Imperial College London

Individual Project

Imperial College London

Department of Computing

Towards a Taxonomy of Inducing Variables for Sparse Gaussian Processes

Author:
Harry Jake Cunningham

Supervisor:
Mark van der Wilk

Submitted in partial fulfillment of the requirements for the MSc degree in Computing (Artificial Intelligence and Machine Learning) of Imperial College London

September 12, 2021

Abstract

This work begins to classify different ways in which we can construct interdomain inducing variables for sparse Gaussian processes. Gaussian processes suffer from a $\mathcal{O}(N^3)$ computational complexity when performing exact inference. As a result, in recent times there has been a wealth of research drawing on connections between Gaussian processes, stochastic differential equations (SDEs) and the reproducing kernel Hilbert space (RKHS) in order to perform approximate inference. A particularly successful method has been using RKHS projections to map a Gaussian process onto a set of orthogonal basis functions. However, deriving useful forms of RKHS inner products is not trivial. As a result we wish to begin to classify when and where we can perform such projections. In this project we therefore aim to identify what choices we have when constructing interdomain inducing variables via projections that use the RKHS inner product, focusing in particular on the dimensionality and boundness of the domain in which the Gaussian process is defined. Our work utilises the SDE representation of Gaussian processes, whilst also drawing on deep connections between stochastic processes and the function spaces in which they are associated, to derive functional form RKHS inner products using methods of solving deterministic partial differential equations. In particular we introduce the variational formulation of boundary value problems as a direct approach to constructing RKHS inner products on bounded and unbounded one-dimensional domains and unbounded high-dimensional domains. We also stress a no-go approach to constructing RKHS inner products on high-dimensional bounded domains using regular methods of solving deterministic PDEs, due to the incomplete theory of stochastic elliptic regularity.

Acknowledgements

I would like to thank my supervisor Mark van der Wilk for his never ending enthusiasm and support in this project. His love of research and energy is inspiring and I look forward to working with him in the future. I would also like to thank Nicolas Durrande for his help in answering many a question on the intricacies of the RKHS.

Contents

1 Introduction

Gaussian process regression is a non-parametric Bayesian machine learning method that incorporates prior beliefs in order to predict outcomes with uncertainty. Prediction using Gaussian processes is performed using probabilistic inference, returning a posterior distribution conditioned on the training data and the test point input. However, computing Gaussian process posteriors becomes computationally intractable for large datasets ($N > 10,000$ data points) due to the $\mathcal{O}(N^3)$ time complexity to inverse the $N \times N$ covariance matrix during inference. In response, there has been a wealth of research investigating methods of approximate inference in order to reduce time complexity. In particular *sparse* methods approximate the Gaussian process posterior by instead conditioning on a number of M inducing variables or features, where $M \ll N$, reducing the computational cost to $\mathcal{O}(M^2N)$. Sparse approximations utilise variational inference in order to find the optimal approximate posterior that is closest to the exact Gaussian process posterior by minimising the Kullback-Leibler divergence.

Interdomain Gaussian processes are used to construct a representative and computationally efficient set of inducing features that lie in a different domain to the Gaussian process. Specifying interdomain inducing features relies on transforming the Gaussian process of interest to a different domain according to some linear operator $\mathcal{L} : \mathbb{R}^{\mathcal{X}} \to \mathbb{R}$. This allows for the incorporation of prior knowledge about the properties of functions that we are interested in and also enables us to build inducing variables that have global influence.

However, our primary interest in using interdomain transformations is to design inducing features that construct computationally efficient structure in the approximate Gaussian process posterior. By specifying the inducing features in a different domain it detaches the functional form of the exact Gaussian process covariance function with that of the approximate interdomain Gaussian process. By extension this enables us to represent approximate interdomain Gaussian process posteriors by a set of basis functions determined by the choice of interdomain transformation.

Many sparse Gaussian process methods have been proposed that utilise interdomain transformations to find orthogonal subspaces when conditioning on inducing variables. Indeed, in order to reduce the computational bottleneck caused by inverting the $\mathbf{K_{uu}}$ covariance matrix, we wish to condition on a set of M orthogonal inducing variables $\{\mathbf{u}_i\}_{i=1}^M$ such that the $\mathbf{K_{uu}}$ matrix is diagonal, removing the $\mathcal{O}(M^3)$ cost of inversion. As an example, Hensman et al. [1] defined *Variational Fourier Features* (VFF) by projecting a Gaussian process with a stationary one-dimensional half-integer Matérn kernel onto a Fourier basis, constructing a diagonal plus low-rank feature covariance matrix $\mathbf{K_{uu}}$.

However, for multi-dimensional input spaces VFF requires a new set of basis constructed by taking the outer product of the one-dimensional orthogonal basis. This causes the number of inducing variables to grow exponentially with input dimensionality. As an improvement, Dutordoir et al. [2] constructed inducing variables using a basis of spherical harmonics with implicit ordering such that the M most influential features are picked for a given computational budget. This enables the spherical harmonic model to reduce the impact of the curse of dimensionality.

Consequently, in order to control the orthogonality of the $\mathbf{K_{uu}}$ matrix when conditioning a sparse Gaussian process, we have 2 main decisions to make: the choice of kernel and the type of interdomain inducing variables. The choice of kernel and type of interdomain transformation are however often not independent decisions. Many interdomain inducing features, such as VFF, leverage properties of the kernel in order to project a function onto a set of orthogonal

basis. Such properties of a Gaussian process kernel include stationarity, representation as a *stochastic differential equation* (SDE) and access to the *reproducing kernel Hilbert space* (RKHS) inner product. As a result the choice of kernel places constraints on the type of interdomain transformation and therefore properties of interdomain inducing variables.

In this work, we would like to begin to classify ways in which we can construct interdomain inducing variables. In particular we are interested in the choices we have when we restrict ourselves to only consider stationary Gaussian processes with a rational spectral density function. Further, given the success of RKHS projection methods, we would like to consider the choices we have when we wish to project a Gaussian process onto a set of basis functions using an RKHS inner product. To do this, we focus on the construction of RKHS inner products under given conditions, specifically the dimensionality and boundness of the domain in which we work. We can visualise the relationships that we wish to define by the square in Figure 1.1, which seeks to map different choices available to use when constructing RKHS inner products on different domains.

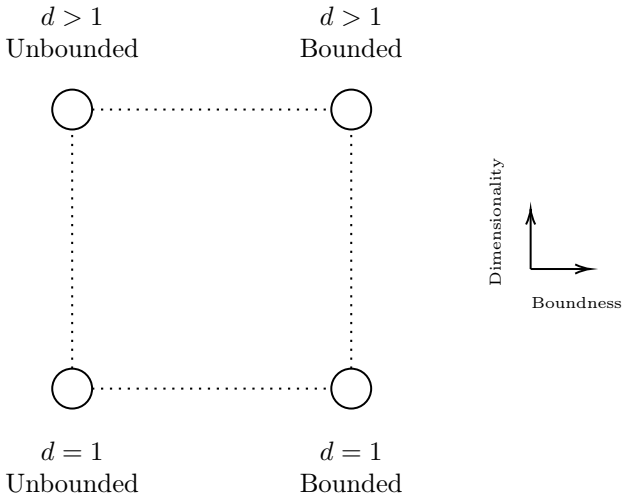

Figure 1.1: Square showing the relationships between different domains and the ability to construct an RKHS inner product that we wish to explore in this project. The y axis represents the dimensionality of the domain and the x axis the boundness. d is the dimensionality of the domain

Traditionally RKHS inner products are given as an infinite series expansion [3]. However, in this work we wish to derive functional form RKHS inner products that don't require expressing the inner product via some basis expansion. In order to do so, we make use of the connection between Gaussian processes and stochastic partial differential equations (SPDEs). This enables us to simplify the expressions by deriving RKHS inner products in an integral formulation. Hence this project also explores connections between Gaussian processes and SPDEs and SPDEs and the RKHS.

We note that some preliminary work done on the relationships between Gaussian process, SPDEs and the RKHS was done in Cunningham [4], submitted as an Independent Study Option (ISO) as part of the requirements for the MSc degree in Computing (Artificial Intelligence and Machine Learning). As a result all results derived in Cunningham [4] will be properly referenced as such.

2 Gaussian Processes

In this chapter we will begin by defining Gaussian processes as stochastic processes, before introducing their use as *distributions over functions* in the context of Gaussian process regression. We will also discuss the limiting computational cost associated with exact Gaussian process inference and sparse techniques used to address this issue. In particular we will motivate the use of approximate inference techniques that make use of RKHS projections in order to construct computationally efficient structure in the Gaussian process posterior.

2.1 Gaussian Processes

Let us start by first defining a random variable. A random variable is a real-valued quantity dependent on the outcome of a random event. More formally, if a random event is modelled by some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, then a random variable can be understood as a *measurable* function $X : \Omega \to \mathbb{R}$ that maps from the set of possible values Ω to the set of real numbers $\mathbb{R}$. The outcome of a random event therefore means choosing the outcome $\omega \in \Omega$ at random according to the probability measure $\mathbb{P}$ such that $X(\omega)$ is the value of the random variable corresponding to the outcome ω.

Definition 2.1.1 (Random Variable)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A random variable is a measurable function $X : \Omega \to \mathbb{R}$*

A stochastic process is any process that develops in time or space according to probabilistic laws rather than deterministic ones. Mathematically, a stochastic process is defined as a collection of random variables $\{X(t, \omega), t \in T\}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ indexed by some parameter t in the non-empty index set T [5, 6]. This means that for every $t \in T$ there exists a corresponding random variable $X(t, \omega)$ whose value is dependent on the random outcome $\omega \in \Omega$ which in turn is dependent on some probability measure $\mathbb{P}$. For simplicity we drop the dependence on ω in our notation and simply denote a stochastic process indexed at $t \in T$ as $X(t)$.

Definition 2.1.2 (Stochastic Process)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and T a non-empty index set. A stochastic process is the collection of random variables $\{X(t), t \in T\}$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ indexed by the non-empty set T.*

One way of defining a stochastic process is too specify the joint probability law of all finite collections of n random variables $X(t_1), ..., X(t_n)$ indexed by $\{t_1, ..., t_n\} \subset T$. Kolmogorov's extension theorem tells us that we can construct a stochastic process if there exists a collection of suitably *consistent* probability measures $\mathcal{P} = \{P_{t_1,...,t_n} : n \in \mathbb{N}, t_1, ..., t_n \in T\}$ such that $P_{t_1,...,t_n}$ is a probability measure on $\mathbb{R}^n$. If this is true then there exists a stochastic process $\{X(t) : t \in T\}$ whose finite-dimensional marginal distributions are given by the collection $\mathcal{P}$. This means that for every $n \in \mathbb{N}$ and every $\{t_1, ..., t_n\} \in \mathbb{N}$ the finite-dimensional joint distribution of the random variables $X(t_1), ..., X(t_n)$ is given by $P_{t_1,...,t_n}$.

A Gaussian process is a stochastic process such that any finite collection of random variables are jointly normally distributed. Given a Gaussian process, one obtains complete knowledge of the joint probability law of the process from a knowledge of the mean value function $\mu(\cdot) : \mathcal{X} \to \mathbb{R}$ and covariance function $k(\cdot, \cdot) : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. We denote a Gaussian process $f(\cdot)$ as $f(\cdot) \sim \mathcal{GP}(\mu(\cdot), k(\cdot, \cdot'))$.

Definition 2.1.3 (Gaussian Process)**.** *Let $\mathcal{X}$ be a non-empty set, $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a positive definite kernel and $\mu : \mathcal{X} \to \mathbb{R}$ be any real-valued function. A stochastic process $\{f(\mathbf{x}), \mathbf{x} \in \mathcal{X}\}$ is a Gaussian process, with mean function $\mu(\cdot)$ and kernel $k(\cdot, \cdot)$, where the sample paths $f : \mathcal{X} \to \mathbb{R}$ are denoted by $f(\cdot) \sim \mathcal{GP}(\mu(\cdot), k(\cdot, \cdot))$, if for all $n \in \mathbb{N}$ any finite set $X = \{\mathbf{x}_1, ..., \mathbf{x}_n\} \subset \mathcal{X}$, the random vector,*

$$f_X = [f(\mathbf{x}_1), \cdots f(\mathbf{x}_n)]^T \in \mathbb{R}^n, \tag{2.1}$$

follows the multivariate normal distribution $\mathcal{N}(\boldsymbol{\mu}_X, \mathbf{K}_{XX})$ with covariance matrix $\mathbf{K}_{XX} = [k(\mathbf{x}_i, \mathbf{x}_j)]_{i,j=1}^n \in \mathbb{R}^{n \times n}$ and mean vector $\boldsymbol{\mu}_X = [\mu(\mathbf{x}_1), \cdots \mu(\mathbf{x}_n)]^T$.

Remark 2.1.1. We are able to define a Gaussian process by it's mean and covariance function as the family of multivariate normal distributions X associated with a given mean function $\mu(\cdot)$ and kernel $k(\cdot, \cdot)$ satisfy the *consistency* conditions of the Kolmogorov extension theorem and therefore can be used to construct a stochastic process.

Remark 2.1.2. Being able to define a Gaussian process by it's mean function $\mu(\cdot)$ and covariance function $k(\cdot, \cdot)$ is very useful. Firstly, it is usually much easier to find both $\mu(\cdot)$ and $k(\cdot, \cdot)$ of a stochastic process than it is to find its complete probability law. Secondly, many important questions about a stochastic process, such as Bayesian inference, can be answered based on the knowledge of only its mean and covariance function.

The mean function $\mu(\cdot)$ returns the expected value of the Gaussian process $f(\cdot)$ at a given point, $\mathbb{E}[f(\mathbf{x})]$. The covariance function or kernel $k(\cdot, \cdot)$ defines the similarity between two random variables belonging to the Gaussian process, such that

$$\mathbb{C}_f[f(\mathbf{x}_i), f(\mathbf{x}_j)] = \mathbb{E}_f[(f(\mathbf{x}_i) - \mu(\mathbf{x}_i)), (f(\mathbf{x}_j) - \mu(\mathbf{x}_j))] = k(\mathbf{x}_i, \mathbf{x}_j). \tag{2.2}$$

In order for a function to be a valid kernel, it must be both symmetric and positive definite.

Definition 2.1.4 (Positive Definite Kernels)**.** *Let $\mathcal{X}$ be a non-empty set. A symmetric function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a positive definite kernel if for any $n \in \mathbb{N}$, $\{c_1, ..., c_n\} \subset \mathbb{R}$ and $\{\mathbf{x}_1, ..., \mathbf{x}_n\} \subset \mathcal{X}$,*

$$\sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j k(\mathbf{x}_i, \mathbf{x}_j) \geq 0. \tag{2.3}$$

The choice of kernel places structure on the random variables that exist on $\mathcal{X}$ by defining notions of similarity. Specifically, a kernel defines an inner product on the space of random variables, providing a measure of the strength of relationship between different random variables in the input space.

In this project we are particularly interested in stationary kernels.

Definition 2.1.5. *Let $\mathcal{X}$ be a non-empty set. A kernel is considered stationary if it is a function of $\boldsymbol{\tau} = \|\mathbf{x} - \mathbf{x}'\|$, $\forall \mathbf{x}, \mathbf{x}' \in \mathcal{X}$*

Remark 2.1.3. A stationary covariance function can be represented as the Fourier transform of a positive finite measure known as the spectral density. Hence stationary kernels have many convenient properties which we wish to use to construct stochastic partial differential equations.

An example of a popular class of stationary kernel is the *Matérn* class of kernels,

Example 2.1.1 (Matérn Kernels)**.** *The Matérn class of covariance kernel is given as,*

$$k_v(\boldsymbol{\tau}) = \sigma^2 \frac{2^{1-v}}{\Gamma(v)} \left(\frac{\sqrt{2v}\boldsymbol{\tau}}{l} \right)^v K_v \left(\frac{\sqrt{2v}\boldsymbol{\tau}}{l} \right), \tag{2.4}$$

where Γ is the gamma function, K_v the modified Bessel function, v and l are positive shape parameters and $\boldsymbol{\tau} = \|\mathbf{x}_i - \mathbf{x}_j\|$.

Remark 2.1.4. A one-dimensional Gaussian process with Matérn covariance is k times differentiable in the mean-square sense, if and only if $v > k$. Hence, v controls the smoothness of a Matérn Gaussian process and l the lengthscale.

Remark 2.1.5. The Matérn covariance function simplifies considerably when v is a half-integer: $v = p + 1/2$ where p is a positive integer and d is the dimensionality of the inputs, such that the kernel can be constructed as the product of an exponential and polynomial of order ,

$$k_{p+1/2}(\boldsymbol{\tau}) = \sigma^2 \exp\left(\frac{\sqrt{2v}\boldsymbol{\tau}}{l}\right) \frac{\Gamma(p+1)}{\Gamma(2p+1)} \sum_{i=0}^{p} \frac{(p+i)!}{i!(p-i)!} \left(\frac{\sqrt{8v}\boldsymbol{\tau}}{l}\right)^{p-i}, \tag{2.5}$$

In machine learning we are most interest in the cases where $v = 1/2$, $v = 3/2$ and $v = 5/2$,

$$k_{1/2}(\boldsymbol{\tau}) = \sigma^2 \exp\left(\frac{-\boldsymbol{\tau}}{l}\right), \tag{2.6}$$

$$k_{3/2}(\boldsymbol{\tau}) = \sigma^2 \left(1 + \frac{\sqrt{3}\boldsymbol{\tau}}{l}\right) \exp\left(\frac{-\sqrt{3}\boldsymbol{\tau}}{l}\right), \tag{2.7}$$

$$k_{5/2}(\boldsymbol{\tau}) = \sigma^2 \left(1 + \frac{\sqrt{5}\boldsymbol{\tau}}{l} + \frac{5\boldsymbol{\tau}^2}{3l^2}\right) \exp\left(\frac{-\sqrt{5}\boldsymbol{\tau}}{l}\right). \tag{2.8}$$

$$\tag{2.9}$$

As an example, Figure 2.1 shows samples from two different Gaussian priors with Matérn 1/2 and 3/2 kernels, highlighting the difference in smoothness.

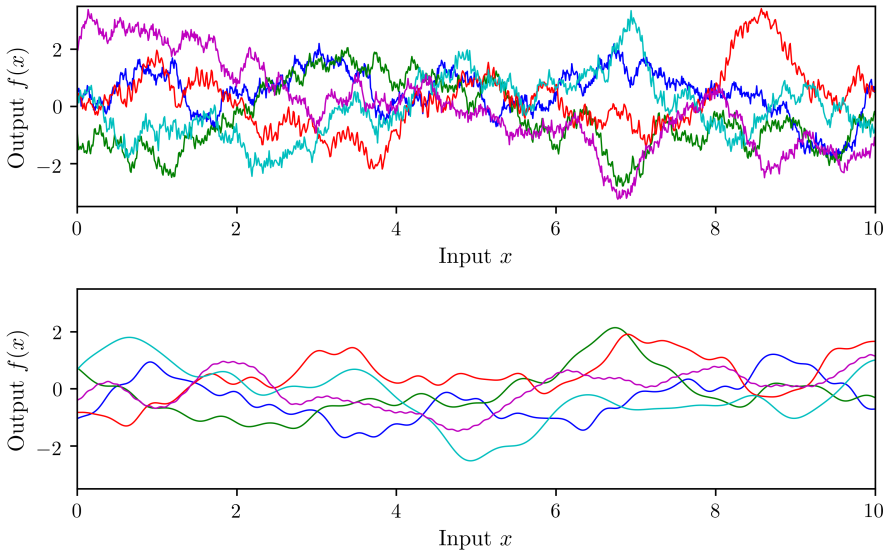

Figure 2.1: Samples from a Gaussian process prior with Matérn 1/2 kernel (Top) and Matérn 3/2 kernel (Bottom), with lengthscales $l = 1$

Finally, we reinforce the notation that samples path of a stochastic process distributed according to a Gaussian process are denoted as,

$$f(\cdot) \sim \mathcal{GP}\left(\mu(\cdot), k(\cdot,\cdot)\right). \tag{2.10}$$

In the next section, we will consider how stochastic processes are used in machine learning. In particular, we will examine how Gaussian processes are used as a prior distribution over functions in Bayesian inference

2.2 Gaussian Process Regression

In *machine learning* stochastic processes are commonly used to model some function $f(\cdot)$ given a set of noisy observations $y = f(\cdot) + \textit{noise}$. In particular we are interested in using Gaussian processes as a prior distribution over functions in a supervised regression task.

Let us first motivate why we want to use Gaussian processes as priors over functions. Consider the regression problem,

$$y_n = \phi(\mathbf{x}_n)^T \boldsymbol{\theta} + \epsilon_n, \tag{2.11}$$

where we seek to *learn* a parameter vector $\boldsymbol{\theta}$ for a set of basis functions $\phi(\cdot)$ that best represent an underlying process $f(\cdot)$, such that $f(\cdot) = \phi(\cdot)^T \boldsymbol{\theta}$. In Bayesian linear regression, we model uncertainty in our predictions by placing a prior distribution on the possible values of $\boldsymbol{\theta}$. Implicitly, by placing a distribution on the possible parameters $\boldsymbol{\theta}$ we are placing a distribution on possible function values $\phi(\cdot)^T \boldsymbol{\theta}$ and hence we place a prior over functions.

However, to properly capture uncertainty and have a flexible enough model, we often require many or *infinite* basis functions. Consequently, when $\boldsymbol{\theta}$ is an infinite dimensional *random* vector, such that there are an infinite number of basis functions, from the previous section $\boldsymbol{\theta}$ is now a stochastic process. Hence, just like in the finite dimensional case where we use the random vector $\boldsymbol{\theta}$ to place a prior distribution on functions, we would like to use stochastic processes as a prior distribution over functions.

The advantage of using stochastic processes as a distribution over functions, is that we no longer place a prior on possible function values via the prior on parameters, but instead we directly place a prior on all function values. Gaussian processes, enable us to place a prior over all function values according to a mean and covariance function, such that the marginal distribution of a function value $f(\mathbf{x})$ at a given input $\mathbf{x}$ is normally distributed. This is advantageous, as the Gaussian process prior removes the dependency on a *suitably descriptive* set of basis functions, which can restrict model expressiveness, by instead assigning placing probability mass on the space of all continuous functions, enabling Gaussian processes to estimate any continuous function arbitrarily well.

We now wish to perform regression using Gaussian processes. Given a dataset $\mathcal{D} = \{(y_n, \mathbf{x}_n)\}_{n=1}^N$ of N noisy observations $y_n \in \mathbb{R}$ and corresponding feature vectors $\mathbf{x}_n \in \mathcal{X}$, we can construct a Gaussian process regression model as,

$$y_n = f(\mathbf{x}_n) + \epsilon_n, \tag{2.12}$$

$$f(\cdot) \sim \mathcal{GP}(0, k(\cdot, \cdot)), \tag{2.13}$$

$$\epsilon \sim \mathcal{N}(0, \sigma_n^2 \mathbf{I}), \tag{2.14}$$

by placing a Gaussian process prior over possible functions $f(\cdot)$ and modelling observation noise ϵ as a zero-mean normal distribution with variance $\sigma_n^2 \mathbf{I}$.

To make predictions we perform inference, by conditioning the Gaussian process $f(\cdot)$ on the observations $\mathbf{y}$ to form the predictive posterior distribution $p(f(\cdot) \mid \mathbf{y}, \mathbf{X})$, where $\mathbf{y} = \{y_n\}_{n=1}^N$ and $\mathbf{X} = \{\mathbf{x}_n\}_{n=1}^N$. As a note, this project will only consider the case where the likelihood function is Gaussian, $p(y_n \mid f(\mathbf{x}_n)) = \mathcal{N}(y_n; f(\mathbf{x}_n), \sigma^2)$, such that our posterior distribution is also Gaussian by the rules of Gaussian conditioning. To find the posterior we first write the joint distribution $p(f(\cdot), \mathbf{y})$ between the noisy observations $\mathbf{y}$ and the function values $f(\cdot)$ under the Gaussian process prior, as,

$$p(f(\cdot), \mathbf{y} \mid \mathbf{X}) = \mathcal{N}\left(\begin{bmatrix} f(\cdot) \\ y \end{bmatrix}; 0, \begin{bmatrix} k(\cdot, \cdot) & \mathbf{k}_{\cdot \mathbf{f}} \\ \mathbf{k}_{\mathbf{f} \cdot} & \mathbf{K}_{\mathbf{ff}} + \sigma_n^2 \mathbf{I} \end{bmatrix} \right), \tag{2.15}$$

where $\mathbf{K}_{\mathbf{ff}}[m,n] = k(\mathbf{x}_m, \mathbf{x}_n)$ and $\mathbf{k}_{\mathbf{u}\cdot}[n] = \mathbf{k}_{\mathbf{u}\cdot}{}^T[n] = \mathbb{C}_f[f(\cdot), f(\mathbf{x}_n)] = k(\cdot, \mathbf{x}_n)$. Conditioning the joint distribution on the observations $\mathbf{y}$, we fit the Gaussian process to the data. By the rules of Gaussian conditioning, the predictive Gaussian process posterior is,

$$p(f(\cdot) \mid \mathbf{y}, \mathbf{X}) = \mathcal{N}(f(\cdot); \quad \mathbf{k}_{\cdot\mathbf{f}}(\mathbf{K}_{\mathbf{ff}} + \sigma^2\mathbf{I})^{-1}\mathbf{y}, \quad k(\cdot,\cdot') - \mathbf{k}_{\cdot\mathbf{f}}(\mathbf{K}_{\mathbf{ff}} + \sigma^2\mathbf{I})^{-1}\mathbf{k}_{\mathbf{f}\cdot'}). \tag{2.16}$$

This is an elegant result that allows us to construct a closed-form expression for the predictive posterior, that represents a distribution over functions that explain the observed data. Figure 2.2 gives an example of Gaussian process regression for a toy dataset.

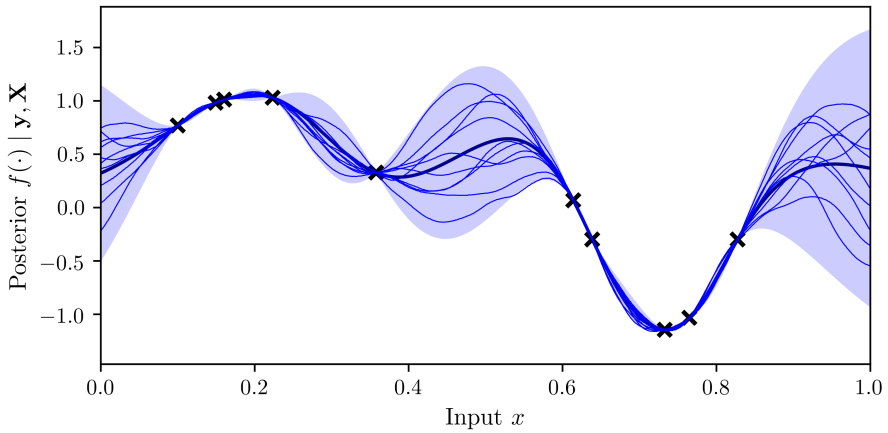

Figure 2.2: Gaussian Process regression for a toy dataset using the Matérn 5/2 covariance function. The thick dark blue line is the mean function, and the thinner lines samples from the posterior. The shaded blue region represents the 95% confidence interval. Observing the confidence intervals, it is clear that the model becomes more sure in its predictions where lots of data is present and less sure the further away from the data it is.

However, Gaussian processes famously suffer from an $\mathcal{O}(N^3)$ complexity cost, that grows cubically with the size of the data set. This is due to the inversion of the $(\boldsymbol{K}_{\mathbf{ff}} + \sigma^2\boldsymbol{I})$ matrix at inference, rendering Gaussian processes computationally intractable for comparatively small datasets with $N > 10,000$ training observations.

2.3 Sparse Gaussian Processes

To overcome the $\mathcal{O}(N^3)$ time complexity of exact Gaussian process inference many approximate and spare methods have been proposed in the literature [7, 8, 9, 1, 10, 11, 12, 13, 14, 2]. The goal of these sparse approximations is to reduce the time complexity for inference by constructing a computationally tractable approximate Gaussian process posterior that is *as close as possible* to the exact Gaussian process posterior, defined by minimising the Kullback-Leibler (KL) divergence between the approximate and exact posterior.

Sparse Gaussian processes achieve this by limiting the amount of data that is used to represent a Gaussian process posterior by constructing a conditional distribution based on a set of M inducing variables $\mathbf{u}_m = f(z_m)$ at inducing input locations $\mathbf{Z} = \{\mathbf{z}_m\}_{m=1}^M$ that lie in the initial input space $\mathcal{X}$ [15, 16], where $M \ll N$. This reduces the time complexity for inference from $\mathcal{O}(N^3)$ to $\mathcal{O}(NM^2)$. The intuition behind using a smaller set of inducing variables to represent a larger Gaussian process is that there is a lot of redundant information when observing locally many noisy observations, which can be well represented by a single point.

We can derive the approximate inducing point posterior by first considering the joint distribution,

$$p(f(\cdot), \mathbf{u}) = \mathcal{N}\left(\begin{bmatrix} f(\cdot) \\ \mathbf{u} \end{bmatrix}; 0, \begin{bmatrix} k(\cdot,\cdot') & \mathbf{k}_{\cdot\mathbf{u}} \\ \mathbf{k}_{\mathbf{u}\cdot'} & \mathbf{K}_{\mathbf{uu}} \end{bmatrix}\right), \tag{2.17}$$

where $\mathbf{K_{uu}}$ is formed by evaluating the covariance function at all pairs of inducing input locations, $\mathbf{K_{uu}}[m,n] = k(\mathbf{z}_m, \mathbf{z}_n)$, and $\mathbf{k}_{\cdot\mathbf{u}}$ is formed by evaluating the covariance function between the data input points and inducing input locations, $\mathbf{k}_{\cdot\mathbf{u}}[m] = \mathbf{k}_{\mathbf{u}\cdot'}{}^T[m] = k(\cdot, \mathbf{z}_m)$. Next we write the conditional distribution of $f(\cdot)$ conditioned on $\mathbf{u}$, using the standard conditioning rules for a multivariate Gaussian and the joint distribution from equation (2.17),

$$p(f(\cdot) \mid \mathbf{u}) = \mathcal{N}(f(\cdot); \quad \mathbf{k}_{\cdot\mathbf{u}}\mathbf{K}_{\mathbf{uu}}^{-1}\mathbf{u}, \quad k(\cdot,\cdot') - \mathbf{k}_{\cdot\mathbf{u}}\mathbf{K}_{\mathbf{uu}}^{-1}\mathbf{k}_{\mathbf{u}\cdot'}). \tag{2.18}$$

Now we introduce a marginal distribution over all the inducing variables $\mathbf{u}$ as $q(\mathbf{u}) = \mathcal{N}(\mathbf{u}; \mathbf{m_u}, \mathbf{S_{uu}})$. Here we specify both the mean $\mathbf{m_u} \in \mathbb{R}^M$ and variance $\mathbf{S_{uu}} \in \mathbb{R}^{M\times M}$ as free variational parameters which we can optimise over. Using the marginal distribution $q(\mathbf{u})$ we integrate out $\mathbf{u}$ from the conditional distribution in equation (2.18), to form our approximate posterior $q(f(\cdot))$,

$$q(f(\cdot)) = \int p(f(\cdot) \mid \mathbf{u}) q(\mathbf{u}) d\mathbf{u}. \tag{2.19}$$

We are now able to derive the family of approximate sparse Gaussian process posteriors, parameterised by the free parameters $\mathbf{m_u}$ and $\mathbf{S_{uu}}$ as derived by Titsias [7],

$$q(f(\cdot)) = \mathcal{N}(f(\cdot); \quad \mathbf{k}_{\cdot\mathbf{u}}\mathbf{K}_{\mathbf{uu}}^{-1}\mathbf{m_u}, \quad k(\cdot,\cdot') - \mathbf{k}_{\cdot\mathbf{u}}\mathbf{K}_{\mathbf{uu}}^{-1}(\mathbf{K_{uu}} - \mathbf{S_{uu}})\mathbf{K}_{\mathbf{uu}}^{-1}\mathbf{k}_{\mathbf{u}\cdot'}), \tag{2.20}$$

and the approximate posterior process is given as,

$$f(\cdot) \sim \mathcal{GP}(\mathbf{k}_{\cdot\mathbf{u}}\mathbf{K}_{\mathbf{uu}}^{-1}\mathbf{m_u}, \quad k(\cdot,\cdot') - \mathbf{k}_{\cdot\mathbf{u}}\mathbf{K}_{\mathbf{uu}}^{-1}(\mathbf{K_{uu}} - \mathbf{S_{uu}})\mathbf{K}_{\mathbf{uu}}^{-1}\mathbf{k}_{\mathbf{u}\cdot'}). \tag{2.21}$$

Note how the computational complexity for computing the inverse matrix of $\mathbf{K_{uu}}$ is $\mathcal{O}(M^3)$, which is no longer dependent on the number of data points N, but the controllable number of inducing points M. However, the number of inducing points M also determines the expressiveness of the sparse Gaussian process; the more points meaning more inducing inputs and hence a more accurate approximate representation of the exact posterior. Consequently, it is clear that when $M = N$ the true posterior exists within the family of possible approximate posteriors [17]. Naturally therefore, the larger the number of inducing points the greater the computational complexity and therefore in practise we utilise a low M.

Using a low M can however cause issues. Given a high dimensional input space, due to inducing variables having only a small local influence, as defined by the kernel length scale, the number of inducing inputs required to cover the input space grows exponentially with the input dimensions [18, 2]. This is an example of the curse of dimensionality and in practise requires a large M to obtain an accurate approximate posterior, reducing the benefit of the sparse approximation.

It should also be made aware that the choice of a Gaussian distribution $q(\mathbf{u})$ and subsequent Gaussian form of the approximate posterior family $q(f(\cdot))$ allows for the KL divergence between $q(f(\cdot))$ and $p(f(\cdot) \mid \mathbf{y}, X)$ to be finite and calculable in closed form. In practise this enables us to phrase the approximate inference problem as an optimisation problem using variational inference [7, 8, 9].

2.4 Interdomain Gaussian Processes

Standard inducing point methods focus on finding a small representative dataset of inducing variables from the same domain as the input data, reducing inference time complexity from $\mathcal{O}(N^3)$ to $\mathcal{O}(M^2N)$. However, representing the full Gaussian process with $M \ll N$ inducing points can be difficult, especially when the input data is high dimensional or for short kernel length scales that cause the sparse Gaussian process posterior to collapse to the prior mean and variance [2].

Interdomain Gaussian processes can be used to find a more representative set of inducing features that lie in a different domain to the input data, but with the same computational speedup [19, 17]. By transforming the inducing points into a more informative domain we are able to incorporate prior knowledge about the data directly into our inducing features.

We define an interdomain transformation by some linear operator $\mathcal{L} : \mathbb{R}^{\mathcal{X}} \to \mathbb{R}$ which takes a Gaussian process $f(\mathbf{x})$ with $x \in \mathbb{R}^D$ and returns a real value $\mathcal{L}f(\mathbf{x})$. Since $\mathcal{L}f(\cdot)$ is obtained by a a linear transformation of the Gaussian process $f(\cdot)$ then $\mathcal{L}f(\cdot)$ is also a Gaussian process. Consequently, we can define an interdomain inducing variable u_m by the linear operator $\mathcal{L}_m$,

$$u_m = \mathcal{L}_m f(\cdot). \tag{2.22}$$

The choice of linear operator $\mathcal{L}$ is practically important. Commonly, interdomain inducing variables are obtained by considering $\mathcal{L}$ to be an integral operator with an inducing function $\phi_m(\mathbf{x})$ [19],

$$u_m = \int_{\mathbb{R}^D} \phi_m(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}. \tag{2.23}$$

The inducing variables obtained via the integral operator can be seen as projections of the target Gaussian process $f(\mathbf{x})$ on the inducing function $\phi_m(\mathbf{x})$ over the entire input domain $\mathbb{R}^D$. Hence, each inducing feature now represents information about $f(\mathbf{x})$ everywhere, making each interdomain inducing point more informative than their previously local feature.

In the same way that we derived the inducing point posterior in (2.20), we can derive the interdomain inducing point posterior from the joint distribution $p(f(\cdot), \mathcal{L}f(\cdot))$,

$$p(f(\cdot), \mathcal{L}f(\cdot)) = \mathcal{N}\left(\begin{bmatrix} f(\cdot) \\ \mathcal{L}f(\cdot) \end{bmatrix}; 0, \begin{bmatrix} k(\cdot,\cdot') & \mathbf{k}_{\cdot\mathcal{L}} \\ \mathbf{k}_{\mathcal{L}\cdot'} & \mathbf{K}_{\mathcal{L}\mathcal{L}} \end{bmatrix}\right), \tag{2.24}$$

with covariances, $\mathbf{k}_{\mathcal{L}\cdot}[m] = \mathbb{C}[\mathcal{L}_m f(\cdot), f(\cdot)]$ and $\mathbf{K}_{\mathcal{L}\mathcal{L}}[m,n] = \mathbb{C}[\mathcal{L}_m f(\cdot), \mathcal{L}_n f(\cdot)]$. Applying the rules for the linear transformations of Gaussian processes we can expand the covariances by taking expectations [17, 11],

$$\mathbb{C}[\mathcal{L}_m f(\cdot), f(\mathbf{x}')] = \mathrm{E}_{f(\cdot)}[\mathcal{L}_m f(\cdot) f(\mathbf{x}')] = \mathcal{L}_m \mathrm{E}_{f(\cdot)}[f(\cdot), f(\mathbf{x}')] = \mathcal{L}_m k(\cdot, \mathbf{x}'), \tag{2.25}$$

$$\mathbb{C}[\mathcal{L}_m f(\cdot), \mathcal{L}_n f(\cdot)] = \mathrm{E}_{f(\cdot)}[\mathcal{L}_m f(\cdot) \mathcal{L}_n f(\cdot)] = \mathcal{L}_m k(\cdot,\cdot') \mathcal{L}_{n'}^T, \tag{2.26}$$

where $\mathcal{L}_{n'}^T$ is the linear operator applied to the second argument $\cdot'$ of the kernel. The transpose here is just a normal matrix transpose.

Consequently, we can write the approximate inducing point posterior for interdomain Gaussian processes similarly as in (2.20),

$$q(f(\cdot)) = \mathcal{N}(f(\cdot); \quad \mathbf{k}_{\cdot\mathcal{L}} \mathbf{K}_{\mathcal{L}\mathcal{L}}^{-1} \mathbf{m}_{\mathbf{u}}, \quad k(\cdot,\cdot') - \mathbf{k}_{\cdot\mathcal{L}} \mathbf{K}_{\mathcal{L}\mathcal{L}}^{-1} (\mathbf{K}_{\mathcal{L}\mathcal{L}} - \mathbf{S}_{\mathbf{uu}}) \mathbf{K}_{\mathcal{L}\mathcal{L}}^{-1} \mathbf{k}_{\mathcal{L}\cdot'}). \tag{2.27}$$

Another use of interdomain transformations is to try and induce computational speedup. This has been achieved by transforming the domain of the inducing variables such that the $\mathbf{K}_{\mathbf{uu}}$ matrix is a (near) diagonal matrix eradicating the $\mathcal{O}(N^3)$ cost of inverting a matrix at inference. The following section will review two examples of recent work that exploits the RKHS representation of a positive-definite Gaussian processes covariance function in order to derive interdomain inducing variables that produce computationally beneficial structure in the $\mathbf{K}_{\mathbf{uu}}$ matrix.

2.4.1 Spectral Inducing Point Methods

Several methods exist in the literature for defining inducing variables that lead to (near) diagonal $\mathbf{K}_{\mathbf{uu}}$ matrices by using a RKHS projection between a Gaussian processes of interest and a set of basis functions. The following section will review two methods in particular: Variational Fourier Features and Spherical Harmonic Features.

Variational Fourier Features

The Variational Fourier Features (VFF) approximation [1] combines variational inference with interdomain inducing points. They define the inducing variables by projecting the Gaussian process f onto a truncated Fourier basis ϕ_m using a Matérn RKHS inner product defined over a one-dimensional bounded domain,

$$u_m = \langle f, \phi_m \rangle_{\mathcal{H}} = \mathcal{P}_{\phi_m}(f), \tag{2.28}$$

where $\phi_0 = 1$, $\phi_{m+1} = \cos(\omega_m(x-a))$, $\phi_{2m+1} = \sin(\omega_m(x-a))$ and $\omega_m = \frac{2\pi m}{b-a}$ for an input space $[a, b]$. Similarly, like before the covariance and cross covariance of the inducing variables with the function values are calculated as,

$$\mathbb{C}[u_m, f(\mathbf{x})] = \mathbb{E}[u_m f(\mathbf{x})] = \mathbb{E}[\mathcal{P}_{\phi_m}(f) f(\mathbf{x})] = \mathcal{P}_{\phi_m}(k(\mathbf{x}, \cdot)) = \phi_m(\mathbf{x}), \tag{2.29}$$

$$\mathbb{C}[u_m, u_n] = \mathbb{E}[u_m u_n] = \mathbb{E}[\mathcal{P}_{\phi_m}(f)\mathcal{P}_{\phi_n}(f))] = \mathcal{P}_{\phi_m}(\phi_n) = \langle \phi_m, \phi_n \rangle_{\mathcal{H}}. \tag{2.30}$$

Consequently, this leads to,

$$\mathbf{k}_{\cdot\mathbf{u}}(\mathbf{x}) = [\phi_i(\mathbf{x})]_{i=0}^{M-1}, \quad \mathbf{K}_{\mathbf{uu}} = [\langle \phi_i, \phi_j \rangle_{\mathcal{H}}]_{i,j=0}^{M-1}. \tag{2.31}$$

This structure produces several computational benefits. Firstly, the $\mathbf{k}_{\cdot\mathbf{u}}$ vector is a function of the Fourier basis, independent of the kernel parameters and therefore can be easily precomputed. Secondly, for Matérn kernels, where the inner product has an analytical form [20], the $\mathbf{K}_{\mathbf{uu}}$ matrix is the sum of a diagonal matrix plus low rank matrices. This property reduces the computational bottleneck associated with computing $\mathbf{K}_{\mathbf{uu}}^{-1}$ at inference from $\mathcal{O}(NM^2)$ to $\mathcal{O}(NM)$ using the Woodbury Identity [1].

Spherical Harmonic Features

There are however, several flaws with the VFF approach [2]. The most significant flaw is that VFF generalises poorly to high dimensional inputs spaces, as the number of inducing point increases exponentially with dimensionality D.

Dutordoir et al. [2] propose a solution, by instead projecting the input data onto a unit hypersphere $\mathbb{S}^{d-1}$ in order to use spherical harmonic representations. The projection onto the hypersphere is defined by the linear mapping $(\mathbf{x}, y) \mapsto (\mathbf{x}/\|\mathbf{x}\|, y/\|\mathbf{x}\|)$. The equivalent to stationary kernels on the input space are zonal kernels on the hypersphere. It can also be shown that the eigenfunctions of the Laplacian operator on the input space, are equivalent to the eigenfunctions of the Laplace-Beltrami operator on the hypersphere (See [2] for greater detail). This property is very useful as it is commonly known that the eigenfunctions of the the Laplace-Beltrami operator are given by spherical harmonics. Spherical harmonics defined on a hypersphere form a complete set of orthogonal basis functions forming a higher-dimensional analogy to the Fourier series [2, 21, 22]. Hence, Dutordoir et al. [2] expand the kernel according to Mercer's theorem [See Definition 3.4.2],

$$k(\mathbf{x}, \mathbf{x}') = \sum_{l=0}^{\infty} \sum_{k=1}^{N_l^d} \alpha_{l,k} \phi_{l,k}(\mathbf{x}) \phi_{l,k}(\mathbf{x}'), \tag{2.32}$$

where $\mathbf{x}, \mathbf{x}' \in \mathbb{S}^{d-1}$, $\alpha_{l,k}$ are positive coefficients, $\phi_{l,k}$ are the spherical harmonic basis in $\mathbb{S}^{d-1}$ corresponding the to the zonal kernels eigenfunctions and N_l^d is the number of spherical harmonics for a given level l. The positive coefficients $\alpha_{l,k}$ have been derived for both the Arc-Cosine kernel [2] and the Matérn family of kernels [13]. Similarly to VFF, given the Mercer representation, we can therefore define the inducing variables u_m by the RKHS inner product between the Gaussian process f and the spherical harmonics ϕ_m,

$$u_m = \langle f, \phi_m \rangle_{\mathcal{H}}. \tag{2.33}$$

The corresponding covariances are,

$$\mathbb{C}[u_m, f(\mathbf{x})] = \mathbb{E}[u_m f(\mathbf{x})] = \langle k(\mathbf{x}, \cdot), \phi_m \rangle_{\mathcal{H}} = \phi_m(\mathbf{x}), \tag{2.34}$$

$$\mathbb{C}[u_m, u_n] = \mathbb{E}[u_m u_n] = \langle \phi_m, \phi_n \rangle_{\mathcal{H}} = \frac{\delta_{m,n}}{\alpha_m}, \tag{2.35}$$

where $\mathbf{K_{uu}}$ is a diagonal matrix with elements $\mathbf{K_{uu}} = [1/\alpha_{m,n}]_{m,n=0}^{M-1}$ Consequently, spherical harmonic inducing features have the same benefits of VFF, being able to precompute $\mathbf{k_{\cdot u}}$, diagonal $\mathbf{K_{uu}}$ matrix, but they also scale to higher dimensions as the number of spherical harmonics for a specific dimension d and level l is given by,

$$N_l^d = (2l + d - 2)\frac{\Gamma(l + d - 2)}{\Gamma(l + 1)\Gamma(d - 2)}, \tag{2.36}$$

where d is the dimensionality plus a bias term of the input space, l the level of the expansion and Γ the gamma function. For example, using VFF and the Kronecker construction, for an input with 8 dimensions, using only 4 frequencies per dimension, results in $M = 4^8 = 65,536$ inducing variables. [2], show that the equivalent spherical harmonic approximation up to degree 3, which uses $M = \sum_{l=0}^{3} N_l^8 = 264$ inducing points, minimises the negative log predictive density and improves computation time when compared to VFF and SVGPs. Further, the benefits of spherical harmonics over VFF motivates deriving and working with RKHS projections on higher-dimensional domains.

2.5 Discussion

In this chapter we have introduced Gaussian processes as stochastic processes and their use in machine learning as a non-parametric Bayesian approach to regression tasks. We have also seen how sparse approximate methods are used to reduce the $\mathcal{O}(N^3)$ cost of exact inference. In particular we introduce a class of interdomian inducing points that utilise RKHS projections in order to map a Gaussian process onto a set of orthogonal basis functions. This is beneficial as constructs near diagonal $\mathbf{K_{uu}}$ matrices removing the $\mathcal{O}(NM^2)$ cost of sparse Gaussian process inference.

The successful use of RKHS projections in approximate inference is the motivation for this work. In this project we are interested in beginning to classify when are where we can use interdomain transformations to construct computationally efficient sparse Gaussian processes. In particular we are focused on understanding when it is possible to derive RKHS inner products for chosen Gaussian process kernels, with a focus on trying to derive higher-dimensional RKHS inner products. To do this, this project will explore the relationships between Gaussian process, stochastic partial differential equations and the reproducing kernel Hilbert space, building extensively from the preliminary work found in Cunningham [4].

3 Reproducing Kernel Hilbert Space

In this section we will introduce some fundamental theory from functional analysis that will be used throughout this work. Section 3.1, 3.2 and 3.3 will define notions of function spaces, linear operators and the RKHS respectively. In the latter sections we will also begin to make connections between stochastic processes and the RKHS. In particular in Section 3.5 we will introduce Loeve's isometry theorem which defines a one-to-one correspondence between stochastic process and an associated RKHS.

3.1 Vector Spaces and Function Spaces

Underpinning the theory of functional analysis, is the study of topological vector spaces and the linear functions defined on such spaces and their inherited properties.

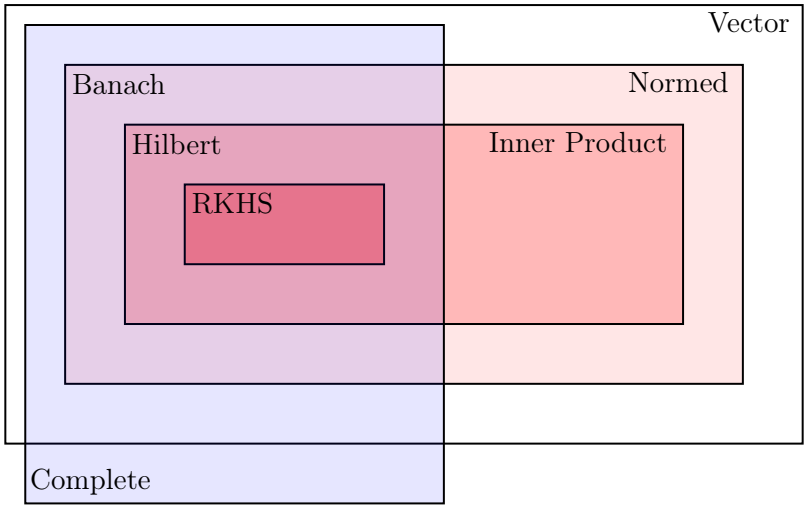

Figure 3.1: Hierarchy of mathematical spaces. The reproducing kernel Hilbert space is a subset of the Hilbert space, where as the Banach space is a superset of the Hilbert space.

Let us first define a vector space as a collection of objects, known as vectors, that abide by the rules of addition and scalar multiplication.

Definition 3.1.1 (Vector Space)**.** *A vector space $\mathcal{V}$ over a field F, is defined as the set of $\mathcal{V}$ together with the following two operations, $\forall \mathbf{x}, \mathbf{y} \in \mathcal{V}$:*

1. *Addition, $(\mathbf{x}, \mathbf{y}) \to \mathbf{x} + \mathbf{y}$*
2. *Scalar multiplication, $(\alpha, \mathbf{x}) \to \alpha\mathbf{x}$, where α is a scalar*

such that the following conditions hold:

- *Addition is associative and commutative*

$$(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z}), \tag{3.1}$$

$$\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}. \tag{3.2}$$

- *Scalar multiplication is associative, commutative, distributive and satisfies the identity property*

$$\alpha(\beta\mathbf{x}) = (\alpha\beta)\mathbf{x}, \tag{3.3}$$
$$\alpha\mathbf{x} = \mathbf{x}\alpha, \tag{3.4}$$
$$\alpha(\mathbf{x}+\mathbf{y}) = \alpha\mathbf{x} + \alpha\mathbf{y}, \tag{3.5}$$
$$(\alpha+\beta)\mathbf{x} = \alpha\mathbf{x} + \beta\mathbf{x}, \tag{3.6}$$
$$1\mathbf{x} = \mathbf{x}. \tag{3.7}$$

- *There exists a unique element* 0 *such that,*

$$\mathbf{x} + 0 = \mathbf{x}, \tag{3.8}$$
$$\mathbf{x} + (-\mathbf{x}) = 0, \tag{3.9}$$
$$0\mathbf{x} = 0. \tag{3.10}$$

Remark 3.1.1. In this work we assume the field $F = \mathbb{R}$

Further, we can define a vector space as *complete* if every Cauchy sequence of it's elements, whose terms become arbitrarily close together, converge in the limit to a point within the vector space. This property is not guaranteed for all linear vector spaces.

Definition 3.1.2 (Complete Space)**.** *A vector space $\mathcal{V}$ is complete if every Cauchy sequence $\{v_n\}_{n=1}^{\infty} \in \mathcal{V}$ converges such that it has a limit and this limit is in $\mathcal{V}$*

Remark 3.1.2. All square-integrable functions $f \in \mathcal{V}$ for which,

$$\int_{-\infty}^{\infty} \|f(\mathbf{x})\|^2 d\mathbf{x} < \infty, \tag{3.11}$$

form a complete vector space, in which all cauchy limits exist within $\mathcal{V}$

Given a vector space we can also define additional structure on it. A norm is a real valued function $\|\cdot\|$ that maps from a vector space $\mathcal{V}$ to the positive real numbers $\mathbb{R}^+$, evaluating the size of vectors in $\mathcal{V}$ and hence defining a notion of distance.

Definition 3.1.3 (Norm)**.** *Let $\mathcal{V}$ be a vector space. A function $\|\cdot\| : \mathcal{V} \to \mathbb{R}$ that maps from a vector space $\mathcal{V}$ into $\mathbb{R}^+$ is called a norm if $\forall \mathbf{x}, \mathbf{y} \in \mathcal{V}$ the following conditions hold,*

1. *$\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = 0$,*
2. *$\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$ for $\alpha \in \mathbb{F}$,*
3. *$\|\mathbf{x}+\mathbf{y}\| \leq \|x\| + \|y\|$ (known as the triangle inequality).*

Hence, a vector space equipped with a norm is a *normed vector space.* If the normed vector space is complete it is denoted a *Banach space* $\mathcal{B}$.

Definition 3.1.4 (Banach Space)**.** *A Banach space $\mathcal{B}$ is a complete vector space equipped with a norm $\|\cdot\| : \mathcal{B} \to \mathbb{R}$*

We can also define an inner product on a vector space $\mathcal{V}$. An inner product is as a function $\langle\cdot,\cdot\rangle$ that maps a pair of vectors to the positive real numbers $\mathbb{R}^+$, introducing further geometrical notions, such as distance and angle, on $\mathcal{V}$.

Definition 3.1.5 (Inner Product)**.** *Let $\mathcal{V}$ be a vector space. A mapping $\langle\cdot,\cdot\rangle : \mathcal{V} \times \mathcal{V} \to \mathbb{R}$ is called an inner product in $\mathcal{V}$ if $\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{V}$ and $\forall \alpha, \beta \in \mathbb{F}$, the following conditions hold,*

1. *The inner product is symmetric,*

$$\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle, \tag{3.12}$$

2. *The inner product is linear,*

$$\langle \alpha\mathbf{x} + \beta\mathbf{y}, \mathbf{z} \rangle = \alpha\langle \mathbf{x}, \mathbf{z} \rangle + \beta\langle \mathbf{y}, \mathbf{z} \rangle, \tag{3.13}$$

3. *The inner product of an element with itself is positive definite,*

$$\langle \mathbf{x}, \mathbf{x} \rangle \geq 0 \text{ and } \langle \mathbf{x}, \mathbf{x} \rangle = 0 \text{ if and only if } \mathbf{x} = 0. \tag{3.14}$$

Hence, a vector space equipped with an inner product is an *inner product space*. Further, every inner product gives rise to a norm $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$ and therefore every inner product space is a normed vector space. The converse is not however true, as not all norms can be expressed as an inner product, for example every L^p norm apart from the L^2 norm.

If an inner product space is complete it is known as a *Hilbert space* $\mathcal{H}$.

Definition 3.1.6 (Hilbert Space)**.** *A Hilbert space $\mathcal{H}$ is a complete vector space equipped with an inner product $\langle \cdot, \cdot \rangle : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$*

We can also define a Hilbert space of functions. If a function space satisfies the same axioms as a vector space, such as addition and multiplication, then a function space is also a vector space. Indeed, the addition of functions can be naturally thought of in terms of a vector space, as simply the point-wise addition of function values.

Definition 3.1.7 (Function Space)**.** *Let $\mathcal{X}$ be a non-empty set and $\mathcal{V}$ a vector space. $\mathcal{F} : \mathcal{X} \to \mathcal{V}$ is the space of all functions that maps from $\mathcal{X}$ into $\mathcal{V}$. Hence, $\mathcal{F}$ is a vector space if it supports addition and scalar multiplication, such that $\forall f, g \in \mathcal{F}$ and $\mathbf{x} \in \mathcal{X}$,*

$$(f + g)(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x}), \tag{3.15}$$

$$(\alpha f)(\mathbf{x}) = \alpha f(\mathbf{x}). \tag{3.16}$$

As a result, for a Hilbert space of functions $\mathcal{H}$ defined on some domain $\mathcal{D}$, we can define an explicit inner product of functions,

$$\langle f, g \rangle_{\mathcal{H}} = \int_{\mathcal{D}} f(\mathbf{x}) g(\mathbf{x}) d\mu(\mathbf{x}), \tag{3.17}$$

where $\mathbf{x} \in \mathcal{D}$, $f, g \in \mathcal{H}$ and $d\mu(\mathbf{x})$ is some measure of integration.

In the next section we will consider the existence and application of linear operators and their relationships to vector and functions spaces.

3.2 Linear Operators

A linear operator $\mathcal{L} : V \to V'$ is a mapping from elements in a vector space V to elements in another vector space V', preserving the operations of vector addition and scalar multiplication.

Definition 3.2.1 (Linear Operator)**.** *Let V and V' be vector spaces. An operator $\mathcal{L} : V \to V'$ is a linear operator if, and only if, $\forall \mathbf{x}, \mathbf{y} \in V$ and scalars α, β,*

$$\mathcal{L}(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha\mathcal{L}\mathbf{x} + \beta\mathcal{L}\mathbf{y}. \tag{3.18}$$

In almost every setting we consider in this project the operators are considered *bounded*.

Definition 3.2.2 (Bounded Linear Operator)**.** *Let $\mathcal{F}$ and $\mathcal{F}'$ be normed vector spaces with norms $\|\cdot\|_{\mathcal{F}}$ and $\|\cdot\|_{\mathcal{F}'}$ respectively. A linear operator $\mathcal{L}:\mathcal{F}\to\mathcal{F}'$ is bounded if there exists a constant α such that $\forall\mathbf{x}\in\mathcal{F}$,*

$$\|\mathcal{L}\mathbf{x}\|_{\mathcal{F}'}\leq\alpha\|\mathbf{x}\|_{\mathcal{F}}. \tag{3.19}$$

Remark 3.2.1. Note that a bounded linear operator $\mathcal{L}:\mathcal{F}\to\mathcal{F}'$ is *continuous* everywhere in $\mathcal{F}$ and therefore all *continuous* operators are bounded operators [23]

Bounded linear operators can also be defined on a Hilbert space $\mathcal{H}$. In order to do so we must first set out the topological dual of a normed vector space.

Definition 3.2.3 (Topological Dual)**.** *Let $\mathcal{F}$ be a normed vector space. The space $\mathcal{F}^*$ of continuous linear functionals $\mathcal{L}:\mathcal{F}\to\mathbb{R}$ is called the topological dual of $\mathcal{F}$.*

When the vector space $\mathcal{F}$ is a Hilbert space $\mathcal{H}$, according to the *Riesz representation theorem*, every continuous linear operator $\mathcal{L}\in\mathcal{H}^*$ has a representation $f_{\mathcal{L}}\in\mathcal{H}$, called the Riesz representation of $\mathcal{L}$.

Theorem 3.2.1 (Riesz Representation Theorem)**.** *Let $\mathcal{H}$ be a Hilbert space and $\mathcal{H}^*$ it's topological dual. For every bounded linear operator $\mathcal{L}\in H^*$, there exists a unique $f_{\mathcal{L}}\in\mathcal{H}$, such that $\forall g\in\mathcal{H}$,*

$$\mathcal{L}g=\langle f_{\mathcal{L}},g\rangle. \tag{3.20}$$

Proof. See Yosida [24, Theorem III.6] □

A consequence of the Riesz representation theorem is the existence of the *adjoint* (or *Hermitian*) operator of a bounded linear operator defined on a Hilbert space.

Definition 3.2.4 (Adjoint Operator)**.** *Let $\mathcal{H}$ and $\mathcal{H}'$ be Hilbert spaces with inner products $\langle\cdot,\cdot\rangle_{\mathcal{H}}$ and $\langle\cdot,\cdot\rangle_{\mathcal{H}'}$ respectively. Given a bounded linear operator $\mathcal{L}:\mathcal{H}\to\mathcal{H}'$ there exists a unique operator $\mathcal{L}^*:\mathcal{H}'\to\mathcal{H}$, known as the adjoint operator of $\mathcal{L}$ defined $\forall f,g\in\mathcal{H}$ by,*

$$\langle\mathcal{L}f,g\rangle_{\mathcal{H}}=\langle f,\mathcal{L}^*g\rangle_{\mathcal{H}}. \tag{3.21}$$

Remark 3.2.2. We note that if $\mathcal{L}=\mathcal{L}^*$, then $\mathcal{L}$ is a self-adjoint operator.

We can prove the existence and uniqueness of the adjoint of a bounded linear operator using the Riesz representation theorem.

Proof. For $\mathcal{L}^*$ to exist, $\forall f,g\in\mathcal{H}$ there is a bounded linear functional $h\in\mathcal{H}$ such that,

$$\langle\mathcal{L}f,g\rangle_{\mathcal{H}}=\langle f,h\rangle. \tag{3.22}$$

For a fixed g, let us define the linear operator $\phi_g\in\mathcal{H}^*$ as,

$$\phi_g(f)=\langle\mathcal{L}f,g\rangle. \tag{3.23}$$

According to the Riesz representation theorem there exists a unique $h\in\mathcal{H}$ such that,

$$\phi_g(f)=\langle h,f\rangle. \tag{3.24}$$

Hence, the Riesz representation theorem satisfies Equation (3.22) where $h=\mathcal{L}^*g$. Uniqueness follows from the definition of the Riesz representation theorem. □

A class of linear operator that we are particularly interested in are linear elliptic differential operators, due to their connections with Green's functions and variational formulation of boundary value problems.

Definition 3.2.5 (Linear Elliptic Differential Operator)**.** *Let $\mathcal{L}$ be a second order linear operator, on some domain $\mathcal{D} \in \mathbb{R}^n$, be defined as,*

$$\mathcal{L} = \sum_{i,j=1}^{n} a_i j \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^{n} b_i \frac{\partial}{\partial x_i} + c, \tag{3.25}$$

where the coefficients are measurable functions on $\mathcal{D}$. $\mathcal{L}$ is elliptic if the coefficient matrix $A = [a_{ij}]$ is positive definite for all $x \in \mathcal{D}$

3.3 Reproducing Kernel Hilbert Space

A reproducing kernel Hilbert space (RKHS) further extends the notion of a Hilbert space of functions. The RKHS states that the point evaluation of functions belonging to the Hilbert space $\mathcal{H}$, by a positive-definite *reproducing kernel* must be a continuous mapping, such that $k(\cdot, \mathbf{x}) : f \to f(\mathbf{x})$. This gives the property that for an RKHS defined on a non-empty set $\mathcal{X}$, if two functions $f \in \mathcal{H}$ and $g \in \mathcal{H}$ are close together in terms of the norm of $\mathcal{H}$, then their point wise values $f(\mathbf{x})$ and $g(\mathbf{x})$ are also close together $\forall x \in \mathcal{X}$. That is if $\|f - g\|_{\mathcal{H}}$ for $f, g \in \mathcal{H}$ is small, then so to is $\|f(\mathbf{x}) - g(\mathbf{x})\|_{L^2}$, $\forall \mathbf{x} \in \mathcal{X}$.

Let use now build a mathematical description of the RKHS. Since we are considering a Hilbert space of functions on $\mathcal{X}$, there exits for all $\mathbf{x} \in \mathcal{X}$ a unique functional on $\mathcal{H}$ that evaluates each $f \in \mathcal{H}$ at $\mathbf{x}$, known as the evaluation functional,

Definition 3.3.1 (Evaluation Functional)**.** *Let $\mathcal{H}$ be a Hilbert space of functions defined on some non-empty set $\mathcal{X}$. For a fixed $\mathbf{x} \in \mathcal{X}$, the map $\delta_{\mathbf{x}} : \mathcal{H} \to \mathbb{R}$ is given by the evaluation functional $\delta_{\mathbf{x}} : f \mapsto f(\mathbf{x})$ at $\mathbf{x}$.*

Evaluation functionals must also be linear, such that for $f, g \in \mathcal{H}$ and $\alpha, \beta \in \mathbb{R}$, $\delta_{\mathbf{x}}(\alpha f + \beta g) = \alpha f(\mathbf{x}) + \beta g(\mathbf{x})$, hence $\delta_{\mathbf{x}}$ is a linear operator. To construct an RKHS we are interested in the case where the evaluation functional is a continuous, and therefore bounded, linear positive-definite *reproducing kernel* k, such that $\delta_{\mathbf{x}} = k(\cdot, \mathbf{x})$. Consequently, we can define the RKHS as a Hilbert space of functions whose evaluation functional is given by an associated *reproducing kernel.*

Definition 3.3.2 (Reproducing Kernel Hilbert Space)**.** *Let $\mathcal{X}$ be a non-empty set and $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ a positive definite kernel on $\mathcal{X}$. A Hilbert space $\mathcal{H}$ of functions on $\mathcal{X}$ equipped with an inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}_k} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a reproducing kernel Hilbert space $\mathcal{H}_k$ with positive definite reproducing kernel k, if the following conditions hold,*

1. *$\forall \mathbf{x} \in \mathcal{X}$, there exists $k(\cdot, \mathbf{x}) \in \mathcal{H}_k$,*
2. *$\forall \mathbf{x}, \mathbf{y} \in \mathcal{X}$ and for all $f \in \mathcal{H}_k$,*

$$f(\mathbf{x}) = \langle f, k(\cdot, \mathbf{x}) \rangle_{\mathcal{H}_k}, \tag{3.26}$$
$$k(\mathbf{x}, \mathbf{y}) = \langle k(\cdot, \mathbf{x}), k(\cdot, \mathbf{y}) \rangle_{\mathcal{H}_k}, \tag{3.27}$$

*known as the **reproducing property***

Given that the reproducing kernel is a bounded linear operator, we can use the Riesz representation theorem to establish existence and uniqueness of the reproducing kernel $k(\cdot, \mathbf{x})$ associated with an RKHS $\mathcal{H}_k$.

Proposition 3.3.1 (Existence of the reproducing kernel)**.** *$\mathcal{H}_k$ is a reproducing kernel Hilbert space, if and only if $\mathcal{H}_k$ has a reproducing kernel.*

Proof. Let $\mathcal{H}$ be a Hilbert space defined on a non-empty set $\mathcal{X}$ with a reproducing kernel $k(\cdot, \mathbf{x})$ that satisfies the reproducing property $\langle f, k(\cdot, \mathbf{x})\rangle_{\mathcal{H}} = f(\mathbf{x})$. In order to apply the Riesz representation theorem we must first prove that the evaluation functional is a bounded linear operator $\forall \mathbf{x} \in \mathcal{X}$,

$$
\begin{aligned}
|\delta_{\mathbf{x}}| &= |f(\mathbf{x})|, && (3.28)\\
&= |\langle f, k(\cdot, \mathbf{x})\rangle_{\mathcal{H}_k}, && (3.29)\\
&\leq \|k(\cdot, \mathbf{x})\|_{\mathcal{H}_k} \|f\|_{\mathcal{H}_k} \quad \text{(Cauchy-Schwarz Inequality)}, && (3.30)\\
&= \langle k(\cdot, \mathbf{x}), k(\cdot, \mathbf{x})\rangle_{\mathcal{H}_k}^{1/2} \|f\|_{\mathcal{H}_k}, && (3.31)\\
&= k(\mathbf{x}, \mathbf{x})^{1/2} \|f\|_{\mathcal{H}_k}. && (3.32)
\end{aligned}
$$

Consequently, f($\mathbf{x}$) is a continuous linear functional for all $\mathbf{x} \in \mathcal{X}$. Applying the Riesz representation theorem, there exists a functional $f_{\delta_{\mathbf{x}}} \in \mathcal{H}$ such that,

$$\delta_{\mathbf{x}} f = \langle f_{\delta_{\mathbf{x}}}, f\rangle_{\mathcal{H}}, \tag{3.33}$$

where by we define $f_{\delta_{\mathbf{x}}} = k(\cdot, \mathbf{x}) \in \mathcal{H}$ and the reproducing property is given as $\langle f, k(\cdot, \mathbf{x})\rangle_{\mathcal{H}} = \delta_{\mathbf{x}} f = f(\mathbf{x})$. Thus $k(\cdot, \mathbf{x})$ is the reproducing kernel of $\mathcal{H}$. □

Proposition 3.3.2 (Uniqueness of the reproducing kernel)**.** *If it exists, the reproducing kernel $k(\cdot, \mathbf{x})$ is unique.*

Proof. Assume $\mathcal{H}$ has two reproducing kernels k_1 and k_2. Hence,

$$
\begin{aligned}
\langle f, k_1(\cdot, \mathbf{x}) - k_2(\cdot, \mathbf{x})\rangle_{\mathcal{H}_k}, &= \langle f, k_1(\cdot, \mathbf{x})\rangle_{\mathcal{H}_k} - \langle f, k_2(\cdot, \mathbf{x})\rangle_{\mathcal{H}_k}, && (3.34)\\
&= f(\mathbf{x}) - f(\mathbf{x}), && (3.35)\\
&= 0. && (3.36)
\end{aligned}
$$

Hence, each RKHS has a uniquely defined reproducing kernel □

Having shown existence and uniqueness of a reproducing kernel belonging to an RKHS, the Moore-Aronszajn theorem further states that every positive definite function is a reproducing kernel that uniquely defines an associated RKHS.

Theorem 3.3.1 (Moore-Aronszajn Theorem)**.** *Let $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be positive definite. There exists a unique RKHS $\mathcal{H}_k$ defined on $\mathcal{X}$ with reproducing kernel $k(\cdot, \mathbf{x})$ and inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}_k}$ defined by,*

$$\langle f, g\rangle_{\mathcal{H}_k} = \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \beta_j k(\mathbf{x}_i, \mathbf{x}_j), \tag{3.37}$$

where,

$$f(\cdot) = \sum_{i=1}^{n} \alpha_i k(\cdot, \mathbf{x}_i), \quad g(\cdot) = \sum_{j=1}^{m} \beta_j k(\cdot, \mathbf{x}_j). \tag{3.38}$$

Proof. See Aronszajn [3] □

Remark 3.3.1. This inner product satisfies the *reproducing property* for the associated reproducing kernel $k(\cdot, \mathbf{x})$,

$$\langle f, k(\cdot, \mathbf{x})\rangle_{\mathcal{H}_k} = \sum_{i=1}^{n} \alpha_i k(\mathbf{x}, \mathbf{x}_i) = f(\mathbf{x}). \tag{3.39}$$

The Moore-Aronszajn theorem is remarkable in how it connects all positive definite functions via a one-to-one relationship to an RKHS, such that for each kernel k there is a uniquely associated RKHS $\mathcal{H}_k$. The Moore Aronszajn theorem further states that functions $f \in \mathcal{H}_k$ can be constructed as a linear combination of the kernel function (3.38) [25]. Consequently, it is obvious that functions $f \in \mathcal{H}_k$ inherit properties of the reproducing kernel k. As an example, if a kernel k is n-times differentiable for $n \in \mathbb{N}$, then the functions $f \in \mathcal{H}_k$ are also n-times differentiable [Steinwart and Christmann [26], Corollary 4.36].

Having now defined Gaussian processes, linear operators and the RKHS, the next few sections will draw on connections between Gaussian processes and the RKHS. To do so we will first introduce some ideas from the spectral theory of compact linear operators on a Hilbert space, and how we can represent both stochastic processes and functions in the RKHS by a series expansion of an orthonormal eigenbasis.

3.4 Spectral Representation Theory

Spectral theory connects the theory of linear operators on a Hilbert space of functions to their orthonormal decomposition in terms of their respective eigenfunctions and eigenvalues. In this section we explore the important overlap between representations of stochastic process and the RKHS representations via the spectral decomposition of a kernel integral operator into it's eigenfunctions and eigenvalues. This section therefore begins a development of the relationships between stochastic processes and the RKHS.

Let us first consider the representation of functions in the RKHS via a spectral decomposition. To do so, we must define a compact kernel integral operator $\mathcal{K}$ on a Hilbert space $\mathcal{H}$, otherwise known as a Hilbert-Schmidt integral operator or *covariance operator*.

Definition 3.4.1 (Kernel Integral Operator)**.** *Let $\mathcal{X}$ be some non-empty set with a finite Borel measure μ on $\mathcal{X}$, $\mathcal{H}$ a Hilbert space with L^2 inner product on $\mathcal{X}$ with respect to the Borel measure μ and $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ a positive definite kernel. The integral operator $\mathcal{K} : \mathcal{H} \to \mathcal{H}$, is defined $\forall f \in \mathcal{H}$ by,*

$$\mathcal{K}f(\cdot) = \int_{\mathcal{X}} k(\cdot, \mathbf{x}) f(\mathbf{x}) d\mu(\mathbf{x}), \tag{3.40}$$

is continuous, compact and self-adjoint.

According to the spectral theorem, we can decompose a Hilbert-Schmidt integral operator into an orthonormal system of its eigenfunctions and eigenvalues. Indeed, the spectral theorem for compact self-adjoint operators guarantees the existence of an at most countable eigensystem $\{\phi_i, \lambda_i\}_{i \in I}$ of eigenfunctions $\{\phi\}_{i \in I}$ and eigenvalues $\{\lambda_i\}_{i \in I}$ converging to 0, such that $|\lambda_1| > |\lambda_2| > ... > 0$, such that $\mathcal{K}f(\cdot)$ is represented $\forall f \in \mathcal{H}$ by,

$$\mathcal{K}f(\cdot) = \sum_{i \in I} \lambda_i \langle \phi_i, f(\cdot) \rangle_{\mathcal{H}} \phi_i. \tag{3.41}$$

Remark 3.4.1. We note that the eigensystem $\{\phi_i, \lambda_i\}_{i \in I}$ constitutes a complete set of solutions to the Fredholm-type integral equation,

$$\mathcal{K}\phi_i(\cdot) = \int_{\mathcal{X}} k(\cdot, \mathbf{x}) \phi_i(\mathbf{x}) d\mu(\mathbf{x}) = \lambda_i \phi_i(\mathbf{x}), \tag{3.42}$$

forming an orthonormal system in $\mathcal{H}$ such that,

$$\langle \phi_i, \phi_j \rangle_{\mathcal{H}} = \int_{\mathcal{X}} \phi_i(\mathbf{x}) \phi_j(\mathbf{x}) d\mu(\mathbf{x}) = \delta_{ij}, \tag{3.43}$$

where δ_{ij} is the Kronecker delta such that $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ otherwise.

Mercer's theorem states that a positive-definite kernel $k \in \mathcal{H}$ can be represented in terms of the eigensystem $\{\phi_i, \lambda_i\}_{i \in I}$ defined by the Spectral theorem (3.41).

Theorem 3.4.1 (Mercer's Theorem)**.** *Let $\mathcal{X}$ be some non-empty set with a finite Borel measure μ on $\mathcal{X}$, $\mathcal{H}$ a Hilbert space with L^2 inner product on $\mathcal{X}$ with respect to the Borel measure μ, $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ a positive definite kernel and $\mathcal{K} : \mathcal{H} \to \mathcal{H}$ a compact self-adjoint integral operator. Let $\{\phi_i, \lambda_i\}$ be the eigen-system of $\mathcal{K}$, then the kernel $k \in \mathcal{H}$ can be represented $\forall \mathbf{x} \in \mathcal{X}$ and $\forall \phi_i \in \mathcal{H}$ as,*

$$k(\cdot, \mathbf{x}) = \sum_{i \in I} \lambda_i \phi_i(\cdot) \phi_i(\mathbf{x}), \tag{3.44}$$

where the convergence is absolute and uniform.

Remark 3.4.2. Mercer's expansion is dependent on the choice of measure μ, since the Fredholm-type integral Equation (3.42) is dependent on the Hilbert-Schmidt integral operator $\mathcal{K}$ which is defined by the choice of Borel measure μ. Consequently, different choices of measure μ result in a different eigensystem expansion of the same kernel k.

We can also use the spectral theorem to define a series representation of functions f in the RKHS $\mathcal{H}_k$, known as Mercer's representation.

Definition 3.4.2 (Mercer's Representation)**.** *Let $\mathcal{X}$ be some non-empty set with a finite Borel measure μ on $\mathcal{X}$, $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ a positive definite kernel, $\mathcal{H}_k$ a RKHS with reproducing kernel k, and $\{\phi_i, \lambda_i\}$ the eigensystem as given by the Spectral theorem (3.41). We can represent functions $f \in \mathcal{H}_k$ by the expansion,*

$$f = \sum_{i \in I} \alpha_i \sqrt{\lambda_i} \phi_i, \tag{3.45}$$

where $\{\sqrt{\lambda_i}\phi_i\}_{i \in I}$ forms an orthonormal basis of $\mathcal{H}_k$.

Remark 3.4.3. Letting $g = \sum_{i \in I} \beta_i \sqrt{\lambda_i} \phi_i$, the RKHS inner product corresponding to Mercer's representation theorem is given by,

$$\langle f, g \rangle_{\mathcal{H}_k} = \sum_{i \in I} \frac{\alpha_i \beta_i}{\lambda_i}, \tag{3.46}$$

such that the reproducing property holds,

$$\langle f(\cdot), k(\cdot, \mathbf{x}) \rangle_{\mathcal{H}_k} = \left\langle \sum_{i \in I} \alpha_i \sqrt{\lambda_i} \phi_i(\cdot), \sum_{i \in I} \lambda_i \phi_i(\cdot) \phi_i(\mathbf{x}) \right\rangle = \sum_{i \in I} \frac{\alpha_i \sqrt{\lambda_i} \lambda_i \phi_i(\mathbf{x})}{\lambda_i} = f(\mathbf{x}). \tag{3.47}$$

Remark 3.4.4. Like Mercer's theorem, the eigensystem is dependent on the choice of measure μ. Therefore, different choices of μ result in different eigensystems $\{\phi_i, \lambda_i\}_{i \in I}$ of the same RKHS $\mathcal{H}_k$.

Similar to the Mercer representation of functions in the RKHS, the Karhunen-Loeve expansion represents a stochastic process $\{X(t), t \in T\}$ as a series of independent random variables $\{\xi_i\}_{i \in N}$ and an orthogonal basis $\{\sqrt{\lambda_i}\phi_i\}_{i \in N}$

Theorem 3.4.2 (Karhunen-Loeve Theorem)**.** *Let T be some non-empty set with a finite Borel measure μ on T, $k : T \times T$ a positive-definite kernel and $\{\phi_i, \lambda_i\}$ the eigensystem given as before by Equation (3.41). Let $\{X(t), t \in T\}$ be a real-valued zero-mean second-order stochastic process with covariance function $k(t, t') = \mathbb{E}[X(t)X(t')]$, then $X(t)$ can be represented $\forall t \in T$ and $\forall \phi_i \in \mathcal{H}$ as,*

$$X(t) = \sum_{i \in I} \xi_i \sqrt{\lambda_i} \phi_i(t), \tag{3.48}$$

where $\{\xi_i\}_{i\in I}$ *are a family of independent random variables given as,*

$$\xi_i = \frac{1}{\sqrt{\lambda_i}} \int_T X(t)\phi_i(t)dt, \tag{3.49}$$

such that $\mathbb{E}[\xi_i] = 0$ *and* $\mathbb{E}[\xi_i\xi_j] = \delta_{ij}$

Remark 3.4.5. For a zero-mean Gaussian process prior, $\xi_i \sim \mathcal{N}(0, \lambda_i)$ and are independent from one another. Consequently, the variance of the Gaussian process $X(t)$ is the sum of the eigenvalues. This lends to the use of a *truncated* KL expansion, where we can approximate the process using the first N-modes of the eigendecomposition. It is important to note though that the random variables $\{\xi_i\}_{i\in I}$ are dependent on the Gaussian process prior as seen in Equation (3.49).

Consequently, we have seen in this section how we can represent both functions in an RKHS $\mathcal{H}_k$ and stochastic processes $\{X(t), t \in T\}$ via an eigendecomposition related to a compact self-adjoint integral operator. Indeed, the kernel of such an operator is both the reproducing kernel of $\mathcal{H}_k$ and the covariance function of the $\{X(t), t \in T\}$. We will explore this intimate relationship between stochastic processes and the RKHS in greater detail in the next section. Specifically we will explore the isometry between Hilbert spaces generated by stochastic processes and a corresponding RKHS, focusing on the sample path properties of stochastic processes in the RKHS.

3.5 Gaussian Processes and the RKHS

In this section we would like to develop the relationship between zero-mean Gaussian process priors and the RKHS. In particular we are interested in the sample paths $f \sim \mathcal{GP}(0, k)$ of Gaussian processes and their properties in the RKHS with corresponding reproducing kernel k.

To begin let us further investigate the relationship between stochastic processes and the RKHS. Let $\{X(t), t \in T\}$ be a real-valued zero-mean second-order stochastic process. The covariance function $k : T \times T \to \mathbb{R}$ is defined as the positive definite function $k(t, t') = \mathbb{E}[X(t)X(t')]$. Hence, there exists on T a RKHS $\mathcal{H}_k$ for which k is the reproducing kernel. We can also construct a Hilbert space $\mathcal{H}$ as the closure in L^2 of the space *spanned* by all $\{X(t)\}_{t\in T}$ [27, 28]. Then, $\mathcal{H}$ consists of all random variables expressed as a linear combination of $\{X(t)\}_{t\in T}$ of the form,

$$U = \sum_i a_i X(t_i) \quad \text{for } t_i \in T, \tag{3.50}$$

where $\{a_i\} \in \mathbb{R}$ with L^2 inner product,

$$\langle U, U' \rangle_{L^2} = \mathbb{E}[UU']. \tag{3.51}$$

According to the Loeve isometry theorem, the Hilbert space $\mathcal{H}$ of a second order stochastic process $\{X(t), t \in T\}$ can be represented by the RKHS $\mathcal{H}_k$ whose reproducing kernel k is equal to the covariance of the process $\mathbb{E}[X(t)X(t')]$. See Berlinet and Thomas-Agnan [29, Theorem 35].

Theorem 3.5.1 (Loeve Isometry)**.** *Let* $\{X(t), t \in T\}$ *be a real-valued zero-mean second-order stochastic process with covariance function* $k(t, t') = \mathbb{E}[X(t)X(t')]$*. The Hilbert space* $L^2(X)$ *defined by the* L^2 *closure of the linear space spanned by* $\{X(t)\}_{t\in T}$*, is isometrically isomorphic to the RKHS* $\mathcal{H}_k$ *with reproducing kernel* k*.*

Proof. Define the map $\psi : L^2(X) \to \mathcal{H}_k$ as $\psi(X(t)) = k(\cdot, t)$. Hence $\forall X(t), X(t') \in L^2(X)$,

$$\langle X(t), X(t') \rangle_{L^2(X)} = \langle \psi(X(t)), \psi(X(t')) \rangle_{\mathcal{H}_k} = \langle k(\cdot, t), k(\cdot, t') \rangle_{\mathcal{H}_k}, \tag{3.52}$$

and therefore the Hilbert space $L^2(X)$ and RKHS $\mathcal{H}_k$ are isometrically isomorphic. □

Hence, we can write the one-to-one correspondence between the Hilbert space $L^2(X)$ and the RKHS $\mathcal{H}_k$ as,

$$\langle X(t), X(t')\rangle_{L^2(X)} = \mathbb{E}[X(t)X(t')] = k(t,t') = \langle k(\cdot,t), k(\cdot,t')\rangle_{\mathcal{H}_k}. \tag{3.53}$$

Let us now consider an orthonormal basis expansion of $f \in \mathcal{H}_k$. Let $\{e_i\}_{i\in I}$ be an orthonormal basis of $\mathcal{H}_k$, then $\forall f \in \mathcal{H}_k$,

$$f(t) = \sum_{i\in I} \langle f(t), e_i\rangle e_i = \sum_{i\in I} e_i(t) e_i. \tag{3.54}$$

Using Loeve's Isometry theorem, we can represent the stochastic process $X(t) \in L^2(X)$ via the orthonormal expansion,

$$X(t) = \psi^{-1}(f(t)) = \sum_{i\in I} e_i(t)\psi^{-1}(e_i) = \sum_{i\in I} e_i(t) X_i, \tag{3.55}$$

where $\{X_i\}_{i\in I}$ is a family of independent random variables $X_i = \psi^{-1}(e_i)$. If we let each finite dimensional marginal distribution $X(t), t \in T$ be normally distributed, such that the process is a Gaussian process, then $\{X_i\}_{i\in I}$ is a family of independent zero-mean Gaussian random variables [Janson et al. [30], Theorem 8.22]. As a result, $X(t), t \in T$ can be represented by the orthonormal basis expansion,

$$X(t) = \sum_{i=1}^{N} X_i e_i(t), \quad X_i \sim \mathcal{N}(0, \lambda_i). \tag{3.56}$$

Remark 3.5.1. Note here that if we let our orthonormal basis be $\{\sqrt{\lambda_i}\phi_i\}_{i\in I}$ derived from the spectral theorem, then we recover the Karhunen-Loeve expansion (3.48).

We would now like to highlight another important relationship between stochastic processes and the RKHS. Namely, that sample paths of $X(t), t \in T$ do not belong in $\mathcal{H}_k$ with probability 1. To do so let us consider sample paths $f \sim \mathcal{GP}(0, k(\cdot,\cdot))$ from a Gaussian process, which we have represented by the orthonormal basis function expansion in Equation (3.56). Sampling the coefficients X_i from the expansion, we can write the expectation of the norm squared as,

$$\mathbb{E}[\|X(\cdot)\|^2_{\mathcal{H}}] = \sum_{i=1}^{N} \frac{\mathbb{E}[X_i^2]}{\lambda_i} = N \to \infty \text{ as } N \to \infty. \tag{3.57}$$

Therefore, the expansion does not converge and hence $f \notin \mathcal{H}_k$ with probability 1 as the expected value of the RKHS norm is infinite. Note that this is not a proof as convergence of the KL expansion is weaker than that of the RKHS norm [25]. Hence we point the reader to Steinwart and Christmann [26] for a more detailed discussion on convergence and Driscoll [31] for conditions for Gaussian process samples paths $f \sim \mathcal{GP}(0,k)$ to belong to a RKHS $\mathcal{H}_r$ with probability 1 or 0. This does however stress the important point that $f \in \mathcal{H}_k$ and $f \sim \mathcal{GP}(0,k)$ are *not* the same.

3.5.1 Dual Representation of Gaussian Processes in the RKHS

However, whilst sample paths f don't belong in the RKHS with probability 1, the *posterior mean function* of a Gaussian process conditioned on some training data does.

We can show this by constructing a dual representation of a Gaussian process posterior $\mathcal{GP}(\mu(\cdot), k(\cdot,\cdot))$ in the RKHS by treating $\mu(\cdot)$ as an object in the RKHS [32, 33]

Definition 3.5.1 (Dual Representation of a Gaussian Process)**.** *Let $\mathcal{GP}(0, k)$ be a zero-mean Gaussian process prior defined on the non-empty set $\mathcal{X}$, with positive-definite kernel $k(\cdot, \cdot) : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ and let $\mathcal{H}_k$ be the RKHS defined on $\mathcal{X}$ with reproducing kernel $k(\cdot, \cdot) = \langle \phi(\cdot), \phi(\cdot) \rangle_{\mathcal{H}_k}$. A Gaussian process posterior $\mathcal{GP}(\hat{\mu}, \hat{k})$ conditioned on some data, has a dual formulation in the RKHS $\mathcal{H}_k$ such that $\forall \mathbf{x}, \mathbf{x}' \in \mathcal{X}$ and $\phi(\cdot) \in \mathcal{H}_k$,*

$$\hat{\mu}(\mathbf{x}) = \langle \phi(\mathbf{x}), m \rangle_{\mathcal{H}_k}, \qquad \hat{k}(\mathbf{x}, \mathbf{x}') = \langle \phi(\mathbf{x}), \Sigma \phi(\mathbf{x}') \rangle_{\mathcal{H}_k}, \tag{3.58}$$

where $m \in \mathcal{H}_k$ is the realisation of the posterior mean in $\mathcal{H}_k$, $\Sigma : \mathcal{H} \to \mathcal{H}$ a bounded positive definite self-adjoint operator and $\phi : \mathcal{X} \to \mathcal{H}$ some feature map.

Given the dual formulation we now wish to show that $\hat{\mu}(\mathbf{x})$ does belong to the RKHS $\mathcal{H}_k$.

Proposition 3.5.1. *The Gaussian process posterior mean $\hat{\mu}(\mathbf{x})$ associated with a zero-mean Gaussian process prior $\mathcal{GP}(0, k)$, belongs to the RKHS $\mathcal{H}_k$ with reproducing kernel k*

Proof. Let $\phi(\mathbf{x}) = k(\cdot, \mathbf{x})$ such that the reproducing holds

$$k(\mathbf{x}, \mathbf{x}') = \langle k(\cdot, \mathbf{x}), k(\cdot, \mathbf{x}') \rangle_{\mathcal{H}_k}. \tag{3.59}$$

Hence, the posterior mean $\hat{\mu}(\mathbf{x})$ is given by,

$$\hat{\mu}(\mathbf{x}) = \langle k(\cdot, \mathbf{x}), m \rangle_{\mathcal{H}_k}, \tag{3.60}$$

where m is the representation of $\hat{\mu}$ in the RKHS and by the reproducing property $\hat{\mu}(\cdot) \in \mathcal{H}_k$ □

Remark 3.5.2. Whilst the posterior mean $\hat{\mu}(\mathbf{x})$ belongs to the RKHS $\mathcal{H}_k$ the sample paths $f \sim \mathcal{GP}(\hat{\mu}, \hat{k})$ almost certainly do not.

We note that this connection between Gaussian process posteriors and the RKHS is also present in the Spline literature, which involves the reconstruction of a function from noisy data as a regularised optimisation problem over a RKHS. Indeed, by minimising an RKHS norm, the RKHS minimum variance estimate is equivalent to the posterior mean of a Gaussian process, with covariance proportional to the kernel associated with the RKHS [34, 27, 25].

3.6 Discussion

In this chapter we have introduced some theory from functional analysis that we will use throughout this project. Further we have also made some initial connections between stochastic processes and the RKHS. Of particular note is Loeve's Isometry theorem, which enables us to determine that sample paths from a Gaussian process prior do not belong to its RKHS.

4 Gaussian Processes and Stochastic (Partial) Differential Equations

Previously we noted that a zero-mean Gaussian process prior is a stochastic process defined by its covariance function. Consequently, we would like to generate sample paths of a Gaussian process prior by forcing a dynamical system with Gaussian white noise such that solutions inherit the same properties, such as smoothness and lengthscale, as a chosen covariance function.

In this chapter we detail the construction of stochastic differential equations and stochastic partial differential equations corresponding to covariance functions that have a rational spectral density function. In particular we explore ways in which we can construct different SDEs and SPDEs corresponding to the same spectral density function. This is a feature of representing covariance functions as differential equations that is largely undiscussed in the literature, however both forms are different and favour themselves for use in different situations.

4.1 Stochastic Partial Differential Equations

So far in this work we have explored the relationship between stochastic processes and the RKHS. In particular we have looked at zero-mean Gaussian processes, stochastic processes defined entirely by a positive-definite covariance function k, and their uniquely associated RKHS with reproducing kernel k. Consequently, our construction of Gaussian processes and the RKHS has been focused on defining kernels.

Instead of defining Gaussian processes via kernels, we can equivalently define them as affine maps of Gaussian white noise processes. Such mappings result in Gaussian processes, whose covariance functions are defined implicitly by the mapping itself. These mappings between processes are known as *stochastic partial differential equations.*

Consequently, in this section we will introduce SPDEs, from two perspectives, from the sample path or process view of solutions and the function space view of solutions.

4.1.1 Process Perspective

Most similar to deterministic partial differential equations, we can define SPDEs by a dynamical system forced by a random noise process,

$$\mathcal{L}f(\mathbf{x}) = w(\mathbf{x}), \tag{4.1}$$

where $\mathcal{L}$ is an mth-order linear differential operator and w a noise process. Typically we denote the left hand side the *deterministic part* and the right hand side, the *stochastic part.*

As stochastic equations, we can classify SPDEs according to the type of random noise, the way in which noise is injected into the system and the type of stochastic integral we use. In this work we are only interested in the case of additive Gaussian white noise, as solutions to linear SPDEs driven by additive Gaussian white noise are Gaussian processes, due to the linearity of Gaussian distributions.

Definition 4.1.1 (Gaussian White Noise)**.** *Let $w(\mathbf{x})$ be a Gaussian white noise process with the following properties,*

1. *The values* $w(\mathbf{x})$ *and* $w(\mathbf{x}')$ *are independent if* $\mathbf{x} \neq \mathbf{x}'$,

2. $\mathbb{E}[w(\mathbf{x})] = 0$,

3. $\mathbb{E}[w(\mathbf{x})w(\mathbf{x}')] = \delta(\mathbf{x} - \mathbf{x}')Q$,

where δ *is a dirac delta function and* Q *is the spectral density of the process.*

Remark 4.1.1. We note that the sample paths $\mathbf{x} \mapsto w(\mathbf{x})$ are discontinuous almost everywhere

Given that Gaussian white noise is a discontinuous process we cannot apply the traditional theory differential equations. Hence, we are required to use stochastic integrals. In this work we will therefore assume the Ito integral, whenever we integrate with respect to a stochastic process, such that we preserve properties similar to deterministic equations, such as the order of the equation.

Further by the linearity of Gaussian distributions, we know that the solution $f(\mathbf{x})$ to the stochastic differential equation will also be a Gaussian process defined as,

$$f(\mathbf{x}) = \mathcal{L}^{-1}w(\mathbf{x}), \tag{4.2}$$

where $\mathcal{L}^{-1}$ is a bounded linear operator representing the inverse of the differential operator $\mathcal{L}$. It is important to note the mean-square differentiability of solutions to SPDEs. Processes derived from an mth order SDE are only $m-1$ times differentiable. This is due to the nature of stochastic integrals where we integrate against a white noise process which is nowhere differentiable.

As partial differential equations, we can classify SPDEs according to the traditional properties of deterministic PDEs such as, the order of the equation, linearity, and the type of initial and boundary conditions applied. In this work we are only interested in linear elliptic SPDEs as defined by Definition 3.2.5, whose deterministic part has the same properties as regular deterministic elliptic PDEs. Discussion of boundary conditions and initial conditions is also the same as deterministic PDEs, except that the conditions can now be specified by random elements.

4.1.2 Function Space Perspective

An alternative representation to the process view is the function space perspective of SPDEs, where instead of mapping between stochastic processes, we define SPDEs by some bounded linear differential operator $\mathcal{L}$ that maps between Hilbert spaces. Hence, we wish to design an operator $\mathcal{L}$, such that it maps from some Hilbert space into an RKHS with a chosen reproducing kernel, corresponding to the Gaussian process solution.

To do this we introduce the idea of a *zero-mean generalised Gaussian field* f, or *generalised Gaussian process prior*, over a Hilbert space $\mathcal{H}$, that maps from $\mathcal{H}$ to the space of random variables.

Definition 4.1.2 (Zero-mean Generalised Gaussian Field)**.** *A zero-mean generalised Gaussian field* f *over a Hilbert space* $\mathcal{H}$ *is a collection of Gaussian random variables* $\{f(g), g \in \mathcal{H}\}$ *with the properties,*

1. $\mathbb{E}[f(g)] = 0 \ \forall g \in \mathcal{H}$

2. *There exists a bounded linear self-adjoint covariance operator* $\mathcal{K}$ *on* $\mathcal{H}$ *such that*

$$\mathbb{E}[f(h)f(g)] = \langle \mathcal{K}h, g\rangle_{\mathcal{H}}, \qquad \forall h, g \in \mathcal{H}. \tag{4.3}$$

Remark 4.1.2. We note that a generalised Gaussian field $\hat{F}$ is generated by a Gaussian process f. Like for Loeve's Isometry theorem, let $\mathcal{H}$ be the closure in L^2 of the space *spanned* by all Gaussian

random variables $\{f(\mathbf{x})\}_{x\in X}$, such that $f(\cdot) \in \mathcal{H}$. Consequently by the Riesz representation theorem we can represent a generalised Gaussian field as a bounded linear operator acting on $\mathcal{H}$,

$$\hat{F}(g) = \langle f, g\rangle_{\mathcal{H}}, \tag{4.4}$$

for all $g \in \mathcal{H}$

Remark 4.1.3. Given that a Gaussian process f generates a Gaussian random field $\hat{F}$, when $\mathcal{H} = L^2(f)$, we can also show that the covariance operator $\mathcal{K}$ of $\hat{F}$ is an integral operator on $\mathcal{H}$ with kernel $k(\mathbf{x}, \mathbf{y}) = \mathbb{E}[f(\mathbf{x})f(\mathbf{y})]$,

$$\begin{aligned}
\langle \mathcal{K}h, g\rangle_{\mathcal{H}} &= \mathbb{E}[\hat{F}(h)\hat{F}(g)], && (4.5)\\
&= \mathbb{E}[\langle f, h\rangle_{\mathcal{H}}\langle f, g\rangle_{\mathcal{H}}], && (4.6)\\
&= \mathbb{E}\left[\int_{\mathcal{X}} f(\mathbf{x})h(\mathbf{x})d\mathbf{x}\int_{\mathcal{X}} f(\mathbf{y})g(\mathbf{y})d\mathbf{y}\right], && (4.7)\\
&= \int_{\mathcal{X}}\left(\int_{\mathcal{X}} \mathbb{E}[f(\mathbf{x})f(\mathbf{y})]h(\mathbf{x})d\mathbf{x}\right) g(\mathbf{y})d\mathbf{y}, && (4.8)\\
&= \int_{\mathcal{X}}\left(\int_{\mathcal{X}} k(\mathbf{x}, \mathbf{y})h(\mathbf{x})d\mathbf{x}\right) g(\mathbf{y})d\mathbf{y}. && (4.9)
\end{aligned}$$

Hence,

$$\mathcal{K}h(\cdot) = \int_{\mathcal{X}} k(\cdot, \mathbf{x})h(\mathbf{x})d\mathbf{x}, \tag{4.10}$$

where $k(\cdot, \mathbf{x})$ is the kernel of a zero-mean Gaussian process f and $\mathcal{H} = L^2(\mathcal{X})$.

If a zero-mean generalised Gaussian field $\hat{W}$ is Gaussian white noise field on $\mathcal{H}$, then the covariance operator $\mathcal{K}$ is identity and we have, $\mathbb{E}[\hat{W}(g)\hat{W}(h)] = \langle f, g\rangle_{\mathcal{H}}$. We let the right hand side of our stochastic differential equation equal a zero-mean generalised Gaussian white noise field. Now let us define the solution of a stochastic partial differential equation using generalised Gaussian fields, such that we map between Hilbert spaces.

Definition 4.1.3 (Solution of an SPDE)**.** *Let $\hat{W}$ be a zero-mean generalised Gaussian field over the Hilbert space $\mathcal{G}$ and $\mathcal{L} : \mathcal{H} \to \mathcal{G}$ a bounded linear operator. The zero-mean generalised Gaussian process $\hat{F}$ over the Hilbert space $\mathcal{H}$ is a solution to the SPDE,*

$$\mathcal{L}\hat{F} = \hat{W}, \tag{4.11}$$

if $\forall g \in \mathcal{G}$,

$$\hat{F}(\mathcal{L}^* g) = \hat{W}(g). \tag{4.12}$$

Remark 4.1.4. From this definition of a solution to an SPDE, if $\mathcal{L}$ is invertible, $\forall h \in \mathcal{H}$,

$$\hat{F}(h) = \hat{W}(\mathcal{L}^{-1^*} h). \tag{4.13}$$

is the unique solution of Equation (4.11)

This form avoids the construction of pathwise solutions. Instead it provides general existence of a Hilbert space $\mathcal{H}$, that the random variables of the Gaussian process f, that generates $\hat{F}$, lie in. By the Loeve Isometry theorem this Hilbert space is isometrically isomoprhic to an RKHS $\mathcal{H}_k$ whose reproducing kernel is given by the kernel of the Gaussian process f. This function space is determined by the transformation due to the inverse differential operator $\mathcal{L}^{-1}$ (This will become clearer in our discussion of Green's functions in Chapter 5).

In this section we have therefore shown how we can define an SPDE as a mapping between Hilbert spaces using generalised Gaussian fields defined on a Hilbert space. We have further shown that a

generalised Gaussian field is generated by a classical Gaussian process. We motivate that we can find function spaces in which the random variables of the Gaussian process, that classically solves the SPDE, lie in. By Loeve's Isometry theorem this Hilert space has a one-to-one correspondence with the RKHS generated by the Gaussian process. Hence we can find the RKHS of our solution Gaussian process without considering pathwise solutions to the SPDE.

Consequently, we wish to construct linear differential operators $\mathcal{L}$, such that we map into an RKHS with a desired reproducing kernel. We note that this is the same linear operator $\mathcal{L}$ that we use in the process perspective, since a Gaussian process with covariance function k, encoded by $\mathcal{L}$, generates an RKHS with reproducing kernel k. The following sections will outline how we move from covariance functions to linear partial differential operators in order to map into desired RKHSs.

4.2 Covariance Functions and Spectra

Having defined what we mean by an SPDE, we would now like to construct linear operators $\mathcal{L}$ that enable us to map to Gaussian processes and function spaces with desired covariance functions. To do this we consider Bochner's theorem.

Bochner's theorem characterises all continuous *stationary* positive definite functions on $\mathbb{R}^d$ as the Fourier transform of a positive finite Borel measure.

Theorem 4.2.1 (Bochner's Theorem)**.** *Let $k(\boldsymbol{\tau})$ be a continuous complex-valued function on $\mathbb{R}^d$. Then k is positive definite covariance function of a weakly stationary mean-square continuous complex-valued stochastic process on $\mathbb{R}^d$, if and only if there is a finite positive Borel measure μ on $\mathbb{R}^d$ such that,*

$$k(\boldsymbol{\tau}) = \int_{\mathbb{R}^d} e^{i\boldsymbol{\omega}\cdot\boldsymbol{\tau}} d\mu(\boldsymbol{\omega}). \tag{4.14}$$

Remark 4.2.1. Equation (4.14) places a positive power onto each frequency $\boldsymbol{\omega}$, and is analogous to placing only positive probability mass in the prior covariance function [35].

Bochner's result is powerful as it says that we can construct any positive definite function by simply choosing a corresponding measure. In most cases of interest, the measure $\mu(\cdot)$ has a Lebesgue density, known as the spectral density $S(\cdot)$,

$$\mu(B) = \int_B S(x)dx, \tag{4.15}$$

where B is any Borel set and $S \in L_1(\mathbb{R}^d)$ is a continuous positive definite function.

The spectral density $S(\boldsymbol{\omega})$ of a stochastic process $f(\mathbf{x})$ describes the energy in the process at a given frequency $\boldsymbol{\omega}$. Hence, the spectral density of $f(\mathbf{x})$ is defined as the point wise magnitude of $F(i\boldsymbol{\omega})$,

$$S(\boldsymbol{\omega}) = |F(i\boldsymbol{\omega})|^2 = F(i\boldsymbol{\omega})F(-i\boldsymbol{\omega}), \tag{4.16}$$

where $F(i\boldsymbol{\omega})$ is the Fourier transform of $f(\mathbf{x})$.

The Wiener-Khintchine theorem describes the special case of Bochner's theorem, where the measure has a Lebesgue density equal too the spectral density function. Indeed, the Wiener-Khintchine theorem introduces the Fourier dual between the spectral density and covariance function of a stochastic process.

Theorem 4.2.2 (Wiener-Khintchine Theorem)**.** *Let $k(\boldsymbol{\tau})$ be a continuous positive definite covariance function on $\mathbb{R}^d$. A weakly stationary mean-square continuous stochastic process with*

covariance function $k(\boldsymbol{\tau})$ has a real valued Fourier transform, with spectral $S(\boldsymbol{\omega})$. Then $k(\boldsymbol{\tau})$ and $S(\boldsymbol{\omega})$ are the Fourier duals of each other,

$$k(\boldsymbol{\tau}) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S(\boldsymbol{\omega})e^{i\boldsymbol{\omega}\cdot\boldsymbol{\tau}}d\boldsymbol{\omega}, \quad S(\boldsymbol{\omega}) = \int_{-\infty}^{\infty} k(\boldsymbol{\omega})e^{-i\boldsymbol{\omega}\cdot\boldsymbol{\tau}}d\boldsymbol{\tau}. \tag{4.17}$$

Remark 4.2.2. Substituting $\boldsymbol{\tau} = |\mathbf{x} - \mathbf{x}'|$,

$$k(\boldsymbol{\tau}) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S(\boldsymbol{\omega})e^{i\boldsymbol{\omega}\cdot|\mathbf{x}-\mathbf{x}'|}d\boldsymbol{\omega} = \frac{1}{2\pi}\int_{-\infty}^{\infty} S(\boldsymbol{\omega})e^{i\boldsymbol{\omega}\cdot\mathbf{x}}\left(e^{i\boldsymbol{\omega}\cdot\mathbf{x}'}\right)^{*}d\boldsymbol{\omega}, \tag{4.18}$$

the complex exponentials $e^{i\boldsymbol{\omega}\cdot\mathbf{x}}$ are the eigenfunctions of the stationary kernel k with respect to the Lebesgue measure [35]. Hence, the spectral density $S(\boldsymbol{\omega})$ is approximately the amount of power placed on each eigenfunction with frequency $\boldsymbol{\omega}$.

In turns out that the local behaviour of a stochastic function is intimately related to the behaviour of the spectral density. Generally, the quicker the spectral density decreases as $|\omega|$ increases, the smoother the stochastic process (See Figure 4.1)

As an example we consider the Matérn covariance function and it's corresponding spectral density function.

Example 4.2.1 (Matérn Spectral Density)**.** *The spectral density of a stochastic process with Matérn covariance defined on $\mathbb{R}^d$ is the d-dimensional Fourier transform of the Matérn kernel (2.4), given explicitly by Stein [36],*

$$S(\boldsymbol{\omega}) = \frac{2\sigma^2\sqrt{\pi}\Gamma(\nu+\frac{d}{2})(2\nu)^{\nu}}{\Gamma(\nu)l^{2\nu}}\left(\frac{2\nu}{l^2}+\boldsymbol{\omega}^2\right)^{-(\nu+\frac{d}{2})}. \tag{4.19}$$

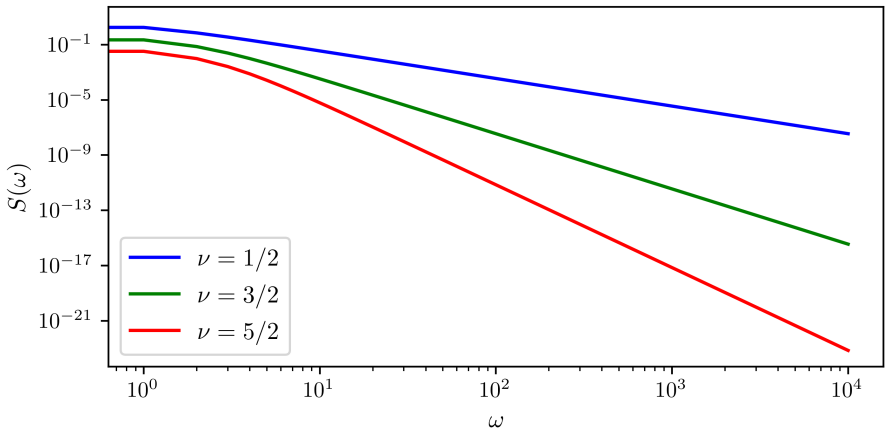

Figure 4.1: Rate of spectral density function $S(\omega)$ decay for Matérn 1/2, 3/2 and 5/2 kernels for $d = l = \sigma = 1$. The spectral density of smoother functions, such as a Gaussian process with Matérn 5/2 kernel, are faster decaying. Hence, the spectral density function provides important information about its corresponding stochastic process

.

4.3 Spectra and Stochastic Differential Equations

One of the most useful results existing due to the Wiener-Khintchine theorem, is that we can formulate stochastic differential equations that have the same spectral density function as the Fourier transform of a chosen Gaussian process covariance function.

In the seminal work by Doob [5], a stationary Gaussian process covariance function can be represented *exactly* by a stochastic differential equation if and only if it has a rational spectral density function of the form,

$$S(\boldsymbol{\omega}) = \frac{q}{|\sum_{k=0}^{n} a_k (i\boldsymbol{\omega})^k|^2} = \frac{\text{constant}}{\text{polynomial in } (i\boldsymbol{\omega})}. \tag{4.20}$$

Given a rational spectral density function we now wish to find a transfer function $\hat{\mathcal{L}}(i\boldsymbol{\omega})$, whose point-wise magnitude, when fed with Gaussian white noise $W(i\boldsymbol{\omega})$ with spectral density q, is equal to our desired spectral density function, where,

$$F(i\boldsymbol{\omega}) = \hat{\mathcal{L}}(i\boldsymbol{\omega}) W(i\boldsymbol{\omega}). \tag{4.21}$$

However, given assumptions about the form of $F(i\boldsymbol{\omega})$ it is possible to derive different transfer functions and therefore different linear operators that produce the same spectral density function and therefore correspond to the same covariance function.

4.3.1 Stochastic Differential Equation

Let us first assume that $F(i\omega)$ is a one-dimensional complex valued function. In this case the squared magnitude of $F(i\omega)$ is given as the product of $F(i\omega)$ and its complex conjugate $F(-i\omega)$. Therefore we can factorise the spectral density $S(i\omega)$ as,

$$S(\omega) = F(i\omega) F(-i\omega), \tag{4.22}$$

ensuring that all poles belonging to $F(i\omega)$ are in the upper left s-plane and the corresponding complex conjugate pole pairs belonging to $F(i\omega)$ are in the lower left s-plane. We do this so that we can construct a stable causal Markov process using $F(i\omega)$ [37], such that the impulse response, given by the inverse Fourier transform of the transfer function, can be represented by a physical system [38].

As an example we consider the Matérn SDE.

Theorem 4.3.1 (Matérn SDE)**.** *The stochastic differential equation corresponding to the Matérn spectral density function, assuming $F(i\omega)$ is a 1D complex function such that $f \in \mathbb{R}$, is,*

$$\left(\lambda + \frac{d}{dt}\right)^{\nu+1/2} f(t) = w(t). \tag{4.23}$$

Proof. The Matérn spectral density function (4.19) can be given as,

$$S(\omega) \propto (\lambda^2 + \omega^2)^{-(\nu+\frac{1}{2})}, \tag{4.24}$$

where $\lambda^2 = 2\nu/l^2$. Factorising $S(\omega)$,

$$S(\omega) \propto (\lambda + i\omega)^{-(\nu+\frac{1}{2})} (\lambda - i\omega)^{-(\nu+\frac{1}{2})}, \tag{4.25}$$

and plotting the poles in Figure 4.2, we extract the causal transfer function $F(i\omega)$,

$$F(i\omega) = \hat{\mathcal{L}}(i\omega) W(i\omega) = (\lambda + i\omega)^{-(\nu+\frac{1}{2})} W(i\omega), \tag{4.26}$$

where,

$$W(i\omega) = \frac{2\sigma^2 \sqrt{\pi} \Gamma(\nu + \frac{d}{2}) (2\nu)^\nu}{\Gamma(\nu) l^{2\nu}}. \tag{4.27}$$

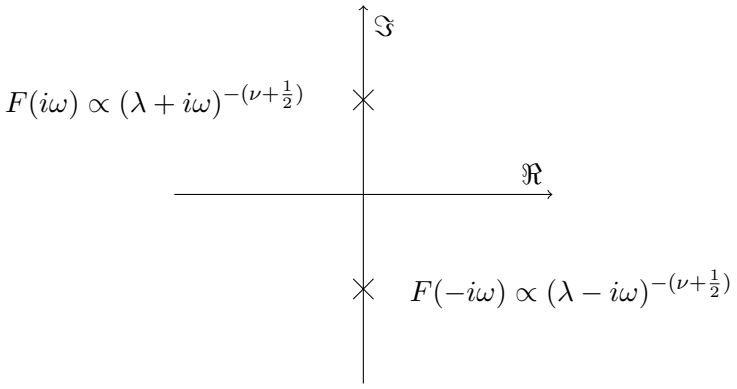

Figure 4.2: Pole-Zero plot for Matérn spectral density function

such that the marginal variance of the Gaussian process is encoded into the SDE via the marginal variance of the Gaussian white noise. Rearranging and taking the inverse Fourier transform,

$$\mathcal{F}^{-1}\left[(\lambda + i\omega)^{(\nu+\frac{1}{2})}\right] f(t) = w(t), \tag{4.28}$$

the Matérn SDE representation is,

$$\left(\lambda + \frac{d}{dt}\right)^{\nu+1/2} f(t) = w(t). \tag{4.29}$$

□

4.3.2 Stochastic Partial Differential Equation

Alternatively we can assume that $F(i\boldsymbol{\omega})$ is an n-dimensional real valued function. Hence, we can factorise the spectral density function $S(\boldsymbol{\omega})$ as,

$$S(\boldsymbol{\omega}) = \sqrt{F(i\boldsymbol{\omega})}. \tag{4.30}$$

Like previously, let us consider the Matérn SPDE.

Theorem 4.3.2 (Matérn SPDE)**.** *The stochastic partial differential equation corresponding to the Matérn spectral density function, assuming $F(i\omega)$ is a d-dimensional real-valued function such that $f \in \mathbb{R}^d$, is,*

$$\left(\lambda^2 - \Delta\right)^{\frac{\nu+d/2}{2}} f(\mathbf{x}) = w(\mathbf{x}). \tag{4.31}$$

Proof. The Matérn spectral density function (4.19) can be given as,

$$S(\boldsymbol{\omega}) \propto (\lambda^2 + \boldsymbol{\omega}^2)^{-(\nu+\frac{d}{2})}, \tag{4.32}$$

where $\lambda^2 = 2\nu/l^2$. Letting $F(i\boldsymbol{\omega})$ be the Fourier transform of a self-adjoint operator,

$$F(i\boldsymbol{\omega}) = \sqrt{S(\boldsymbol{\omega})} \propto (\lambda^2 + \boldsymbol{\omega}^2)^{-\frac{\nu+d/2}{2}}. \tag{4.33}$$

Consequently,

$$F(i\boldsymbol{\omega}) = \hat{\mathcal{L}}(i\boldsymbol{\omega})W(i\boldsymbol{\omega}) = (\lambda^2 + \boldsymbol{\omega}^2)^{-\frac{\nu+d/2}{2}} W(i\boldsymbol{\omega}). \tag{4.34}$$

Rearranging and taking the inverse Fourier transform,

$$\mathcal{F}^{-1}\left[(\lambda^2 + \boldsymbol{\omega}^2)^{\frac{\nu+d/2}{2}}\right] f(t) = w(t), \tag{4.35}$$

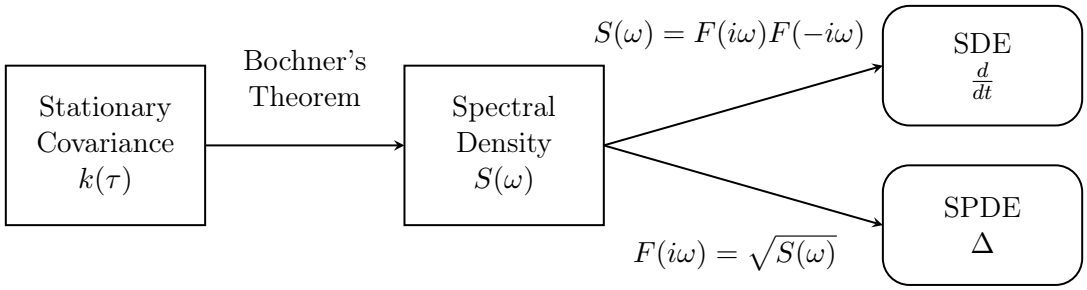

Figure 4.3: Process of converting from a stationary covariance function to a stochastic (partial) differential equation

the Matérn SPDE representation is,

$$\left(\lambda^2 - \Delta\right)^{\frac{\nu+d/2}{2}} f(t) = w(t). \tag{4.36}$$

□

Remark 4.3.1. In the 1D case, *under the same stability assumptions* we can show equality between the Matérn SDE and Matérn SPDE forms. Ensuring all poles belong in the upper left $s-plane$, for $d = 1$,

$$\left(\lambda^2 - \frac{d^2}{dt}\right)^{\frac{\nu+1/2}{2}} = \left[\left(\lambda + \frac{d}{dt}\right)^{\nu+1/2}\left(\lambda - \frac{d}{dt}\right)^{\nu+1/2}\right]^{1/2}, \tag{4.37}$$

$$= \left(\lambda + \frac{d}{dt}\right)^{\nu+1/2}. \tag{4.38}$$

4.4 Discussion

Domain	Matérn S(P)DE	Approximation
$t \in \mathbb{R}$	$\mathcal{L} = \left(\lambda + \frac{d}{dt}\right)^{\nu+1/2}$	Kalman Filtering [37] Interdomain Inducing Points [1, 2]
$\mathbf{x} \in \mathbb{R}^d$	$\mathcal{L} = \left(\lambda^2 - \Delta\right)^{\frac{\nu+1/2}{2}}$	GMRF[39]

Table 4.1: Comparison between the Matérn SDE and SPDE and approximations that use each representation

We have shown that for a given kernel with rational spectral density function that it is possible to derive two different dynamical systems. A summary of the different derivations for each system is given in Figure 4.3.

Pointwise, it should be noted that for a given input, both systems provide different solutions. However, by placing a distribution on the forcing function, such as the case of Gaussian white noise, both dynamical systems place the same distribution on the solution f. This occurs as both systems share the same spectral density function, which by Bochner's theorem defines the same stationary covariance function. Hence, inputting Gaussian white noise, by the linear transformation of Gaussian random variables, the solutions to both the SDE and SPDE are Gaussian processes with the same covariance function.

As mentioned previously, we can simulate the sampling from a Gaussian process prior by solving a stochastic differential equation. Given the wealth of literature on solving SDEs and SPDEs this

makes the dynamical system representation of a Gaussian process highly suitable for performing approximate inference. However, the SDE and SPDE formulations do have fundamentally different properties which are useful in different situations.

Most notably, the SDE form is restricted to only one-dimensional domains. This has benefits however, as allows us to represent a Gaussian process as a linear time invariant SDE (LTI SDE) for which inference can be approximated in linear time using Kalman filters [37]. On the other hand, the SPDE form can be defined on higher dimensional domains for $d \geq 2$. This property was exploited by Lindgren et al. [39] who approximated the solution to the SPDE via the finite element method in order to approximate a multi-dimensional Gaussian process as a Gaussian Markov Random Field. A comparative summary between the SDE and SPDE representations of Gaussian process is given in Table 4.1.

In the following sections we would like to consider how we can use both the SDE and SPDE representation of a Gaussian process and their solutions in order to construct RKHS inner products for chosen covariance functions. In particular our focus will be trying to classify when it is possible to construct an RKHS inner product. To do so we will be using methods for solving partial differential equations, namely Green's functions and the variational formulation of boundary value problems.

5 RKHS Inner Products and Green's Functions

Green's functions are an integral operator kernel, associated with some linear differential operator $\mathcal{L}$, that enable us to represent solutions to a differential equation via an integral formulation. Having developed a theory for the SPDE representations of Gaussian processes, in this section we consider the application of Green's functions in order to construct functional form RKHS inner products.

This section relates to the key results from Cunningham [4], in particular how Green's functions relate to the RKHS and how to construct an RKHS inner product via a Green's functions associated to a particular linear differential operator. However, Cunningham [4] lacked a detailed discussion on the reproducing property, choice of domain and boundary conditions. Consequently, this section will introduce the key results from Cunningham [4] before extending the discussion to consider the domain of interest, how this affects the usability of the RKHS inner product and a discussion on boundary conditions.

This chapter begins by first introducing Green's identities, from which Green's functions are derived, before developing relationships between Green's functions and RKHS inner products.

5.1 Green's Identities

In order to introduce Green's functions we will first define Green's three identities. Let us first recall the divergence theorem in d-dimensions.

Theorem 5.1.1 (Divergence Theorem)**.** *Let $\mathcal{D}$ be some smooth d-dimensional region $D \subset \mathbb{R}^d$ and $\mathbf{F}$ a vector field over $\mathbb{R}^d$. If the vector field $\mathbf{F}$ has continuous first order partial derivatives such that $\mathbf{F} \in C^1(\mathcal{D})$ then,*

$$\int_{\mathcal{D}} (\nabla \cdot \mathbf{F})\, dV = \oint_{\partial \mathcal{D}} (\mathbf{F} \cdot \hat{\mathbf{n}})\, dS, \tag{5.1}$$

where $\partial\mathcal{D}$ represents the boundary of the domain $\mathcal{D}$ and $\mathbf{n}$ the unit outward normal vector to the surface element dS.

Green's first identity is a special case of the divergence theorem. Let $\mathbf{F} = f(\nabla g)$, where $g \in C^2(\mathcal{D})$ is twice continuously differentiable, and using the identity,

$$\nabla \cdot (f\mathbf{F}) = \nabla f \cdot \mathbf{F} + f \nabla \cdot \mathbf{F}, \tag{5.2}$$

the divergence theorem states,

$$\int_{\mathcal{D}} (\nabla f)(\nabla g) dV + \int_{\mathcal{D}} f(\nabla^2 g) dV = \oint_{\partial \mathcal{D}} f(\nabla g) \cdot \hat{\mathbf{n}} dS. \tag{5.3}$$

Equation (5.3) is known as **Green's first identity**, and is the higher dimensional equivalent to integration by parts.

Now define both $f, g \in C^2(\mathcal{D})$ such that they are both twice continuously differentiable on $\mathcal{D}$. Let $\mathbf{F} = f(\nabla g) - g(\nabla f)$, and defining the normal derivative as,

$$\frac{\partial}{\partial n} \equiv \nabla \cdot \hat{\mathbf{n}}, \tag{5.4}$$

the divergence theorem states,

$$\int_{\mathcal{D}} \left[f(\Delta g) - g(\Delta f) \right] dV = \oint_{\partial \mathcal{D}} \left[f \frac{\partial g}{\partial n} - g \frac{\partial f}{\partial n} \right] dS. \tag{5.5}$$

Equation (5.5) is known as **Green's second identity**. We often generalise Green's second identity to consider linear second-order elliptic differential operators $\mathcal{L}$,

$$\int_{\mathcal{D}} \left[f(\mathcal{L} g) - g(\mathcal{L} f) \right] dV = \oint_{\partial \mathcal{D}} \left[f \frac{\partial g}{\partial n} - g \frac{\partial f}{\partial n} \right] dS. \tag{5.6}$$

We now construct Green's third identity from the second identity by introducing *Green's functions* G. The primary use of Green's function is to construct solutions to linear non-homogeneous boundary value problems. For a linear (partial) differential operator $\mathcal{L}$, the Green's function G is defined as the unique solution of the linear differential equation,

$$\left.\begin{array}{ll} \mathcal{L} G(\mathbf{x}, \mathbf{z}) = \delta(\mathbf{x} - \mathbf{z}) & \text{in } \mathcal{D}, \\ G(\mathbf{x}, \mathbf{z}) = 0 & \text{on } \partial \mathcal{D}, \end{array}\right\} \tag{5.7}$$

where δ is a dirac delta function and the Green's function G satisfies homogeneous Dirichlet boundary conditions, disappearing at the edge of the domain $\partial \mathcal{D}$.

Letting $g = G(\mathbf{x}, \mathbf{z})$, and substituting into Green's second identity,

$$\begin{aligned} \int_{\mathcal{D}} [f(\mathcal{L} g) - g(\mathcal{L} f)] dV &= \int_{\mathcal{D}} [f(\mathbf{x})(\mathcal{L} G(\mathbf{x}, \mathbf{z})) - G(\mathbf{x}, \mathbf{z}) \mathcal{L} f(\mathbf{x})] dV(\mathbf{x}), \\ &= \int_{\mathcal{D}} [f(\mathbf{x}) \delta(\mathbf{x} - \mathbf{z}) - G(\mathbf{x}, \mathbf{z}) \mathcal{L} f(\mathbf{x})] dV(\mathbf{x}), \\ &= f(\mathbf{z}) - \int_{\mathcal{D}} G(\mathbf{x}, \mathbf{z}) \mathcal{L} f(\mathbf{x}) dV(\mathbf{x}). \end{aligned} \tag{5.8}$$

and on the boundary $\partial \mathcal{D}$,

$$\begin{aligned} \oint_{\partial \mathcal{D}} \left[f \frac{\partial g}{\partial n} - g \frac{\partial f}{\partial n} \right] dS &= \oint_{\partial \mathcal{D}} \left[f(\mathbf{x}) \frac{\partial G(\mathbf{x}, \mathbf{z})}{\partial n} - G(\mathbf{x}, \mathbf{z}) \frac{\partial f(\mathbf{x})}{\partial n} \right] dS(\mathbf{x}), \\ &= \oint_{\partial \mathcal{D}} \left[f(\mathbf{x}) \frac{\partial G(\mathbf{x}, \mathbf{z})}{\partial n} - \underbrace{\cancel{G(\mathbf{x}, \mathbf{z})} \frac{\partial f(\mathbf{x})}{\partial n}}_{G(\mathbf{x}, \mathbf{z}) = 0 \text{ on } \partial \mathcal{D}} \right] dS(\mathbf{x}), \\ &= \oint_{\partial \mathcal{D}} f(\mathbf{x}) \frac{\partial G(\mathbf{x}, \mathbf{z})}{\partial n} dS(\mathbf{x}). \end{aligned} \tag{5.9}$$

such that,

$$f(\mathbf{z}) - \int_{\mathcal{D}} G(\mathbf{x}, \mathbf{z}) \mathcal{L} f(\mathbf{x}) dV(\mathbf{x}) = \oint_{\partial \mathcal{D}} f(\mathbf{x}) \frac{\partial G(\mathbf{x}, \mathbf{z})}{\partial n} dS(\mathbf{x}). \tag{5.10}$$

Equation (5.10) is known as **Green's third identity** and is fundamental in constructing solutions to boundary value problems.

As an example let us consider the poisson equation subject to non-homogeneous boundary conditions,

$$\left.\begin{array}{ll} \Delta f(\mathbf{x}) = w(\mathbf{x}) & \text{in } \mathcal{D}, \\ f(\mathbf{x}) = b(\mathbf{x}) & \text{on } \partial \mathcal{D}, \end{array}\right\} \tag{5.11}$$

where $b(\mathbf{x})$ is some continuous function on the boundary of the domain $\partial \mathcal{D}$.

The general solution to (5.11) is composed of the solution to the non-homogeneous problem $\Delta f(\mathbf{x}) = w(\mathbf{x})$ that is homogeneous at the boundary, f_p, with a solution to the homogeneous problem $\Delta f(\mathbf{x}) = 0$ that satisfies the non-homogeneous boundary conditions, f_h,

$$f(\mathbf{x}) = f_h(\mathbf{x}) + f_p(\mathbf{x}). \tag{5.12}$$

Using Green's third identity, we can construct the general solution as,

$$f(\mathbf{x}) = \int_{\mathcal{D}} G(\mathbf{x}, \mathbf{z}) w(\mathbf{z}) d\mathbf{z} + \oint_{\partial \mathcal{D}} \frac{\partial G(\mathbf{x}, \mathbf{z})}{\partial n} b(\mathbf{z}) dS(\mathbf{z}). \tag{5.13}$$

Remark 5.1.1. Crucially, Green's third identity allows us to write the solution to a boundary value problem as the summation of two integral equations. Observing the first integral, a Green's function can be viewed as the impulse response of a linear differential operator to an arbitrary forcing function. Further the second integral, the derivative of the Green's function, is also a kernel, known as the Poisson kernel.

Remark 5.1.2. We note that (5.13) is a *general solution*, composed of the solution to the non-homogeneous problem $\mathcal{L} f(\mathbf{x}) = w(\mathbf{x})$ that is homogeneous at the boundary, f_p, with a solution to the homogeneous problem $\mathcal{L} f(\mathbf{x}) = 0$ that satisfies the non-homogeneous boundary conditions, f_h, where,

$$f_p(\mathbf{x}) = \int_{\mathcal{D}} G(\mathbf{x}, \mathbf{z}) w(\mathbf{z}) d\mathbf{z}, \quad f_h(\mathbf{x}) = \oint_{\partial \mathcal{D}} \frac{\partial G(\mathbf{x}, \mathbf{z})}{\partial n} b(\mathbf{z}) dS(\mathbf{z}). \tag{5.14}$$

As an example, for homogeneous boundary conditions $f(\mathbf{x}) = 0$ on $\partial \mathcal{D}$, the second term on the RHS disappears and we are left with a solution to the non-homogeneous problem with homogeneous boundary conditions.

Let us now prove that Equation (5.13) is indeed a solution to the boundary value problem defined by Equation (5.11):

Proof.

$$\begin{aligned}
f(\mathbf{x}) &= \int_{\mathcal{D}} G(\mathbf{x}, \mathbf{z}) w(\mathbf{z}) d\mathbf{z} + \oint_{\partial \mathcal{D}} \frac{\partial G(\mathbf{x}, \mathbf{z})}{\partial n} b(\mathbf{z}) dS(\mathbf{z}), && (5.15)\\
&= \int_{\mathcal{D}} G(\mathbf{x}, \mathbf{z}) w(\mathbf{z}) d\mathbf{z} + \oint_{\partial \mathcal{D}} \frac{\partial G(\mathbf{x}, \mathbf{z})}{\partial n} b(\mathbf{z}) dS(\mathbf{z}) - \oint_{\partial \mathcal{D}} G(\mathbf{x}, \mathbf{z}) \frac{\partial f(\mathbf{z})}{\partial n} dS(\mathbf{z}), && (5.16)\\
&= -\int_{\mathcal{D}} \nabla G(\mathbf{x}, \mathbf{z}) \cdot \nabla f(\mathbf{z}) d\mathbf{z} + \frac{\partial G(\mathbf{x}, \mathbf{z})}{\partial n} b(\mathbf{z}) dS(\mathbf{z}), && (5.17)\\
&= \int_{\mathcal{D}} \nabla^2 G(\mathbf{x}, \mathbf{z}) f(\mathbf{z}) d\mathbf{z}, && (5.18)\\
&= \int_{\mathcal{D}} \delta(\mathbf{x} - \mathbf{z}) f(\mathbf{z}) d\mathbf{z}, && (5.19)\\
&= f(\mathbf{x}). && (5.20)
\end{aligned}$$

□

Having defined Green's functions we would now like to consider their relationship to stochastic processes and in particular positive definite kernels that define zero-mean Gaussian process priors.

5.2 Green's Functions and Kernels

This section will introduce some of the key results from Cunningham [4] that draw on connections between Green's functions and positive-definite kernels. These results are introduced here and their proofs are provided in Appendix A.

Firstly, we define the relationship between a kernel k belonging to a Gaussian process f, defined as the solution to a linear SPDE with linear differential operator $\mathcal{L}$, and the Green's function G associated with $\mathcal{L}$.

Definition 5.2.1 (ISO Result 1)**.** *Let $f \sim GP(0, k)$ be a Gaussian process defined over some domain $\mathcal{D}$, $k : \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ a positive definite kernel, $\mathcal{L}$ a linear differential operator and $G : \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ such that,*

$$\mathcal{L}G(\mathbf{x}, \mathbf{z}) = \delta(\mathbf{x} - \mathbf{z}). \tag{5.21}$$

We can define the kernel k in terms of the Green's functions G as,

$$k(\cdot, \mathbf{x}) = \int_{\mathcal{D}} G(\cdot, \mathbf{z})G(\mathbf{x}, \mathbf{z})d\mathbf{z}. \tag{5.22}$$

Proof. See Appendix A.1 □

We can derive this same result by considering a Green's operator $\mathcal{L}^{-1}$ as the inverse of a linear differential operator $\mathcal{L}$,

$$\mathcal{L}^{-1}f(\cdot) = \int_{\mathcal{D}} G(\cdot, \mathbf{z})f(\mathbf{z})d\mathbf{z}, \tag{5.23}$$

where $\mathcal{L}^{-1}$ is a Hilbert-Schmidt integral operator with respect to the Lebesgue measure and G the Green's function associated with the linear differential operator $\mathcal{L}$. Consequently, we can view $\mathcal{L}^{-1}$ as a mapping between Hilbert spaces, enabling us to utilise the function space perspective of SPDEs developed in Section 4.1.2.

$$\begin{aligned}
\langle \mathcal{K}f, g\rangle_{\mathcal{H}} &= \mathbb{E}[\hat{F}(g)\hat{F}(h)], && (5.24)\\
&= \mathbb{E}[\hat{W}(\mathcal{L}^{-1}g)\hat{W}(\mathcal{L}^{-1}h)], && (5.25)\\
&= \langle \mathcal{L}^{-1}g, \mathcal{L}^{-1}\rangle_{\mathcal{H}}, && (5.26)\\
&= \langle g, \mathcal{L}^{-1}(\mathcal{L}^{-1}h)\rangle_{\mathcal{H}}, && (5.27)\\
&= \left\langle g, \int_{\mathcal{D}} G(\cdot, \mathbf{x}) \left(\int_{\mathcal{D}} G(\mathbf{x}, \mathbf{z})h(\mathbf{z})d\mathbf{z} \right) d\mathbf{x} \right\rangle_{\mathcal{H}}, && (5.28)\\
&= \left\langle g, \int_{\mathcal{D}} \left[\int_{\mathcal{D}} G(\cdot, \boldsymbol{x})G(\boldsymbol{z}, \boldsymbol{x})d\boldsymbol{x} \right] h(\mathbf{z})d\mathbf{z} \right\rangle_{\mathcal{H}}, && (5.29)
\end{aligned}$$

therefore,

$$\mathcal{K}h(\cdot) = \int_{\mathcal{D}} \left[\int_{\mathcal{D}} G(\cdot, \boldsymbol{z})G(\boldsymbol{x}, \boldsymbol{z})d\boldsymbol{z} \right] h(\mathbf{x})d\mathbf{x}, \tag{5.30}$$

and hence,

$$k(\cdot, \boldsymbol{x}) = \int_{\mathcal{D}} G(\cdot, \boldsymbol{z})G(\boldsymbol{x}, \boldsymbol{z})d\boldsymbol{z}. \tag{5.31}$$

The second important result from Cunningham [4] relates a kernel k belonging to a Gaussian process f, defined as the solution to a linear SPDE with linear differential operator $\mathcal{L}$, the Green's function G associated with $\mathcal{L}$ and the linear operator $\mathcal{L}$ itself.

Definition 5.2.2 (ISO Result 2)**.** *Let $f \sim GP(0, k)$ be a Gaussian process defined over some domain $\mathcal{D}$, $k : \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ a positive definite kernel, $\mathcal{L}$ a linear differential operator and $G : \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ defined as previously. We can define the Green's functions G as,*

$$\mathcal{L}k(\cdot, \boldsymbol{x}) = G(\cdot, \boldsymbol{x}). \tag{5.32}$$

Proof. See Appendix A.2 □

These results are significant as enable us to relate a bounded linear differential operator $\mathcal{L}$ and it's associated Green's function G with a reproducing kernel k that defines an RKHS $\mathcal{H}_k$. Indeed, Green's functions provide a link between the theory of RKHSs and SPDEs offering a powerful way of expressing reproducing kernels in terms of the solution to some dynamical system.

5.3 Green's Functions and the RKHS

In the previous section we introduced and built upon results from Cunningham [4] that considered the relationships between Green's functions and stochastic processes. We showed that Green's functions are intimately related to positive-definite kernels, represented by some linear differential operator $\mathcal{L}$. Consequently, in this section we would like to develop connections between Green's functions and the RKHS, building upon the work in Cunningham [4].

In particular we are interested in deriving functional forms of the RKHS inner product. Functional forms are useful as enable us to perform cheap inference by projecting a Gaussian process onto a set of orthogonal basis functions.

5.3.1 Unbounded Domain

First of all let us derive an RKHS inner product using Green's functions on an *unbounded* domain $D = \mathbb{R}^d$. From previous sections we know that we can represent a Gaussian process prior $f \sim \mathcal{GP}(0, k)$ as a solution to a linear non-homogeneous SPDE defined by $\mathcal{L}$, where the covariance k is defined by the operator $\mathcal{L}$. Consequently, let us define a non-homogeneous SPDE over an unbounded domain $\mathcal{D} \in \mathbb{R}^d$ as,

$$\mathcal{L}f(\boldsymbol{x}) = w(\boldsymbol{x}) \quad \text{in } \mathbb{R}^d, \ \} \tag{5.33}$$

where $w(\cdot)$ is a Gaussian white noise process.

The effect of letting the domain $\mathcal{D}$ occupy the full space $\mathbb{R}^d$, is that there are no boundaries and therefore we do not prescribe boundary conditions. We would now like to use Green's functions to define a solution to (5.33). However, in order for the Green's function to be uniquely defined, in the absence of boundary conditions, we must define decay conditions. Consequently, the Green's function G of $\mathcal{L}$ defined over the full-space $\mathbb{R}^d$ is given as the unique solution to the linear differential equation,

$$\left.\begin{aligned} &\mathcal{L}G(\mathbf{x}, \mathbf{z}) = \delta(\mathbf{x} - \mathbf{z}), \quad \text{in } \mathbb{R}^d, \\ &\lim_{|\mathbf{x}| \to \infty} G(\mathbf{x}, \mathbf{z}) = 0, \end{aligned}\right\} \tag{5.34}$$

This has the affect that when defining a solution to Equation (5.33) using Green's third identity (5.10), that we no longer have a boundary value term. Hence we can write a solution to (5.33) as,

$$f(\mathbf{x}) = \int_{\mathbb{R}^d} G(\mathbf{x}, \mathbf{z}) \mathcal{L} f(\mathbf{z}) d\mathbf{z}, \tag{5.35}$$

Equipped with a solution (5.35) to the SPDE (5.33) on a unbounded domain, we can now use the results from Section 5.2 to first derive a reproducing kernel associated with $\mathcal{L}$ in terms of

Green's functions and then an RKHS $\mathcal{H}_k$ with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}_k}$. This derivation is however, a formalisation of the results from Section 4.2 in Cunningham [4], which did not include a discussion on the type of domain the SPDE is defined over. Hence we formalise the result in Definition 5.3.1.

Definition 5.3.1 (ISO Result 3)**.** *Let $\mathcal{D} \in \mathbb{R}^d$ be the full-space domain, $\mathcal{H}_k$ a RKHS on $\mathcal{D}$ with reproducing kernel k and $\mathcal{L}$ a linear differential operator associated with k according to Bochner's theorem. The RKHS inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}_k}$ on $\mathcal{D}$ is given $\forall f, g \in \mathcal{H}_k$ as,*

$$\langle f, g \rangle_{\mathcal{H}_k} = \int_{\mathbb{R}^d} \mathcal{L}f(\mathbf{x}) \mathcal{L}g(\mathbf{x}) d\mathbf{x}. \tag{5.36}$$

Proof. See Appendix A.3. Here we will prove that the reproducing property holds, which was lacking from Cunningham [4].

$$\begin{aligned}
\langle f, k(\cdot, \mathbf{x}') \rangle_{\mathcal{H}_k} &= \int_{\mathbb{R}^d} \mathcal{L}f(\mathbf{x}) \mathcal{L}k(\mathbf{x}, \mathbf{x}') d\mathbf{x}, && (5.37)\\
&= \int_{\mathbb{R}^d} \mathcal{L}f(\mathbf{x}) \left(\int_{\mathbb{R}^d} \mathcal{L}G(\mathbf{x}, \mathbf{z}) G(\mathbf{x}', \mathbf{z}) d\mathbf{z} \right) d\mathbf{x}, && (5.38)\\
&= \int_{\mathbb{R}^d} \mathcal{L}f(\mathbf{x}) \left(\int_{\mathbb{R}^d} \delta(\mathbf{x} - \mathbf{z}) G(\mathbf{x}', \mathbf{z}) d\mathbf{z} \right) d\mathbf{x}, && (5.39)\\
&= \int_{\mathbb{R}^d} \mathcal{L}f(\mathbf{x}) G(\mathbf{x}, \mathbf{x}') d\mathbf{x}, && (5.40)\\
&= f(\mathbf{x}'). && (5.41)
\end{aligned}$$

Hence, the inner product (5.36) satisfies the reproducing property and is therefore an RKHS inner product. □

Consequently, we have shown that for unbounded domains we can derive functional form expressions for the RKHS inner product in arbitrary dimensions.

5.3.2 Bounded Domain

Juxtaposing the unbounded domain, we can instead define an RKHS inner product using Green's functions over a *bounded domain* $\mathcal{D} \subset \mathbb{R}^d$. Consequently, in order to ensure that our Green's function is unique, we must define linear homogeneous *boundary conditions*. Hence we consider Gaussian process priors as solutions to *stochastic non-homogeneous boundary value problems*.

We define a stochastic non-homogeneous boundary value problem via an SPDE over a bounded domain $\mathcal{D} \subset \mathbb{R}^d$ with accompanying boundary conditions as,

$$\left.\begin{array}{ll} \mathcal{L}f(\mathbf{x}) = w(\mathbf{x}) & \text{in } \mathcal{D}, \\ f(\mathbf{x}) = b(\mathbf{x}) & \text{on } \partial\mathcal{D}, \end{array}\right\} \tag{5.42}$$

Just like before, we now wish to use Green's functions to define a solution to (5.42). In this case however, we are working on a bounded domain and therefore we apply linear homogeneous boundary conditions to our Green's function. Hence, the Green's function G is defined as the unique solution to the linear differential equation,

$$\left.\begin{array}{ll} \mathcal{L}G(\mathbf{x}, \mathbf{z}) = \delta(\mathbf{x} - \mathbf{z}) & \text{in } \mathcal{D}, \\ G(\mathbf{x}, \mathbf{z}) = 0 & \text{on } \partial\mathcal{D}, \end{array}\right\} \tag{5.43}$$

Using the Green's function as defined by (5.43), recall the general solution to the boundary value problem (5.42) is given by Green's third identity (5.10),

$$f(\mathbf{x}) = \int_{\mathcal{D}} G(\mathbf{x},\mathbf{z})\mathcal{L}f(\mathbf{z})d\mathbf{z} + \oint_{\partial\mathcal{D}} f(\mathbf{z})\frac{\partial G(\mathbf{x},\mathbf{z})}{\partial n}dS(\mathbf{z}), \tag{5.44}$$

$$= \underbrace{\int_{\mathcal{D}} G(\mathbf{x},\mathbf{z})w(\mathbf{x})d\mathbf{z}}_{\text{Particular -}f_p} + \underbrace{\oint_{\partial\mathcal{D}} b(\mathbf{z})\frac{\partial G(\mathbf{x},\mathbf{z})}{\partial n}dS(\mathbf{z})}_{\text{Homogeneous -}f_h}, \tag{5.45}$$

where the homogeneous and particular solutions have been highlighted.

Consequently, we now wish to construct an RKHS $\mathcal{H}_k$ that has the same reproducing kernel k as the Gaussian process f defined as the solution to some stochastic boundary value problem. To do so let us understand how the general solution to the stochastic boundary value problem exists in the RKHS.

Firstly, it is important to note that by definition f_h and f_p define orthogonal function spaces. This is true since a solution to the homogeneous problem can never be a solution to the non-homogeneous problem, and vice versa. Consequently, both sets of solutions exist in separate orthogonal Hilbert spaces. From Gu [40, Theorem 2.5],

Theorem 5.3.1 (Orthogonal Hilbert Spaces)**.** *If the reproducing kernel k belonging to the RKHS $\mathcal{H}_k$, defined on some domain $\mathcal{D}$, can be decomposed into $k = k_0 + k_1$, where $\forall x, y \in \mathcal{D}$:*

1. *k_0 and k_1 are both positive definite,*
2. *$k_0(x,\cdot), k_1(x,\cdot) \in \mathcal{H}_k$,*
3. *$\langle k_0(x,\cdot), k_1(y,\cdot)\rangle = 0$,*

then the spaces $\mathcal{H}_{k0}$ and $\mathcal{H}_{k1}$ corresponding to k_0 and k_1 form an orthogonal decomposition of $\mathcal{H}_k$. Conversely, if k_0 and k_1 are both positive definite and $\mathcal{H}_{k0} \cap \mathcal{H}_{k1} = 0$, then,

$$\mathcal{H}_k = \mathcal{H}_{k0} \oplus \mathcal{H}_{k1}, \tag{5.46}$$

has a reproducing kernel $k = k_0 + k_1$.

Proposition 5.3.1. *For a general solution (5.45) to a stochastic non-homogeneous boundary value problem (5.42) decomposed into a particular solution f_p and a homogeneous solution f_h, f_p and f_h define orthogonal Hilbert spaces.*

Proof. The general solution for the non-homogeneous boundary value problem using Green's function is a summation of the solutions to the non-homogeneous (particular) problem with homogeneous boundary conditions,

$$\left.\begin{array}{ll} \mathcal{L}f_p(\boldsymbol{x}) = w(\boldsymbol{x}) & \text{in } \mathcal{D}, \\ f_p(\boldsymbol{x}) = 0 & \text{on } \partial\mathcal{D}, \end{array}\right\} \tag{5.47}$$

and the homogeneous problem with non-homogeneous boundary conditions,

$$\left.\begin{array}{ll} \mathcal{L}f_h(\boldsymbol{x}) = 0 & \text{in } \mathcal{D}, \\ f_h(\boldsymbol{x}) = b(\boldsymbol{x}) & \text{on } \partial\mathcal{D}. \end{array}\right\} \tag{5.48}$$

Heuristically we can see that a particular solution f_p and homogeneous solution f_h define orthogonal Hilbert spaces. □

Since the homogeneous and particular solutions are associated with orthogonal RKHSs, this makes drawing connections between the general solution for the non-homogeneous boundary value problem and the RKHS more straightforward as we can consider the reproducing kernel and inner product for each space separately. Sadly however, this is where the straightforwardness ends.

Let us first derive the RKHS inner product using Green's functions for the case of a one-dimensional bounded domain.

One-Dimensional Case

We note that in the literature there are many methods and examples of RKHS inner products defined over one-dimensional bounded domains [41, 42, 43, 44, 45, 46]. These methods vary however and often the scope of the work is either too broad or different to our aims. As a result, we will use and adapt results from the literature to suit the aims of this work: *for a specific reproducing kernel, and bounded domain, construct a functional form RKHS inner product.*

In order to define a unique solution to an mth-order boundary value problem, we must define m boundary conditions. In order to specify the boundary conditions we introduce the notion of boundary operator $\boldsymbol{\mathcal{B}}$ [45, 46, 43]. The boundary operator $\boldsymbol{\mathcal{B}}$ simply evaluates the unknown function f and its derivatives in m different ways [43, Definition 3.3].

Definition 5.3.2 (Boundary Value Operator)**.** *Let $\mathcal{D}$ be a regular bounded domain, such that it satisfies strong Lipschitz conditions, with a regular boundary $\partial\mathcal{D}$. Let $\mathcal{H}(\mathcal{D})$ be a Hilbert space on $\mathcal{D}$ and $L^2(\partial\mathcal{D})$ an L^2 space restricted to the boundary $\partial\mathcal{D}$. For an mth-order boundary value problem, the boundary value operator $\boldsymbol{\mathcal{B}}$ is defined as the vector operator,*

$$\boldsymbol{\mathcal{B}}f = \begin{bmatrix} B_1, & \cdots, & B_m \end{bmatrix}^T, \tag{5.49}$$

of length m, formed of a finite number of boundary operators $B_i : \mathcal{H}(\mathcal{D}) \to L^2(\partial\mathcal{D})$

Remark 5.3.1. Applying the same methodology as Fasshauer and Ye [43], the restricted boundary space $L^2(\partial\mathcal{D})$ is defined as,

$$L^2(\partial\mathcal{D}) = \{f : \partial\mathcal{D} \to \mathbb{R} : \|f\|_{\partial\mathcal{D}} < \infty\}, \tag{5.50}$$

with the inner product $\forall f, g \in L^2(\partial\mathcal{D})$,

$$\langle f, g\rangle_{L^2(\partial\mathcal{D})} = \oint_{\partial\mathcal{D}} f(\mathbf{x})g(\mathbf{x})dS(\mathbf{x}), \tag{5.51}$$

implying that the functions f are integrable on the boundary $\partial\mathcal{D}$ where dS is the surface area element.

Remark 5.3.2. In the special case of a one-dimensional domain, the restricted boundary space is given as,

$$L^2(\partial\mathcal{D}) = \{f : \partial\mathcal{D} = \{a, b\} \to \mathbb{R}\}. \tag{5.52}$$

such that the measure is only defined at the endpoints, $S(a) = S(b) = 1$ [43] and the inner product is given by,

$$\langle f, g\rangle_{L^2(\partial\mathcal{D})} = f(a)g(a) + f(b)g(b). \tag{5.53}$$

Remark 5.3.3. In the 1D case the boundary value operator $\boldsymbol{\mathcal{B}}$ for an mth-order differential equation is given similar to Ramsay [46] as,

$$\boldsymbol{\mathcal{B}}f = \begin{bmatrix} f(\cdot) \\ f'(\cdot) \\ \vdots \\ f^{(m-2)/2}(\cdot) \end{bmatrix}, \tag{5.54}$$

except that it is applied to both endpoints via the measure on $L^2(\partial\mathcal{D})$ and not directly via the operator.

We can now define a stochastic non-homogeneous boundary value problem over a one-dimensional bounded region $\mathcal{D} = [a, b]$ as,

$$\left.\begin{array}{ll} \mathcal{L}f(x) = w(x) & \text{in } \mathcal{D}, \\ \boldsymbol{\mathcal{B}}f(x) = \mathbf{b} & \text{on } \partial\mathcal{D}, \end{array}\right\} \tag{5.55}$$

where $\mathcal{L}$ is a bounded linear differential operator, and $\boldsymbol{\mathcal{B}}$ the boundary value operator (5.54). We can equivalently define a Green's function as the unique solution to the linear differential equation,

$$\left.\begin{array}{ll} \mathcal{L}G(x, z) = \delta(x - z) & \text{in } \mathcal{D}, \\ \boldsymbol{\mathcal{B}}G(x, z) = \mathbf{0} & \text{on } \partial\mathcal{D}, \end{array}\right\} \tag{5.56}$$

such that the Green's function G satisfies homogeneous boundary conditions.

We would now like to partition the general solution to (5.55) as,

$$f = f_p + f_h, \tag{5.57}$$

where f_p is the particular solution of (5.55), and the solution to,

$$\left.\begin{array}{ll} \mathcal{L}f_p(\boldsymbol{x}) = w(\boldsymbol{x}) & \text{in } \mathcal{D}, \\ \boldsymbol{\mathcal{B}}f_p(\boldsymbol{x}) = \mathbf{0} & \text{on } \partial\mathcal{D}, \end{array}\right\} \tag{5.58}$$

and f_h is the homogeneous solution of (5.55), and the solution to,

$$\left.\begin{array}{ll} \mathcal{L}f_h(\boldsymbol{x}) = 0 & \text{in } \mathcal{D}, \\ \boldsymbol{\mathcal{B}}f_h(\boldsymbol{x}) = \mathbf{b} & \text{on } \partial\mathcal{D}. \end{array}\right\} \tag{5.59}$$

As mentioned the function spaces $\mathcal{H}_{kp}$ and $\mathcal{H}_{kh}$ containing the solutions f_p and f_h respectively, define *orthogonal Hilbert spaces*, the summation of which gives us the Hilbert space of our general solution $\mathcal{H}_k$. As a result the RKHS inner product $\langle\cdot,\cdot\rangle_{\mathcal{H}_k}$ which we want to define for some given kernel k, is given as the summation,

$$\langle\cdot,\cdot\rangle_{\mathcal{H}_k} = \langle\cdot,\cdot\rangle_{\mathcal{H}_{kp}} + \langle\cdot,\cdot\rangle_{\mathcal{H}_{kh}}. \tag{5.60}$$

Let us first consider the particular solution f_p. The Hilbert space $\mathcal{H}_{kp}$ is the space of functions f_p such that $\mathcal{H}_{kp} = null(\boldsymbol{\mathcal{B}})$, that is all functions f_p (and their $(m-2)/2$ derivatives if $m > 2$ and m is even) are equal 0 at the boundary. Remembering that $w(\cdot)$ is Gaussian white noise process and hence f_p is a Gaussian process with covariance function k defined by the mapping $\mathcal{L}$, $\mathcal{H}_{kp}$ is then an RKHS generated by f_p over the bounded domain D. Consequently, like previously for unbounded domains, we can construct an RKHS inner product on $\mathcal{H}_{kp}$ using Green's functions defined by (5.56), $\forall f, g \in \mathcal{H}_{kp}$ as,

$$\langle f, g\rangle_{\mathcal{H}_{kp}} = \int_a^b \mathcal{L}f(\mathbf{x})\mathcal{L}g(\mathbf{x})d\mathbf{x}. \tag{5.61}$$

Next let us consider the homogeneous solution f_h, where the Hilbert space $\mathcal{H}_{kh}$ is the m-dimensional space of functions f_h such that $\mathcal{H}_{kh} = null(\mathcal{L})$. We would now like to define a finite dimensional RKHS inner product on $\mathcal{H}_{kh}$ using the boundary value operator.

Applying the boundary value operator to functions in $\mathcal{H}_{kh}$ maps them to the restricted L^2 space on the boundary. Indeed, given our knowledge that f is a Gaussian process and given we have already specified its covariance function k in setting up the problem, the action of $\boldsymbol{\mathcal{B}}$ on $f \in \mathcal{H}_{kh}$ is to return an $m-$dimensional vector of random variables. We now wish to apply Loeve's Isometry theorem to these random variables defined on $L^2(\partial\mathcal{D})$ in order to construct a *finite dimensional RKHS* that defines an inner product. To do so let us first define a finite dimensional RKHS.

Corollary 5.3.1.1 (Finite Dimensional RKHS)**.** *Let $\mathcal{H}_k \subset \mathbb{R}^d$ be a finite dimensional RKHS, with reproducing kernel defined by the unique symmetric positive-definite matrix $\mathbf{K}$, where,*

$$\mathbf{K} = \{k(\mathbf{x}_i, \mathbf{x}_j)\}_{i,j=1,\dots,N}, \tag{5.62}$$

and inner product,

$$\langle f, g \rangle_{\mathcal{H}_k} = f^T \mathbf{K}^{-1} g. \tag{5.63}$$

Proof. To prove that Equation (5.63) is in fact a finite dimensional RKHS inner product we must show that it satisfies the reproducing property. To do so fix $f \in \mathcal{H}_k$. Since $f \in \mathcal{H}_k$ and $\mathcal{H}_k$ is spanned by $k(\cdot, x_i)$, there is a vector α such that,

$$f = \mathbf{K}\alpha. \tag{5.64}$$

Given that,

$$k(\cdot, x_i) = \mathbf{K} e_i, \tag{5.65}$$

where e_i denotes the vector whose elements are zero except for the ith entry which is unity,

$$\langle f, k(\cdot, x_i) \rangle_{\mathcal{H}_k} = \langle \mathbf{K}\alpha, \mathbf{K}e_i \rangle = e_i^T \mathbf{K}\mathbf{K}^{-1}\mathbf{K}\alpha = e_i^T \mathbf{K}\alpha = e_i^T f, \tag{5.66}$$

$$\langle k(\cdot, x_i), k(\cdot, x_j) \rangle_{\mathcal{H}_k} = \langle \mathbf{K}e_i, \mathbf{K}e_j \rangle_{\mathcal{H}_k} = e_i^T \mathbf{K}\mathbf{K}^{-1}\mathbf{K}e_j = e_i^T \mathbf{K}e_j = \mathbf{K}_{ij}. \tag{5.67}$$

and the reproducing property is satisfied. □

Hence, to construct a finite dimensional RKHS inner product, it is necessary to define the matrix $\mathbf{K}$. Noting that the random variables defined on $L^2(\partial\mathcal{D})$ by the vector boundary value operator $\boldsymbol{\mathcal{B}}$ are evaluations of f and its derivatives, by the linearity of derivatives and that Gaussian process covariance functions are typically differentiable, the covariance between a derivative and function observation is,

$$\mathbb{E}\left[f'(x) f(y)\right] = \frac{\partial k(x, y)}{\partial x}, \tag{5.68}$$

$$\mathbb{E}\left[f'(x) f'(y)\right] = \frac{\partial^2 k(x, y)}{\partial x \partial y}. \tag{5.69}$$

. Consequently, we can construct elements of $\mathbf{K}$ as,

$$\mathbf{K}_{ij} = \left. \frac{\partial^{i+j}}{\partial x^i \partial y^j} k(x, y) \right|_{x=y \in \{a,b\}}. \tag{5.70}$$

Example 5.3.1. *For the Matérn class of kernel, $\mathbf{K}$ is given as,*

$$\mathbf{K}_{1/2} = \sigma^2, \tag{5.71}$$

$$\mathbf{K}_{3/2} = \sigma^2 \begin{bmatrix} 1 & 0 \\ 0 & \frac{3}{l^2} \end{bmatrix}, \tag{5.72}$$

$$\mathbf{K}_{5/2} = \sigma^2 \begin{bmatrix} 1 & 0 & -\frac{5}{3l^2} \\ 0 & \frac{5}{3l^2} & 0 \\ -\frac{5}{3l^2} & 0 & \frac{25}{l^4} \end{bmatrix}. \tag{5.73}$$

Noting that the covariance between a Gaussian process and its odd derivatives is always 0.

Inverting $\mathbf{K}$ we can now derive the RKHS inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}_{kh}}$ $\forall f, g \in \mathcal{H}_{kh}$, as,

$$\langle f, g \rangle_{\mathcal{H}_{kh}} = (\boldsymbol{\mathcal{B}} f)^T \mathbf{K}^{-1} \boldsymbol{\mathcal{B}} g. \tag{5.74}$$

Example 5.3.2. *For the Matérn class of kernel, $\langle f, g\rangle_{\mathcal{H}_{kh}}$ is given as*

$$\langle f, g\rangle_{\mathcal{H}_{k1/2}} = \frac{1}{\sigma^2} f(a)g(a), \tag{5.75}$$

$$\langle f, g\rangle_{\mathcal{H}_{k3/2}} = \begin{bmatrix} f(a) & f'(a) \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{l^2}{3\sigma^2} \end{bmatrix} \begin{bmatrix} g(a) \\ g'(a) \end{bmatrix}, \tag{5.76}$$

$$\langle f, g\rangle_{\mathcal{H}_{k5/2}} = \begin{bmatrix} f(a) & f'(a) & f''(a) \end{bmatrix} \begin{bmatrix} \frac{9}{8\sigma^2} & 0 & \frac{3l^2}{40\sigma^2} \\ 0 & \frac{3l^2}{5\sigma^2} & 0 \\ \frac{3l^2}{40\sigma^2} & 0 & \frac{9l^2}{200\sigma^2} \end{bmatrix} \begin{bmatrix} g(a) \\ g'(a) \\ g''(a) \end{bmatrix}. \tag{5.77}$$

Finally therefore we can construct the RKHS inner product for a one-dimensional bounded domain as the sum of $\langle \cdot, \cdot\rangle_{\mathcal{H}_{kp}} + \langle \cdot, \cdot\rangle_{\mathcal{H}_{kh}}$.

Theorem 5.3.2 (RKHS Inner Product on a 1D Bounded Domain)**.** *Let $\mathcal{D} = [a, b]$ be a one-dimensional bounded domain, $\mathcal{L}$ an mth order linear differential operator and $\boldsymbol{\mathcal{B}}$ the vector boundary value operator of length m. The RKHS generated by the solution to the stochastic (partial) differential equation,*

$$\left.\begin{array}{ll} \mathcal{L}f(x) = w(x) & \text{in } \mathcal{D}, \\ \boldsymbol{\mathcal{B}}f(x) = \mathbf{b} & \text{on } \partial\mathcal{D}, \end{array}\right\} \tag{5.78}$$

is given as,

$$\langle f, g\rangle_{\mathcal{H}_k} = \int_a^b \mathcal{L}f(\mathbf{x})\mathcal{L}g(\mathbf{x})d\mathbf{x} + (\boldsymbol{\mathcal{B}}f)^T \mathbf{K}^{-1} \boldsymbol{\mathcal{B}}g, \tag{5.79}$$

where the elements of K are,

$$\mathbf{K} = \left\{ \frac{\partial^{i+j}}{\partial x^i \partial y^j} k(x, y)\bigg|_{x=y} \right\}_{i,j=0,\ldots,m-1}. \tag{5.80}$$

Remark 5.3.4. This result is almost exactly equivalent to that of Parzen et al. [41] however, we derive the solution using a boundary value operator used by Ramsay and Dalzell [45] and Fasshauer and Ye [43].

Example 5.3.3. *For the Matérn class of kernels, we construct the RKHS inner proudct using the Matérn SDE,*

$$\left(\lambda + \frac{d}{dx}\right)^{\nu+1/2} f(x) = w(t). \tag{5.81}$$

The resulting RKHS inner products over a bounded domain are given by,

$$\langle f, g\rangle_{\mathcal{H}_{k1/2}} = \frac{l}{2\sigma^2} \int_b^a \left(\frac{1}{l}f + f'\right)\left(\frac{1}{l}g + g'\right) dx + \frac{1}{\sigma^2} f(a)g(a), \tag{5.82}$$

$$\begin{aligned} \langle f, g\rangle_{\mathcal{H}_{k3/2}} =& \frac{l^3}{12\sqrt{3}\sigma^2} \int_b^a \left(\frac{\sqrt{3}}{l}f + f'\right)^2 \left(\frac{\sqrt{3}}{l}g + g'\right)^2 dt + \frac{1}{\sigma^2} f(a)g(a) \\ &+ \frac{l^2}{3\sigma^2} f'(a)g'(a) \end{aligned}, \tag{5.83}$$

$$\begin{aligned} \langle f, g\rangle_{\mathcal{H}_{k5/2}} =& \frac{3l^5}{400\sqrt{5}\sigma^2} \int_a^b \left(\frac{\sqrt{5}}{l}f + f'\right)^3 \left(\frac{\sqrt{5}}{l}g + g'\right)^3 dx + \frac{9}{8\sigma^2} f(a)g(a) \\ &+ \frac{9l^4}{200\sigma^2} f(a)''g(a)'' + \frac{3l^2}{5\sigma^2} f'(a)g'(a) + \frac{3l^2}{40\sigma^2}\left(f''(a)g(a) + f(a)g''(a)\right) \end{aligned}. \tag{5.84}$$

Remark 5.3.5. The constant terms in front of the integral signs arise due to our assumption that the Matérn SPDE is proportional to the spectral density function in (4.19), and hence the constant reflects the value $1/q$ given by,

$$\frac{1}{q} = \left[\frac{2\sigma^2 \sqrt{\pi} \Gamma(\nu + \frac{d}{2})(2\nu)^\nu}{\Gamma(\nu) l^{2\nu}} \right]^{-1}. \tag{5.85}$$

Remark 5.3.6. These expressions are the same as presented in Durrande et al. [20] and Hensman et al. [1], derived using Green's functions. Hence, for proof of the reproducing property for the Matérn class of kernels we point the reader to Hensman et al. [1, Section 3.3]

Consequently, we have been able to derive closed form expressions for the RKHS inner product defined on a one-dimensional bounded domain.

High-Dimensional Case

Applying the Green's function method to construct high-dimensional RKHS inner products from *stochastic partial differential equations* is hard.

Firstly, in high dimensions such as the unit disk or unit ball, where $d \geq 2$, the space of functions on the boundary is no longer finite dimensional. This is due to the change of measure dS that defines the restricted space $L^2(\partial\mathcal{D})$. For $d \geq 2$, dS is the surface area element. Consequently, we are forced to consider in more detail the idea of a *boundary trace mapping* which restricts the derivative of a function $f \in \mathcal{H}^m(\mathcal{D})$ to the boundary trace $\partial\mathcal{D}$.

For any m times weakly differentiable function, such that it belongs to the Sobolev space $\mathcal{H}^m(\mathcal{D})$, the restriction of f and its derivatives on the boundary of the domain $\partial\mathcal{D}$ is called the trace of f on $\partial\mathcal{D}$. However, if $f \in \mathcal{H}^m(\mathcal{D})$, we can only assign boundary values to to the derivatives of f of order $m - 1/2$, since the mth order derivatives are simply L^2 functions which do not possess a trace [47, 48]. Thus, we can define a boundary trace operator as [48, Theorem 12.76],

Definition 5.3.3 (Boundary Trace Operator)**.** *Let $\mathcal{D} \subset \mathbb{R}^d$ be a bounded domain with Lipschitz boundary. There exists a surjective bounded linear operator τ called the boundary trace operator, defined as,*

$$\tau : \mathcal{H}^m(\mathcal{D}) \to \prod_{l=0}^{m-1} \mathcal{H}^{m-|l|-1/2}(\partial\mathcal{D}), \tag{5.86}$$

that associates every continuous function $f \in \mathcal{H}^m(\mathcal{D})$ with its restriction on the boundary $f_{\partial\mathcal{D}}$ such that $\forall f \in \mathcal{H}^m(\mathcal{D})$,

$$\tau f = f|_{\partial\mathcal{D}}. \tag{5.87}$$

Remark 5.3.7. We specify that the domain must be Lipschitz in order to promote a natural notion of a surface area element dS [49],

Consequently, for all cases where $d \geq 2$, the space of functions that exist on the boundary of the domain become increasingly abstract and difficult to work with as there is a *loss of* $1/2$ *a derivative*. For example, a function in $\mathcal{H}^1(\mathcal{D})$, takes on boundary values in the space $\mathcal{H}^{1/2}(\partial\mathcal{D})$ and we are left working with spaces of functions that have fractional derivatives. This makes specifying boundary conditions in order to construct RKHS inner products particularly difficult.

As a result, in this work we don't consider the case for constructing RKHS inner products on high-dimensional bounded domains using the Green's function method, deeming it a *no-go approach*

5.4 Discussion

In this section we have used connections between Green's functions, positive definite kernels and the RKHS in order to construct RKHS inner products for kernels of interest. We have also explored the domain in which we can define RKHS inner products on using Green's functions and SPDEs.

We showed that we can construct high-dimensional RKHS inner product from *stochastic partial differential equations* on an unbounded domain. Working with unbounded domains is naturally much easier than the bounded ones as we don't have to consider the effect of boundary conditions and therefore the space of functions which satisfy them. We also showed that we could construct RKHS inner products on a one-dimensional bounded domain $\mathcal{D} = [a, b]$, due to the measure of integration in the restricted space $L^2(\partial\mathcal{D})$ being defined only at the endpoints, such that $S(a) = S(b) = 1$. On the other hand, we also motivated the difficulty in specifying boundary conditions for domains where the dimensionality $d \geq 2$, due to the existence of fractional derivatives and increasingly abstract function spaces.

In the following chapter we will explore a second way of constructing RKHS inner products on bounded domains using the Variational formulation of SPDEs that allow us to write down specific boundary conditions such that we generate an RKHS inner product.

6 RKHS Inner Products and the Variational Formulation

In the previous chapter we detailed using Green's functions to solve SPDEs for a Gaussian process of interest. This approach took a *process perspective*, defining solutions to SPDEs as Gaussian processes via an integral equation. These integral forms then allowed us to construct functional form RKHS inner products using the Gaussian process solution to the SPDE. Indeed, this approach, constructing function spaces from Gaussian processes, worked well for unbounded domains. However when applied to bounded domains, application of the boundary conditions quickly becomes complicated. In particular, it was often very unclear what the actual boundary conditions were and how one should select the correct boundary conditions in order to derive an inner product that satisfies the reproducing property.

As introduced in Section 4.1.2, we can also approach solving SPDEs from a *function space perspective*, where the solution is instead an RKHS generated by an affine transformation of a zero-mean generalised Gaussian white noise field. Further, by Loeve's Isometry theorem, there is a one-to-one correspondence between the stochastic process solution and the RKHS found by viewing SPDEs from the function space persepctive. Hence, given the aim of this work is to construct RKHS inner products, it makes sense where possible to work from the function space perspective.

In this chapter we will derive RKHS inner products using the *variational formulation of elliptic boundary value problems*. The variational formulation relaxes the classical notion of a solution to a differential equation, allowing for solutions which aren't classically differentiable. We will first introduce some fundamental theory before applying the variational formulation to construct Matérn RKHS inner products. We note that the variational method results in an identical formulation of the RKHS inner product as the Green's function approach in Example 5.3.3 but provides a more obvious application of boundary conditions.

6.1 Distributions

Before we begin any discussion of the solution of boundary value problems via the variational formulation, it is necessary that we introduce the idea of *generalised functions* or *distributions*. Typically, we specify real-valued functions $f : \mathcal{X} \to \mathbb{R}$ as acting on points over some domain $\mathcal{X}$. However, the basic idea of distributions is to describe functions through their actions on a space of smooth test functions ϕ. The introduction of test functions enables us to significantly broaden our understanding of calculus.

Hence, let us first establish what we mean by test functions. Test functions ϕ are usually interpreted as belonging to the space of infinitely differentiable functions $C_0^\infty(\mathcal{D})$ with compact support [49, Definition 7.1].

Definition 6.1.1 (Test Function Space $C_0^\infty(\mathcal{D})$)**.** *Let $\mathcal{D} \in \mathbb{R}^d$ be an open domain. We denote the space of infinitely differentiable functions with compact (closed and bounded) support in $\mathcal{D}$ $C_0^\infty(\mathcal{D})$ as,*

$$C_0^\infty(\mathcal{D}) = \left\{ f \in C^\infty(\mathcal{D}) : supp(f) = \overline{\{\mathbf{x} \in \mathcal{D} : f(\mathbf{x}) \neq 0\}} \subset \mathcal{D} \right\}, \tag{6.1}$$

so that f and all its partial derivatives are zero outside of it's support.

Many standard functions and mathematical operations can be repostulated as acting on test functions (instead of their usual pointwise interpretation) by *integrating against a test function.* For example a continuous function $f : \mathbb{R} \to \mathbb{R}$ can be reinterpreted by *integrating against a test function* ϕ,

$$D(f) = \int_{\mathbb{R}} f(\mathbf{x})\phi(\mathbf{x})d\mathbf{x}, \tag{6.2}$$

as a mapping D whose domain is the space of test functions.

A *distribution* is defined as a linear functional on the space of test functions $C_0^\infty(\mathcal{D})$ [49, Definition 7.9]

Definition 6.1.2 (Distribution)**.** *Let $\mathcal{D} \in \mathbb{R}^d$ be an open domain and $C_0^\infty(\mathcal{D})$ the space of continuously differentiable test functions ϕ. A distribution on $\mathcal{D}$ is a linear functional on $C_0^\infty(\mathcal{D})$ such that $C_0^\infty(\mathcal{D})^*$ is the space of distributions on $\mathcal{D}$.*

Remark 6.1.1. Two distributions D_1 and D_2 are identical when their action on every test function is the same, such that $\forall \phi \in C_0^\infty(\mathcal{D})$,

$$\langle D_1, \phi \rangle = \langle D_2, \phi \rangle. \tag{6.3}$$

Having defined both test functions and distributions, we would now like to introduce the notion of *weak (or distributional) derivaties*, that are central to the theory of variational solutions to boundary value problems.

The main idea of weak derivatives is to transfer derivatives onto the test functions, as if one were to perform integration by parts. The formula for integration by parts for differentiable functions f and ϕ on a 1D bounded domain $\mathcal{D} = [a, b]$ is,

$$\int_a^b f(x)\phi'(x)dx = [u(x)\phi(x)]_a^b - \int_a^b f'(x)\phi(x)dx. \tag{6.4}$$

Substituting, $u \in C^1(\mathcal{D})$ and $\phi \in C_0^\infty(\mathcal{D})$ a test function, the boundary terms disappear as ϕ is compactly supported and hence vanishes near the boundary $\partial\mathcal{D}$. Hence,

$$\int_a^b f(x)\phi'(x)dx = -\int_a^b f'(x)\phi(x)dx, \tag{6.5}$$

where $f'(x)$ is defined as the weak derivative of $f(x)$ (In this case the weak derivative is also the strong derivative). More generally we define weak derivatives as [49, Definition 7.14],

Definition 6.1.3 (Weak Derivatives)**.** *Let $\mathcal{D} \in \mathbb{R}^d$ be an open domain and $C_0^\infty(\mathcal{D})$ the space of continuously differentiable test functions ϕ. Let $D \in C_0^\infty(\mathcal{D})^*$ be a distribution. The distributional derivative $\partial_x D$ is the distribution defined $\forall \phi \in C_0^\infty(\mathcal{D})$ as,*

$$\langle \partial_x D, \phi \rangle = -\langle F, \partial_x \phi \rangle, \tag{6.6}$$

generalising to distributional derivatives of any order $\forall \phi \in C_0^\infty(\mathcal{D})$ as,

$$\langle D^\alpha F, \phi \rangle = (-1)^{|\alpha|} \langle F, D^\alpha \phi \rangle, \tag{6.7}$$

where D^α is defined as,

$$D^\alpha \phi = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}. \tag{6.8}$$

Remark 6.1.2. If a function f is classically differentiable on $\mathcal{D}$ then the classical solution is also the weak derivative

In this section we have introduced weak distributions and weak derivatives which as we will see underpin the variational formulation of Boundary value problems. In the following section we present Sobolev spaces, another basic tool used to define the variational formulation of boundary value problems.

6.2 Lebesgue Spaces and Sobolev Spaces

In order to make precise the notion of a weak solution, we use Sobolev spaces, which are a fundamental tool in the treatment of boundary value problems. We define the Sobolev space $W^{m,p}$ as the space of functions who have weak derivatives of order m, belonging to various Lebesgue spaces L^p.

A Lebesgue space L^p consists of all measurable functions whose absolute value raised to the pth power has a finite integral [50, Definition 4.2].

Definition 6.2.1 (Lebesgue Space L^p)**.** *Let $\mathcal{D} \in \mathbb{R}^d$ be an open domain and $p \in [1, \infty)$. The Lebesgue space $L^p(\mathcal{D})$ is the space of functions which are lebesgue integrable to the power of p denoted by,*

$$L^p(\mathcal{D}) = \left\{ f : \int_{\mathcal{D}} |f|^p(x) d\mu(x) < \infty \right\}, \tag{6.9}$$

with respect to some measure μ, and equipped with the norm,

$$\|f\|_{L^p(\mathcal{D})} = \left(\int_{\mathcal{D}} |f|^p(x) d\mu(x) \right)^{1/p}. \tag{6.10}$$

Remark 6.2.1. We almost always consider the special case when $p = 2$, such that $L^2(\mathcal{D})$ is a Hilbert space. Note that $L^2(\mathcal{D})$ is the only Hilbert space among the L^p spaces. Functions belonging to $L^2(\mathcal{D})$ are often called square-integrable functions. The inner product defined by $L^2(\mathcal{D})$ is,

$$\langle f, g \rangle_{L^2} = \int_{\mathcal{D}} f(x) g(x) d\mu(x). \tag{6.11}$$

A Sobolev space $W^{m,p}$ is a Banach space equipped with a norm that is some combination of L^p-norms [50, Definition 8.2]

Definition 6.2.2 (Sobolev Space $W^{m,p}$)**.** *Let $\mathcal{D} \in \mathbb{R}^d$ be an open domain, m a positive integer and $p \in [1, \infty)$. The Sobolev space $W^{m,p}(\mathcal{D})$ is defined as,*

$$W^{m,p}(\mathcal{D}) = \{ u \in L^p(\mathcal{D}) : D^\alpha u \in L^p(\mathcal{D}) \, \forall \alpha \text{ with } |\alpha| \leq m \}, \tag{6.12}$$

equipped with the norm,

$$\|u\|_{m,p} = \left(\sum_{|\alpha| \leq m} \|D^\alpha u\|_{L^p(\mathcal{D})} \right)^{1/p}, \quad \text{if } 1 \leq p < \infty, \tag{6.13}$$

$$\|u\|_{m,\infty} = \max_{|\alpha| \leq m} \|D^\alpha u\|_{L^\infty(\mathcal{D})}. \tag{6.14}$$

Remark 6.2.2. When $p = 2$, the Sobolev space $H^m(\mathcal{D}) = W^{m,2}(\mathcal{D})$ is a Hilbert space equipped with inner product,

$$\langle f, g \rangle_{H^m(\mathcal{D})} = \langle f, g \rangle_{L^2} + \sum_{\alpha=1}^{m} \langle D^\alpha f, D^\alpha g \rangle_{L^2}, \tag{6.15}$$

$$= \int_{\mathcal{D}} f(\mathbf{x}) g(\mathbf{x}) d\mathbf{x} + \sum_{\alpha=1}^{m} \int_{\mathcal{D}} D^\alpha f(\mathbf{x}) D^\alpha g(\mathbf{x}) d\mathbf{x}, \tag{6.16}$$

with the associated norm,

$$\|f\|_{H^m} = \sqrt{\sum_{|\alpha| \leq m} \|D^\alpha u\|_{L^2(\mathcal{D})}}, \tag{6.17}$$

such that a function $f \in L^2(\mathcal{D})$ belongs to $f \in H^m(\mathcal{D})$ if all the derivatives of f up to order m, in the weak sense, belong to $L^2(\mathcal{D})$.

Of particular interest are the 1D Sobolev spaces $H^1(\mathcal{D})$ and $H^2(\mathcal{D})$, defined over over the interval $\mathcal{D} = [a, b] \subset \mathbb{R}$, equipped with inner products,

$$\langle f, g \rangle_{H^1} = \langle f, g \rangle_{L^2} + \langle f', g' \rangle_{L^2}, \tag{6.18}$$

$$= \int_a^b \big(f(x)g(x) + f'(x)g'(x) \big)\, dx, \tag{6.19}$$

$$\langle f, g \rangle_{H^2} = \langle f, g \rangle_{L^2} + \langle f', g' \rangle_{L^2} + \langle f'', g'' \rangle_{L^2}, \tag{6.20}$$

$$= \int_a^b \big(f(x)g(x) + f'(x)g'(x) + f''(x)g''(x) \big)\, dx. \tag{6.21}$$

We also introduce an important set of subspaces to $W^{m,p}(\mathcal{D})$, that are useful in solving boundary value problems.

Definition 6.2.3 (Sobolev Space $W_0^{m,p}(\mathcal{D})$)**.** *Let $\mathcal{D}$ be some domain, m a positive integer and $p \in \mathbb{R}$ with $1 \leq p \leq \infty$. Denote $W_0^{m,p}(\mathcal{D})$ as the closure of $C_0^\infty(\mathcal{D})$ in the norm of $W^{m,p}(\mathcal{D})$*

Remark 6.2.3. When $p = 2$, the Sobolev space $W_0^{m,2}(\mathcal{D}) = H_0^m(\mathcal{D}) \subset H^m(\mathcal{D})$ is defined as the closure of the space of infinitely differentiable functions $C_0^\infty(\mathcal{D})$ in $H^m(\mathcal{D})$.

For domains $\mathcal{D}$ with a Lipschitz boundary[1], $H_0^m(\mathcal{D})$ is the set of functions in $H^m(\mathcal{D})$ that are homogeneous (disappear) at the boundary $\partial\mathcal{D}$. The below result gives a basic characterization of functions in $W_0^{1,p}(\mathcal{D})$ from Brezis [50, Theorem 8.12]

Theorem 6.2.1. *Let $\mathcal{D}$ be a domain with Lipschitz boundary and $f \in W^{1,p}(\mathcal{D})$. Then $f \in W_0^{1,p}(\mathcal{D})$ if and only if $f = 0$ on $\partial\mathcal{D}$*

Consequently, we will use the definitions of both $\mathcal{H}^m(\mathcal{D})$ and $\mathcal{H}_0^m(\mathcal{D})$ in the following sections to precisely construct the variational formulation of boundary value problems.

6.3 Linear Variational Problems

So far we have introduced Sobolev spaces and weak derivatives which are our basic tool in making precise notions of weak solutions to boundary value problems. In this section we will introduce the *linear variational problem* which we wish to recast many classical boundary value problems as. The fundamental result however, that enables us to prove the existence and uniqueness of weak solutions to elliptic boundary value problems is the Lax-Milgram theorem.

The Lax-Milgram theorem is a type of representation theorem for bounded linear functionals on a Hilbert space $\mathcal{H}$. Indeed, to define the linear variational problem and the Lax-Milgram theorm we use *bilinear forms* on a Hilbert space $\mathcal{H}$.

Definition 6.3.1 (Bilinear form)**.** *Let $\mathcal{H}$ be a Banach space and $b(\cdot, \cdot) : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ a bilinear form on $\mathcal{H}$.*

1. *The bilinear form b is continuous if,*

$$|b(f, g)| \leq \alpha \|f\|_{\mathcal{H}} \cdot \|g\|_{\mathcal{H}} \quad \forall f, g \in \mathcal{H} \text{ and } \alpha > 0, \tag{6.22}$$

[1] An important feature of Lipschitz boundaries is that the outer normal vector at the surface is defined almost everywhere and is almost everywhere continuous. This allows us to apply Green's second identity when constructing the variational formulation

2. *The bilinear form is coercive if,*

$$b(f,g) \geq \beta \|f\|_{\mathcal{H}}^2 \quad \forall f \in \mathcal{H} \text{ and } \beta > 0. \tag{6.23}$$

Using the definition of a bilinear, form the fundamental linear variational problem is given as,

Definition 6.3.2 (Abstract Linear Variational Problem)**.** *Find $f \in \mathcal{H}$ such that,*

$$b(f,g) = \phi(g), \quad \forall g \in \mathcal{H}, \tag{6.24}$$

where b is a bilinear form and ϕ a linear functional.

In order to prove the existence and uniqueness of a solution $f \in \mathcal{H}$ we use the Lax-Milgram theorem, which can be seen as a generalisation of the Riesz representation theorem for *continuous* and *coercive* bilinear forms.

Theorem 6.3.1 (Lax-Milgram)**.** *Let $\mathcal{H}$ be a Hilbert space and $b(\cdot,\cdot) : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ a continuous coercive bilinear form on $\mathcal{H}$. For any bounded linear functional $\phi : \mathcal{H} \to \mathbb{R}$ there exists a unique $f \in \mathcal{H}$ such that,*

$$b(f,g) = \phi(g), \quad \forall g \in \mathcal{H} \tag{6.25}$$

. Moreover, if b is symmetric, then f is characterised by the property,

$$\frac{1}{2} b(f,f) - \phi(f) = \min_{g \in \mathcal{H}} \left[\frac{1}{2} b(g,g) - \phi(g) \right], \quad \forall g \in \mathcal{H}. \tag{6.26}$$

known as Dirichlet's principle.

Remark 6.3.1. Note that the bilinear form b does not need to be symmetric and therefore it doesn't necessarily define an inner product on $\mathcal{H}$. This is the difference between the Lax-Milgram theorem and the Riesz representation theorem. Consequently, given the Riesz representation theorem, when the bilinear form $b(\cdot,\cdot)$ is symmetric it defines an inner product on $\mathcal{H}$ and,

$$b(\cdot,\cdot) = \langle \cdot, \cdot \rangle_{\mathcal{H}}. \tag{6.27}$$

Therefore all *symmetric* bilinear forms on a Hilbert space $\mathcal{H}$ define an inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ on $\mathcal{H}$

Remark 6.3.2. Dirichlet's principle is the assumption that the solution f to a boundary value problem is the minimiser of an energy functional, where the bilinear form represents internal work done and the linear functional work done by external forces. The variational formulation in Equation (6.3.1) expresses that these forces must balance and hence the variational formulation is a form of the *principle of virtual work* [49]

The Lax-Milgram theorem is essential to solving partial differential equations for their weak solutions and we will use it to establish existence of solution and inner products on $\mathcal{H}$.

6.4 Solution Types

The running theme of this work is *how can we construct an RKHS inner product, corresponding to a chosen covariance function.* So far we have examined the relationship between Green's function and reproducing kernels in order to construct RKHS inner products in functional forms. Our method relied on constructing an integral kernel (Green's function) associated with some linear differential operator, to solve a *non-homogeneous boundary value problem.*

However, there exist many different ways of solving boundary value problems with varying notions of solution. In this section we will briefly discuss the *meaning of solution* before introducing the

variational formulation as an alternative method to solving boundary value problems that is useful in constructing RKHS inner products.

To introduce different types of solution, let us first define a general boundary value problem as,

$$\left.\begin{array}{ll} \mathcal{L}f = u & \text{in } \mathcal{D}, \\ f = v & \text{on } \partial\mathcal{D}, \end{array}\right\} \tag{6.28}$$

where $\mathcal{L}$ is a mth order elliptic partial differential operator, $u : \mathcal{D} \to \mathbb{R}$ is a real valued function inside the domain and $v : \partial\mathcal{D} \to \mathbb{R}$ is a real valued function on the boundary of the domain.

Firstly, although notions of solution differ, we can unambiguously define what it means to *solve* the above boundary value problem. We say a boundary value problem is *well-posed* if we can show *existence, regularity* and *stability* of the solution [49, 51]. These desirable properties don't however define what we mean by a *solution*.

Indeed, every problem can be constructed in several different ways, with each way having a different notion of solution. We will briefly consider different types of solution to the generic second-order elliptic boundary value problem defined above by Equation (6.41) [49]:

- **Classical** solutions are twice continuously differentiable functions satisfying (6.41), $f \in C^2(\mathcal{D})$, and the differential equation and boundary conditions are satisfied pointwise everywhere.
- **Strong** solutions are twice weakly differentiable functions satisfying (6.41), $f \in \mathcal{H}_0^2(\mathcal{D})$, relaxing the requirement that f be continuous everywhere, such that solutions belong to the Sobolev space $\mathcal{H}_0^2(\mathcal{D})$ and possess weak derivatives up to the second order in $L^2(\mathcal{D})$.
- **Weak** solutions further relax the requirements on f, such that the derivatives may not all exist. Weak solutions therefore belong to the Sobolev space $\mathcal{H}^{m-1}(\mathcal{D})$, i.e. $f \in \mathcal{H}^{m-1}(\mathcal{D})$.

Practically when solving boundary value problems, the classical and strong formulations are often too rigid, and many physical problems of interest don't have such solutions.

As a motivating example we consider the inviscid Burgers' equation. Burger's equation is a fundamental partial differential equation that can develop discontinuities, given by,

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = 0, \quad u(x, 0) = u_0, \tag{6.29}$$

such that the speed of the wave u at a given height is dependent on that height. Looking at the development of the solution u to Burger's equation over time subject to sinusoidal initial conditions, in Figure 6.29, we observe that the top of the wave moves faster than the bottom. For $0 \leq t < 0.15$ the solution is a classical solution as u is twice differentiable everywhere. However, at $t = 0.15$ the peak of the curve is directly above the bottom right edge, such that the wave front is vertical, creating a discontinuity where gradients no longer exist. Further, for $t > 0.15$, as the peak of the wave overtakes the bottom, the solution becomes multi valued and therefore isn't a function at all. This motivates the idea that we can find solutions to partial differential equations that are not always differentiable and not always functions.

Consequently, we would like to relax our solution requirements to consider a broader range of solutions. To do this we define weak solutions as solutions to the variational form of the partial differential equation. By relaxing the constraint on solutions being differentiable, weak solutions can be distributions whose derivatives don't exist in a classical sense, but are locally integrable.

Considering a second order PDE, we can also show using the Sobolev embedding theorem, that every classical solution is also a weak solution since,

$$\mathcal{H}^1(\mathcal{D}) \subset \mathcal{H}^2(\mathcal{D}) \subset C_2^\infty(\mathcal{D}). \tag{6.30}$$

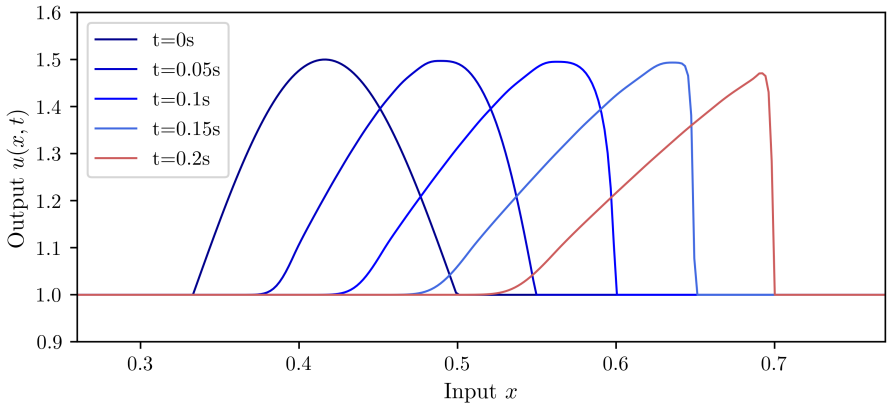

Figure 6.1: Burger's Equation

Hence every classical solution is a strong solution and every strong solution is a weak solution. Consequently a test for regularity of weak solutions is to ensure that the weak solution of a second order PDE is of class $C^2(\mathcal{D})$ and that we can recover a classical solution by showing that any weak solution $\in C^(\mathcal{D})$ is also a classical solution. Figure 6.2 visualises the hierarchy of solution spaces for a second order PDE.

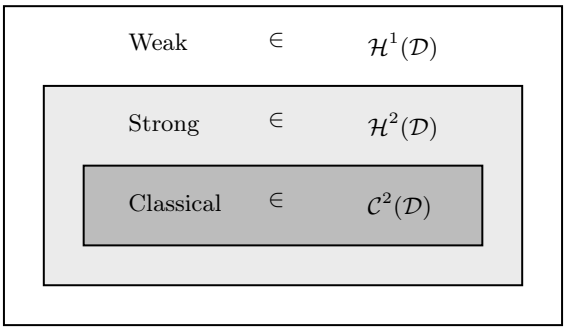

Figure 6.2: Hierarchy of solution function spaces according to the Sobolev embedding theorem

In the next section we will detail how to construct the variational formulation of a boundary value problem.

6.5 Variational Formulation of Boundary Value Problems

Having motivated the notion of weak solutions we would now like to recast general boundary value problems as *abstract linear variational problems* in order to find weak solutions.

To construct the variational formulation of a boundary value problem we follow the below steps [49, Page 509]:

1. Multiply the differential equation by a smooth test function, that is compatible with the boundary conditions, and integrate over the domain $\mathcal{D}$.

2. Transfer derivatives onto the test functions using integration by parts and the boundary conditions.

3. Ensure that the choice of test functions and variational formulation are correct by recovering the original formulation by integrating by parts in reverse. This checks that the classical solutions of both formulations are equivalent.

4. Interpret the variational formulation as an abstract linear variational problem 6.3.2 in a suitable Hilbert space. Generally the Hilbert space is a Sobolev space.

Having constructed the variational formulation, one can then solve the abstract linear variational problem using the Lax-Milgram theorem 6.3.1.

As an example we will consider the Sturm-Lioville boundary value problem.

6.5.1 Sturm-Liouville Boundary Value Problem

Consider the general 2nd order Sturm-Liouville boundary value problem over the finite interval $\mathcal{D} = [a, b]$ [52, Page 302],

$$\left.\begin{aligned} -(pf')' + qf &= w && \text{in } \mathcal{D}, \\ \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \begin{bmatrix} f(a) \\ pf'(a) \end{bmatrix} + \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} \begin{bmatrix} f(b) \\ pf'(b) \end{bmatrix} &= 0 && \text{on } \partial\mathcal{D}, \end{aligned}\right\}, \tag{6.31}$$

where $p \in C^1(\mathcal{D})$, $q \in C(\mathcal{D})$, $f \in L^2(\mathcal{D})$ and the boundary conditions are given as a system of linear equations such that,

$$\operatorname{rank} \begin{bmatrix} \alpha_{11} & \alpha_{12} & \beta_{11} & \beta_{12} \\ \alpha_{21} & \alpha_{22} & \beta_{21} & \beta_{22} \end{bmatrix} = 2. \tag{6.32}$$

where the α's and β's are real.

A classical solution of Equation (6.31) is a function $f \in C^2(\mathcal{D})$. We now wish to construct the variational formulation following the above steps:

1. First we define the space of test functions as the closed convex set K in the space $H^1(\mathcal{D})$,

$$K = \left\{ g \in H^1(\mathcal{D}); \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \begin{bmatrix} v(a) \\ pv'(a) \end{bmatrix} + \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} \begin{bmatrix} v(b) \\ pv'(b) \end{bmatrix} = 0 \right\}, \tag{6.33}$$

such that the test functions have continuous first order derivatives and satisfy the boundary conditions. Let $g \in K$. Multiplying the differential equation by g and integrating over the domain $\mathcal{D}$,

$$\int_a^b \left[-(pf')' + qf \right] g = \int_a^b wg. \tag{6.34}$$

2. We now expand the first term using integration by parts,

$$\int_a^b -(pf')'g = \int_a^b pf'g' - pf'(x)g(x)\Big|_a^b, \tag{6.35}$$

introducing the boundary conditions to obtain,

$$\int_a^b pf'g' + \int_a^b qfg - pf'(x)g(x)\Big|_a^b = \int_a^b wg. \tag{6.36}$$

We can rewrite,

$$-\,pf'(x)g(x)\Big|_a^b = \begin{bmatrix} g(a) & g(b) \end{bmatrix} \begin{bmatrix} pf'(a) \\ -pf'(b) \end{bmatrix}. \tag{6.37}$$

Rearranging the boundary conditions and introducing a minus sign in front of the $pf'(b)$ term we can write,

$$\begin{bmatrix} \alpha_{11} & \beta_{11} \\ \alpha_{21} & \beta_{21} \end{bmatrix} \begin{bmatrix} f(a) \\ f(b) \end{bmatrix} = \begin{bmatrix} -\alpha_{12} & \beta_{12} \\ -\alpha_{22} & \beta_{22} \end{bmatrix} \begin{bmatrix} pf'(a) \\ -pf'(b) \end{bmatrix}. \tag{6.38}$$

Assuming the RHS coefficient matrix is non-singular and without loss of generality that its determinant $\alpha_{22}\beta_{12} - \alpha_{12}\beta_{22} = 1$, then,

$$\begin{bmatrix} pf'(a) \\ -pf'(b) \end{bmatrix} = \begin{bmatrix} \alpha_{11}\beta_{22} - \alpha_{21}\beta_{12} & \beta_{11}\beta_{22} - \beta_{12}\beta_{22} \\ \alpha_{11}\alpha_{22} - \alpha_{12}\alpha_{21} & \beta_{11}\alpha_{22} - \beta_{21}\alpha_{12} \end{bmatrix} \begin{bmatrix} f(a) \\ f(b) \end{bmatrix}. \tag{6.39}$$

Substituting the boundary conditions we can now state the variational formulation of the Sturm-Liouville boundary value problem,

$$\begin{aligned} \int_a^b pf'g' + \int_a^b qfg + [g(a)\ g(b)] \begin{bmatrix} \alpha_{11}\beta_{22} - \alpha_{21}\beta_{12} & \beta_{11}\beta_{22} - \beta_{12}\beta_{22} \\ \alpha_{11}\alpha_{22} - \alpha_{12}\alpha_{21} & \beta_{11}\alpha_{22} - \beta_{21}\alpha_{12} \end{bmatrix} \begin{bmatrix} f(a) \\ f(b) \end{bmatrix} \\ = \int_a^b wg \quad \forall g \in H^1(\mathcal{D}) \end{aligned}. \tag{6.40}$$

Remark 6.5.1. Here we see the key concept to the weak formulation. We have shown that when solving a second order differential equation in f (6.31), we can reformulate the problem by introducing a test function g, such that, via integration by parts, the solutions f don't have to be twice continuously differentiable. The weak formulation therefore generalises (6.31) to include solutions which are not necessarily classical solutions.

3. Assuming Equation (6.40) is correct, integrating by parts in reverse we return to Equation (6.34). Hence, for classical solutions both formulations (6.33) and (6.40) are equivalent.

4. The final step is to define the linear variational problem,

$$\left.\begin{array}{l} \textit{Determine } u \in \mathcal{H}^1(\mathcal{D}) \textit{ such that} \\ b(f,g) = \phi(g), \quad \forall g \in \mathcal{H}^1(\mathcal{D}), \end{array}\right\} \tag{6.41}$$

where the bilinear form is given as,

$$b(f,g) = \int_a^b pf'g' + \int_a^b qfg + [g(a)\ g(b)] \begin{bmatrix} \alpha_{11}\beta_{22} - \alpha_{21}\beta_{12} & \beta_{11}\beta_{22} - \beta_{12}\beta_{22} \\ \alpha_{11}\alpha_{22} - \alpha_{12}\alpha_{21} & \beta_{11}\alpha_{22} - \beta_{21}\alpha_{12} \end{bmatrix} \begin{bmatrix} f(a) \\ f(b) \end{bmatrix}, \tag{6.42}$$

and the linear functional,

$$\phi(g) = \int_a^b wg. \tag{6.43}$$

As a specific example, consider the case for Dirichlet Boundary conditions where $\alpha_{11} = \beta_{21} = 1$ and all other terms equal 0. The bilinear form is given by,

$$b(f,g) = \int_a^b pf'g' + \int_a^b qfg. \tag{6.44}$$

Note that the bilinear form is symmetric, and is therefore an inner product on $\mathcal{H}^1(\mathcal{D})$,

$$\langle f, g \rangle_{H^1(\mathcal{D})} = \int_a^b pf'g' + \int_a^b qfg, \tag{6.45}$$

for functions that vanish at a and b.

6.6 Variational Formulation and the RKHS

We would now like to use the variational formulation of boundary value problems to construct RKHS inner products. To do so we must establish two things: 1) The existence and uniqueness of solutions to stochastic boundary value problems and 2) connections between the variational formulation and the RKHS under these conditions.

Consider the boundary value problem,

$$\left.\begin{aligned} &\lambda f - \Delta u = w \quad \text{in } \mathcal{D}, \\ &f(a) = 0, \\ &f(b) = 0, \end{aligned}\right\} \tag{6.46}$$

where w is a Gaussian white noise process. For samples of w the variational form of Equation (6.46) is to find a function $f \in \mathcal{H}_0^1(\mathcal{D})$ that satisfies,

$$\int_a^b f'g'dx + \lambda \int_a^b fg\, dx = \int_a^b wg\, dx, \tag{6.47}$$

where w is a Gaussian white noise process. However, applying Lax-Milgram as defined in Theorem 6.3.1 to Equation (6.47) will provide us with a unique solution $f \in \mathcal{H}_0^1(\mathcal{D})$ for every unique sample of w and therefore Lax-Milgram's theorem does not guarantee uniqueness of solution for a stochastic forcing function, where we expect our solution to be a uniquely defined stochastic process. Indeed, the availability of general existence and uniqueness theorems are hard to come by for SPDEs [23].

We can however, still define uniqueness of solution by considering the one-to-one mapping between Gaussian processes and generalized fields. By considering generalized Gaussian fields we are able to use results from section 4.1.2. The first step to establishing uniqueness and existence of solutions to the stochastic boundary value problem (6.47) is to define what we mean by zero-mean generalized Gaussian white noise in this context [23, Proposition 4.2.4].

Proposition 6.6.1. *Let $\mathcal{D} \subset \mathbb{R}^d$ be a bounded domain with sufficiently regular boundary and $\hat{W}$ a generalized Gaussian white noise field on $L^2(\mathcal{D})$. Then,*

$$\hat{W} \in L^2\left(\mathcal{H}^{-\gamma}(\mathcal{D})\right), \tag{6.48}$$

if and only if $\gamma > d/2$.

Remark 6.6.1. From this definition we can rewrite the RHS of (6.47) as,

$$\int_a^b w(x)g(x)\, dx = \int_a^b g(x)\, dW(x) = \hat{W}(g). \tag{6.49}$$

Hence, the Sobolev space regularity of the solution to Equation (6.47), as derived by Lototsky and Rozovsky [23, Theorem 4.2.5], is given as,

Theorem 6.6.1. *Let $\mathcal{L}$ be a linear second-order partial differential operator on $\mathcal{D} \subset \mathbb{R}^d$. Suppose that, for every $w \in \mathcal{H}^\gamma$, the deterministic equation $\mathcal{L}g = w$ has a unique solution $g \in \mathcal{H}^{\gamma+2}$. Then equation $\mathcal{L}f = \hat{W}$ has a unique solution,*

$$f \in L^2\left(\mathcal{H}^{-\gamma+2}(D)\right). \tag{6.50}$$

Consequently, we have established uniqueness of solution that allows the solution to be some stochastic process in $L^2(\mathcal{H}^{-\gamma+2}(\mathcal{D}))$

We would now like to consider the case where the bilinear form b is an RKHS inner product such that,

$$b(f,g) = \langle f, g\rangle_{\mathcal{H}_k} = \hat{F}(\mathcal{T}g) = \hat{W}(g), \tag{6.51}$$

according to Equation (4.4), where $\mathcal{T}$ is the map from $\mathcal{H}^{-\gamma}(\mathcal{D})$ to $\mathcal{H}^{-\gamma+2}$. Given that the variational formulation reformulates the differential equation as an integral equation against a chosen space of test functions, the RKHS that corresponds to the weak solution of interest is

given by $\mathcal{T}g \in \mathcal{H}^{-\gamma+2}$. Consequently, we have shown that if the bilinear form is symmetric, such that it is an inner product on $\mathcal{H}$, and if for some reproducing kernel it satisfies the reproducing property, then it is a unique weak solution to the linear variational problem.

As a result, we wish to construct the variational formulation subject to particular boundary conditions such that the bilinear form is symmetric and satisifies the reproducing property for a chosen kernel. To do this we construct the variational formulation for a linear partial differential operator $\mathcal{L}$ associated with a reproducing kernel k via Bochner's theorem, choosing a specific space of test functions for which the bilinear form exists on.

We note that this method, unlike the Green's function approach, provides a far more intuitive application of the boundary conditions, to the point that we aren't required to specify a Boundary value operator or construct a finite-dimensional RKHS.

A technicality of this method however, is that in order for the bilinear form to be symmetric, the space of test functions must be half as differentiable as the classical solutions to the SPDE. This happens because by transferring derivatives from the unknown Gaussian process f onto the test functions g, for an mth order differential operator, we reduce the requirement that f be m times continuously differentiable. A result of this effect is, for example, to find the Matérn 1/2 RKHS inner product, we instead construct the variational formulation for the Matérn 3/2 SPDE.

The remainder of this section will construct RKHS inner products using the variational formulation of the Matérn SPDE in one-dimension.

6.6.1 Matérn Class

We remind the reader of the classical form of the Matérn SPDE from Theorem 4.3.2 in one-dimension,

$$\left(\lambda^2 - \frac{d^2}{dx^2}\right)^{\frac{\nu+d/2}{2}} f(x) = w(x). \tag{6.52}$$

Matérn $v = 1/2$

In order to construct the RKHS inner product corresponding to the Matérn 1/2 kernel we begin by considering the Matérn 3/2 boundary value problem on a one-dimensional $\mathcal{D} = [a, b]$,

$$\lambda^2 f - f'' = w \quad \text{in } \mathcal{D}. \ \Big\} \tag{6.53}$$

A Gaussian process with Matérn 1/2 covariance is once differentiable in the mean-square sense. Hence, we define the space of test functions as $g \in \mathcal{H}_1(\mathcal{D})$ subject to some yet to be specified boundary conditions.

The next step is to multiply the Matérn 3/2 SPDE by the test functions g and integrate by parts,

$$\begin{aligned}
\int_a^b \left(\lambda^2 f - f''\right) g \, dx &= -\int_a^b f'' g \, dx + \lambda^2 \int_a^b fg \, dx, && (6.54)\\
&= \int_a^b f' g' \, dx + \lambda^2 \int_a^b + fg \, dx - \left[f'(x) g(x)\right]_a^b, && (6.55)\\
&= \int_a^b f' g' \, dx + \lambda^2 \int_a^b fg \, dx + f'(a) g(a) - f'(b) g(b), && (6.56)
\end{aligned}$$

such that,

$$\int_a^b f' g' \, dx + \lambda^2 \int_a^b fg \, dx + f'(a) g(a) - f'(b) g(b) = \int_a^b wg \, dx. \tag{6.57}$$

We must now define a set of boundary conditions in order for the bilinear form to be symmetric,

$$\left.\begin{array}{ll} f'(a) - \alpha f(a) & = 0, \\ f'(b) + \beta f(b) & = 0. \end{array}\right\} \tag{6.58}$$

Applying the boundary conditions, the variational formulation is symmetric as,

$$\int_a^b f'g'\,dx + \lambda^2 \int_a^b fg\,dx + \alpha f(a)g(a) + \beta f(b)g(b) = \int_a^b wg\,dx, \tag{6.59}$$

and the bilinear form defines an inner product on $\mathcal{H}^1(\mathcal{D})$,

$$\langle f, g\rangle_{\mathcal{H}^1} = \int_a^b f'g'\,dx + \lambda^2 \int_a^b fg\,dx + \alpha f(a)g(a) + \beta f(b)g(b). \tag{6.60}$$

We now would like to define the constants α and β such that the bilinear form satisfies the reproducing property for the Matérn $1/2$ kernel,

$$k_{1/2}(x-y) = \sigma^2 \exp\left(-\lambda|x-y|\right)), \tag{6.61}$$

such that,

$$\alpha = \beta = \frac{1}{2\sigma^2}, \tag{6.62}$$

and the RKHS inner product corresponding to the Matérn $1/2$ kernel is given by,

$$\langle f, g\rangle_{\mathcal{H}_{k1/2}} = \frac{1}{2\sigma^2\lambda}\int_a^b f'g'\,dx + \frac{\lambda}{2\sigma^2}\int_a^b fg\,dx + \frac{1}{2\sigma^2}\left[f(a)g(a) + f(b)g(b)\right], \tag{6.63}$$

where we again remember to multiply by the constant $1/q$.

Remark 6.6.2. We note that this form is completely identical to that found using Green's functions, which we can prove via integration by parts,

$$\begin{aligned} \int_a^b (\lambda f + f')(\lambda g + g')dx &= \int_a^b \left(f'g' + \lambda^2 fg\right)dx + \lambda\int_a^b \left(fg' + f'g\right)dx, && (6.64)\\ &= \int_a^b \left(f'g' + \lambda^2 fg\right)dx + \lambda[fg]_a^b, && (6.65)\\ &= \int_a^b \left(f'g' + \lambda^2 fg\right)dx + \lambda[f(b)g(b) - f(a)g(a)]. && (6.66) \end{aligned}$$

Hence substituting in,

$$\langle f, g\rangle_{\mathcal{H}_{k1/2}} = \frac{1}{2\lambda\sigma^2}\int_a^b \left(\lambda f + f'\right)\left(\lambda g + g'\right)dx + \frac{1}{\sigma^2}f(a)g(a). \tag{6.67}$$

Matérn 3/2

Just like for the previous example, in order to find the Matérn $3/2$ kernel, we begin by considering the Matérn $5/2$ boundary value problem over $\mathcal{D} = [a, b]$

$$\lambda^4 f - 2\lambda^2 f'' + f'''' = w \quad \text{in } \mathcal{D}\Big\}, \tag{6.68}$$

Given that a Gaussian process with Matérn $3/2$ kernel is twice differentiable we select the space of test functions $g \in \mathcal{H}^2(\mathcal{D})$. The variational formulation is constructed by multiplying by g and

integrating by parts, such that,

$$\begin{aligned} b(f,g) = &\int_a^b f''g''\,dx + 2\lambda^2\int_a^b f'g'\,dx + \lambda^4\int_a^b fg\,dx \\ &- 2\lambda^2 f'(b)g(b) + 2\lambda^2 f'(a)g(a) \\ &- f(b)''g'(b) + f(a)''g'(a) \\ &+ f(b)'''g(b) - f(a)'''g(a) \end{aligned} \quad . \tag{6.69}$$

To ensure the bilinear form is symmetric we apply the following boundary conditions,

$$\left.\begin{array}{rl} f(a)'' - \alpha f(a)' & = 0, \\ f(b)'' + \beta f(b)' & = 0, \\ f(b)''' - 2\lambda^2 f'(b) - \gamma f(b) & = 0, \\ f(a)''' - 2\lambda^2 f'(a) + \epsilon f(a) & = 0, \end{array}\right\} \tag{6.70}$$

such that,

$$\begin{aligned} b(f,g) = &\int_a^b f''g''\,dx + 2\lambda^2\int_a^b f'g'\,dx + \lambda^4\int_a^b fg\,dx \\ &+ \begin{bmatrix} f(a) & f'(a) \end{bmatrix} \begin{bmatrix} \epsilon & 0 \\ 0 & \alpha \end{bmatrix} \begin{bmatrix} g(a) \\ g'(a) \end{bmatrix} \\ &+ \begin{bmatrix} f(b) & f'(b) \end{bmatrix} \begin{bmatrix} \gamma & 0 \\ 0 & \beta \end{bmatrix} \begin{bmatrix} g(b) \\ g'(b) \end{bmatrix}. \end{aligned} \tag{6.71}$$

Defining the constants, we can construct the RKHS inner product corresponding to the Matérn 3/2 kernel as,

$$\begin{aligned} \langle f,g\rangle_{\mathcal{H}_{k3/2}} = &\frac{1}{4\lambda^3\sigma^2}\int_a^b f''g''\,dx + \frac{1}{2\lambda\sigma^2}\int_a^b f'g'\,dx + \frac{\lambda}{4\sigma^2}\int_a^b fg\,dx \\ &+ \begin{bmatrix} f(a) & f'(a) \end{bmatrix} \begin{bmatrix} \frac{1}{2\sigma^2} & 0 \\ 0 & \frac{1}{\lambda^2\sigma^2} \end{bmatrix} \begin{bmatrix} g(a) \\ g'(a) \end{bmatrix} \\ &+ \begin{bmatrix} f(b) & f'(b) \end{bmatrix} \begin{bmatrix} \frac{1}{2\sigma^2} & 0 \\ 0 & \frac{1}{\lambda^2\sigma^2} \end{bmatrix} \begin{bmatrix} g(b) \\ g'(b) \end{bmatrix}. \end{aligned} \tag{6.72}$$

Matérn 5/2

To find the Matérn 5/2 RKHS inner product we consider the Matérn 3/2 boundary value problem over $\mathcal{D} = [a,b]$,

$$\left.\lambda^6 f - 3\lambda^4 f^{(2)} + 3\lambda^2 f^{(4)} + \lambda^2 f^{(6)} = w \quad \text{in } \mathcal{D}\right\}, \tag{6.73}$$

We choose the space of test functions to be $\mathcal{H}^3(\mathcal{D})$. The variational formulation is given as,

$$\begin{aligned} b(f,g) = &\int_a^b f'''g'''\,dx + 3\lambda^2\int_a^b f''g''\,dx + 3\lambda^4\int_a^b f'g'\,dx + \lambda^6\int_a^b fg\,dx \\ &- 3\lambda^4 f'(b)g(b) + 3\lambda^4 f'(a)g(a) \\ &+ 3\lambda^2 f'''(b)g(b) - 3\lambda^2 f'''(a)g(a) \\ &- 3\lambda^2 f''(b)g'(b) + 3\lambda^2 f''(a)g'(a) \\ &+ f''''(b)g(b) - f''''(a)g(a) \\ &- f''''(b)g'(b) + f''''(a)g'(a) \\ &+ f'''(b)g''(b) - f'''(a)g''(a), \end{aligned} \tag{6.74}$$

such that we can group terms together as,

$$\begin{aligned} b(f,g) = &\int_a^b f'''g''' + 3\lambda^2 \int_a^b f''g'' + 3\lambda^4 \int_a^b f'g' + \lambda^6 \int_a^b fg \\ &+ 3\lambda^2 f'''(b)g(b) - 3\lambda^2 f'''(a)g(a) \\ &+ \left[f''''(b) - 3\lambda^4 f'(b)\right] g(b) - \left[f''''(a) - 3\lambda^4 f'(a)\right] g(a) \\ &- \left[f''''(b) + 3\lambda^2 f''(b)\right] g'(b) + \left[f''''(a) + 3\lambda^2 f''(a)\right] g'(a) \\ &+ f'''(b)g''(b) - f'''(a)g''(a). \end{aligned} \tag{6.75}$$

In order for the bilinear form to be symmetric we define the boundary conditions as,

$$\left.\begin{aligned} f'''(b) - 3\lambda^2\beta_1 f(b) - \beta_1 f''(b) &= 0, \\ f'''(a) + 3\lambda^2\alpha_1 f(a) + \alpha_1 f''(a) &= 0, \\ f''''(b) + 3\lambda^2 f''(b) + \beta_2 f'(b) &= 0, \\ f''''(a) + 3\lambda^2 f''(a) - \alpha_2 f'(a) &= 0, \\ f''''(b) - 3\lambda^4 f'(b) - \beta_3 f(b) &= 0, \\ f''''(a) - 3\lambda^4 f'(a) + \alpha_3 f(a) &= 0. \end{aligned}\right\} \tag{6.76}$$

Substituting in boundary conditions and constructing the bilinear form,

$$\begin{aligned} b(f,g) = &\int_a^b f'''g''' + 3\lambda^2 \int_a^b f''g'' + 3\lambda^4 \int_a^b f'g' + \lambda^6 \int_a^b fg \\ &+ \begin{bmatrix} g(a) & g'(a) & g''(a) \end{bmatrix} \begin{bmatrix} 9\lambda^4\alpha_1 + \alpha_3 & 0 & 3\lambda^2\alpha_1 \\ 0 & \alpha_2 & 0 \\ 3\lambda^2\alpha_1 & 0 & \alpha_1 \end{bmatrix} \begin{bmatrix} f(a) \\ f'(a) \\ f''(a) \end{bmatrix} \\ &+ \begin{bmatrix} g(b) & g'(b) & g''(b) \end{bmatrix} \begin{bmatrix} 9\lambda^4\beta_1 + \beta_3 & 0 & 3\lambda^2\beta_1 \\ 0 & \beta_2 & 0 \\ 3\lambda^2\beta_1 & 0 & \beta_1 \end{bmatrix} \begin{bmatrix} f(b) \\ f'(b) \\ f''(b) \end{bmatrix}. \end{aligned} \tag{6.77}$$

Defining constants such that the reproducing property holds, the RKHS inner product is given as,

$$\begin{aligned} \langle f,g\rangle_{\mathcal{H}_{k_{5/2}}} = &\frac{3}{16\lambda^5\sigma^2} \int_a^b f'''g''' + \frac{9}{16\lambda^3\sigma^2} \int_a^b f''g'' + \frac{9}{16\lambda\sigma^2} \int_a^b f'g' + \frac{3\lambda}{16\sigma^2} \int_a^b fg \\ &+ \begin{bmatrix} g(a) & g'(a) & g''(a) \end{bmatrix} \begin{bmatrix} \frac{9}{16\sigma^2} & 0 & \frac{3}{16\lambda^2\sigma^2} \\ 0 & \frac{3}{2\lambda^2\sigma^2} & 0 \\ \frac{3}{16\lambda^2\sigma^2} & 0 & \frac{9}{16\lambda^4\sigma^2} \end{bmatrix} \begin{bmatrix} f(a) \\ f'(a) \\ f''(a) \end{bmatrix} \\ &+ \begin{bmatrix} g(b) & g'(b) & g''(b) \end{bmatrix} \begin{bmatrix} \frac{9}{16\sigma^2} & 0 & \frac{3}{16\lambda^2\sigma^2} \\ 0 & \frac{3}{2\lambda^2\sigma^2} & 0 \\ \frac{3}{16\lambda^2\sigma^2} & 0 & \frac{9}{16\lambda^4\sigma^2} \end{bmatrix} \begin{bmatrix} f(b) \\ f'(b) \\ f''(b) \end{bmatrix}. \end{aligned} \tag{6.78}$$

6.7 Higher Dimensional Variational formulation

Unlike the Green's function approach, the variational formulation lends itself to constructing higher dimensional bilinear forms by a more direct application of Green's identities. Indeed, for problems where $d = 2$ or $d = 3$, instead of using integration by parts to transfer derivatives onto test functions we can instead apply Green's 1st Identity (5.3) which can be seen as a higher dimensional equivalent to integration by parts,

$$\int_{\mathcal{D}} (\Delta f) g dV + \int_{\mathcal{D}} (\nabla f)(\nabla g) dV = \oint_{\partial\mathcal{D}} \frac{\partial f}{\partial n} g dS. \tag{6.79}$$

This method is still not perfect however, as introduces a normal derivative which is equally as hard to work with and evaluate.

6.7.1 Matérn Variational Formulation in Higher Dimensions

However, there exists a larger obstacle when considering the differentiability of higher dimensional Matérn kernels.

Firstly, let us define the Matérn $\nu = 1/2$ partial differential operator $\mathcal{L}$ in 3 dimensions,

$$\mathcal{L}f(\mathbf{x}) = (\lambda^2 - \Delta)^{\frac{1}{4}+\frac{3}{4}} f(\mathbf{x}), \tag{6.80}$$
$$= (\lambda^2 - \Delta)^2 f(\mathbf{x}), \tag{6.81}$$
$$= \lambda^4 f(\mathbf{x}) - 2\lambda^2 \Delta f(\mathbf{x}) + \Delta(\Delta f(\mathbf{x})), \tag{6.82}$$

where we define $\Delta(\Delta f(\mathbf{x})$ as the biharmonic operator,

$$\Delta(\Delta f) = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)\left(\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2}\right), \tag{6.83}$$
$$= \frac{\partial^4 f}{\partial x^4} + \frac{\partial^4 f}{\partial y^4} + \frac{\partial^4 f}{\partial z^4} + 2\frac{\partial^4 f}{\partial x^2 \partial y^2} + 2\frac{\partial^4 f}{\partial y^2 \partial z^2} + 2\frac{\partial^4 f}{\partial x^2 \partial z^2}, \tag{6.84}$$
$$= \nabla^4 f. \tag{6.85}$$

Consequently, the Matérn 1/2 SPDE is a 4th order differential equation,

$$\mathcal{L}f(\mathbf{x}) = \lambda^4 f(\mathbf{x}) - 2\lambda^2 \nabla^2 f(\mathbf{x}) + \nabla^4 f(\mathbf{x}) \tag{6.86}$$

This is confusing since a three-dimensional Gaussian process with a Matérn 1/2 kernel is still non-differentiable and therefore the solution to its SPDE should be once differentiable. Further, this contradiction also affects the ability to define an RKHS inner product via the Variational formulation. Namely, that one cannot construct a symmetric bilinear form using the space of test functions that corresponds to the differentiability requirements of the Matérn kernel of interest.

Consequently, this presents yet another roadblock to constructing higher dimensional RKHS inner products on a bounded domain.

6.8 Discussion

In this section we have presented an alternative method to deriving RKHS inner products via the weak solution of boundary value problems. There are several benefits to this approach.

Firstly, it enables us to deal directly with the RKHS inner product when solving for the weak solution, unlike the Green's function approach which requires us to consider the solution from a Gaussian process perspective before constructing an RKHS inner product. Secondly, it enables us to define distinct and obvious boundary conditions, which again differs to the Green's function method where it wasn't obvious what boundary conditions we were applying.

We do note there are some idiosyncrasies however. In particular, the need to work with higher order SPDEs, in order to derive a desired RKHS inner product, as we transfer derivatives onto the space of test functions.

Unfortunately, similar to the Green's function method we run into similar issues when trying to derive higher-dimensional RKHS inner products on bounded domains. Further, we also question the differentiability of solutions to higher dimensional Matérn SPDEs. This raises the question, given an mth order d dimensional SPDE, how many times differentiable in the mean-square sense is it's Gaussian process.

7 Discussion

7.1 Availability of RKHS Inner Product

In this project we have presented a number of relationships between Gaussian processes, SPDEs and the RKHS, with a focus on constructing RKHS inner products. Using both Green's functions and the variational formulation of boundary value problems we have been able to establish functional form RKHS inner products for both one-dimensional unbounded and bounded domains and high-dimensional unbounded domains. Hence we can begin to classify when and where we can use RKHS projections to construct interdomain inducing points, visualised in Figure 7.2.

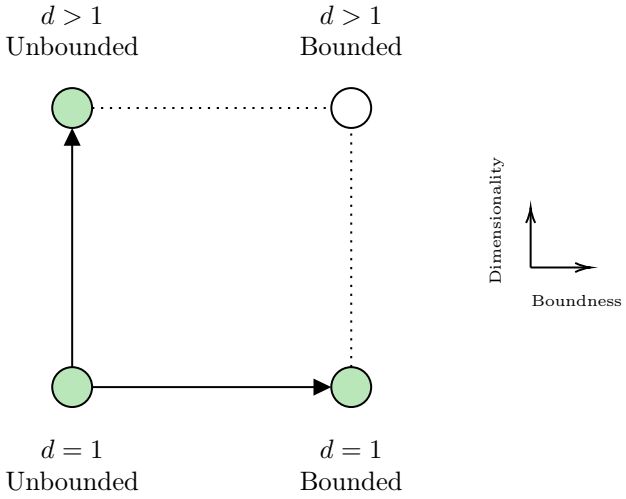

Figure 7.1: Square showing the relationships between different domains and the ability to construct an RKHS inner product. The y axis represents the dimensionality of the domain and the x axis the boundness. Arrows represent where relationships have been found.

However, we haven't been able to construct an RKHS inner product on high-dimensional bounded domains due to difficulty in defining the boundary conditions. This is in part due to limitations in theory of stochastic partial differential equations. Indeed we quote an extract from Lototsky and Rozovsky [23, Page 171] on the existence and uniqueness of stochastic elliptic equations,

> *'A complete theory of stochastic elliptic regularity, similar to the deterministic equations, has not yet been developed even for Gaussian noise. The main difficulties lie in,*
>
> 1. *Description of the noise;*
> 2. *Description of the boundary values of the solution*
>
> *Accordingly, most of the information about square-integrable solutions of stochastic elliptic equations is contained in specific examples.'*

This supports our results that application of methods from deterministic PDEs may be flawed when applied to high-dimensional bounded problems.

In the following section we will assess how we can utilise the relationships we have derived between RKHS inner products and Gaussian processes defined over particular domains.

7.2 Classification of Inducing Variables

As mentioned in the introduction, we wish to use RKHS projections to map a Gaussian process onto an orthogonal set of basis functions. So far we have classified when we are able to perform such a projection, Figure 7.2, however, we would now like to discuss our choice of basis functions given each projection. The choice of basis function however, is fundamentally determined by whether the domain is bounded or unbounded.

Let us first consider using a Fourier basis in an unbounded domain. As discussed in Hensman et al. [1], the Fourier transform at a particular frequency of a Gaussian process defined over the whole real line, is the sum of all infinitely many random variables in the GP, times some sinusoidal weight. This results in a Gaussian random variable with infinite variance. However, in order for an interdomain inducing variable to be valid, Matthews et al. [53] showed that interdomain inducing features must be finite, deterministic projections of the Gaussian process. Consequently, we are unable to use a Fourier basis, or any basis in fact, that spans the width of the real line over an unbounded domain.

The solution to using a Fourier basis such that the projections converge and we get valid inducing features with finite variance, is to *window* the domain [19, 1, 13]. Introduced by Lázaro-Gredilla and Figueiras-Vidal [19], the *only* way to ensure convergence using a Fourier basis is to specify an input density. In Hensman et al. [1] and Solin and Särkkä [13] they define a uniform input density over the 1D interval $[a, b]$ effectively bounding the domain by limiting the limits of integration between a and b. This ensures that the variance of the Gaussian random variables is finite and therefore a valid inducing feature.

However, Fourier features aren't the only choice of basis. We can also consider orthogonal basis functions that don't necessarily span the width of the real line or that decay to zero at infinity. Possibles functions include Hermite polynomials which are orthognoal in $L^2(\mathcal{D})$ and have been used by Burt et al. [54] to construct diagonal $\mathbf{K_{uu}}$ matrices for Gaussian processes defined over $\mathbb{R}$, or potentially orthogonal wavelets which are localised in time and frequency. The use of orthonormal basis defined over $\mathbb{R}^d$ and that produce valid interdomain inducing variables is an area of future research. In particular it could allow one to define higher-dimensional RKHS inner products over $\mathbb{R}^d$ which could provide a useful tool in combating the curse of dimensionality suffered by scaling VFF to higher-dimensions.

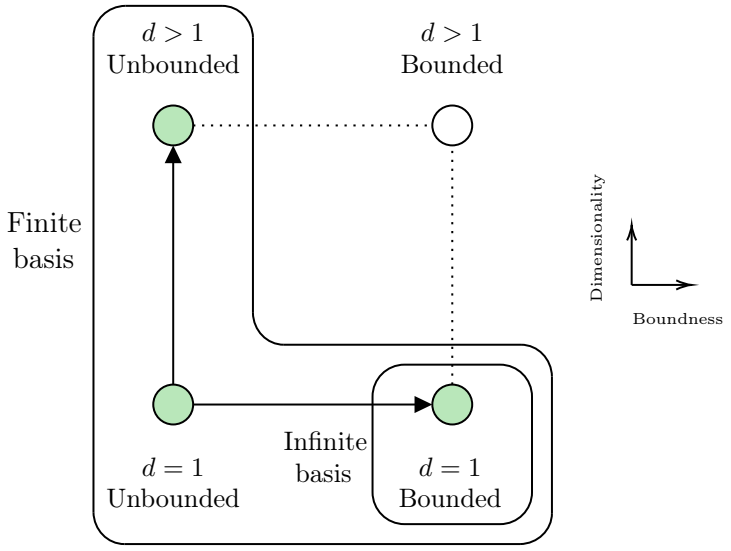

Figure 7.2: Square showing the type of basis we can use given a particular domain. Finite basis refers to basis functions who are finite in width, and infinite those who stretch to infinity

7.3 Manifolds

Another type of domain that we haven't considered in this project is manifolds. A manfiold is a topological space that is locally Euclidean. The interesting case relevant to our work is the idea of compact manifolds without boundaries, such as the unit hypershpere $\mathbb{S}^{d-1}$. The advantage of manifolds is that they can describe high-dimensional spaces without the need to prescribe any boundary conditions.

Recent work by Borovitskiy et al. [55] has defined the Matérn covariance function on compactly supported Riemannian manifolds using the Matérn SPDE by replacing the Laplace operator with the Laplace-Beltrami operator. Further, Dutordoir et al. [2] projected one-dimensional data onto the unit hypersphere to construct orthogonal interdomain inducing features using spherical harmonics.

Consequently, another avenue of future research would be the definition and application of RKHS inner products defined over d-dimensional manifolds as an alternative method to deriving higher dimensional RKHS inner products.

8 Conclusion

In this work we set out to begin to classify methods in which we can construct interdomain inducing variables. In particular we were interested in the case of RKHS projections when we restrict ourselves to stationary Gaussian processes with a rational spectral density function. Consequently, we explored two different choices we have when looking to construct an RKHS inner product for a given stationary kernel; the dimensionality of the domain and the boundness of the domain.

In order to classify these choices, we attempted to construct RKHS inner products by utilising deep connections between Gaussian processes, SPDEs and the RKHS. Indeed, in Chapter 2 we made links between stochastic processes and the RKHS according to Loeve's Isometry theorem. This connection enabled us to specify a finite dimensional RKHS at the end-points when constructing the RKHS inner product on a one-dimensional bounded domain. Most notably, however we utilised the SPDE representation of a Gaussian process in Chapter 4, as defined by Bochner's theorem for covariance functions with rational spectral density.

The SPDE representation enables us to define a zero-mean Gaussian process prior as a solution to a dynamical system subject to additive Gaussian white noise. Consequently, given the wealth of literature concerning methods of solving PDEs, we used the solution to such SPDE representations of Gaussian processes in order to construct functional form RKHS inner products.

We did this in 2 ways, using Green's functions and by introducing the variational formulation of boundary value problems. For both methods we were able to construct functional form RKHS inner product for one-dimensional bounded and unbounded domains and high-dimensional unbounded domains. An advantage of the variational formulation, is that we were able to directly access the RKHS inner product, without having to first consider a solution to the SPDE as a stochastic process, unlike in the Green's function approach. Further, when using Green's function it is often unclear what boundary conditions we are actually applying, where as by constructing symmetrical bilinear forms, the variational formulation allows us to write down the exact boundary conditions that produce an RKHS inner product.

However, we note that both methods were unsuccessful in deriving RKHS inner products for high-dimensional bounded domains. Indeed, this is in part due to the limitations of methods primarily developed in the study of deterministic boundary value problems, when applied to their stochastic counterparts. Hence, we suggest that using Green's function or the variational formulation are no-go approaches to constructing higher dimensional RKHS inner products on bounded domains.

Consequently, given access to the RKHS inner product our second choice in which to make is our choice of basis functions. We are able to classify this choice quite simply depending on whether or not the basis functions produce valid inducing variables which are everywhere deterministic and finite [53]. For a Fourier basis it has been discussed that as they exist everywhere on the real line, their projection in the RKHS does not converge and therefore we have to use strictly bounded domains [1]. However, less research has been done on sets of orthogonal basis whose projection converges when defined on $\mathbb{R}^d$. Such basis would be beneficial as would allow us to utilise the higher-dimensional RKHS inner product on an unbounded domain. This presents a direction for future work, alongside further classification of inducing point methods.

Bibliography

[1] J. Hensman, N. Durrande, A. Solin, et al. Variational Fourier Features for Gaussian Processes. *J. Mach. Learn. Res.*, 18(151–1).

[2] V. Dutordoir, N. Durrande, and J. Hensman. Sparse Gaussian processes with spherical harmonic features. In: *International Conference on Machine Learning*. PMLR. 2020, 2793–2802.

[3] N. Aronszajn. Theory of reproducing kernels. *Transactions of the American mathematical society*, 68(3):337–404, 1950.

[4] H. J. Cunningham. *The Connection Between GPs, SDEs and the RKHS*. Independent Study Option. Imperial College London, 2021.

[5] J. L. Doob. *Stochastic processes*. Vol. 10. New York Wiley, 1953.

[6] E. Parzen. *Stochastic processes*. SIAM, 1999.

[7] M. Titsias. Variational learning of inducing variables in sparse Gaussian processes. In: *Artificial intelligence and statistics*. PMLR. 2009, 567–574.

[8] J. Hensman, N. Fusi, and N. D. Lawrence. Gaussian processes for big data. *arXiv preprint arXiv:1309.6835*, 2013.

[9] J. Hensman, A. Matthews, and Z. Ghahramani. Scalable variational Gaussian process classification. In: *Artificial Intelligence and Statistics*. PMLR. 2015, 351–360.

[10] J. Hartikainen, M. Seppanen, and S. Sarkka. State-space inference for non-linear latent force models with application to satellite orbit prediction. *arXiv preprint arXiv:1206.4670*, 2012.

[11] S. Särkkä. Linear operators and stochastic partial differential equations in Gaussian process regression. In: *International Conference on Artificial Neural Networks*. Springer. 2011, 151–158.

[12] A. Solin and S. Särkkä. Explicit link between periodic covariance functions and state space models. In: *Artificial Intelligence and Statistics*. PMLR. 2014, 904–912.

[13] A. Solin and S. Särkkä. Hilbert space methods for reduced-rank Gaussian process regression. *Statistics and Computing*, 30(2):419–446, 2020.

[14] D. Burt, C. E. Rasmussen, and M. Van Der Wilk. Rates of convergence for sparse variational Gaussian process regression. In: *International Conference on Machine Learning*. PMLR. 2019, 862–871.

[15] M. Seeger, C. Williams, and N. Lawrence. *Fast forward selection to speed up sparse Gaussian process regression*. Tech. rep. 2003.

[16] J. Quinonero-Candela and C. E. Rasmussen. A unifying view of sparse approximate Gaussian process regression. *The Journal of Machine Learning Research*, 6:1939–1959, 2005.

[17] M. van der Wilk, V. Dutordoir, S. John, A. Artemev, V. Adam, and J. Hensman. A framework for interdomain and multioutput Gaussian processes. *arXiv preprint arXiv:2003.01115*, 2020.

[18] M. Bauer, M. van der Wilk, and C. E. Rasmussen. Understanding probabilistic sparse Gaussian process approximations. *arXiv preprint arXiv:1606.04820*, 2016.

[19] M. Lázaro-Gredilla and A. R. Figueiras-Vidal. Inter-domain Gaussian Processes for Sparse Inference using Inducing Features. In: *NIPS*. Vol. 22. Citeseer. 2009, 1087–1095.

[20] N. Durrande, J. Hensman, M. Rattray, and N. D. Lawrence. Detecting periodicities with Gaussian processes. *PeerJ Computer Science*, 2:e50, 2016.

[21] C. Müller. *Spherical harmonics*. Vol. 17. Springer, 2006.

[22] C. Efthimiou and C. Frye. *Spherical harmonics in p dimensions*. World Scientific, 2014.

[23] S. V. Lototsky and B. L. Rozovsky. *Stochastic partial differential equations*. Springer, 2017.
[24] K. Yosida. *Functional analysis*. eng. Classics in mathematics. Berlin: Springer-Verlag, 1995 - 1980. ISBN: 9783540586548.
[25] M. Kanagawa, P. Hennig, D. Sejdinovic, and B. K. Sriperumbudur. Gaussian processes and kernel methods: A review on connections and equivalences. *arXiv preprint arXiv:1807.02582*, 2018.
[26] I. Steinwart and A. Christmann. *Support vector machines*. Springer Science & Business Media, 2008.
[27] G. Wahba. *Spline models for observational data*. SIAM, 1990.
[28] I. Steinwart. Convergence types and rates in generic Karhunen-Loeve expansions with applications to sample path properties. *Potential Analysis*, 51(3):361–395, 2019.
[29] A. Berlinet and C. Thomas-Agnan. *Reproducing kernel Hilbert spaces in probability and statistics*. Springer Science & Business Media, 2011.
[30] S. Janson et al. *Gaussian hilbert spaces*. 129. Cambridge university press, 1997.
[31] M. F. Driscoll. The reproducing kernel Hilbert space structure of the sample paths of a Gaussian process. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 26(4):309–316, 1973.
[32] C.-A. Cheng and B. Boots. Incremental variational sparse Gaussian process regression. *Advances in Neural Information Processing Systems*, 29:4410–4418, 2016.
[33] H. Salimbeni, C.-A. Cheng, B. Boots, and M. Deisenroth. Orthogonally decoupled variational Gaussian processes. *arXiv preprint arXiv:1809.08820*, 2018.
[34] G. S. Kimeldorf and G. Wahba. A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. *The Annals of Mathematical Statistics*, 41(2):495–502, 1970.
[35] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005. ISBN: 026218253X.
[36] M. L. Stein. *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media, 1999.
[37] J. Hartikainen and S. Särkkä. Kalman filtering and smoothing solutions to temporal Gaussian process regression models. In: *2010 IEEE international workshop on machine learning for signal processing*. IEEE. 2010, 379–384.
[38] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan. *Estimation with applications to tracking and navigation: theory algorithms and software*. John Wiley & Sons, 2004.
[39] F. Lindgren, H. Rue, and J. Lindström. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498, 2011.
[40] C. Gu. *Smoothing spline ANOVA models*. Vol. 297. Springer Science & Business Media, 2013.
[41] E. Parzen et al. An approach to time series analysis. *Annals of mathematical statistics*, 32(4):951–989, 1961.
[42] G. E. Fasshauer. Green's functions: Taking another look at kernel approximation, radial basis functions, and splines. In: *Approximation Theory XIII: San Antonio 2010*. Springer, 2012, 37–63.
[43] G. E. Fasshauer and Q. Ye. Reproducing kernels of Sobolev spaces via a Green kernel approach with differential operators and boundary operators. *Advances in Computational Mathematics*, 38(4):891–921, 2013.
[44] G. E. Fasshauer and M. J. McCourt. *Kernel-based approximation methods using Matlab*. Vol. 19. World Scientific Publishing Company, 2015.
[45] J. O. Ramsay and C. Dalzell. Some tools for functional data analysis. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(3):539–561, 1991.
[46] J. O. Ramsay. Functional data analysis. *Encyclopedia of Statistical Sciences*, 4, 2004.

[47] R. A. Adams and J. J. Fournier. *Sobolev spaces*. Elsevier, 2003.
[48] J. K. Hunter and B. Nachtergaele. *Applied analysis*. World Scientific Publishing Company, 2001.
[49] S. Salsa. *Partial differential equations in action: from modelling to theory*. Vol. 99. Springer, 2016.
[50] H. Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Springer Science & Business Media, 2010.
[51] L. C. Evans. *Partial differential equations*. eng. 2nd ed. Graduate studies in mathematics ; v. 19. Providence, R.I: American Mathematical Society, 2010. ISBN: 9780821849743.
[52] A. M. Krall. *Hilbert space, boundary value problems, and orthogonal polynomials*. Vol. 133. Springer Science & Business Media, 2002.
[53] A. G. d. G. Matthews, J. Hensman, R. Turner, and Z. Ghahramani. On sparse variational methods and the Kullback-Leibler divergence between stochastic processes. In: *Artificial Intelligence and Statistics*. PMLR. 2016, 231–239.
[54] D. R. Burt, C. E. Rasmussen, and M. van der Wilk. Variational orthogonal features. *arXiv preprint arXiv:2006.13170*, 2020.
[55] V. Borovitskiy, A. Terenin, P. Mostowsky, and M. P. Deisenroth. Mat\'ern Gaussian processes on Riemannian manifolds. *arXiv preprint arXiv:2006.10160*, 2020.

A Appendix

A.1 ISO Result 1

From Cunningham [4] Section 4.1.1. Let us consider the non-homogeneous SDE defined by the linear differential operator $\mathcal{L}$, with homogeneous boundary conditions,

$$\left.\begin{array}{ll} \mathcal{L}f(\mathbf{x}) = w(\mathbf{x}) & \text{in } \mathcal{D}, \\ f(\mathbf{x}) = 0 & \text{on } \partial\mathcal{D}, \end{array}\right\} \tag{A.1}$$

where $w(\boldsymbol{x})$ is a Gaussian white noise process with covariance $\mathbb{E}[w(\boldsymbol{x})w(\boldsymbol{x}')] = \delta(\boldsymbol{x} - \boldsymbol{x}')$. Using the definition of a Green's function on a homogeneous domain $\mathcal{D}$, we can define the covariance of the Gaussian process f, inside of $\mathcal{D}$ in terms of the Green's functions G,

$$\mathbb{C}_f[f(\boldsymbol{x}), f(\boldsymbol{x}')] = \mathbb{E}\left[\iint_{\mathcal{D}} G(\boldsymbol{x}, \boldsymbol{z})G(\boldsymbol{x}', \boldsymbol{z}')w(\boldsymbol{z})w(\boldsymbol{z}')d\boldsymbol{z}d\boldsymbol{z}'\right] \tag{A.2}$$

$$= \int_{\mathcal{D}} \left(\int_{\mathcal{D}} \mathbb{E}\left[w(\boldsymbol{z})w(\boldsymbol{z}')\right] G(\boldsymbol{x}, \boldsymbol{z})d\boldsymbol{z}\right) G(\boldsymbol{x}', \boldsymbol{z}')d\boldsymbol{z}' \tag{A.3}$$

$$= \int_{\mathcal{D}} \left(\int_{\mathcal{D}} \delta(\boldsymbol{z} - \boldsymbol{z}')G(\boldsymbol{x}, \boldsymbol{z})d\boldsymbol{z}\right) G(\boldsymbol{x}', \boldsymbol{z}')d\boldsymbol{z}' \tag{A.4}$$

$$= \int_{\mathcal{D}} G(\boldsymbol{x}, \boldsymbol{z}')G(\boldsymbol{x}', \boldsymbol{z}')d\boldsymbol{z}' \tag{A.5}$$

Hence,

$$k(\mathbf{x}, \mathbf{x}') = \int_{\mathcal{D}} G(\boldsymbol{x}, \boldsymbol{z}')G(\boldsymbol{x}', \boldsymbol{z}')d\boldsymbol{z}' \tag{A.6}$$

A.2 ISO Result 2

From Cunningham [4] Section 4.1.1. Apply the same linear differential operator $\mathcal{L}$ associated with G to Equation (A.6),

$$\mathcal{L}k(\boldsymbol{x}, \boldsymbol{x}') = \mathcal{L}\int_{\mathcal{D}} G(\boldsymbol{x}, \boldsymbol{z}')G(\boldsymbol{x}', \boldsymbol{z}')d\boldsymbol{z}', \tag{A.7}$$

$$= \int_{\mathcal{D}} \mathcal{L}G(\boldsymbol{x}, \boldsymbol{z}')G(\boldsymbol{x}', \boldsymbol{z}')d\boldsymbol{z}', \tag{A.8}$$

$$= \int_{\mathcal{D}} [\mathcal{L}G(\boldsymbol{x}', \boldsymbol{z}')]G(\boldsymbol{x}, \boldsymbol{z}')d\boldsymbol{z}', \tag{A.9}$$

$$= \int_{\mathcal{D}} \delta(\boldsymbol{x}' - \boldsymbol{z}')G(\boldsymbol{x}, \boldsymbol{z}')d\boldsymbol{z}', \tag{A.10}$$

$$= G(\boldsymbol{x}, \boldsymbol{x}') \tag{A.11}$$

Therefore,

$$\mathcal{L}k(\boldsymbol{x}, \boldsymbol{x}') = G(\boldsymbol{x}, \boldsymbol{x}') \tag{A.12}$$

A.3 ISO Result 3

From Cunningham [4, Section 4.2]. The series representation of an RKHS inner product is given by Aronszajn [3] as,

$$\langle f, g \rangle_{\mathcal{H}_k} = \sum_{ij} \alpha_i \beta_j k(x_i, x_j) \tag{A.13}$$

Substituting the integral form of the reproducing kernel (A.6) from ISO Result 3, into the series representation of the RKHS inner product (A.13),

$$\langle f, g \rangle_{\mathcal{H}_{k_p}} = \sum_{ij} \alpha_i \beta_j \int_{\mathcal{D}} G(\boldsymbol{x}_i, \boldsymbol{z}) G(\boldsymbol{x}_j, \boldsymbol{z}) d\boldsymbol{z}, \tag{A.14}$$

Substituting equation (A.12) into equation (A.14),

$$\begin{aligned}
\langle f, g \rangle_{\mathcal{H}_{k_p}} &= \sum_{ij} \alpha_i \beta_j \int_{\mathcal{D}} G(\boldsymbol{x}_i, \boldsymbol{z}) G(\boldsymbol{x}_j, \boldsymbol{z}) d\boldsymbol{z}, && \text{(A.15)} \\
&= \sum_{ij} \alpha_i \beta_j \int_{\mathcal{D}} \mathcal{L}k(\boldsymbol{z}, \boldsymbol{x}_i) \mathcal{L}k(\boldsymbol{z}, \boldsymbol{x}_j) d\boldsymbol{z}, && \text{(A.16)} \\
&= \int_{\mathcal{D}} \left[\mathcal{L} \sum_i \alpha_i k(\boldsymbol{z}, \boldsymbol{x}_i) \right] \left[\mathcal{L} \sum_j \beta_j k(\boldsymbol{z}, \boldsymbol{x}_j) \right] d\boldsymbol{z} && \text{(A.17)} \\
&= \int_{\mathcal{D}} \mathcal{L}f(\boldsymbol{z}) \mathcal{L}g(\boldsymbol{z}) d\boldsymbol{z} && \text{(A.18)}
\end{aligned}$$

Hence, the RKHS inner product for $\mathcal{H}_{k_p}$ in functional form is,

$$\langle f, g \rangle_{\mathcal{H}_{k_p}} = \int_{\mathcal{D}} \mathcal{L}f(\boldsymbol{x}) \mathcal{L}g(\boldsymbol{x}) d\boldsymbol{x} \tag{A.19}$$