $\boldsymbol{\Pi}'_{n,g}$ to indicate $\boldsymbol{\Pi}_{n,g}$ with the last row and column omitted.

Theorem 48 (Rader factorisation) *Given prime n and a matrix $\boldsymbol{\Omega}_n$ with structure as in* (5.5) *there is a generator g of $\mathcal{Z}_n$ and a permutations $\boldsymbol{\Pi}_{n,g}$ and $\boldsymbol{\Pi}_{n,g^{-1}}$ such that*

$$\boldsymbol{\Pi}'_{n,g^{-1}}\,\boldsymbol{\Omega}_n\,\boldsymbol{\Pi}^{T}_{n,g} = \begin{bmatrix} \mathbf{C}_n & \omega_n \mathbf{1}_{n-1} \end{bmatrix} := \mathbf{C}^{\text{ext}}_n,$$

and furthermore $\boldsymbol{\Pi}'_{n,g}\,\boldsymbol{\omega}_n = \boldsymbol{c}_n$.

It is well known that multiplication by a circulant matrix $\mathbf{C}_n$ can in fact be performed by Fast Fourier Transform (FFT). We see that

$$\mathbf{C}_n\boldsymbol{x} = \boldsymbol{F}_n\boldsymbol{\Lambda}(\boldsymbol{c}_n)\boldsymbol{F}_n\boldsymbol{x}$$

where we have written $\boldsymbol{F}_n$ for the discrete Fourier transform matrix, and $\boldsymbol{\Lambda}(\boldsymbol{c}_n)$ for the matrix with the Fourier components of $\boldsymbol{c}_n$ along the diagonal, that is $\boldsymbol{\Lambda}(\boldsymbol{c}_n) = \text{diag}(\boldsymbol{F}_n\boldsymbol{c}_n)$. All matrix-vector multiplications involving $\boldsymbol{F}_n$ can be performed using FFT, which is well

known to be $\mathcal{O}(n \log n)$ in computational cost, thus the overall matrix-vector multiplication is $\mathcal{O}(n \log n)$ as well. Note that multiplication by $\mathbf{C}_n^{\text{ext}}$ does not cost significantly more as the extra column only adds $\mathcal{O}(n)$ operations to the overall matrix-vector multiplication.

Note that permutation matrices are orthogonal, hence we have that

$$\boldsymbol{\Pi}'^{T}_{n,g^{-1}} \, \mathbf{C}_n^{\text{ext}} \, \boldsymbol{\Pi}_{n,g} = \boldsymbol{\Omega}_n$$

Thus, noting also that as n is prime we have $\varphi(n) = n - 1$, we can write (5.4) as follows,

$$\boldsymbol{e}_{d+1}^2 \;=\; [e_{d,n}^{\text{sh}}(\boldsymbol{z})]^2 \, \mathbf{1}_{n-1} \;+\; \boldsymbol{\Omega}_{n,d+1} \boldsymbol{q} \qquad \text{where} \qquad \boldsymbol{q} \;:=\; \frac{\gamma_{d+1}}{n} \left(\sum_{\ell=1}^{d+1} \tau_\ell \boldsymbol{q}_{d,\ell-1} \right) .$$

The matrix-vector multiplication can be performed as follows,

$$\boldsymbol{\Omega}_{n,d+1} \boldsymbol{q} = \boldsymbol{\Pi}'^{T}_{n,g^{-1}} \, \mathbf{C}_n^{\text{ext}} \, \boldsymbol{\Pi}_{n,g} \, \boldsymbol{q} \, .$$

If we use the permuted vectors $\boldsymbol{q}^P = \boldsymbol{\Pi}_{n,g} \boldsymbol{q}$ and $[\boldsymbol{e}_d^P]^2 = \boldsymbol{\Pi}'_{n,g^{-1}} \, \boldsymbol{e}_d^2$, then if we multiply (5.4) by $\boldsymbol{\Pi}'_{n,g^{-1}}$, we find

$$[\boldsymbol{e}_{d+1}^P]^2 \;=\; [e_{d,n}^{\text{sh}}(\boldsymbol{z})]^2 \, \mathbf{1}_{n-1} \;+\; \mathbf{C}_n^{\text{ext}} \, \boldsymbol{q}^P \, ,$$

noting that the permutation makes no difference to the vector $\mathbf{1}_{n-1}$. Thus the operation can be made to take $\mathcal{O}(n \log n)$ time. To find the next component of the generating vector, z_{d+1}, we find the smallest value of $\boldsymbol{e}_{d+1}^P$, which say is at index z_{d+1}^P, then simply take $z_{d+1} = g^{-(z_{d+1}^P - 1)} \bmod n$.

The next step is to generate the vectors $\boldsymbol{q}_{d+1,\ell}$ for each $\ell = 1, \ldots, d+1$, which, under the recursion (5.3), would normally be of the form

$$\boldsymbol{q}_{d+1,\ell} \;=\; \boldsymbol{q}_{d,\ell} + \gamma_{d+1} \, \tau_\ell \, \boldsymbol{\Omega}_{n,d+1}(z_{d+1}) \;.*\; \boldsymbol{q}_{d,\ell-1},$$

where $\boldsymbol{\Omega}_{n,d+1}(z_{d+1})$ is the row of the matrix $\boldsymbol{\Omega}_{n,d+1}^P$ that corresponds to the chosen z_{d+1}, and the operator $.*$ denotes element-wise vector-vector multiplication. However it is better maintain the permuted ordering of elements, to save on computational cost and storage. Thus, using the notation $\boldsymbol{q}_{d,\ell}^P = \boldsymbol{\Pi}_{n,g} \boldsymbol{q}_{d,\ell}$ we have the following recursion

$$\boldsymbol{q}_{d+1,\ell}^P \;=\; \boldsymbol{q}_{d,\ell}^P + \gamma_{d+1} \, \tau_\ell \, \mathbf{C}_{n,d+1}^{\text{ext}}(z_{d+1}^P) \;.*\; \boldsymbol{q}_{d,\ell-1}^P,$$

where now we multiply element-wise by the row of $\mathbf{C}_{n,d+1}$ that corresponds to the index of the minimum of $\boldsymbol{e}_{d+1}^P$.

We see that we must maintain storage of the vectors $\boldsymbol{q}_{d+1,\ell}^P$ for $\ell = 1, \ldots, d$ at each iteration $d+1$, however we can overwrite $\boldsymbol{q}_{d,\ell}^P$ with $\boldsymbol{q}_{d+1,\ell}^P$ at each step, hence we require $\mathcal{O}(s \, n)$ storage for the algorithm. Note that there is no need to permute the vectors $\boldsymbol{q}_{d+1,\ell}^P$ since in the first dimension all components are initialised to the same value 1.

Therefore this procedure has a "search" cost of $\mathcal{O}(n \log n)$ operations which corresponds to the use of FFT for the matrix-vector multiplication, and there is an "update" cost of $\mathcal{O}(d\,n)$ operations at step d which is needed for calculating the vectors $\boldsymbol{q}_{d,\ell}^P$. The overall construction cost is therefore

$$\sum_{d=1}^{s} \mathcal{O}(n \log n + d\,n) \;=\; \mathcal{O}(s\,n \log n + s^2 n) \quad \text{operations.}$$

5.1.2 Fast CBC construction for POD weights in the anchored space

For the anchored space, POD weights are not preserved by the auxiliary weights which are used for the implementation of the CBC algorithm, making the computational cost prohibitive.

A remedy has been proposed in [17] for POD weights of the special form (5.1) which arise from PDE applications. The corresponding auxiliary weights (3.33) can be bounded as follows

$$\begin{aligned}
\widetilde{\gamma}_{s,\mathfrak{v}} \;&=\; \sum_{\mathfrak{v}\subseteq\mathfrak{u}\subseteq\{1:s\}} (|\mathfrak{u}|!)^a \left(\prod_{j\in\mathfrak{u}} \gamma_j\right)\left(\prod_{j\in\mathfrak{u}\setminus\mathfrak{v}} C_{0,j}\right) \\
&=\; (|\mathfrak{v}|!)^a \left(\prod_{j\in\mathfrak{v}} \gamma_j\right) \sum_{\mathfrak{w}\subseteq\{1:s\}\setminus\mathfrak{v}} \left(\frac{(|\mathfrak{v}|+|\mathfrak{w}|)!}{|\mathfrak{v}|!}\right)^a \prod_{j\in\mathfrak{w}} (C_{0,j}\,\gamma_j) \\
&\le\; (|\mathfrak{v}|!)^a \left(\prod_{j\in\mathfrak{v}} \gamma_j\right) \sum_{\mathfrak{w}\subseteq\{1:s\}\setminus\mathfrak{v}} \left(|\mathfrak{w}|!\,2^{|\mathfrak{v}|+|\mathfrak{w}|}\right)^a \prod_{j\in\mathfrak{w}} (C_{0,j}\,\gamma_j) \;\le\; \widetilde{\widetilde{\gamma}}_{\mathfrak{v}}\, c_{s,\boldsymbol{\gamma}},
\end{aligned}$$

where

$$\widetilde{\widetilde{\gamma}}_{\mathfrak{v}} \;:=\; (|\mathfrak{v}|!)^a \prod_{j\in\mathfrak{v}} (2^a\,\gamma_j) \quad \text{and} \quad c_{s,\boldsymbol{\gamma}} \;:=\; \sum_{\mathfrak{w}\subseteq\{1:s\}} (|\mathfrak{w}|!)^a \prod_{j\in\mathfrak{w}} (2^a\,C_{0,j}\,\gamma_j).$$

Using this we can see, from (3.35), that

$$[e_{s,n}^{\mathrm{sh}}(\boldsymbol{z})]^2 \;\le\; c_{s,\boldsymbol{\gamma}} \sum_{\emptyset\ne\mathfrak{v}\subseteq\{1:s\}} \frac{\widetilde{\widetilde{\gamma}}_{\mathfrak{v}}}{n} \sum_{k=1}^{n} \prod_{j\in\mathfrak{v}} \left(\theta_j\left(\left\{\frac{kz_j}{n}\right\}\right) - C_{0,j}\right).$$

The expression on the right-hand side, without the $c_{s,\boldsymbol{\gamma}}$ factor, can be used as the search criterion in the CBC algorithm and we can obtain a similar error bound to that in Theorem 20. Since the new weights $\widetilde{\widetilde{\gamma}}_{\mathfrak{v}}$ are of POD form, the algorithm can be implemented as in the case of the unanchored space.

5.1.3 Computing θ_j

In order to be able to compute the shift-averaged worst-case error given by (3.21) or (3.32), we must be able to compute $\theta_j(i/n)$ as defined in (3.20), (3.28) or (3.29) for $i = 0, \ldots, n-1$. For simplicity we consider here the domain $D = \mathbb{R}$ and anchor $c = 0$. We also assume that ϕ and ψ_j are symmetric about 0. Thus, for the unanchored space

we see that for $i \le \lfloor n/2 \rfloor$, (3.29) becomes

$$\begin{aligned}\theta_j\left(\tfrac{i}{n}\right) &= 2\int_{\Phi^{-1}(i/n)}^{0} \frac{\Phi(t) - i/n}{\psi_j^2(t)}\,\mathrm{d}t - 2\int_{-\infty}^{0} \frac{\Phi^2(t)}{\psi_j^2(t)}\,\mathrm{d}t \\ &= 2\int_{i/n}^{1/2} \frac{x - i/n}{\psi_j^2(\Phi^{-1}(x))\,\phi(\Phi^{-1}(x))}\,\mathrm{d}x - 2\int_{0}^{1/2} \frac{x^2}{\psi_j^2(\Phi^{-1}(x))\,\phi(\Phi^{-1}(x))}\,\mathrm{d}x, \qquad (5.6)\end{aligned}$$

where we used the substitution $x = \Phi(t)$. We also have $C_{0,j} = 0$ and $C_{1,j} = \theta_j(0)$. For $i > \lfloor n/2 \rfloor$, we use $\theta_j(i/n) = \theta_j((n-i)/n)$ due to symmetry.

The integrals in (5.6) may now be computed using a one-dimensional quadrature. However, there is a singularity at $x = 0$ for both integrands, thus we make use of the tanh-sinh transform first proposed in [65], see also [5, 4]. For the first integral in (5.6) we use the substitution

$$x \;=\; v(t) \;=\; \left(\frac{1}{2} - \frac{i}{n}\right)\tanh\left(\frac{\pi}{2}\sinh(t)\right) + \frac{1}{2}.$$

which maps the interval $(-\infty, 0]$ to $(i/n, 1/2]$. A similar substitution can be used for the second integral. We then approximate the integrals by the sum $h\sum_{k=-m}^{0} \Upsilon(v(kh))v'(kh)$, where $\Upsilon(t)$ is our integrand, m is the number of quadrature points, and h is the mesh-size which is chosen here to be $h = \frac{2}{m}\log(\pi m)$ to balance the truncation and discretization errors (see [43] for details). Note that $v(-t) = -v(t)$, so indeed the sum approximates the transformed integral on the half real line $(-\infty, 0]$.

This quadrature for calculating θ_j, $C_{0,j}$ and $C_{1,j}$ evidently requires $\mathcal{O}(m\,n)$ operations. We must choose m to balance the the quadrature error with other sources of error. In general, we may also need to approximate Φ^{-1} numerically.

5.2 Results of the CBC algorithm

Here we implement the CBC algorithm in the unanchored space. We explore this setting in the abstract, with parameters that are specified without a connection to a practical problem. We take the weights to be of POD type, given by

$$\gamma_{s,\mathfrak{u}} = \gamma_{\mathfrak{u}} = \left((|\mathfrak{u}|!)^2 \prod_{j\in\mathfrak{u}} \frac{\kappa}{j^\eta}\right)^{1/(1+\lambda)},$$

for some $\kappa > 0$, $\eta > 2$, and $1/2 < \lambda \le 1$. We consider three combinations of probability densities ϕ and weight functions $\psi_j = \psi$:

Combination 1 $\phi(y) = e^{-x^2/2}/\sqrt{2\pi}$ and $\psi(y) = e^{-|x|/\alpha}$.
Combination 2 $\phi(y) = e^{-x^2/2}/\sqrt{2\pi}$ and $\psi(y) = e^{-x^2/(2\alpha)}$.
Combination 3 $\phi(y) = e^{-|x|}/2$ and $\psi(y) = 1$.

In particular, we take a selection of parameters

$$\eta = 3.1,\, 5.0, \quad \kappa = 0.01,\, 0.1, \quad \lambda = 0.51,\, 0.75 \quad \text{and} \quad \alpha = 4,\, 16,$$

and implement the CBC algorithm for $n = 1009, 2003, 4001, 8009, 16001, 32003$, up to $s = 100$ dimensions. The corresponding shift-averaged worst-case errors $e^{\mathrm{sh}}_{s,n}(\boldsymbol{z})$ are presented in Tables 5.1–5.5, together with an estimate on the observed rate of convergence $\mathcal{O}(n^{-r})$, found by performing a linear-least squares fit on the log-log plot of the results.

Table 5.1: $e_{s,n}^{\mathrm{sh}}$ for Combination 1, $\lambda = 0.51$

	$\eta = 3.1$				$\eta = 5$			
	$\alpha = 4$		$\alpha = 16$		$\alpha = 4$		$\alpha = 16$	
n	$\kappa = 0.01$	$\kappa = 0.1$	$\kappa = 0.01$	$\kappa = 0.1$	$\kappa = 0.01$	$\kappa = 0.1$	$\kappa = 0.01$	$\kappa = 0.1$
1009	5.73e-04	2.71e-03	4.59e-04	2.12e-03	3.76e-04	1.06e-03	3.05e-04	8.44e-04
2003	3.14e-04	1.63e-03	2.48e-04	1.26e-03	1.97e-04	5.72e-04	1.58e-04	4.49e-04
4001	1.71e-04	9.63e-04	1.33e-04	7.31e-04	1.03e-04	3.11e-04	8.21e-05	2.41e-04
8009	9.36e-05	5.74e-04	7.21e-05	4.31e-04	5.40e-05	1.68e-04	4.25e-05	1.29e-04
16001	5.12e-05	3.43e-04	3.89e-05	2.52e-04	2.82e-05	9.08e-05	2.19e-05	6.84e-05
32003	2.83e-05	2.04e-04	2.12e-05	1.49e-04	1.47e-05	4.91e-05	1.14e-05	3.66e-05
r	0.869	0.749	0.888	0.766	0.937	0.887	0.950	0.906

Table 5.2: $e_{s,n}^{\mathrm{sh}}$ for Combination 1, $\lambda = 0.75$

	$\eta = 3.1$				$\eta = 5$			
	$\alpha = 4$		$\alpha = 16$		$\alpha = 4$		$\alpha = 16$	
n	$\kappa = 0.01$	$\kappa = 0.1$	$\kappa = 0.01$	$\kappa = 0.1$	$\kappa = 0.01$	$\kappa = 0.1$	$\kappa = 0.01$	$\kappa = 0.1$
1009	1.05e-03	5.03e-03	8.33e-04	3.93e-03	5.25e-04	1.48e-03	4.23e-04	1.17e-03
2003	6.00e-04	3.14e-03	4.71e-04	2.42e-03	2.79e-04	8.20e-04	2.22e-04	6.38e-04
4001	3.38e-04	1.92e-03	2.62e-04	1.46e-03	1.48e-04	4.57e-04	1.17e-04	3.48e-04
8009	1.93e-04	1.18e-03	1.48e-04	8.87e-04	7.86e-05	2.53e-04	6.11e-05	1.92e-04
16001	1.10e-04	7.33e-04	8.31e-05	5.42e-04	4.16e-05	1.40e-04	3.20e-05	1.04e-04
32003	6.33e-05	4.53e-04	4.72e-05	3.31e-04	2.19e-05	7.72e-05	1.67e-05	5.70e-05
r	0.808	0.697	0.827	0.715	0.917	0.853	0.933	0.872

Table 5.3: $e_{s,n}^{\mathrm{sh}}$ for Combination 2, $\lambda = 0.55$

	$\eta = 3.1$		$\eta = 5$	
	$\alpha = 16$		$\alpha = 16$	
n	$\kappa = 0.01$	$\kappa = 0.1$	$\kappa = 0.01$	$\kappa = 0.1$
1009	6.14e-04	2.63e-03	4.24e-04	1.14e-03
2003	3.40e-04	1.57e-03	2.29e-04	6.24e-04
4001	1.87e-04	9.29e-04	1.23e-04	3.44e-04
8009	1.04e-04	5.56e-04	6.63e-05	1.90e-04
16001	5.75e-05	3.30e-04	3.57e-05	1.04e-04
32003	3.21e-05	1.98e-04	1.92e-05	5.75e-05
r	0.853	0.748	0.894	0.862

Table 5.4: $e_{s,n}^{\mathrm{sh}}$ for Combination 2, $\lambda = 0.75$

	$\eta = 3.1$				$\eta = 5$			
	$\alpha = 4$		$\alpha = 16$		$\alpha = 4$		$\alpha = 16$	
n	$\kappa = 0.01$	$\kappa = 0.1$	$\kappa = 0.01$	$\kappa = 0.1$	$\kappa = 0.01$	$\kappa = 0.1$	$\kappa = 0.01$	$\kappa = 0.1$
1009	1.70e-03	6.79e-03	1.07e-03	4.72e-03	1.03e-03	2.56e-03	5.82e-04	1.55e-03
2003	1.01e-03	4.30e-03	6.11e-04	2.93e-03	5.96e-04	1.50e-03	3.16e-04	8.62e-04
4001	5.99e-04	2.70e-03	3.46e-04	1.79e-03	3.46e-04	8.88e-04	1.71e-04	4.85e-04
8009	3.58e-04	1.70e-03	1.99e-04	1.10e-03	2.01e-04	5.26e-04	9.33e-05	2.71e-04
16001	2.14e-04	1.08e-03	1.14e-04	6.79e-04	1.17e-04	3.10e-04	5.06e-05	1.52e-04
32003	1.28e-04	6.81e-04	6.57e-05	4.20e-04	6.79e-05	1.83e-04	2.74e-05	8.52e-05
r	0.747	0.665	0.805	0.699	0.785	0.761	0.882	0.837

Table 5.5: $e^{\text{sh}}_{s,n}$ for Combination 3

	$\lambda = 0.51$				$\lambda = 0.75$			
	$\eta = 3.1$		$\eta = 5$		$\eta = 3.1$		$\eta = 5$	
n	$\kappa = 0.01$	$\kappa = 0.1$	$\kappa = 0.01$	$\kappa = 0.1$	$\kappa = 0.01$	$\kappa = 0.1$	$\kappa = 0.01$	$\kappa = 0.1$
1009	6.91e-04	3.30e-03	4.49e-04	1.29e-03	1.26e-03	6.08e-03	6.31e-04	1.81e-03
2003	3.82e-04	2.01e-03	2.37e-04	7.02e-04	7.29e-04	3.83e-03	3.38e-04	1.01e-03
4001	2.12e-04	1.20e-03	1.25e-04	3.86e-04	4.17e-04	2.37e-03	1.81e-04	5.66e-04
8009	1.16e-04	7.21e-04	6.57e-05	2.11e-04	2.40e-04	1.48e-03	9.66e-05	3.19e-04
16001	6.42e-05	4.34e-04	3.44e-05	1.15e-04	1.39e-04	9.20e-04	5.14e-05	1.78e-04
32003	3.59e-05	2.63e-04	1.82e-05	6.27e-05	8.05e-05	5.79e-04	2.75e-05	9.96e-05
r	0.855	0.733	0.925	0.872	0.793	0.681	0.904	0.837

The numbers presented demonstrate that our theory holds reasonably well, however, there is a strong dependency on certain parameters of the weights. We focus our attention primarily on the rate of convergence.

The parameter κ does not affect the theoretical convergence rate, however we see that κ makes an impact on the observed numerical convergence rates. This is most probably due to the strong effect of κ on the scaling of $e^{\text{sh}}_{s,n}$. Larger κ make $e^{\text{sh}}_{s,n}$ quite a lot larger, and likely make the problem harder, as we are probably not at the asymptotic regime for the worst-case error for the range of n considered.

In some cases for ψ_j and ϕ the parameter α may have an effect on the theoretical convergence rate, as summarised Table 3.2. Combination 3 has no parameter α, and for Combination 1 there is no theoretical dependence of the convergence rate on α. In both cases we have $\mathcal{O}(n^{-1+\delta})$ convergence in the theory, regardless of the choice of α. This is reflected in the numerics, where we see the convergence rates have a weaker dependence on α than on η or κ. Examining Table 3.2 however we see that Combination 2 expects a theoretical convergence rate of $1 - 1/\alpha$, thus we are limited for the smaller case of $\lambda = 0.55$ to $\alpha = 16$. While we do not observe this rate of convergence precisely, there is a noticeable dependence of the observed rate on α.

Finally, we observe that the parameter η does not have an explicit impact on the rates of convergence for a fixed s. However it does have an impact on the condition (3.46), and hence affects implicitly what range of λ are possible if we want our bound of $e^{\text{sh}}_{s,n}$ to be independent of s. In any case, we see that η affects the scaling of $e^{\text{sh}}_{s,n}$.

5.2.1 Scaling the weights

Consider weights of the generic POD form (2.23). Let us normalise the product part of the weights, that is, we write $\tau_j = \gamma_j/\gamma_1$ such that $\tau_1 = 1$, and if we let $\gamma_1 = c$, then we can write $\gamma_j = c\tau_j$. We find that

$$\gamma_{\mathfrak{u}} \;=\; c^{|\mathfrak{u}|}\,\Gamma_{|\mathfrak{u}|}\prod_{j\in\mathfrak{u}}\tau_j\,, \tag{5.7}$$

Evidently, poor choices of c can potentially lead to extreme scaling of the weights for $\mathfrak{u}$ of larger cardinality.

Table 5.6: $\gamma_{\mathfrak{u}}$ for first 10 complete sets $\mathfrak{u}$.

	$\kappa = 0.01$	$\kappa = 1$	$\kappa = 4$
$\gamma_{\{1\}}$	0.01	1	4
$\gamma_{\{1,2\}}$	5.0×10^{-5}	0.5	8
$\gamma_{\{1:3\}}$	1.67×10^{-7}	0.16	10.67
$\gamma_{\{1:4\}}$	4.17×10^{-10}	4.17×10^{-2}	10.67
$\gamma_{\{1:5\}}$	8.33×10^{-13}	8.33×10^{-3}	8.53
$\gamma_{\{1:6\}}$	1.39×10^{-15}	1.39×10^{-4}	5.69
$\gamma_{\{1:7\}}$	1.98×10^{-18}	1.98×10^{-4}	3.25
$\gamma_{\{1:8\}}$	2.48×10^{-21}	2.48×10^{-5}	1.63
$\gamma_{\{1:9\}}$	2.76×10^{-24}	2.76×10^{-6}	0.72
$\gamma_{\{1:10\}}$	2.76×10^{-27}	2.76×10^{-7}	0.29

Initially in the literature it was assumed that the weights are scaled such that $\gamma_{\{1\}} = 1$. In [19] it was argued that this may not be appropriate, in fact that scaling the weights may come to our advantage in minimising the overall error bound. Indeed the weights, in (4.51), chosen for the porous-flow problem, certainly do not obey $\gamma_{\{1\}} = 1$, and in fact we had some problems with adjusting the scale of the weights. We discuss the issue here in an abstract setting.

It is instructive to further refine our example above and inspect some numerical outcomes. Take the weights to be of a similar form to that in the previous section, that is

$$\gamma_{s,\mathfrak{u}} = \gamma_{\mathfrak{u}} = (|\mathfrak{u}|!)^a \prod_{j \in \mathfrak{u}} \frac{\kappa}{j^\eta}\,. \tag{5.8}$$

If say we let $a = 2$ and $\eta = 3$ (likely choices in our application problems), then consider the sequence of weights $\gamma_{\{1:d\}}$, that is the weight for the set of all components up to d, for $d = 1, \ldots, 10$. In Table 5.6 we present the weights for three choices of $\kappa = 0.01, 1, 4$.

Numerically our problem lies with the worst-case error. Take the unbounded unanchored space with ϕ and ψ_j as in Combination 1 of the previous section. We set the weights exactly the same as above, with $\alpha = 1$, and the parameters $a = 2$ and $\eta = 3$. In Table 5.7 we present the resulting worst-case errors obtained from the CBC algorithm. In the first 3 columns are the results for the unanchored and unbounded space for various choices of κ in the weights. We see quite clearly that there is a strong sensitivity of the worst-case error at $s = 100$ to κ. It is not hard to reach numerical overflow if one makes κ a little larger.

In the last three columns of Table 5.7 we present worst-case errors from the CBC algorithm applied in the unanchored Sobolev space, with exactly the same weights as the unbounded space (and the corresponding choices of κ). This dependency on κ is not so pronounced for the unanchored weighted Sobolev spaces of §2.6, and in general we see that the results are much smaller. We conclude that allowing for unbounded integrands really makes things quite a lot more difficult. This is primarily because the one-dimensional shift-averaged kernel is far steeper in the unbounded space than the Sobolev space, as we

Table 5.7: $e^{\text{sh}}_{s,n}(\boldsymbol{z})$ for $s = 1, \ldots, 10$ and 100.

	Unbounded Space			Sobolev Space		
s	$\kappa = 0.01$	$\kappa = 1$	$\kappa = 4$	$\kappa = 0.01$	$\kappa = 1$	$\kappa = 4$
1	9.27×10^{-5}	0.01	3.71×10^{-2}	1.03×10^{-7}	1.03×10^{-5}	4.13×10^{-5}
2	1.15×10^{-4}	0.12	1.78×10^{0}	1.17×10^{-7}	2.38×10^{-5}	2.41×10^{-4}
3	1.23×10^{-4}	0.43	1.89×10^{1}	1.22×10^{-7}	3.24×10^{-5}	5.15×10^{-4}
4	1.26×10^{-4}	1.01	1.33×10^{2}	1.24×10^{-7}	3.75×10^{-5}	7.48×10^{-4}
5	1.27×10^{-4}	1.94	7.41×10^{2}	1.24×10^{-7}	4.04×10^{-5}	8.96×10^{-4}
6	1.28×10^{-4}	3.27	3.45×10^{3}	1.25×10^{-7}	4.23×10^{-5}	9.99×10^{-4}
7	1.29×10^{-4}	5.01	1.39×10^{4}	1.25×10^{-7}	4.37×10^{-5}	1.07×10^{-3}
8	1.29×10^{-4}	7.14	5.01×10^{4}	1.26×10^{-7}	4.45×10^{-5}	1.12×10^{-3}
9	1.30×10^{-4}	9.61	1.63×10^{5}	1.26×10^{-7}	4.52×10^{-5}	1.16×10^{-3}
10	1.30×10^{-4}	12.35	4.86×10^{5}	1.26×10^{-7}	4.57×10^{-5}	1.18×10^{-3}
⋮						
100	1.31×10^{-4}	88.92	4.90×10^{14}	1.26×10^{-7}	4.80×10^{-5}	1.31×10^{-3}

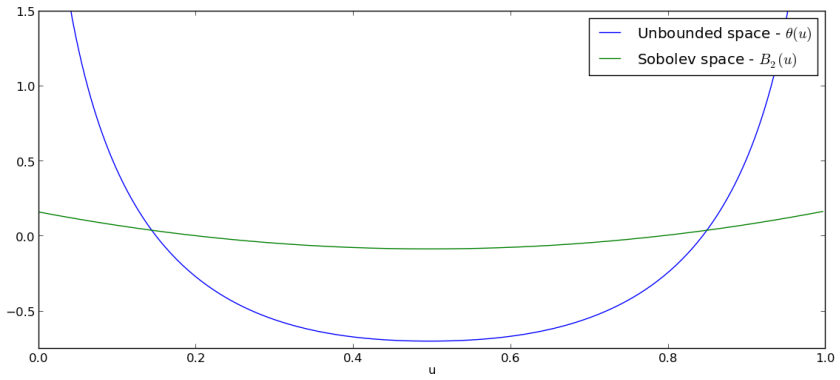

Figure 5.1: One-dimensional shift-averaged kernels in the unanchored unbounded space and unanchored Sobolev space.

can see quite clearly in the plot of θ and B_2 in Figure 5.1. One way of quantifying the difference between the two is comparing the following quantities

$$\frac{1}{n}\sum_{k=1}^{n} \theta\left(\frac{k}{n}\right) = 1.177 \qquad \text{and} \qquad \frac{1}{n}\sum_{k=1}^{n} B_2\left(\frac{k}{n}\right) = 1.312 \times 10^{-3}\,.$$

The quantities above represent the worst-case error in one dimension, see (2.34) and (3.32). Note that we have dropped the j subscript on θ as there is no coordinate dependency in this model example. Although this does not rigorously explain why the higher-dimensional worst-case errors are quite different between the spaces, inspecting the equations (2.34) and (3.32) shows that these quantities make a good guideline for what we may expect of the worst-case errors in their respective spaces. As the first sum above is quite a bit larger, it is no surprise that we get the much larger worst-case errors for the unbounded space.

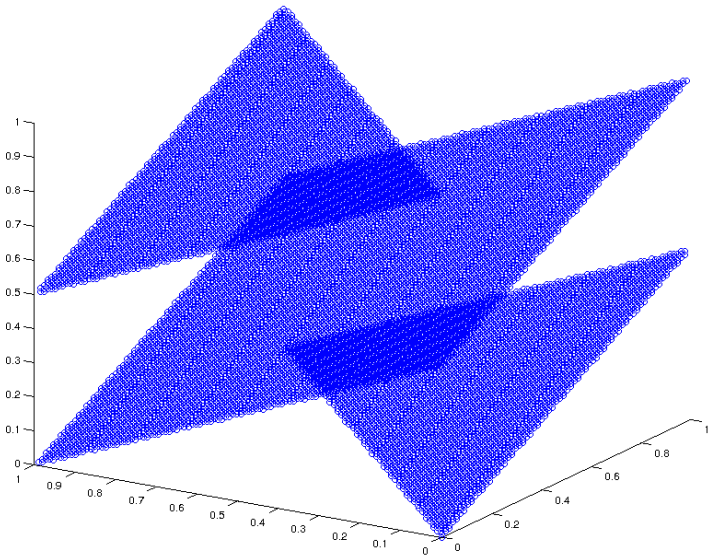

Figure 5.2: A lattice rule with a bad projection in coordinates $\{2, 3, 4\}$, caused by the weight $\gamma_{\{2,3,4\}}$ being too small.

This issue is also exacerbated when the product part of the weights γ_j does not immediately follow the polynomial decay: as seen in applications, γ_j may have a "plateau" for the first few coordinates. In this case the weights can be extremely large. For example, if our weights are given by $\gamma_{\mathfrak{u}} = |\mathfrak{u}|! \prod_{j \in \mathfrak{u}} \gamma_j$, and we have say $\gamma_j \sim 1$ for $j \leq 10$, before the γ_j exhibit some polynomial decay, then $\gamma_{\{1:10\}} \sim 10!$, and hence it may be quite easy for to stumble on numerical overflow in the CBC algorithm, giving us nonsensical results and unusable generating vectors.

Problems also occur if κ is too small, though this is not just to do with numerical underflow. Rather, we observed in practice is that $\gamma_{\mathfrak{u}}$ may be too small for $\mathfrak{u}$ of larger cardinality. In these cases the CBC will choose lattices that do not properly take into account these projections, that is, the lattice will make a very poor covering of the subspace spanned by the coordinates $\mathfrak{u}$. This is best shown visually, and in Figure 5.2 we show an example of a lattice rule with $n = 16001$ points, with a bad projection in coordinates 2, 3 and 4, caused by $\gamma_{\{2,3,4\}} \sim 10^{-14}$ being too small. This lattice rule was constructed for the porous-flow problem. We see that the lattice points makes distinct planes, and do not "fill" the space well. Compare this to Figure 2.1 where the cube is covered more consistently, even with only $n = 127$ points.

5.3 Numerical results of the porous flow problem

We present here a numerical study of the algorithm described above over a range of parameters. Theorem 47 provides us with a theoretical bound for the error in the method, and we examine here whether we see, in the numerics, the behaviour predicted by the theory.

We solve (4.6) with spatial dimension $d = 1$ on $D = [0, 1]$, with a forcing term $f(x) = 1$. We take the truncated expansion of the field a^s of (4.5) with $s = 400$, so that $\boldsymbol{y} \in \mathbb{R}^{400}$. The strong form of the problem we are solving is the parametrised ODE

$$-\frac{\mathrm{d}}{\mathrm{d}x}\left(a^s(x, \boldsymbol{y})\frac{\mathrm{d}u^s(x, \boldsymbol{y})}{\mathrm{d}x}\right) = 1, \tag{5.9}$$

with homogeneous Dirichlet boundary conditions, $u(0, \boldsymbol{y}) = u(1, \boldsymbol{y}) = 0$. We solve (5.9) using the piecewise-linear finite element method with uniform meshes of diameter $h = 1/M$ to get the approximate solution $u_h^s(\cdot, \boldsymbol{y})$. The tridiagonal systems which arise are solved in $\mathcal{O}(M)$ time by the Thomas algorithm. In the numerical experiments that follow, we set $M = 1024$ and compute the entries of the tridiagonal system using the composite mid-point rule applied element-wise.

The quantity of interest is here taken to be $\mathbb{E}[\mathcal{G}(u_h^s)]$, where the functional $\mathcal{G}$ is taken to be point evaluation at 1/3, i.e.

$$F(\boldsymbol{y}) = \mathcal{G}(u_h^s(\cdot, \boldsymbol{y})) = u_h^s(1/3, \boldsymbol{y}).$$

To specify a^s we take $a_* \equiv 0$ and $a_0 \equiv 1$ in (4.2). To specify Z we need to choose a set of basis functions ξ_j and parameters μ_j as in equation (4.3). Although this choice can be arbitrary, in many applications one would take μ_j and ξ_j as the Karhunen–Loève decomposition of a given covariance kernel, as discussed briefly in Chapter 4. Here our choice of $\{(\mu_j, \xi_j)\}_{j\geq 1}$ come from the Karhunen–Loève expansion of the Matérn class of covariance functions, as this offers a real-world scenario of interest to practitioners in the field. The family of Matérn covariance functions are given by

$$\rho(r) \;=\; \rho_\nu(r) \;:=\; \sigma^2\,\frac{2^{1-\nu}}{\Gamma(\nu)}\,(r/\widetilde{\lambda})^\nu\,K_\nu(r/\widetilde{\lambda})\,, \tag{5.10}$$

with $\widetilde{\lambda} = \lambda_{\mathcal{C}}/(2\sqrt{\nu})$. Here Γ is the gamma function and K_ν is the modified Bessel function of the second kind. The parameter $\nu > 1/2$ is a smoothness parameter, σ^2 is the variance and $\lambda_{\mathcal{C}}$ is a length scale parameter. We do not expand on the theory of the Karhunen–Loève expansion and the Matérn class of covariance here, as it is somewhat beyond the scope of this thesis. However, further details of this theory, as well as the description how we find (μ_j, ξ_j) numerically, can be found in our paper [24].

To define the weighted space $\mathcal{W}_s$ in (4.34) and to perform the CBC algorithm for calculating the generating vector $\boldsymbol{z}$, we must choose the weight parameters $\gamma_{\mathfrak{u}}$ and weight functions ψ_j. First we specify the various parameters of the weights, then we go on to specify the parameters for two separate choices of ψ_j.

We define the weight parameters $\gamma_{\mathfrak{u}}$ as in (4.51), after first setting the parameter λ_*. By (4.52), this parameter λ_* is related to p, which in turn was introduced in Assumption A1, and depends on the parameters (μ_j, ξ_j), in particular the rate of convergence of $b_j = \sqrt{\mu_j}\|\xi_j\|_{\boldsymbol{\mathcal{C}}^0(\overline{D})}$. We see that if $b_j = \mathcal{O}(j^{-\vartheta})$, for some $\vartheta > 1$, then we will require $p > 1/\vartheta$ to satisfy Assumption A1. Recall that in these experiments the (μ_j, ξ_j) come

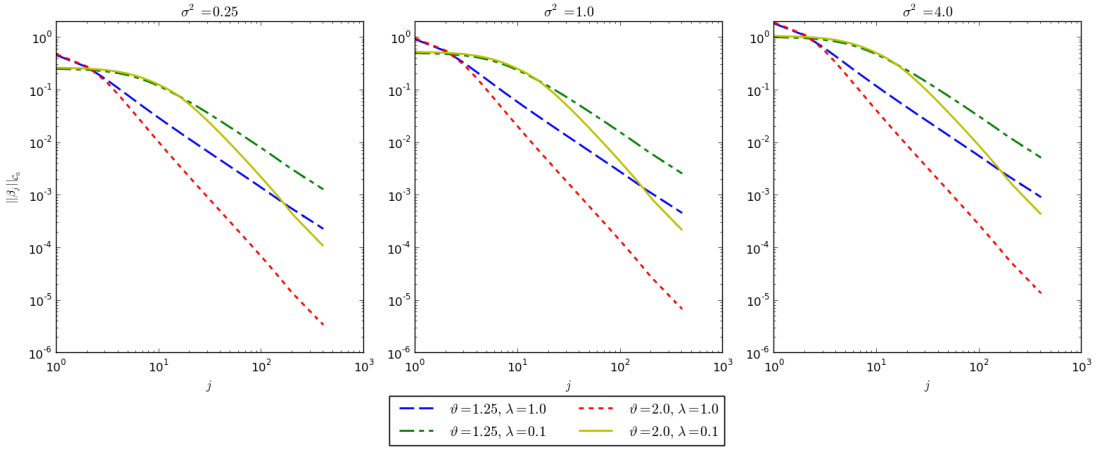

Figure 5.3: Log-log plot of b_j for the various choices of ϑ, σ^2 and $\lambda_{\mathcal{C}}$.

from the Karhunen–Loève expansion of the Matérn covariance. It so happens that there is a direct link between the choice of smoothness parameter ν and ϑ. This link is explored further in [24, Corollary 5], however here we state our choice of parameters in terms of ϑ, for ease of the exposition.

We find from (4.52) that for any small $q > 0$ we have the following relationship between ϑ and λ_*,

$$\lambda_* = \begin{cases} \frac{1}{2\vartheta - 1} + q, & \text{if } 1 < \vartheta < 3/2 \\ \frac{1}{2} + q, & \text{if } \vartheta \geq 3/2. \end{cases} \tag{5.11}$$

This choice of λ_* implies, see Theorem 40, that we obtain theoretical QMC convergence close to $\mathcal{O}(n^{-\min(\vartheta - 1/2, 1)})$. Note that the choice of q involves a trade-off. Smaller values of q lead to a faster convergence, but also to a larger value for $C_{\boldsymbol{\gamma}}(\lambda_*)$ in (4.47). In fact, for $\vartheta \geq 3/2$, we see that $q \to 0$ is equivalent to $\delta \to 0$ in (4.52). This in turn implies $C_{\boldsymbol{\gamma}}(\lambda_*) \to \infty$, by way of (4.47) and (4.60) or (4.63). For $\vartheta \in (1, 3/2)$, we see that $q \to 0$ is equivalent to $p \to 1/\vartheta$, so that the sum in Assumption A1 grows without bound, leading, as in the proof of Theorem 43, again to $C_{\boldsymbol{\gamma}}(\lambda_*) \to \infty$. Here we choose $q = 0.05$.

Recalling (3.5), we write $Q_i = Q_{s,n}(\boldsymbol{\Delta}_i; F)$, where $\boldsymbol{\Delta}_i$ is the i-th independent random shift, uniformly distributed on $[0,1]^s$. Denoting by $\bar{Q}$ the mean of the Q_i, we have the following unbiased estimator with R random shifts of the mean-square error (with respect to the shifts):

$$\frac{1}{R}\frac{1}{R-1}\sum_{i=1}^{R}(Q_i - \bar{Q})^2 \;\approx\; \mathbb{E}^{\boldsymbol{\Delta}}|I_s(F) - Q_{s,n}(\cdot; F)|^2 \;. \tag{5.12}$$

The square-root of the left-hand side of (5.12) is an estimate of the "standard error".

In the following experiments we estimate this standard error for the problem where the set $\{(\mu_j, \xi_j)\}_{j \geq 1}$ is obtained from the Karhunen–Loève expansion of the Matérn covariance, as described earlier. We obtain the Karhunen–Loève expansion for the following

selection of parameters,

$$\vartheta = 2,\ 1.25 \text{ (equivalent to } \nu = 1.5,\ 0.75) \quad \sigma^2 = 0.25,\ 1.0,\ 4.0 \quad \lambda_{\mathcal{C}} = 1.0,\ 0.1,$$

where σ^2 and $\lambda_{\mathcal{C}}$ refer to the variance and length-scale parameters for the Matérn covariance in (5.10). These parameters affect the behaviour of the expansion (μ_j, ξ_j), and consequently the behaviour of b_j which much of our theory depends on. The parameter ϑ determines the asymptotic regime of b_j, and in fact, we have that $b_j = \mathcal{O}(j^{-\vartheta})$, meaning that we for $\vartheta = 2$ and 1.25, we take $\lambda_* = 0.5 + q$ and $0.8 + q$ respectively.

The effects of the other two parameters σ^2 and $\lambda_{\mathcal{C}}$ in the Matérn kernel are not immediately obvious. The parameter $\lambda_{\mathcal{C}}$, which we may refer to as the correlation length, determines the pre-asymptotic regime of b_j. For $\lambda_{\mathcal{C}} = 0.1$ we have a "plateau" effect, that is, the first few values remain constant before the asymptotic regime sets in, while for $\lambda_{\mathcal{C}} = 1.0$, the asymptotic regime is reached quickly, for small j. Indeed if we considered even smaller $\lambda_{\mathcal{C}}$, we would see even more of a plateau effect on b_j. Finally, σ^2, which is often called the variance, only has an effect on the scale of b_j, in direct proportion to σ. These behaviours can be observed in Figure 5.3 where we plot b_j for all combinations of choices of ϑ, σ^2, and $\lambda_{\mathcal{C}}$.

Finally, remember that we have fixed the truncation dimension at $s = 400$ and the spatial resolution at $h = 1/1024$. We use $R = 32$ random shifts.

5.3.1 Exponential ψ_j

Here we examine the details of the numerical implementation with the exponential weight functions (4.58). First we need to specify the parameters α_j in (4.58), then we go on to examine numerical results for the worst-case errors from the CBC algorithm, followed by standard errors of the approximation of the ODE problem.

In principle, our weighted function space framework in Section 4.3.2 allows us to adjust the QMC rule to the integrand behaviour with respect to every coordinate via the j-dependent parameters α_j However, as discussed in §5.1.3, allowing a different value of α_j for each j would cause a substantial increase in the cost of the CBC algorithm. To maintain the full efficiency of the CBC construction, the α_j should be coordinate-independent, at least for large blocks of coordinates. In our numerical experiments we found that the use of a single value of α_j for all j led to unsatisfactory results in the exponential case, but that two values (chosen according to the prescription below) led to acceptable results. The reasoning for this is presented further in §5.3.2

Guided by (4.61) and the general condition (4.59), noting that $b_j \to 0$ as $j \to \infty$ as a consequence of Assumption A1, and writing $b_* := \max_{j \geq 1} b_j$, we choose j_0 to be the smallest positive integer such that $b_j < b_*/2$ for all $j \geq j_0$, and define

$$\alpha_j = \begin{cases} \frac{1}{2}\left(b_* + \sqrt{b_*^2 + 1 - 1/(2\lambda_*)}\right) & \text{for } j < j_0 \\ \frac{1}{2}\left(b_{j_0} + \sqrt{b_{j_0}^2 + 1 - 1/(2\lambda_*)}\right) & \text{for } j \geq j_0 \end{cases}, \tag{5.13}$$

with λ_* to be specified below.

CBC Results

Tables 5.8 and 5.9 present the results of the CBC algorithm, that is, the shift-averaged worst-case error $e^{\mathrm{sh}}_{s,n}(\boldsymbol{z}^*)$, at $s = 400$. These results are presented for the various parameters choices, along with estimated values of the rate of convergence r in the error representation cn^{-r}, estimated by linear regression of the negative log of the standard error against $\log n$.

These results are interesting, as we witness large fluctuations in the scale of the worst-case error for the various choices of parameters. This is most probably because both $e^{\mathrm{sh}}_{s,n}$ and $\|F\|_{\mathcal{W}_s}$ have a strong sensitivity to ϑ, σ^2 and $\lambda_{\mathcal{C}}$.

In particular, we see that for the smaller choice of $\vartheta = 1.25$, we get a huge change in $e^{\mathrm{sh}}_{s,n}$ with $\lambda_{\mathcal{C}}$. This is due largely to the plateau effect of the weights. From Figure 5.3 we see that for $\lambda_{\mathcal{C}} = 0.1$, the b_j remain constant for the first few j, meaning we get exactly the problem with plateau weights discussed in §5.2.1.

There is an additional aspect to consider. Notionally, in setting γ_{u} as in Theorem 43 and α_j as prescribed in Corollary 44, we are minimising S_λ by balancing the *bound* of $e^{\mathrm{sh}}_{s,n}$ with the *bound* of $\|F\|_{\mathcal{W}_s}$, thus minimising our error bound. However it may well be that our bounds are not sharp, and hence there is no guarantee that our choice of weights and α_j, chosen to minimise the error *bound*, actually minimise the product of the *true* values of $e^{\mathrm{sh}}_{s,n}$ and $\|F\|_{\mathcal{W}_s}$. As we actually calculate the true value of $e^{\mathrm{sh}}_{s,n}$ in the CBC algorithm, we may indeed be observing fluctuations from weights that may be over-compensating one way or another.

QMC quadrature results

Tables 5.10 and 5.11 present results using the QMC quadrature analysed in Chapter 4, again with estimated values of the rate of convergence r together with its 90% confidence interval. We include the confidence interval in these results as the data points do not converge as smoothly as the worst-case error results from earlier tables. Here we see a strong dependence on the variance σ^2, but a weaker dependence on the choices of $\lambda_{\mathcal{C}}$ and ϑ. While Theorem 43 suggests that the asymptotic behaviour of the root-mean-square error depends on p (and hence ϑ), in practice the observed rates of convergence bear little relation with the prediction. One explanation may be that with the range of n presented we are in a pre-asymptotic regime. This seems especially true for larger values of σ^2, and hence may explain why we see our QMC quadrature performing similarly to standard MC quadrature for $\sigma^2 = 4.0$.

Recall from the theory that we expect a convergence rate close to $\mathcal{O}(n^{-\min(\vartheta-1/2,1)})$. We see, however, that the results converge almost as well for $\vartheta = 1.25$ as for $\vartheta = 2$. This seems to indicate that our theory is not sharp, and that optimal (close to $\mathcal{O}(n^{-1})$) convergence could potentially be demonstrated for ϑ lower than our current cross-over point of $\vartheta = 1.5$ in (5.11). This is also indicated by the fact that our method sometimes converges faster than predicted by the theory, for example for $\vartheta = 1.25$, $\sigma^2 = 0.25$

Table 5.8: Worst-case errors $e^{\mathrm{sh}}_{s,n}$, for $\vartheta = 2$ using exponential ψ_j and POD weights γ_{u} as in (4.51)

	$\sigma^2 = 0.25$		$\sigma^2 = 1.0$		$\sigma^2 = 4.0$	
n	$\lambda_{\mathcal{C}} = 1.0$	$\lambda_{\mathcal{C}} = 0.1$	$\lambda_{\mathcal{C}} = 1.0$	$\lambda_{\mathcal{C}} = 0.1$	$\lambda_{\mathcal{C}} = 1.0$	$\lambda_{\mathcal{C}} = 0.1$
8,009	2.46e-04	2.49e-03	1.98e-04	3.86e-03	1.71e-04	8.92e-04
16,001	1.40e-04	1.63e-03	1.14e-04	2.59e-03	9.58e-05	5.62e-04
32,003	8.00e-05	1.07e-03	6.55e-05	1.74e-03	5.37e-05	3.54e-04
64,007	4.59e-05	7.07e-04	3.80e-05	1.16e-03	3.01e-05	2.23e-04
120,011	2.75e-05	4.84e-04	2.31e-05	8.11e-04	1.78e-05	1.47e-04
240,007	1.59e-05	3.20e-04	1.34e-05	5.43e-04	9.96e-06	9.34e-05
480,013	9.16e-06	2.11e-04	7.81e-06	3.65e-04	5.60e-06	5.97e-05
Rate	0.81	0.61	0.80	0.58	0.83	0.67

Table 5.9: Worst-case errors $e^{\mathrm{sh}}_{s,n}$, for $\vartheta = 1.25$ using exponential ψ_j and POD weights γ_{u} as in (4.51)

	$\sigma^2 = 0.25$		$\sigma^2 = 1.0$		$\sigma^2 = 4.0$	
n	$\lambda_{\mathcal{C}} = 1.0$	$\lambda_{\mathcal{C}} = 0.1$	$\lambda_{\mathcal{C}} = 1.0$	$\lambda_{\mathcal{C}} = 0.1$	$\lambda_{\mathcal{C}} = 1.0$	$\lambda_{\mathcal{C}} = 0.1$
8,009	6.39e-03	2.80e+02	2.75e-02	5.07e+16	5.00e-02	6.32e+34
16,001	4.27e-03	1.98e+02	1.91e-02	3.59e+16	3.52e-02	4.46e+34
32,003	2.85e-03	1.40e+02	1.33e-02	2.54e+16	2.47e-02	3.15e+34
64,007	1.90e-03	9.89e+01	9.22e-03	1.79e+16	1.74e-02	2.23e+34
120,011	1.31e-03	7.22e+01	6.62e-03	1.31e+16	1.26e-02	1.63e+34
240,007	8.80e-04	5.11e+01	4.59e-03	9.26e+15	8.89e-03	1.15e+34
480,013	5.87e-04	3.61e+01	3.19e-03	6.55e+15	6.25e-03	8.14e+33
Rate	0.60	0.50	0.54	0.50	0.53	0.50

and $\lambda_{\mathcal{C}} = 1.0$, where the observed rate of convergence of approximately 0.89 is significantly larger than the predicted rate of 0.75 from Theorem 43, and also higher than the corresponding convergence of worst-case error, as seen in Table 5.9.

Tables 5.12 and 5.13 present the same experiments as Tables 5.10 and 5.11 respectively, but for MC quadrature. The results agree with the usual behavior of MC methods where standard errors converge with approximately $\mathcal{O}(n^{-1/2})$. Figure 5.4 charts all the findings in Tables 5.10 to 5.13. They demonstrate that in all our test cases QMC always does better than MC, especially for small σ^2, where QMC outperforms MC by up to two orders of magnitude.

As a final comparison, in Tables 5.14 and 5.15 we look at the standard errors for a generic lattice rule, which is not specifically designed to fit the problem. We choose here a lattice rule generated for the Sobolev space of mixed first-order derivatives on $[0,1]^s$, with product weight parameters $\gamma_j = 1/j^2$. We see that these lattice rules still behave very well, attaining similar results to the lattice rules constructed for our specific problem.

5.3.2 Issues with setting α_j

Some difficulty was encountered in finding a good scheme for setting α_j. We required a scheme that would produce satisfactory results for all our choices of the various parameters σ, $\lambda_{\mathcal{C}}$ and ϑ. This was not easy, and bad choices of α_j lead to the sort of numerical overflow problems highlighted in §5.2.1 due to the sensitivity of Ψ_j and ϱ_j to α_j. The

Table 5.10: QMC standard errors for $\vartheta = 2$, using exponential ψ_j and POD weights $\gamma_{\mathfrak{u}}$ as in (4.51)

	$\sigma^2 = 0.25$		$\sigma^2 = 1.0$		$\sigma^2 = 4.0$	
n	$\lambda_{\mathcal{C}} = 1.0$	$\lambda_{\mathcal{C}} = 0.1$	$\lambda_{\mathcal{C}} = 1.0$	$\lambda_{\mathcal{C}} = 0.1$	$\lambda_{\mathcal{C}} = 1.0$	$\lambda_{\mathcal{C}} = 0.1$
8,009	2.69e-05	1.77e-05	1.90e-04	1.00e-04	1.12e-02	3.22e-03
16,001	1.38e-05	8.12e-06	1.02e-04	7.52e-05	5.47e-03	2.44e-03
32,003	8.85e-06	6.22e-06	6.79e-05	5.11e-05	3.83e-03	1.20e-03
64,007	4.49e-06	3.02e-06	3.33e-05	3.49e-05	2.36e-03	7.02e-04
120,011	2.66e-06	1.79e-06	2.46e-05	1.79e-05	3.18e-03	7.87e-04
240,007	1.43e-06	9.95e-07	1.48e-05	9.80e-06	1.74e-03	4.42e-04
480,013	7.82e-07	6.72e-07	9.17e-06	8.41e-06	7.68e-04	2.91e-04
Rate	0.86	0.80	0.73	0.66	0.55	0.58
90% Interval	[0.89, 0.83]	[0.87, 0.74]	[0.78, 0.69]	[0.75, 0.57]	[0.71, 0.40]	[0.68, 0.48]

Table 5.11: QMC standard errors for $\vartheta = 1.25$, using exponential ψ_j and POD weights $\gamma_{\mathfrak{u}}$ as in (4.51)

	$\sigma^2 = 0.25$		$\sigma^2 = 1.0$		$\sigma^2 = 4.0$	
n	$\lambda_{\mathcal{C}} = 1.0$	$\lambda_{\mathcal{C}} = 0.1$	$\lambda_{\mathcal{C}} = 1.0$	$\lambda_{\mathcal{C}} = 0.1$	$\lambda_{\mathcal{C}} = 1.0$	$\lambda_{\mathcal{C}} = 0.1$
8,009	2.80e-05	1.80e-05	1.76e-04	1.12e-04	8.97e-03	2.01e-03
16,001	1.37e-05	7.37e-06	1.25e-04	7.10e-05	7.25e-03	1.69e-03
32,003	8.37e-06	5.78e-06	5.72e-05	3.98e-05	2.42e-03	1.26e-03
64,007	4.36e-06	2.93e-06	3.39e-05	2.70e-05	1.72e-03	8.35e-04
120,011	2.58e-06	1.82e-06	2.00e-05	1.82e-05	1.43e-03	5.63e-04
240,007	1.32e-06	9.56e-07	1.14e-05	1.31e-05	1.57e-03	2.64e-04
480,013	7.06e-07	5.57e-07	6.31e-06	7.52e-06	5.60e-04	2.05e-04
Rate	0.89	0.82	0.83	0.64	0.63	0.60
90% Interval	[0.91, 0.86]	[0.89, 0.76]	[0.88, 0.79]	[0.68, 0.61]	[0.81, 0.44]	[0.70, 0.50]

scheme proposed in Corollary 44 was not obvious to us, and did not come about until after many other attempts at finding a good scheme for α_j.

We also discovered the need for multiple distinct α_j through experimentation. Previous papers such as [71] and [39] considered the unbounded setting but with the weight functions ψ_j taken to be the same over all coordinates. In this PDE problem, however, we found it necessary to offer the ability to at least provide a regime of two distinct ψ_j (or rather its parameter α_j), one fit to the earlier coordinates, and one fit to the tail coordinates. We found this necessary as we have the condition (4.59), where we require $\alpha_j > b_j$, to ensure $\Psi_j < \infty$. The b_j can converge quite quickly, if we have one only choice that is used in all coordinate directions, α_* say, then it will be too big for later coordinates, which happens to lead to weights that are too small, especially for higher-order coordinate collections. This gives us precisely the problem discussed in §5.2.1.

To demonstrate this we can examine the weights in a specific case, both with and without the regime of multiple α_j. Consider the case $\vartheta = 1.25$, $\sigma^2 = 4.0$ and $\lambda_{\mathcal{C}} = 1.0$. We see from Figure 5.3 that the b_j in this case are initially large but converge quickly. In Table 5.16 we present the weights $\gamma_{\mathfrak{u}}$ for two cases. In the first column the weights have one choice of α_j, chosen as per (4.61), but only set for the first coordinate, then left constant. The second column of weights takes in two distinct choices of α_j, as per (5.13). In this example we find the criterion of $b_j < b_*/2$ occurs at $j = 3$. Thus we have $\alpha_1 = \alpha_2$, but then a new choice is made for α_3 and beyond.

Table 5.12: MC standard errors for $\vartheta = 2$

	$\sigma^2 = 0.25$		$\sigma^2 = 1.0$		$\sigma^2 = 4.0$	
n	$\lambda_{\mathcal{C}} = 1.0$	$\lambda_{\mathcal{C}} = 0.1$	$\lambda_{\mathcal{C}} = 1.0$	$\lambda_{\mathcal{C}} = 0.1$	$\lambda_{\mathcal{C}} = 1.0$	$\lambda_{\mathcal{C}} = 0.1$
8,009	7.24e-04	4.19e-04	2.21e-03	1.11e-03	2.70e-02	6.62e-03
16,001	3.98e-04	2.58e-04	1.15e-03	7.22e-04	1.42e-02	4.98e-03
32,003	2.97e-04	1.52e-04	9.73e-04	4.45e-04	1.65e-02	3.56e-03
64,007	1.87e-04	1.07e-04	6.21e-04	3.08e-04	1.02e-02	2.43e-03
120,011	1.25e-04	7.59e-05	4.11e-04	2.17e-04	5.78e-03	1.65e-03
240,007	9.40e-05	6.19e-05	2.97e-04	1.50e-04	4.02e-03	8.78e-04
480,013	7.06e-05	4.16e-05	2.12e-04	9.75e-05	2.79e-03	5.06e-04
Rate	0.56	0.55	0.56	0.59	0.55	0.63
90% Interval	[0.62, 0.51]	[0.61, 0.49]	[0.61, 0.50]	[0.61, 0.57]	[0.65, 0.44]	[0.71, 0.55]

Table 5.13: MC standard errors for $\vartheta = 1.25$

	$\sigma^2 = 0.25$		$\sigma^2 = 1.0$		$\sigma^2 = 4.0$	
n	$\lambda_{\mathcal{C}} = 1.0$	$\lambda_{\mathcal{C}} = 0.1$	$\lambda_{\mathcal{C}} = 1.0$	$\lambda_{\mathcal{C}} = 0.1$	$\lambda_{\mathcal{C}} = 1.0$	$\lambda_{\mathcal{C}} = 0.1$
8,009	6.89e-04	4.01e-04	2.05e-03	1.07e-03	2.30e-02	6.54e-03
16,001	3.82e-04	2.47e-04	1.08e-03	6.90e-04	1.21e-02	4.85e-03
32,003	2.81e-04	1.45e-04	9.05e-04	4.23e-04	1.35e-02	3.43e-03
64,007	1.78e-04	1.02e-04	5.76e-04	2.93e-04	8.51e-03	2.36e-03
120,011	1.20e-04	7.20e-05	3.85e-04	2.07e-04	5.00e-03	1.62e-03
240,007	9.91e-05	5.95e-05	3.11e-04	1.52e-04	3.84e-03	1.24e-03
480,013	6.92e-05	4.08e-05	2.11e-04	9.60e-05	3.85e-03	8.15e-04
Rate	0.55	0.55	0.54	0.58	0.45	0.51
90% Interval	[0.61, 0.49]	[0.61, 0.48]	[0.59, 0.48]	[0.61, 0.55]	[0.56, 0.34]	[0.53, 0.49]

In Table 5.16, we clearly see that in the single-α_j case, the weights become too small, whereas for the double-α_j case, the higher order weights are many orders of magnitude larger, and are of reasonably proportion to other weights. The small weights in the single-α_j case, especially for example the weight $\gamma_{\{2,3,4\}} = 1.83 \times 10^{-14}$, lead to exactly the lower-order projection problem highlighted in Figure 5.2. The larger weights in the double-α_j case are due to the change in α_j from $j = 3$ onwards, in fact we see $\gamma_{\{3\}}$ jumping to a larger value. While it may seem strange that $\gamma_{\{1\}} < \gamma_{\{3\}}$, particularly as we'd expect the 3rd coordinate to contribute less than the 1st to the overall variance of the problem, this is theoretically balanced by the fact that we have changed the dependence on the 3rd coordinate in the norm $\|F\|_{\mathcal{W}_s}$.

Having two distinct α_j is a compromise. Ideally we would follow the scheme in Corollary 44 and have a unique α_j for every coordinate, minimising S_λ. However as discussed the computational cost in using that scheme is excessive. We did not, however, see the need for any more than 2 distinct α_j, primarily as the results did not improve remarkably with more choices, both for the worst-case error results and for the standard error of the QMC quadrature results.

Part of our problems may indeed lie with the quantity $C_{2,j}$ which is used in the bound of $\hat{\theta}_j$, (3.38). We can see from (4.60) that $C_{2,j}$ contains a term of the order of $e^{\alpha_j^2}$, which is evidently quite sensitive to α_j, and likely to fluctuate with α_j more than $\hat{\theta}_j$ itself. We

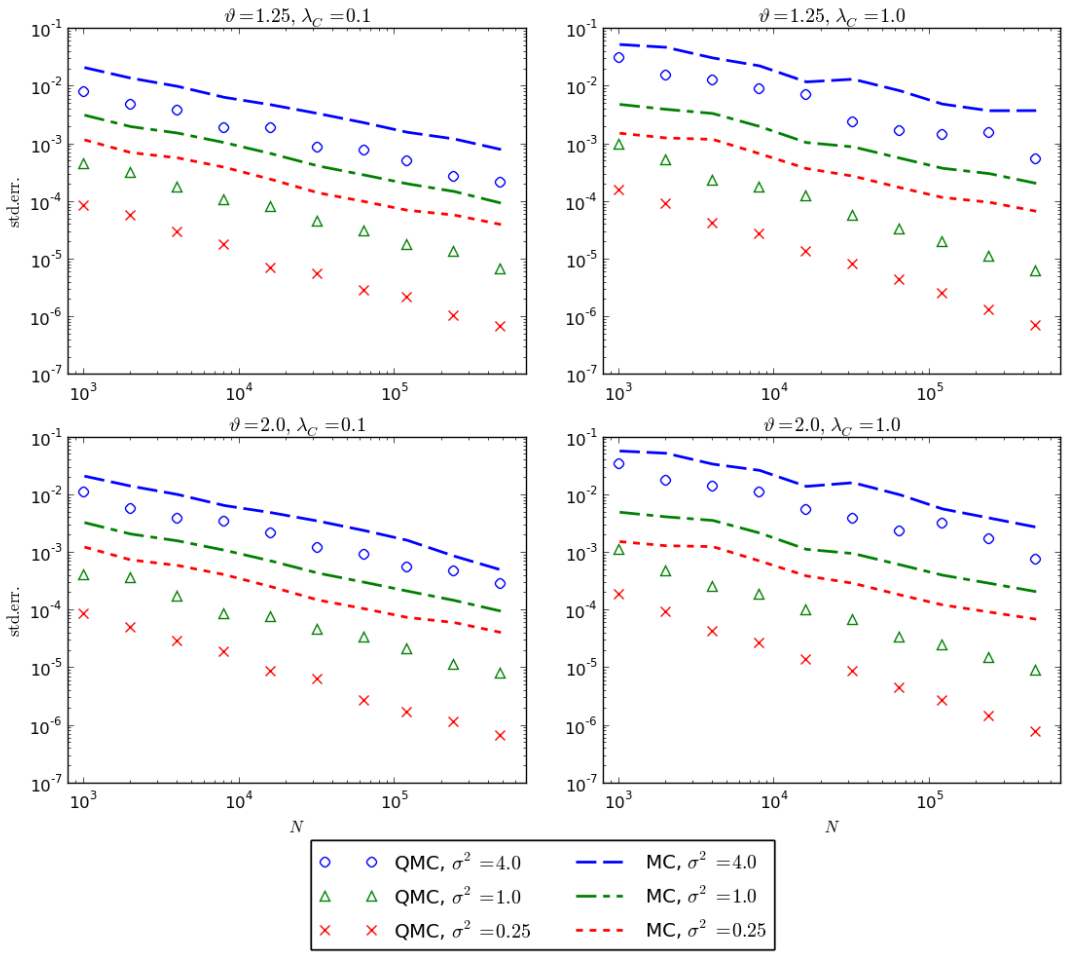

Figure 5.4: Standard errors from Tables 1 to 4 for QMC and MC plotted against n.

conclude that perhaps sharper bounds of the form of (3.38), if we could find them, would make the numerical situation much easier to manage.

5.3.3 Gaussian ψ_j

Here we detail the numerical results with the Gaussian weight function (4.62). Once again we must specify our scheme for setting α_j. Unfortunately however we do not have a closed form specification for α_j as for the exponential case. Instead we minimised the expression $[\varrho_j(\lambda)]^{1/\lambda}\Psi_j$ numerically for $0 < \alpha < 1 - 1/(2\lambda)$, by a simple minimum search on a fine mesh. Unlike in the exponential case, we did not find that multiple α_j made too much of a difference on the results, probably as we did not have the hard constraint (4.59). Hence we present results here where a single α_j was used.

CBC results

We present the results of the CBC algorithm in Table 5.17, again with estimated values of the rate of convergence r. Unfortunately, in this case of Gaussian ψ_j, the method only worked for the combinations of $\sigma^2 = 0.25$ with $\lambda_{\mathcal{C}} = 1.0$ and 0.1, and the combination of $\sigma^2 = 1.0$ with $\lambda_{\mathcal{C}} = 1.0$. For all other cases, e.g. when $\sigma^2 = 4.0$, nonsense generating

Table 5.14: QMC standard errors for $\vartheta = 2$ using generic lattice rules

	$\sigma^2 = 0.25$		$\sigma^2 = 1.0$		$\sigma^2 = 4.0$	
n	$\lambda_{\mathcal{C}} = 1.0$	$\lambda_{\mathcal{C}} = 0.1$	$\lambda_{\mathcal{C}} = 1.0$	$\lambda_{\mathcal{C}} = 0.1$	$\lambda_{\mathcal{C}} = 1.0$	$\lambda_{\mathcal{C}} = 0.1$
8,009	3.38e-05	1.90e-05	2.48e-04	1.24e-04	1.29e-02	2.81e-03
16,001	1.68e-05	9.91e-06	1.23e-04	7.41e-05	7.00e-03	1.78e-03
32,003	8.27e-06	6.48e-06	6.47e-05	5.56e-05	4.05e-03	1.19e-03
64,007	4.32e-06	4.60e-06	3.55e-05	4.05e-05	2.88e-03	9.36e-04
120,011	2.46e-06	2.01e-06	2.59e-05	2.08e-05	4.33e-03	6.41e-04
240,007	1.77e-06	1.41e-06	2.07e-05	1.31e-05	3.54e-03	3.76e-04
480,013	6.84e-07	6.32e-07	7.10e-06	7.26e-06	7.60e-04	2.27e-04
Rate	0.92	0.80	0.80	0.68	0.52	0.59
90% Interval	$[0.99, 0.84]$	$[0.88, 0.72]$	$[0.91, 0.68]$	$[0.76, 0.61]$	$[0.78, 0.25]$	$[0.65, 0.54]$

Table 5.15: QMC standard errors for $\vartheta = 1.25$ using generic lattice rules

	$\sigma^2 = 0.25$		$\sigma^2 = 1.0$		$\sigma^2 = 4.0$	
n	$\lambda_{\mathcal{C}} = 1.0$	$\lambda_{\mathcal{C}} = 0.1$	$\lambda_{\mathcal{C}} = 1.0$	$\lambda_{\mathcal{C}} = 0.1$	$\lambda_{\mathcal{C}} = 1.0$	$\lambda_{\mathcal{C}} = 0.1$
8,009	3.25e-05	1.80e-05	2.27e-04	1.21e-04	1.06e-02	2.82e-03
16,001	1.61e-05	9.42e-06	1.12e-04	6.77e-05	5.52e-03	1.66e-03
32,003	7.87e-06	5.92e-06	5.71e-05	5.13e-05	3.36e-03	1.16e-03
64,007	4.06e-06	4.21e-06	3.11e-05	3.64e-05	2.18e-03	8.49e-04
120,011	2.41e-06	1.81e-06	2.34e-05	1.84e-05	3.13e-03	5.77e-04
240,007	1.69e-06	1.32e-06	1.81e-05	1.25e-05	2.40e-03	3.71e-04
480,013	6.60e-07	5.96e-07	6.69e-06	7.19e-06	6.44e-04	2.40e-04
Rate	0.91	0.81	0.80	0.68	0.54	0.58
90% Interval	$[0.99, 0.84]$	$[0.88, 0.73]$	$[0.91, 0.69]$	$[0.74, 0.61]$	$[0.76, 0.31]$	$[0.61, 0.55]$

vectors were returned as the weights became unstable. To illustrate this Table 5.18 presents the values of some early weights for the different of σ^2, with $\vartheta = 2$ and $\lambda_{\mathcal{C}} = 1.0$ fixed. Upon inspecting the weights for $\sigma^2 = 4.0$ it becomes clear why the CBC does not work for this case – the weights are excessively large and nonsensical, For example we see that $\gamma_{\{1,2\}} = 9.72 \times 10^{14}$ while $\gamma_{\{3\}} = 7.81 \times 10^{-3}$, due mostly to the fact that $\Psi_1 = 1.67 \times 10^6$ while $\Psi_3 = 36.75$.

The poor performance of the Gaussian choice for ψ_j is likely to be a reflection on its unsuitability towards this problem. It is the bound on the norm $\|F\|_{\mathcal{W}_s}$ (4.42) that gives rise to Ψ_j, specifically the $1/\check{a}(\boldsymbol{y})$ term that comes from (4.29) from Theorem 38. It seems that we would be unlikely to improve on this bound.

QMC quadrature results

In Table 5.19 we present the results for the QMC quadrature of the PDE problem with estimated rate of convergence r. We see for those few cases for which the CBC algorithm has worked, these lattice rules perform similarly to the lattice rules built with the exponential ψ_j, as well as the results for the generic lattice rules.

5.4 Conclusion

In Chapter 3 we described weighted spaces for unbounded integrands with general weights, and proved that the CBC construction yields good lattice rules that allow us to perform numerical integration with fast convergence for integrands contained in these spaces.

Table 5.16: $\gamma_{\mathfrak{u}}$ for various $\mathfrak{u}$ and two different regimes for α_j, with exponential ψ_j.

$\mathfrak{u}$	$\gamma_{\mathfrak{u}}$ for 1 choice of α_j	$\gamma_{\mathfrak{u}}$ for 2 distinct choices of α_j
$\{1\}$	2.80e-04	2.80e-04
$\{2\}$	2.78e-05	2.78e-05
$\{1,2\}$	1.73e-08	1.73e-08
$\{3\}$	1.21e-05	1.50e-01
$\{1,3\}$	7.52e-09	9.34e-05
$\{2,3\}$	7.47e-10	9.29e-06
$\{1,2,3\}$	7.40e-13	9.19e-09
$\{4\}$	6.91e-06	4.93e-02
$\{1,4\}$	4.29e-09	3.06e-05
$\{2,4\}$	4.26e-10	3.05e-06
$\{1,2,4\}$	4.22e-13	3.02e-09
$\{3,4\}$	1.86e-10	1.65e-02
$\{1,3,4\}$	1.84e-13	1.63e-05
$\{2,3,4\}$	1.83e-14	1.62e-06
$\{1,2,3,4\}$	2.52e-17	2.23e-09

Table 5.17: Worst-case errors $e^{\mathrm{sh}}_{s,n}$, using Gaussian ψ_j and POD weights $\gamma_{\mathfrak{u}}$ as in (4.51)

	$\vartheta = 2$			$\vartheta = 1.25$		
	$\sigma^2 = 0.25$		$\sigma^2 = 1.0$	$\sigma^2 = 0.25$		$\sigma^2 = 1.0$
n	$\lambda_{\mathcal{C}} = 1.0$	$\lambda_{\mathcal{C}} = 0.1$	$\lambda_{\mathcal{C}} = 1.0$	$\lambda_{\mathcal{C}} = 1.0$	$\lambda_{\mathcal{C}} = 0.1$	$\lambda_{\mathcal{C}} = 0.1$
8,009	1.07e-03	1.39e-02	5.72e-02	7.64e-03	1.89e+03	8.02e-02
16,001	6.14e-04	9.57e-03	3.35e-02	5.09e-03	1.34e+03	5.76e-02
32,003	3.57e-04	6.58e-03	2.07e-02	3.38e-03	9.44e+02	3.99e-02
64,007	2.03e-04	4.52e-03	1.23e-02	2.25e-03	6.67e+02	2.68e-02
120,011	1.24e-04	3.22e-03	8.05e-03	1.56e-03	4.87e+02	1.93e-02
240,007	7.17e-05	2.21e-03	5.08e-03	1.04e-03	3.45e+02	1.34e-02
480,013	4.17e-05	1.52e-03	2.89e-03	6.89e-04	2.44e+02	9.67e-03
Rate	0.79	0.54	0.73	0.59	0.5	0.52

Subsequently in Chapter 4 we investigated the PDE problem with random coefficients, known as the porous-flow problem, and showed that integrands arising from this problem were contained in our new weighted spaces. Thus our theory of good lattice rules applies for this problem, and hence we have shown that QMC methods applied to the PDE problem converged quickly, and in some cases optimally, with close to $\mathcal{O}(n^{-1})$ convergence. Finally, in this chapter we investigated the details of implementation of the techniques in both the previous chapters, and tested the performance of our tailored lattice rules for a model problem, leading to interesting numerical challenges in fitting the various parameters of the space to contain the problem.

We have successfully demonstrated good convergence of our QMC finite element method for this class of PDE problems numerically. Our results demonstrate that QMC rules comfortably beat MC rules in most cases, or in the cases of large σ^2 perform no worse. Furthermore we see that this is the case even for arbitrarily chosen lattice rules, as is demonstrated in Tables 5.14 and 5.15, despite the fact that the theory for these lattice rules does not apply to this problem.

Table 5.18: $\gamma_{\mathfrak{u}}$ for various $\mathfrak{u}$, with $\lambda_{\mathcal{C}} = 1.0$, $\vartheta = 2$, and with Gaussian ψ_j.

$\mathfrak{u}$	$\gamma_{\mathfrak{u}}$ for $\sigma^2 = 0.25$	$\gamma_{\mathfrak{u}}$ for $\sigma^2 = 1.0$	$\gamma_{\mathfrak{u}}$ for $\sigma^2 = 4.0$
$\{1\}$	1.38e+00	4.14e+02	2.89e+11
$\{2\}$	1.44e-01	1.31e+00	1.37e+03
$\{1,2\}$	4.86e-01	1.33e+03	9.72e+14
$\{3\}$	3.35e-02	9.06e-02	7.81e-01
$\{1,3\}$	1.13e-01	9.17e+01	5.53e+11
$\{2,3\}$	1.18e-02	2.90e-01	2.62e+03
$\{1,2,3\}$	6.71e-02	4.96e+02	3.13e+15
$\{4\}$	1.20e-02	2.27e-02	5.35e-02
$\{1,4\}$	4.03e-02	2.30e+01	3.78e+10
$\{2,4\}$	4.21e-03	7.26e-02	1.80e+02
$\{1,2,4\}$	2.40e-02	1.24e+02	2.15e+14
$\{3,4\}$	9.79e-04	5.02e-03	1.02e-01
$\{1,3,4\}$	5.57e-03	8.58e+00	1.22e+11
$\{2,3,4\}$	5.81e-04	2.71e-02	5.79e+02
$\{1,2,3,4\}$	4.80e-03	6.72e+01	1.00e+15

Table 5.19: QMC standard errors, using Gaussian ψ_j and POD weights $\gamma_{\mathfrak{u}}$ as in (4.51)

	$\vartheta = 2$			$\vartheta = 1.25$		
	$\sigma^2 = 0.25$		$\sigma^2 = 1.0$	$\sigma^2 = 0.25$		$\sigma^2 = 1.0$
n	$\lambda_{\mathcal{C}} = 1.0$	$\lambda_{\mathcal{C}} = 0.1$	$\lambda_{\mathcal{C}} = 1.0$	$\lambda_{\mathcal{C}} = 1.0$	$\lambda_{\mathcal{C}} = 0.1$	$\lambda_{\mathcal{C}} = 1.0$
8,009	2.95e-05	1.87e-05	1.98e-04	3.02e-05	1.83e-05	3.05e-04
16,001	1.42e-05	1.17e-05	1.05e-04	1.47e-05	1.29e-05	1.33e-04
32,003	7.75e-06	6.47e-06	6.55e-05	8.66e-06	7.65e-06	5.80e-05
64,007	4.32e-06	2.99e-06	3.50e-05	4.45e-06	3.76e-06	4.93e-05
120,011	2.50e-06	1.90e-06	2.21e-05	2.50e-06	2.32e-06	3.05e-05
240,007	1.53e-06	1.20e-06	1.91e-05	1.62e-06	9.35e-07	1.74e-05
480,013	7.69e-07	7.22e-07	7.86e-06	7.12e-07	8.65e-07	9.84e-06
Rate	0.86	0.8	0.76	0.86	0.79	0.78

Our numerical results do not quite match up with our theoretical predictions. For example, our theory predicts convergence rates of our QMC error that is dependent primarily on the rate of decay of b_j, specified here by the ϑ parameter. In the numerics however we see a much larger dependence on σ^2 and $\lambda_{\mathcal{C}}$ than on ϑ. If anything, this is evidence that there is further work that could be done towards making the theory sharper.

It is important to note that in these experiments neither the MC nor the QMC rules are enhanced using any variance reduction techniques such as the use of antithetic variates. This way we have a fair comparison between two unflavoured implementations.

Evidently there is scope for further work to sharpen our error bounds, as demonstrated by our numerics. Nevertheless, the results show that QMC finite element methods provide an excellent solution to the lognormal porous flow problem, and present a marked improvement over MC methods. Certainly as an "out-of-the-box" solution, implementing the machinery of the CBC algorithm for building custom lattice rules, built to fit this specific problem, may be beyond the scope of the practitioner. However, as has been seen, regular lattice rules, the generating vectors for which are widely available, work very well

for the problem. Our process of proving convergence of the problem in the weighted function space can be considered a theoretical achievement, with the added bonus of good numerical results.

The porous flow problem has been an exciting new setting in which to apply QMC methods. We have managed to analyse the PDE in a new weighted function space setting, leading to a process of finding tailored weights $\gamma_{\mathfrak{u}}$ and weight functions ψ_j, that leads to proven theoretical convergence. This is a novel discovery in the world of QMC.

References

[1] P. A. Acworth, M. Broadie, and P. Glasserman, *A comparison of some Monte Carlo and quasi Monte Carlo techniques for option pricing*, Monte Carlo and Quasi-Monte Carlo Methods 1996 (H. Niederreiter, P. Hellekalek, G. Larcher, and P. Zinterhof, eds.), Lecture Notes in Statistics, vol. 127, Springer New York, 1998, pp. 1–18 (English).

[2] R.J. Adler, *The geometry of random fields*, Classics in applied mathematics, Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104), 1981.

[3] N. Aronszajn, *Theory of reproducing kernels*, Trans. Amer. Math. Soc. **68** (1950), no. 3, 337–404.

[4] D. H. Bailey, *Tanh-sinh high-precision quadrature*, January 2006.

[5] D. H. Bailey and J. M. Borwein, *Effective error bounds in euler-maclaurin based quadrature schemes*, June 2005.

[6] R. E. Caflisch, W. Morokoff, and A. B. Owen, *Valuation of mortgage-backed securities using brownian bridges to reduce effective dimension*, Journal of Computational Finance **1** (1997), 27–46.

[7] J. Charrier, *Strong and weak error estimates for elliptic partial differential equations with random coefficients*, SIAM J. Numer. Anal. **50** (2012), no. 1, 216–246.

[8] J. Charrier and A. Debussche, *Weak truncation error estimates for elliptic pdes with lognormal coefficients*, Stochastic Partial Differential Equations: Analysis and Computations **1** (2013), no. 1, 63–93 (English).

[9] J. Charrier, R. Scheichl, and A. Teckentrup, *Finite element error analysis of elliptic PDEs with random coefficients and its application to multilevel Monte Carlo methods*, Tech. report, University of Bath, January 2013.

[10] K. A. Cliffe, I. G. Graham, R. Scheichl, and L. Stals, *Parallel computation of flow in heterogeneous media modelled by mixed finite elements*, Journal of Computational Physics **164** (2000), no. 2, 258 – 282.

[11] R. Cools, F. Y. Kuo, and D. Nuyens, *Constructing embedded lattice rules for multivariate integration*, SIAM J. Sci. Comp **28** (2006), no. 6, 2162–2188.

[12] G. Da Prato and J. Zabczyk, *Stochastic equations in infinite dimensions*, Encyclopedia of Mathematics and its Applications, Cambridge University Press, 2008.

[13] G. Dagan, *Solute transport in heterogeneous porous formations*, Journal of Fluid Mechanics **145** (1984), 151–177.

[14] R. A. Davis, Y. Wang, and W. T. M. Dunsmuir, *Modelling time series of count data*, Asymptotics, Nonparametrics, and Time Series (S. Ghosh, ed.), Statistics: Textbooks and Monographs, vol. 158, Dekker.

[15] R. A. Davis and G. Yodriguez-Yam, *Estimation for state-space models based on a likelihood approximation*, Statistica Sinica **15** (2005), 381–406.

[16] J. Dick, *On the convergence rate of the component-by-component construction of good lattice rules*, Journal of Complexity **20** (2004), no. 4, 493 – 522.

[17] J. Dick, F. Y. Kuo, and I. H. Sloan, *High-dimensional integration: The quasi-Monte Carlo way*, Acta Numerica **22** (2013), 133–288.

[18] J. Dick and F. Pillichshammer, *Discrepancy theory and quasi-Monte Carlo integration*, Cambridge University Press, Cambridge, 2010.

[19] J. Dick, I. H. Sloan, X. Wang, and H. Woźniakowski, *Liberating the weights*, Journal of Complexity **20** (2004), no. 5, 593 – 623, Dagstuhl 2002 - Festschrift for the 70th Birthday of Joseph F. Traub.

[20] J. Dick, I. H. Sloan, X. Wang, and H. Woźniakowski, *Good lattice rules in weighted korobov spaces with general weights*, Numer. Math. **103** (2006), no. 1, 63–97.

[21] S. Disney and I. H. Sloan, *Error bounds for the method of good lattice points*, Mathematics of Computation **56** (1991), no. 193, pp. 257–266 (English).

[22] D.A. Gilbarg and N.S. Trudinger, *Elliptic partial differential equations of second order*, Classics in mathematics, Springer-Verlag GmbH, 2001.

[23] M. B. Giles, F. Y. Kuo, I. H. Sloan, and B. J. Waterhouse, *Quasi-Monte Carlo for finance applications*, Proceedings of the 14th Biennial Computational Techniques and Applications Conference, CTAC-2008 (G. N. Mercer and A. J. Roberts, eds.), ANZIAM J., vol. 50, November 2008, pp. C308–C323.

[24] I. G. Graham, F. Y. Kuo, J. A. Nichols, R. Scheichl, C. Schwab, and I. H. Sloan, *Quasi-Monte Carlo finite element methods for elliptic PDEs with log-normal random coefficients*, Submitted for publication (2013).

[25] I. G. Graham, F. Y. Kuo, D. Nuyens, R. Scheichl, and I. H. Sloan, *Quasi-Monte Carlo methods for elliptic pdes with random coefficients and applications*, Journal of Computational Physics **230** (2011), no. 10, 3668 – 3694.

[26] M. Griebel, F. Y. Kuo, and I. H. Sloan, *The smoothing effect of the ANOVA decomposition*, J. Complex. **26** (2010), no. 5, 523–551.

[27] M. Griebel, F. Y. Kuo, and I. H. Sloan, *The smoothing effect of integration in $\mathbb{R}^d$ and the ANOVA decomposition*, Math. Comp. **82** (2013), 383–400.

[28] F. J. Hickernell, *Lattice rules: How well do they measure up?*, Technical report (Hong Kong Baptist University. Dept. of Mathematics), Department of Mathematics, Hong Kong Baptist University, 1998.

[29] F.J. Hickernell and H. Woźniakowski, *Integration and approximation in arbitrary dimensions*, Advances in Computational Mathematics **12** (2000), no. 1, 25–58.

[30] E. Hlawka, *Zur angenäherten berechnung mehrfacher integrale*, Monatshefte fr Mathematik **66** (1962), no. 2, 140–151 (German).

[31] N. M. Korobov, *The approximate computation of multiple integrals*, Doklady Akademii Nauk SSSR **124** (1959), 1207–1210 (Russian).

[32] L. Kuipers and H. Niederreiter, *Uniform distribution of sequences*, Dover Books on Mathematics, Dover Publications, 2006.

[33] F. Y. Kuo, *Component-by-component constructions achieve the optimal rate of convergence for multivariate integration in weighted korobov and sobolev spaces*, Journal of Complexity **19** (2003), no. 3, 301 – 320, Oberwolfach Special Issue.

[34] F. Y. Kuo, W. T. M. Dunsmuir, I. H. Sloan, M. P. Wand, and R. S. Womersley, *Quasi-Monte Carlo for highly structured generalised response models*, Methodology and Computing in Applied Probability **10** (2008), 239–275 (English).

[35] F. Y. Kuo and Stephen Joe, *Component-by-component construction of good lattice rules with a composite number of points*, Journal of Complexity **18** (2002), no. 4, 943 – 976.

[36] F. Y. Kuo, Ch. Schwab, and I. H. Sloan, *Quasi-Monte Carlo finite element methods for a class of elliptic partial differential equations with random coefficient*, **50** (2012), no. 6, 3351–3374.

[37] F. Y. Kuo, Ch. Schwab, and I. H. Sloan, *Quasi-Monte Carlo methods for high-dimensional integration: the standard (weighted hilbert space) setting and beyond*, ANZIAM Journal **53** (2012), no. 0.

[38] F. Y. Kuo, Ch. Schwab, and I. H. Sloan, *Multi-level quasi-Monte Carlo finite element methods for a class of elliptic partial differential equations with random coefficient*, (submitted).

[39] F. Y. Kuo, I. H. Sloan, G. W. Wasilkowski, and B. J. Waterhouse, *Randomly shifted lattice rules with the optimal rate of convergence for unbounded integrands*, J. Complex. **26** (2010), 135–160.

[40] F. Y. Kuo, Grzegorz W. Wasilkowski, and Benjamin J. Waterhouse, *Randomly shifted lattice rules for unbounded integrands*, Journal of Complexity **22** (2006), no. 5, 630 – 651, Special Issue: Information-Based Complexity Workshops FoCM Conference Santander, Spain, July 2005.

[41] G. Larcher, G. Leobacher, and K. Scheicher, *On the tractability of the brownian bridge algorithm*, J. Complex. **19** (2003), no. 4, 511–528.

[42] P. LEcuyer, *Quasi-Monte Carlo methods with applications infinance*, Finance and Stochastics **13** (2009), no. 3, 307–349 (English).

[43] Masatake M., *Discovery of the double exponential transformation and its developments*, Publications of The Research Institute for Mathematical Sciences **41** (2005), 897–935.

[44] R. L. Naff, D. F. Haley, and E. A. Sudicky, *High-resolution Monte Carlo simulation of flow and conservative transport in heterogeneous porous media 1. methodology and flow results*, Water Resour. Res. **34** (1998), 663–677.

[45] ———, *High-resolution Monte Carlo simulation of flow and conservative transport in heterogeneous porous media 2. transport results*, Water Resour. Res. **34** (1998), 679–697.

[46] J. A. Nichols and F. Y. Kuo, *Fast cbc construction of randomly shifted lattice rules achieving $\mathcal{O}(n^{-1+\delta})$ convergence for unbounded integrands in $\mathbb{R}^s$ in weighted spaces with POD weights*, Submitted for publication (2013).

[47] H. Niederreiter, *The existence of efficient lattice rules for multidimensional numerical integration*, Mathematics of Computation **58** (1992), no. 197, pp. 305–314 (English).

[48] ———, *Existence theorems for efficient lattice rules*, Numerical Integration (T. O. Espelid and A. Genz, eds.), NATO ASI Series, vol. 357, Springer Netherlands, 1992, pp. 71–80 (English).

[49] ———, *Random number generation and Quasi-Monte Carlo methods*, CBMS-NSF Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics, 1992.

[50] H. Niederreiter, *Improved error bounds for lattice rules*, Journal of Complexity **9** (1993), no. 1, 60 – 75.

[51] E. Novak and H. Woźniakowski, *Tractability of multivariate problems: Linear information*, EMS Tracts in Mathematics, no. 1, European Mathematical Society, 2008.

[52] ———, *Tractability of multivariate problems: Standard information for functionals*, EMS Tracts in Mathematics, no. 2, European Mathematical Society, 2010.

[53] D. Nuyens and R. Cools, *Fast algorithms for component-by-component construction of rank-1 lattice rules in shift-invariant reproducing kernel hilbert spaces*, Mathematics of Computation **75** (2006), no. 254, pp. 903–920 (English).

[54] D. Nuyens and R. Cools, *Fast component-by-component construction of rank-1 lattice rules with a non-prime number of points*, Journal of Complexity **22** (2006), no. 1, 4 – 28.

[55] R. Scheichl, *Iterative solution of saddle point problems using divergence-free finite elements with applications to groundwater flow*, Ph.D. thesis, University of Bath, 2000.

[56] Ch. Schwab and C.J. Gittelson, *Sparse tensor discretizations of high dimensional and stochastic pdes*, Acta Numerica **20** (2011), 291–467.

[57] V. Sinescu, F. Y. Kuo, and I. H. Sloan, *On the choice of weights in a function space for quasi-Monte Carlo methods for a class of generalised response models in statistics*, Monte Carlo and Quasi-Monte Carlo Methods 2012 (Submitted).

[58] I. H. Sloan, F. Y. Kuo, and S. Joe, *Constructing randomly shifted lattice rules in weighted sobolev spaces*, SIAM J. Numer. Anal. **40** (2002), no. 5, 1650–1665.

[59] ———, *On the step-by-step construction of quasi-Monte Carlo integration rules that achieve strong tractability error bounds in weighted sobolev spaces*, Math. Comput. **71** (2002), 1609–1640.

[60] I. H. Sloan and A. V. Reztsov, *Component-by-component construction of good lattice rules*, Math. Comp **71** (2002), 263–273.

[61] I. H. Sloan, X. Wang, and H. Woniakowski, *Finite-order weights imply tractability of multivariate integration*, Journal of Complexity **20** (2004), no. 1, 46 – 74.

[62] I. H. Sloan and H. Woźniakowski, *When are quasi-Monte Carlo algorithms efficient for high dimensional integrals?*, J. Complex. **14** (1998), 1–33.

[63] I. H. Sloan and H. Wozniakowski, *Tractability of multivariate integration for weighted korobov classes*, Journal of Complexity **17** (2001), no. 4, 697 – 721.

[64] I.H. Sloan and S. Joe, *Lattice methods for multiple integration*, Oxford Science Publications, Clarendon Press, 1994.

[65] H Takahasi and M. Mori, *Double exponential formulas for numerical integration*, Publications of The Research Institute for Mathematical Sciences **9** (1973), no. 3, 721–741.

[66] A. L. Teckentrup, R. Scheichl, M. B. Giles, and E. Ullmann, *Further analysis of multilevel Monte Carlo methods for elliptic pdes with random coefficients*, Numer. Math. (2012).

[67] C. Thomas-Agnan, *Computing a family of reproducing kernels for statistical applications*, Numerical Algorithms **13** (1996), 21–32.

[68] G. W. Wasilkowski and H. Woźniakowski, *Complexity of weighted approximation over* $\mathbb{R}^1$, Journal of Approximation Theory **103** (2000), no. 2, 223 – 251.

[69] G. W. Wasilkowski and H. Woźniakowski, *Tractability of approximation and integration for weighted tensor product problems over unbounded domains*, Monte Carlo and Quasi-Monte Carlo Methods 2000 (K. Fang, H. Niederreiter, and F. J. Hickernell, eds.), Springer Berlin Heidelberg, 2002, pp. 497–522 (English).

[70] G.W Wasilkowski and H Woźniakowski, *Weighted tensor product algorithms for linear multivariate problems*, Journal of Complexity **15** (1999), no. 3, 402 – 447.

[71] B. J. Waterhouse, F. Y. Kuo, and I. H. Sloan, *Randomly shifted lattice rules on the unit cube for unbounded integrands in high dimensions*, J. Complex. **22** (2006), 71–101.

[72] Y. Zhao, J. Staudenmayer, B. A. Coull, and M. P. Wand, *General design Bayesian generalized linear mixed models*, Statistical Science **21** (2006), 35–51.