Conditioning of Gaussian Random Variables

Von der Fakultät Mathematik und Physik der Universität Stuttgart zur Erlangung der Würde eines Doktors der Naturwissenschaften (Dr. rer. nat.) genehmigte Abhandlung

vorgelegt von

Daniel Winkle

aus Ludwigsburg

Hauptberichter:	Prof. Dr. Ingo Steinwart
Mitberichter:	Prof. Dr. Bernard Haasdonk
Tag der mündlichen Prüfung:	06. 02. 2026

Institut für Stochastik und Anwendungen

Universität Stuttgart

2026

Abstract

In probability theory, a natural question is how additional information affects the uncertainty in a given variable, expressed via its probability distribution. This is formalized as the conditioning of the probability distribution. It turns out that conditioning random variables is a challenging problem with many subtleties. However, in the case of Gaussian random variables, the situation simplifies significantly: given two jointly Gaussian random variables, conditioning one on the other yields again a Gaussian random variable. This property makes it straightforward to incorporate additional information into the system. This is one of the reasons why Gaussian random variables are widely used in geostatistics, machine learning, uncertainty quantification, and spatial statistics.

In the context of Gaussian random variables, modern methods allow the conditioned random variable to be computed with increasing amounts of information. This leads to several natural questions: Does the process converge as more information is added? If so, what is the limit of this process? And finally, how fast does this convergence occur?

We address the first two questions in Chapter 3. There, we begin by covering the measure-theoretic background needed to understand conditioning and then introduce an operator M that maps the observed Gaussian random variable to the conditioned Gaussian random variable. Under the assumption that M is bounded, we are able to derive convergence rates for the posterior variance of the conditioned Gaussian random variable. The results are formulated in a general way so that they can be applied to any information selection algorithm. To illustrate this, we employ a greedy algorithm, which iteratively selects the most informative measurements given an observed Gaussian random variable.

Before addressing the third question, we make an interlude to explore kernel methods. When applying kernel methods, a common assumption for deriving convergence rates is that the target function lies in the reproducing kernel Hilbert space, the native space of these functions. The escaping-the-native-space question asks: Are reasonable convergence rates still possible if this assumption is not satisfied? We answer this question in Chapter 4 by utilizing the class of scaled reproducing kernel Hilbert spaces.

In the third main part of the thesis, Chapter 5, we are finally in a position to address the third question and derive convergence rates for the realizations of Gaussian random variables. We begin by deriving a concentration inequality that allows us to bound the norm of Gaussian random variables in scaled reproducing kernel Hilbert spaces. By assembling the results from Chapters 3 and 4, we then derive a concentration inequality for the realizations of Gaussian random variables. This concentration inequality is expressed in terms of the posterior variance. Thus, the results of Chapter 3 directly imply the concentration inequality.

Kurzfassung

In der Wahrscheinlichkeitstheorie stellt sich die Frage, wie zusätzliche Informationen die Unsicherheit einer gegebenen Zufallsvariablen beeinflussen. Dies wird formalisiert durch die Konditionierung der Wahrscheinlichkeitsverteilung. Es zeigt sich jedoch, dass das Konditionieren von Zufallsvariablen ein kniffliges Problem mit vielen Feinheiten ist. Im Fall von Gaußschen Zufallsvariablen vereinfacht sich die Situation jedoch erheblich: Sind zwei Zufallsvariablen gemeinsam Gauß-verteilt, so ist auch die bedingte Zufallsvariable wieder Gauß-verteilt. Diese Eigenschaft ermöglicht es, zusätzliche Informationen auf einfache Weise in das System zu integrieren. Aus diesem Grund finden Gaußsche Zufallsvariablen in der Geostatistik, dem maschinellen Lernen, der Unsicherheitsquantifizierung und der räumlichen Statistik breite Anwendung.

Im Zusammenhang mit Gaußschen Zufallsvariablen erlauben moderne Methoden, die bedingte Zufallsvariable unter zunehmender Informationsmenge zu berechnen. Dies führt zu mehreren natürlichen Fragestellungen: Konvergiert dieser Prozess, wenn immer mehr Information hinzugefügt wird? Falls ja, was ist der Grenzwert dieses Prozesses? Und schließlich: Wie schnell erfolgt diese Konvergenz?

Die ersten beiden Fragen werden in Kapitel 3 behandelt. Dort beginnen wir mit den maßtheoretischen Grundlagen, die zum Verständnis der Konditionierung notwendig sind, und führen anschließend einen Operator M ein, der die beobachtete Gaußsche Zufallsvariable auf die bedingte Gaußsche Zufallsvariable abbildet. Unter der Annahme, dass M beschränkt ist, können wir sogar Konvergenzraten für die a-posteriori Varianzen der bedingten Gaußschen Zufallsvariablen herleiten. Die Resultate sind allgemein formuliert, sodass sie auf beliebige Informationsauswahlalgorithmen angewendet werden können. Zur Veranschaulichung verwenden wir einen greedy Algorithmus, der iterativ die informativsten Messungen einer gegebenen Gaußschen Zufallsvariablen auswählt.

Bevor wir die dritte Frage behandeln, legen wir eine Zwischenetappe ein, um uns mit Kernel-Methoden zu befassen. In diesen Methoden ist es eine gängige Annahme für die Herleitung von Konvergenzraten, dass die Zielfunktion in dem reproduzierenden Kern-Hilbertraum liegt, also im nativen Raum der betrachteten Funktionen. Die sogenannte "escaping-the-native-space" Frage lautet: Sind auch dann noch vernünftige Konvergenzraten möglich, wenn diese Annahme nicht erfüllt ist? Diese Frage beantworten wir in Kapitel 4, indem wir die Klasse der skalierten reproduzierenden Kern-Hilberträume verwenden.

Im dritten Hauptteil der Arbeit, Kapitel 5, sind wir schließlich in der Lage, die dritte Frage anzugehen und Konvergenzraten für die Realisierungen Gaußscher Zufallsvariablen herzuleiten. Wir beginnen mit der Herleitung einer Konzentrationsungleichung, die es ermöglicht, die Norm Gaußscher Zufallsvariablen in skalierten reproduzierenden Kern-

Hilberträumen abzuschätzen. Durch die Kombination der Ergebnisse aus den Kapiteln 3 und 4 leiten wir anschließend eine Konzentrationsungleichung für die Realisierungen Gaußscher Zufallsvariablen her. Diese Konzentrationsungleichung wird in Abhängigkeit von der a-posteriori Varianz formuliert. Daher implizieren, die Resultate aus Kapitel 3 diese Konzentrationsungleichung direkt.

Contents

Chapter 1

Introduction

This chapter is structured as follows: in Section 1.1, I list all the publications I contributed to during my PhD studies and specify my contributions to each. In Section 1.2, I provide a brief historical background on Gaussian random variables and outline the motivation for our main results. In Section 1.3, we describe the structure of the overall thesis.

1.1 Publications

During the time of my PhD studies, I worked on the following publications which are undergoing the review process:

[1] D. Winkle, I. Steinwart and B. Haasdonk, *Convergence Analysis of a Greedy Algorithm for Conditioning Gaussian Random Variables*, *arXiv:2502.10772*, 2025.

Author Contribution: I (D.W.) contributed most of the proofs and ideas, while my supervisors (I.S. and B.H.) assisted with the precise formulation and guided me through the submission process. Additionally, I.S. contributed the idea of the measure theory focused analysis that enabled the handling of the null sets.

[2] D. Winkle, I. Steinwart and B. Haasdonk, *Convergence Rates for Realizations of Gaussian Random Variables*, *arXiv:2508.13940*, 2025.

Author Contribution: I (D.W.) contributed most of the proofs and ideas, while my supervisors (I.S. and B.H.) assisted with the precise formulation and guided me through the submission process.

Additionally I was part of a project that led to the following publication

[3] T. Wenzel, D. Winkle, G. Santin and B. Haasdonk, *Adaptive meshfree approximation for linear elliptic partial differential equations with PDE-greedy kernel methods*, *BIT Numerical Mathematics*, Springer, 2025.

Author Contribution: Most of the work was done by T.W., while I (D.W.) contributed to some of the proofs and introduced the well-posedness conditions of the PDEs to obtain convergence rates.

The results of [1] are covered in Chapter 3, while the results of [2] are discussed in Chapters 4 and 5. Lastly, the ideas I contributed in [3] are used to provide an example in Chapter 6, which connects the results of [1] and [2].

1.2 Motivation

This thesis is mainly concerned with Gaussian distributions. The origins of the Gaussian distribution can be traced back to 1733, when Abraham de Moivre demonstrated that the binomial distribution can be approximated by a bell-shaped curve [4, Chapter 2]. This result was later generalized into what became known as the central limit theorem. In 1809, Carl Friedrich Gauss independently rediscovered this distribution while studying the law of errors. He showed that if observational errors are independent and equally likely to be positive or negative, their most probable distribution is proportional to the exponential of the negative square of the deviation [4, Chapter 5]. This formulation became known as the Gaussian law. Many years later, in 1923, Norbert Wiener introduced the Wiener process, which is now recognized as the first Gaussian process [5]. Since these discoveries preceded Kolmogorov's formal definition of random variables in 1933 [6], the concepts were later rigorously formalized within modern probability theory.

Building on these historical developments, Gaussian distributions evolved from a description of measurement errors into a general framework for modeling uncertainty. This development laid the foundation for modern applications of Gaussian processes in data analysis and statistical inference.

Today, interpolation with Gaussian random variables is a widely used method for analyzing functional data, with applications in computer experiments [7], machine learning [8], and geostatistics [9], where this approach is known as kriging. These approaches naturally encode prior uncertainty and enable coherent probabilistic inference, making them fundamental to Bayesian modeling frameworks [10]. In particular, Gaussian process models play a central role in scientific computing and statistical learning, where they are employed to approximate computationally expensive models [11], reconstruct spatial fields [12], and quantify uncertainty in inverse problems [13].

But my personal motivation for working with Gaussian random variables is simpler and more direct: I find researching them fun. Gaussian random variables connect many different areas of mathematics, which allows to bring tools from numerious different branches of mathematics to the table.

The first connection I highlight is with harmonic analysis [14], established through their relationship with kernels [15]. When a kernel satisfies an invariance condition, for example translation invariance, the associated Gaussian random variable is called stationary. Such invariant kernels are analyzed using harmonic analysis, which allows the application of representation theory [16].

Moreover, since Gaussian random variables are generally Banach space-valued, we can readily apply functional analysis to their study [17]. They are also naturally associated with various Hilbert spaces, such as Gaussian Hilbert spaces [18] and abstract Wiener spaces [19], enabling the use of spectral theory [20]. Gaussian random variables can even be employed to solve nonlinear partial differential equations [21].

Just as there are many different mathematical fields that can be applied to the study of Gaussian random variables, there are also many questions that arise concerning them. In other words, the study of Gaussian random variables spans such a broad range of mathematics that I never ask myself, 'What can I do?'. Instead, I often wonder, 'Given all the possible approaches, which one should I choose?'.

Given all those possible approaches which questions should I ask?

One possible question is: given a set of realizations (X_j) of a Gaussian random variable X, what can be inferred about the covariance operator of X [22]? It turns out that the maximum likelihood approach to this problem is ill-posed [23].

Another question, which we focus on in this work, is the conditioning problem. Specifically, given two jointly Gaussian random variables X and Y, what can be said about $X \mid Y$? If the image space of Y is finite-dimensional, this question has been thoroughly addressed [8]. For Gaussian random variables taking values in separable Hilbert spaces, one approach to solving the conditioning problem employs shorted operators [24]. For separable Banach spaces, Ref. [25] shows that the conditioned Gaussian random variable can be represented by a series. More recently, Ref. [26] addressed the conditioning problem using conditional probability, representing the conditioned Gaussian random variable as the limit of a sequence converging in the Banach space norm.

To adress the second question consider the simplest non-trivial setup. Let us assume a Gaussian random variable $X : (\Omega, \mathcal{A}, \mu) \to C(T)$, where $(\Omega, \mathcal{A}, \mu)$ is a probability space and $C(T)$ denotes the space of continuous functions on a metric space (T, d). Furthermore, we assume that X is observed on a finite set $S_n := \{s_1, \ldots, s_n\} \subseteq T$, meaning we observe $Y_n : \Omega \to \mathbb{R}^n$ defined by $Y_n := (\delta_{s_1} X, \ldots, \delta_{s_n} X)$, where $\delta_t(f)$ denotes the point evaluation of $f \in C(T)$ at $t \in T$.

We can then compute the conditional expectation and posterior variance of X given $Y_n = y_n$ as follows:

$$\begin{aligned}\mathbb{E}(X \mid Y_n = y_n) &= \mathbb{E}(X) + \mathrm{cov}(X, Y_n)\mathrm{cov}(Y_n)^\dagger(y_n - \mathbb{E}Y_n),\\ \mathrm{cov}(X \mid Y_n) &= \mathrm{cov}(X) - \mathrm{cov}(X, Y_n)\mathrm{cov}(Y_n)^\dagger\mathrm{cov}(Y_n, X),\end{aligned}$$

where cov denotes the corresponding covariance operator or matrix, and $\mathrm{cov}(Y_n)^\dagger$ is the Moore-Penrose pseudo-inverse of the $n \times n$ covariance matrix [26].

The main computational bottleneck in this setting arises from evaluating the Moore-Penrose pseudo-inverse, since $\mathrm{cov}(Y_n)$ becomes increasingly ill-conditioned as n grows. One way to alleviate this issue is to employ Kalman filtering [27] or multiresolution analysis [28]. These techniques typically do not compute the exact conditional expectation but instead yield a surrogate that depends on certain hyperparameters, which control the trade-off between approximation accuracy and computational cost. This approach enables the efficient treatment of larger value n. Thus, we consider the case $n \to \infty$. Let us assume that we have infinitely many points $s_j \in T$ and define $S := \overline{\{s_j \mid j \in \mathbb{N}\}}^d$. Moreover, let $Y = X|_S \in C(S)$. Then, by [26, Corollary 4.6] and [26, Theorem 3.3], we have for the conditional expectation

$$\mathbb{E}(X \mid Y = y) = \mathbb{E}(X) + \lim_{n\to\infty} \mathrm{cov}(X, Y_n)\mathrm{cov}(Y_n)^\dagger(y_n - \mathbb{E}Y_n) \tag{1.1}$$

and for the posterior variance

$$\mathrm{cov}(X \mid Y) = \mathrm{cov}(X) - \lim_{n\to\infty} \mathrm{cov}(X, Y_n)\mathrm{cov}(Y_n)^\dagger\mathrm{cov}(Y_n, X), \tag{1.2}$$

where the convergence in (1.1) holds in $C(T)$ and the convergence in (1.2) holds in the nuclear norm for operators.

We call the finite-dimensional approximation $\mathbb{E}(X|Y_n)$ of $\mathbb{E}(X|Y)$ the surrogate. The equations (1.1) and (1.2) can, under additional assumptions on the surrogate, guarantee that the surrogate converges to the true conditional expectation and posterior variance. However, even with such assumptions, we generally have no information about the convergence rate, which may be arbitrarily slow. To address this issue, we investigate conditions under which the conditional expectation can be represented by a bounded operator $M : C(S) \to C(T)$ satisfying $My = \mathbb{E}(X \mid Y = y)$.

This representation allows us, given an approximation f_n of the observation y with convergence rate $c_n > 0$, such that $\|f_n - y\|_{C(S)} \leq c_n$, to obtain $\|\mathbb{E}(X \mid Y = y) - Mf_n\|_{C(T)} \leq \|M\|c_n$. Hence, we also obtain convergence rates for the conditional expectation. We discuss this further in Chapter 3.

A central difficulty is that realizations of infinite-dimensional Gaussian random variables almost surely lie outside their associated reproducing kernel Hilbert space [29, 30]. This well-known phenomenon precludes the direct application of classical kernel approximation results [31] and requires addressing the so-called escaping-the-native-space problem [32]. We address this problem in Chapter 4 and apply the results in Chapter 5.

1.3 Structure of this Thesis

This thesis is structured as follows.

Chapter 2 provides the necessary preliminary information for the subsequent chapters, including basic concepts related to Gaussian random variables, their associated spaces, conditioning, kernels, kernel approximation, and greedy algorithms. Additionally, we introduce the common notation used throughout the thesis.

Chapter 3 begins by establishing the connection between the Gaussian Hilbert space, the abstract Wiener space, and the reproducing covariance space of a Gaussian random variable. We then present general results on conditional expectation. In the following section, we provide the main convergence results and introduce the operator M, which maps the observational Gaussian random variable to its conditional expectation. We also state the corresponding convergence analysis results. The chapter concludes with basic illustrative examples.

In Chapter 4, we address the escaping-the-native-space question. We first present its context and then provide the main result for scalable reproducing kernel Hilbert spaces, allowing us to answer this question positively in that setting. We conclude the chapter with an example involving power spaces, where we reproduce, up to a constant, the same convergence rate known from the power space setting.

The results from Chapter 4 are applied in Chapter 5. There, we prove a simple concentration inequality, which together with the escaping-the-native-space results yields a convergence rate for realizations of Gaussian random variables in a concentration inequality setting.

Finally, in Chapter 6, we present a short example where all our results are combined. We also summarize the main results and provide an outlook on open questions. The appendices A and B contain supporting theorems that are required but do not offer further insight.

Chapter 2

Background

The following sections review the required background for the topics studied in this thesis: Gaussian random variables and kernels. The coverage is not exhaustive, but references to the relevant literature are provided for further details.

We begin with Gaussian random variables, followed by a brief interlude on kernels to highlight the close connection between Gaussian random variables and kernels.

2.1 Gaussian Random Variables

Throughout this work, $(\Omega, \mathcal{A}, \mu)$ is a probability space and J is an index set with at most countable many elements. We set $L^2(\mathcal{A}) := L^2(\Omega, \mathcal{A}, \mu)$ as the Hilbert space of μ-equivalence classes of square integrable functions. Note that here we emphasize the dependence on the σ-algebra $\mathcal{A}$ by taking it as argument of the space. The reason for this is that conditioning changes the σ-algebra, while the measure μ usually remains unchanged in this work. In some cases we want to emphasize the measure, we then write $L^2(\mu) := L^2(\mathcal{A})$. Furthermore, when the σ-algebra is irrelevant to the discussion, we abbreviate the probability space as (Ω, μ).

Moreover, troughout this work, E and F denote Banach spaces and E' denotes the dual. We further write B_E for the closed unit ball. Finally for all $A \subseteq E$ we write $\overline{A}$ for the closure of A.

We call a mapping $X : \Omega \to E$ weakly measurable, if $e'(X)$ is measurable for all $e' \in E'$. In addition, we call X strongly measurable if X is measurable with respect to the Borel-σ-algebra generated by the norm topology on E and $X(\Omega)$ is norm-separable. In the case of separable Banach spaces both definitions coincide, as ensured by the well-known Pettis-measurability theorem [33, Theorem E.9]. We call X a random variable if X is strongly measurable.

Moreover recall that, given $p \geq 1$ and a Banach space E, the Bochner space $L^p(\mathcal{A}, E)$ is the linear space of all μ-equivalence classes of strongly measurable functions $f : \Omega \to E$ such that

$$\|f\|^p_{L^p(\mathcal{A},E)} := \int_\Omega \|f\|^p_E \,\mathrm{d}\mu < \infty.$$

For more details about Bochner spaces we refer to [34, Chapter 1].

We continue by discussing the conditioning of random variables. To this end, let $\mathcal{A}_0$ be a sub-σ-algebra of $\mathcal{A}$ and $X \in L^1(\mathcal{A}, E)$. An $\mathcal{A}_0$-measurable function Z is called a conditional expectation of X under $\mathcal{A}_0$ if

$$\int_A Z \,\mathrm{d}\mu = \int_A X \,\mathrm{d}\mu, \quad \text{for all} \quad A \in \mathcal{A}_0.$$

In this case we write $Z \in \mathbb{E}(X|\mathcal{A}_0)$. The existence and almost-sure uniqueness of conditional expectations is guaranteed by [34, Chapter 2.6]. For this reason we often do not distinguish between Z and $\mathbb{E}(X|\mathcal{A}_0)$. Given a random variable $Y : \Omega \to F$ we write $\sigma(Y)$ for the smallest σ-algebra such that Y is weakly measurable. Moreover we set $\mathbb{E}(X|Y) := \mathbb{E}(X|\sigma(Y))$ and for $\mathfrak{F}' \subset F'$ we define

$$\mathbb{E}(X|\mathfrak{F}' \circ Y) := \mathbb{E}(X|\sigma(\{f'(Y) \,:\, f' \in \mathfrak{F}'\})).$$

We call a random variable $X : \Omega \to \mathbb{R}$ a one dimensional Gaussian random variable if there exist $\mu, \sigma \in \mathbb{R}$, such that $\mathbb{E}\left(\mathrm{e}^{\mathrm{i}tX}\right) = \mathrm{e}^{\mathrm{i}t\mu - \frac{1}{2}\sigma^2 t^2}$ holds true for all $t \in \mathbb{R}$. We then write $X \sim \mathcal{N}(\mu, \sigma^2)$. We note that if $\sigma > 0$ one obtains a normal distribution and if $\sigma = 0$ one obtains a point measure. We call a random variable $X : \Omega \to E$ Gaussian if E is separable and $e'(X)$ is a real-valued one dimensional Gaussian random variable for all $e' \in E'$. Given two random variables $X : \Omega \to E$ and $Y : \Omega \to F$, we call them jointly Gaussian if $(X,Y) : \Omega \to E \times F$, $\omega \mapsto (X(\omega), Y(\omega))$ is a Gaussian random variable. If $X : \Omega \to E$ is a Gaussian random variable we have $X \in L^p(\mathcal{A}, E)$ for all $p \geq 1$ by Fernique's Theorem [35, Theorem 5.3].

We will often make the same assumptions, thus we consolidate them into a single statement:

Assumption A. *We require E, F to be separable Banach spaces, $(\Omega, \mathcal{A}, \mu)$ to be a probability space, and the random variables $X : \Omega \to E$, $Y : \Omega \to F$ to be jointly Gaussian. Also we assume $\mathbb{E}(X) = \mathbb{E}(Y) = 0$ and we set $Z := \mathbb{E}(X|Y)$.*

We remark that, under Assumption A Z is indeed a Gaussian random variable with $\mathbb{E}(Z) = 0$, see [26, Theorem 3.3 vi)].

We now introduce covariance operators, which are an essential tool for analyzing Gaussian random variables, see [36, Chapter 1]. To this end, under Assumption A the covariance operator $\mathrm{cov}(X) : E' \to E$ of a Gaussian random variable X is defined by $\mathrm{cov}(X)e' := \mathbb{E}(e'(X)X)$ for all $e' \in E'$. Given $e' \in E'$ the one-dimensional conditional variance is defined by

$$\mathrm{cov}(e'(X)|Y) := \mathbb{E}\big((e'(X) - \mathbb{E}(e'(X)|Y))^2|Y\big),$$

as in [37, Chapter 4]. We want to generalize this to the infinite dimensional case. To this end, recall that given $e_1, e_2 \in E$ the elementary tensor $e_1 \otimes e_2 : E' \to E$ is given by $e' \to e'(e_1)e_2$. Moreover, the injective tensor product $E \check{\otimes} E$ is one instance of a Banach space containing all elementary tensors, see e.g. [36, Section 4]. Now, given a random variable $X : \Omega \to E$, the pointwise defined map $X \otimes X : \Omega \to E \check{\otimes} E$, $\omega \mapsto X(\omega) \otimes X(\omega)$ is also a random variable, see [36, Theorem 6.7].

Definition 2.1.1. *Under Assumption A we define the* conditional variance $\mathrm{cov}(X|Y) : E' \to E$ *as*

$$\mathrm{cov}(X|Y) := \mathbb{E}\Big(\big(X - \mathbb{E}(X|Y)\big) \otimes \big(X - \mathbb{E}(X|Y)\big)|Y\Big).$$

We define the cross covariance operator $\mathrm{cov}(X,Y) : F' \to E$ of X and Y as

$$\mathrm{cov}(X,Y)f' := \int_\Omega f'(Y)X \, \mathrm{d}\mu.$$

2.2 Kernels

As already mentioned, kernels are closely connected to Gaussian random variables, see [15]. In particular, every Gaussian process is fully characterized by its mean function and covariance kernel. Conversely, any positive definite kernel defines the covariance structure of some Gaussian process. This duality establishes a direct correspondence between kernels and the distributional properties of Gaussian processes, making kernels a central tool in both their theoretical analysis and practical implementation.

There are two ways to define a kernel. The first one is via positive-definite matrices, as presented in [31], the second one uses feature maps and feature spaces, as in [38]. Both definitions are equivalent, as shown in [38, Theorem 4.16]. The choice of definition depends on the intended application. When working with finite-dimensional approximations, it is more natural to use the definition via positive definite matrices, as is common in numerical analysis [31]. However, in our setting, we are primarily interested in the connection with Gaussian random variables and the infinite-dimensional case. For this reason, the feature map and feature space definition is more appropriate.

Definition 2.2.1. *Given a non-empty set T, we call $k : T \times T \to \mathbb{R}$ a kernel if there exist a Hilbert space H_0 and a map $\Phi : T \to H_0$, called feature map, such that*

$$k(t,s) = \langle \Phi(t), \Phi(s) \rangle_{H_0}, \qquad \text{for all } t, s \in T.$$

The natural connection between Gaussian random variables and kernels arises by viewing the Gaussian random variable X as the feature map of the kernel. We set $T = B_{E'}$ and define

$$k_X(e_1', e_2') := \langle e_1'(X), e_2'(X) \rangle_{L^2(\mathcal{A})}.$$

It is clear that the feature space is $L^2(\mathcal{A})$, and the feature map is given by $\Phi : B_{E'} \to L^2(\mathcal{A})$, $e' \mapsto e'(X)$. Hence, k_X is indeed a kernel. We will return to this connection later. In this section, we focus on kernels.

We now introduce reproducing kernel Hilbert spaces, which are in one-to-one correspondence with kernels.

Definition 2.2.2. *Given a non-empty set T and given a Hilbert space of functions $H \subseteq \{f : T \to \mathbb{R}\}$*

i) a function $k : T \times T \to \mathbb{R}$ *is called a* reproducing kernel *of* H *if we have* $k(\,\cdot\,, t) \in H$ *for all* $t \in T$ *and the* reproducing property

$$f(t) = \langle f, k(\,\cdot\,, t)\rangle_H$$

holds for all $f \in H$ *and all* $t \in T$.

ii) the space H *is called a* reproducing kernel Hilbert space (RKHS) *over* T *if for all* $t \in T$ *the Dirac functional* $\delta_t : H \to \mathbb{R}$ *defined by*

$$\delta_t(f) := f(t), \qquad f \in H,$$

is continuous.

The one-to-one correspondence with kernels is described by the following Theorem.

Theorem 2.2.3. *Let* H *be an RKHS over* T. *Then* $k : T \times T \to \mathbb{R}$ *defined by*

$$k(t, t') := \langle \delta_t, \delta_{t'}\rangle_H, \qquad \text{for all } t, t' \in T,$$

is the only reproducing kernel of H. *Moreover, if* $(e_j)_{j\in J}$ *is an orthornomal basis (ONB) of* H, *then for all* $t, t' \in T$ *we have*

$$k(t, t') = \sum_{j\in J} e_j(t)e_j(t').$$

For a proof we refer to [38, Theorem 4.20].

Kernels are commonly used to approximate functions, for example, by interpolation. Given a set of points $T_n := \{t_1, \ldots, t_n\} \subseteq T$, assume that $K := ((k(t_i, t_j))_{i,j=1}^n \in \mathbb{R}^{n\times n}$ is a strictly positive definite matrix. Given a function $f : T \to \mathbb{R}$, the interpolant s_{f,T_n} of f is given by

$$s_{f,n}(t) = \sum_{j=1}^{n} a_j k(t, t_j)$$

where $(a_j)_{j=1}^n \in \mathbb{R}^n$ is obtained by solving

$$Ka = \mathbf{f},$$

with $\mathbf{f} = (f(t_1), \cdots, f(t_n))_{i=1}^n \in \mathbb{R}^n$.

A key property of interpolation in RKHS is that it is an orthogonal projection onto the space $V_n := \mathrm{span}\{k(\,\cdot\,, t_j) \mid t_j \in T_n\}$, see [31, Theorem 13.1]. This will also be emphasized in the next section.

2.2.1 Generalized Interpolation

Interpolation can be extended to a more general setting, where the goal is not only to satisfy point evaluations but also arbitrary linear functionals applied to the target function. We follow the approach of [31, Chapter 16]. Let H be a Hilbert space with dual space H'. Given a set of linearly independent functionals $\{\lambda_1, \cdots, \lambda_n\} \subseteq H'$ and values $(f_j)_{j=1}^n \in \mathbb{R}^n$, the generalized recovery problem seeks a function $s_n \in H$ such that $\lambda_j(s_n) = f_j$ for all $1 \leq j \leq n$. The function s_n is then called a generalized interpolant.

If we additionally require that

$$\|s_n\|_H = \min\{\|s\| \,|\, s \in H,\, \lambda_j(s) = f_j,\, 1 \leq j \leq n\}$$

we call s_n the norm-minimal generalized interpolant. If H is an RKHS we call s_n the generalized kernel interpolant.

Moreover, since H is a Hilbert space, we can identify each $\lambda_j \in H'$ with a unique element $v_j \in H$ via the Riesz representation theorem, i.e., $\lambda_j(s) = \langle s, v_j\rangle_H$ for all $s \in H$.

We summarize [31, Theorem 16.1] and [31, Corollary 16.2] into the following theorem.

Theorem 2.2.4. *Let H be a Hilbert space, $\{\lambda_1, \cdots, \lambda_n\} \subseteq H'$ linearly independent functionals with Riesz representers $\{v_1, \cdots, v_n\} \subseteq H$. The mapping $\Pi_n : H \to H$ given by*

$$\Pi_n f := s_n := \sum_{j=1}^{n} a_j v_j,$$

with $s_n \in H$ being the norm-minimal generalized interpolant, where $(a_j)_{j=1}^n \in \mathbb{R}^n$ are such that $\lambda_j(s_n) = \lambda_j(f)$ for all $1 \leq j \leq n$, is an orthogonal projection onto $V_n := \operatorname{span}\{v_j \,|\, 1 \leq j \leq n\}$.

We want to apply this method to our Gaussian random variable X, but X takes values in a larger Banach space $E \supseteq H$ rather than in a Hilbert space H. Therefore, we do not consider all functionals in H', but only those λ_j with $\lambda_j \in E' \subseteq H'$.

The Hilbert space H to which we apply this generalized interpolation framework is the RKHS associated with the kernel k_X, where k_X is a continuous linear kernel defined on $B_{E'}$. This implies that $H \subseteq E'' \subseteq C(B_{E'})$. Consequently, performing interpolation in the space $C(B_{E'})$ with respect to H corresponds to the generalized interpolation setting, since we directly identify points with their associated functionals in $B_{E'}$.

2.2.2 P-Greedy

Now we introduce the *weak-greedy algorithm* and recall the results from [39] we will need in our further endeavors. Let H be a Hilbert space and let $\mathcal{F} \subseteq B_H$ be a compact subset. For fixed $\gamma \in]0, 1]$, we first choose an element $f_0 \in \mathcal{F}$ such that $\gamma \max_{f\in\mathcal{F}} \|f\|_H \leq \|f_0\|_H$ holds true. Assuming $\{f_0, \cdots, f_{n-1}\}$ and $V_n := \operatorname{span}\{f_0, \cdots, f_{n-1}\}$ have been selected, we then take any $f_n \in \mathcal{F}$ such that

$$\gamma \max_{f\in\mathcal{F}} \operatorname{dist}(f, V_n) \leq \operatorname{dist}(f_n, V_n)_H.$$

Now, we turn to the P-greedy algorithm. Consider an RKHS H with kernel k and domain T. In addition, let $\mathcal{F} := \{k(\,\cdot\,, t) \,|\, t \in T\}$ be compact and $k(t, t) \leq 1$ for all $t \in T$. We select points $(t_0, \cdots, t_{n-1}) \subseteq T$ via an iterative method. First we take any $t_0 \in T$ according to the criterion

$$\gamma \sup_{t\in T} \|k(\,\cdot\,, t)\|_H \leq \|k(\,\cdot\,, t_0)\|_H$$

with $\gamma \in]0,1[$ and we set $T_1{:=}\{t_0\}$. The subsequent point $t_{n-1} \in T$ is then selected according to

$$\gamma \sup_{t\in T} \|k(\,\cdot\,,t) - \Pi_{T_{n-1}} k(\,\cdot\,,t)\|_H \leq \|k(\,\cdot\,,t_{n-1}) - \Pi_{T_{n-1}} k(\,\cdot\,,t_{n-1})\|_H,$$

where $\Pi_{T_{n-1}}$ denotes the orthogonal projection onto the space span $(\{k(\,\cdot\,,t)\,|\,t \in T_{n-1}\})$ and again we set $T_n{:=}T_{n-1} \cup \{t_{n-1}\}$. We call this point selection algorithm weak-P-greedy. The following Corollary is a consequence of the statements in [40, Section 4].

Corollary 2.2.5. *Given an RKHS H with kernel k selecting points via the weak P-greedy one obtains*

$$\sup_{f\in B_H} \|f - \Pi_{T_n} f\|_{C(T)} \leq \sqrt{2}\gamma^{-1} \min_{1\leq m<n} d_m^{\frac{n-m}{n}}(\mathcal{F}).$$

with $d_m(\mathcal{F}) := \min_V \sup_{f\in\mathcal{F}} \|f - \Pi_V f\|_H$ denoting the Kolmogorov width of the set $\mathcal{F}$, where the minimum is taken over all m-dimensional subspaces $V \subseteq H$ and Π_V denotes the orthogonal projection in H onto V.

We emphasize that the weak-greedy algorithm above does not determine a unique sequence of spaces as many elements might satisfy the weak-greedy selection criterion. The inequality holds for any sequence obtained from the weak-greedy algorithm.

We note that there are many more greedy algorithms [41] that could be analyzed in our context. However, for our purposes, the P-greedy algorithm is sufficient.

2.2.3 Power Spaces and Scalable RKHS

This section covers the construction of scalable RKHSs. Scalable RKHSs are a tool that has previously been used to analyze Gaussian random variables, see [42]. We use them not only for this purpose but also to answer the escaping-the-native-space question later in Chapter 4, where this connection will be discussed in more detail.

To motivate the definition of scalable RKHSs, we begin with a more familiar construction of RKHSs, the so-called power spaces.

We do not distinguish between a function and its equivalence class. While this can lead to complications, these and their resolution are thoroughly discussed in [43]. Let (T,λ) be a probability space, and let k be a kernel on T such that $\int_T k(t,t)\,\mathrm{d}\lambda(t) < \infty$. We assume there exists an ONB $(e_j)_{j\in N} \subseteq L^2(\lambda)$ and a positive sequence $(\lambda_j) \in \ell^1(\mathbb{N})$ such that $(\sqrt{\lambda_j}e_j) \subseteq H$ forms an orthonormal system in the RKHS H. Assume further that there exists $\beta > 0$ such that

$$\sum_{j=1}^{\infty} \lambda_j^{\beta} e_j(t)^2 < \infty \qquad \text{for all } t \in T,$$

holds true. We define $k_\beta(t,t') := \sum_{j=1}^{\infty} \lambda_j^{\beta} e_j(t)e_j(t')$, and refer to such kernels as power kernels. The associated RKHS H^β is called a power space. One can derive convergence rates for elements in such power spaces by considering the orthogonal projection Π_n :

$L^2(\lambda) \to L^2(\lambda)$ onto the subspace $\operatorname{span}\{e_j \mid 1 \le j \le n\}$. For $f \in H^\beta$, there exists a sequence $(f_j) \in \ell^2(\mathbb{N})$ such that $f = \sum_{j=1}^\infty f_j \sqrt{\lambda_j}^\beta e_j$. Then we compute

$$\|f - \Pi_n f\|_{L^2(\lambda)}^2 = \left\| \sum_{j=n+1}^\infty f_j \sqrt{\lambda_j}^\beta e_j \right\|_{L^2(\lambda)}^2 = \sum_{j=n+1}^\infty f_j^2 \lambda_j^\beta \le \sup_{j \ge n+1} \lambda_j^\beta \cdot \sum_{j=n+1}^\infty f_j^2.$$

If we additionally assume that (λ_j) is monotonically decreasing, we obtain

$$\|f - \Pi_n f\|_{L^2(\lambda)} \le \lambda_{n+1}^{\beta/2} \|f\|_{H^\beta}. \tag{2.1}$$

This is also the optimal convergence rate for the worst-case error, which can be verified by setting $f = e_{n+1}$.

To define scalable RKHSs, we do not start with an ONB in $L^2(\lambda)$, but instead directly with an ONB in the RKHS H. Let $(v_j) \subseteq H$ be an ONB, and let (a_j) be a positive sequence such that

$$\sum_{j=1}^\infty a_j v_j(t)^2 < \infty, \quad \text{for all } t \in T.$$

We define the scaled kernel

$$k_{a,V}(t, t') = \sum_{j=1}^\infty a_j v_j(t) v_j(t').$$

This is indeed a kernel, and we denote the associated RKHS by $H_{a,V}$, see [42]. Moreover, $(\sqrt{a_j} v_j)$ forms an ONB in $H_{a,V}$, see [42, Proposition 3.2].

In contrast to power spaces, scalable RKHSs allow a broader range of parameter choices, as they are defined for arbitrary positive sequences (a_j) rather than just powers (λ_j^β). They are also independent of the measure λ, and any ONB in H can be used in the construction. Such ONBs can be constructed efficiently via the Gram–Schmidt process, which is particularly useful in the context of the P-greedy algorithm, where the projection space is expanded by one function in each step.

2.3 Spaces associated to Gaussian Random Variables

The following subsections cover the spaces associated with Gaussian random variables. At the end of all the subsections, we explain the purpose of each of these spaces.

2.3.1 Gaussian Hilbert Spaces

We give a short introduction into Gaussian Hilbert spaces. For more details, see [18, Chapters 1 and 9]. A Gaussian linear space G is a linear subspace of $L^2(\mathcal{A})$ such that each $g \in G$ is a Gaussian random variable with $\mathbb{E}(g) = 0$. A Gaussian Hilbert space G is a Gaussian linear space that is closed in $L^2(\mathcal{A})$. Under Assumption A we define $G_X := \overline{\{e'(X) \mid e' \in E'\}}$ as the Gaussian Hilbert space associated with the Gaussian random variable X. Moreover, the norm on G_X is given by $\|g\|_{G_X} = \|g\|_{L^2(\mathcal{A})}$. Under Assumption A, we note that (X, Y) is an $E \times F$-valued Gaussian random variable, and the space

$$G_{(X,Y)} = \overline{\{e'(X) + f'(Y) \,|\, (e', f') \in E' \times F'\}} \tag{2.2}$$

is also a Gaussian Hilbert space. The closure can be necessary, see [18, Example 1.25].

Conditioning in Gaussian Hilbert spaces G reduces to orthogonal projections in $L^2(\mathcal{A})$. Specifically, for a set $\mathcal{G} \subset G$ and $g \in G$ the conditional expectation of g given $\mathcal{G}$ is

$$\mathbb{E}(g \,|\mathcal{G}) := \mathbb{E}(g \,|\sigma(\mathcal{G})) = \Pi_V g, \tag{2.3}$$

where Π_V denotes the orthogonal projection onto the subspace $V := \overline{\operatorname{span}(\mathcal{G})}$, see e.g. [26, Theorem 3.3 vii)].

2.3.2 Abstract Wiener Spaces

We set the abstract Wiener space associated to the Gaussian random variable X as

$$W_X := \left\{ e \in E \,\middle|\, \exists g \in G_X \,\text{with}\, e = \int_\Omega gX \,\mathrm{d}\mu \right\} \tag{2.4}$$

with the norm being given by

$$\|e\|_{W_X} = \inf \left\{ \|g\|_{G_X} \,\middle|\, g \in G_X \,\text{with}\, e = \int_\Omega gX \mathrm{d}\mu \right\}. \tag{2.5}$$

In the literature, abstract Wiener spaces are commonly used to construct Gaussian measures, see [19, Chapter 8]. In any case, W_X is a Hilbert space, see Lemma 3.1.1. One can naturally switch between the spaces G_X and W_X by using a canonical isometric isomorphism $V_X : G_X \to W_X$ which is given by $V_X g := \int_\Omega gX \,\mathrm{d}\mu$ with $g \in G_X$, see Lemma 3.1.1. In particular, the adjoint $V_X^* : W_X \to G_X$ of V_X is given via $V_X^* = V_X^{-1}$, see Lemma 3.1.1. Furthermore if we extend V_X to the whole space $L^2(\mathcal{A})$ by defining $\hat{V}_X : L^2(\mathcal{A}) \to W_X$ via $\hat{V}_X g := \int_\Omega gX \,\mathrm{d}\mu$ for $g \in L^2(\mathcal{A})$, then Lemma 3.1.1 shows $\ker(\hat{V}_X) = G_X^\perp$ and that the adjoint $\hat{V}_X^* : W_X \to L^2(\mathcal{A})$ is given by $\hat{V}_X^* = V_X^{-1}$, as in Lemma 3.1.1. One can utilize $\hat{V}_X^*$ to map W_X into $L^2(\mathcal{A})$ and then apply $\hat{V}_X$ to map it into W_Y. This establishes a natural connection between W_X and W_Y, leading to the definition of the operator $L_W : W_X \to W_Y$ as follows:

$$\begin{array}{ccc} W_X & \xrightarrow{\quad L_W \quad} & W_Y \\ & {\scriptstyle \hat{V}_X^*} \searrow \quad \nearrow {\scriptstyle \hat{V}_Y} & \\ & L^2(\mathcal{A}) & \end{array} \tag{2.6}$$

We note that if $Y = LX + N$ where $L : E \to F$ is a bounded operator and $N : \Omega \to F$ is a Gaussian random variable independent of X, then $L_W = L|_{W_X}$. For details, see Lemma 3.1.3. Furthermore, it can be proven that for all $w_y \in W_Y$, we have $L_W^* w_y \in W_Z$, see Lemma 3.2.1. Moreover, Lemma 3.2.1 shows $\operatorname{ran}(L_W^*) \subseteq W_Z$ and thus we can consider the new operator $M_W : W_Y \to W_Z$ given by

$$M_W w_y := L_W^* w_y = \hat{V}_Z \hat{V}_Y^* w_y. \tag{2.7}$$

Note that M_W is a bounded and linear operator with $\|M_W\| \leq 1$. In addition M_W is surjective since $\hat{V}_Z : L^2(\mathcal{A}) \to W_Z$ is surjective and $\ker(\hat{V}_Z)^\perp = G_Z \subseteq G_Y = \operatorname{Im}(\hat{V}_Y^*)$, see Lemmata 3.1.1 and 3.2.14. Finally, M_W is in general *not* an isometry, see Lemma 3.2.17.

2.3.3 Reproducing Covariance Space

We finally introduce a third Hilbert space H_X associated to a Gaussian random variable X. To this end let $\iota : E \to E''$ be the canonical isometric embedding, that is $\iota(e)(e') = e'(e)$. Since $W_X \subseteq E$ we can thus define $H_X(E') := \iota(W_X)$ and equip $H_X(E')$ with the scalar product of W_X that is $\langle \iota f, \iota g\rangle_{H_X(E')} := \langle f, g\rangle_{W_X}$ with $f, g \in W_X$. The space $(H_X(E'), \langle\,\cdot\,,\,\cdot\,\rangle_{H_X(E')})$ is an RKHS whose kernel is given by $k_{X,E'}(e'_1, e'_2) = \langle \mathrm{cov}(X)e'_1, e'_2\rangle_{E,E'}$, see [44]. Restricting the kernel $k_{X,E'}$ to the unit ball $B_{E'}$ gives the kernel $k_X : B_{E'} \times B_{E'} \to \mathbb{R}$, whose RKHS is $H_X := H_X(B_{E'}) := H_X(E')|_{B_{E'}}$, with the kernel then naturally given by $k_X(e'_1, e'_2) = \langle \mathrm{cov}(X)e'_1, e'_2\rangle_{E,E'}$ for $e'_1, e'_2 \in B_{E'}$, see e.g. [38, Lemma 4.3]. Note that this introduces an isometric isomorphism $\iota_X : W_X \to H_X$ given by $\iota_X(w) = (\iota w)|_{B_{E'}}$, as well as a canonical isometric isomorphism $U_X : H_X \to G_X$

$$U_X w := V_X^* \iota_X^* w = V_X^{-1} \iota_X^{-1} w. \tag{2.8}$$

We can canonically and isometrically switch between all three spaces G_X, W_X and H_X, meaning that e.g. for an orthogonal projection in G_X, we find corresponding orthogonal projections in W_X and H_X.

Each of these spaces highlights different aspects of a Gaussian random variable. The Gaussian Hilbert space emphasizes its probabilistic nature, and the connection with conditioning is most direct in this setting. The abstract Wiener space, on the other hand, provides a more natural framework in which the Gaussian random variable “lives”. Thus, statements about the abstract Wiener space are usually the most interpretable and have the closest connection to the Gaussian random variable. Finally, the reproducing covariance space is similar to the abstract Wiener space but has the advantage of being an RKHS, which allows us to apply the full machinery of kernel methods to analyze the Gaussian random variable. The main drawback of the reproducing covariance space is that it can be viewed as a subspace of the bidual of the Banach space containing the Gaussian random variable, making statements concerning it often technical and, at times, difficult to interpret.

In the upcoming Chapter 3, we will see how the properties of each of these spaces can be leveraged to obtain significant results.

Chapter 3

Conditioning of Gaussian Random Variables

In this chapter we investigate how conditioning of Gaussian random variables behaves and what the implications of this behavior are. The main focus lies in the fact that our observations are not necessarily finite-dimensional, but may instead take values in a separable Banach space F. Despite its importance, conditioning in infinite-dimensional settings is considerably less well understood than in the finite-dimensional case. Nevertheless, several approaches have been developed. For Gaussian random variables taking values in separable Hilbert spaces one approach to solve the conditioning problem is using shorted operators [24]. For separable Banach spaces, Ref. [25] shows that the conditioned Gaussian random variable can be represented by a series. More recently, [26] addressed the conditioning problem via conditional probability, representing the conditioned Gaussian random variable as the limit of a sequence converging in the Banach space norm.

A common feature of these existing results is that they establish convergence but do not provide any rate at which this convergence occurs. The aim of this chapter is to close that gap. Specifically, given two jointly Gaussian random variables X and Y, we consider a sequence of $n \in \mathbb{N}$ linear measurements on Y, denoted as $\{f'_j(Y)\}_{j=0}^{n-1}$. We then derive explicit convergence rates for the operator norm of the conditional covariance operator

$$\left\|\operatorname{cov}(X|Y) - \operatorname{cov}\left(X|\{f'_j(Y)\}_{j=0}^{n-1}\right)\right\| \leq C a_n \tag{3.1}$$

for some sequence (a_n) and $C > 0$, see Theorem 3.3.10 and Corollary 3.3.11 for the precise statements. In particular, when X is a Gaussian process with continuous paths, Equation (3.1) implies a convergence rate for the associated covariance function of $X|Y$ in the supremum norm [26].

We obtain (3.1) by introducing a bounded conditioning operator M such that the observed Gaussian random variable Y is mapped to the conditional expectation $Z := \mathbb{E}(X|Y)$, i.e., $MY = Z$.

This operator enables the transfer of known posterior covariance convergence rates of $\left\|\operatorname{cov}(Y) - \operatorname{cov}\left(Y|\{f'_j(Y)\}_{j=0}^{n-1}\right)\right\|$ [31,40] onto the left-hand side of (3.1). The boundedness of the operator M is crucial for this transfer, making the choice of Banach spaces for X and Y a key consideration, even though these spaces do not directly affect the Gaussian variables [19].

We also emphasize that our analysis is not limited to point evaluations, which are common in the literature [45]. But instead we allow any general linear measurement as in [25].

This chapter is structured as follows: In Section 3.1, we use the spaces G_X, W_X, and H_X to introduce canonical operators that allow us to analyze the conditioning of Gaussian random variables. Section 3.2 presents fundamental results on conditional expectation that are required for the main results in Section 3.3. There, we show that conditional expectation can be interpreted as an orthogonal projection in the spaces G_X, W_X, and H_X, and we introduce the operator $M : F \to E$ such that $MY = Z$. In the final Section 3.4, we provide examples demonstrating that the results from Section 3.3 can be applied directly.

3.1 The Connection between G_X, W_X and H_X

We first show that there exists an isometric isomorphism between G_X and W_X.

Lemma 3.1.1. *Let Assumption A be satisfied and $V_X : G_X \to W_X$ be the map given by*

$$V_X g := \int_\Omega gX \,\mathrm{d}\mu\,.$$

Then the following statements hold true:

i) V_X is an isometric isomorphism and its adjoint V_X^* of V_X is given by $V_X^* = V_X^{-1}$.

ii) W_X is a Hilbert space.

iii) The operator V_X can be extended to an operator $\hat{V}_X : L^2(\mathcal{A}) \to W_X$ given by

$$\hat{V}_X g := \int_\Omega gX \,\mathrm{d}\mu$$

and we have $\ker(\hat{V}_X) = G_X^\perp$. Additionally the adjoint $\hat{V}_X^*$ of $\hat{V}_X$ is given by $\hat{V}_X^* = V_X^{-1}$.

Proof. i) First we note that $\mathrm{Im}(V_X) = W_X$ directly follows from the definition of W_X, as stated in equation (2.4). Moreover, for the proof of the injectivity we fix a $g \in \ker(V_X)$. Then we have

$$0 = V_X g = \int_\Omega gX \,\mathrm{d}\mu,$$

and consequently, for all $e' \in E'$, we obtain

$$0 = e'(V_X g) = \int_\Omega g e'(X) \,\mathrm{d}\mu.$$

This in turn implies $g \in G_X^\perp = \{0\}$, where $G_X^\perp$ is the orthogonal complement of G_X in itself.

Since we have just seen that V_X is bijective, we now conclude that it is isometric by the very definition (2.5) of the norm of W_X.

ii) Directly follows from *i)*.

iii) We now write $G_X^\perp$ for the orthogonal complement of G_X in $L^2(\mathcal{A})$. Then, for all $g \in L^2(\mathcal{A})$, there exist unique $g_X \in G_X$ and $g_X^\perp \in G_X^\perp$ such that

$$g = g_X + g_X^\perp \,. \tag{3.2}$$

Note that for all $e' \in E'$ we have

$$e'\left(\int_\Omega g_X^\perp X \,\mathrm{d}\mu\right) = \int_\Omega g_X^\perp e'(X)\mathrm{d}\mu = 0,$$

and hence Hahn-Banach's theorem shows

$$\int_\Omega g_X^\perp X \,\mathrm{d}\mu = 0\,.$$

Using $g = g_X + g_X^\perp$ we then see that $\hat{V}_X g \in W_X$ and $\ker(\hat{V}_X) = G_X^\perp$.

To prove the last assertion, we fix a $w \in W_X$ and a $g \in L^2(\mathcal{A})$ with (3.2). Then we have

$$\langle g, \hat{V}_X^* w\rangle_{L^2(\mathcal{A})} = \langle \hat{V}_X g, w\rangle_{W_X} = \langle V_X g_X, w\rangle_{W_X} = \langle g, V_X^{-1} w\rangle_{L^2(\mathcal{A})}\,,$$

where in the second step we used $\ker(\hat{V}_X) = G_X^\perp$ and in the last step we used *i)*. □

The isometric isomorphism between W_X and H_X is given by $\iota_X : W_X \to H_X$, which is straightforward. Nevertheless, we include a proof of this statement.

Lemma 3.1.2. *The mapping $\iota_X : W_X \to H_X$, given by $\iota_X f = \iota f|_{B_{E'}}$, is an isometric isomorphism.*

Proof. By the definition of H_X, the mapping ι_X is surjective.

For $0 \neq f \in W_X$, we have by [31, Theorem 10.46] and the definition of $H_X(E'')$ that $\|\iota_X f\|_{H_X} = \|\iota f\|_{H_X(E'')} = \|f\|_{W_X} \neq 0$, hence $\iota_X f \neq 0$. We conclude that ι_X is injective and therefore an isometric isomorphism. □

The next lemma demonstrates that the abstract operator $L_W = \hat{V}_Y \hat{V}_X^*$ can be easily calculated in the common setting of $Y = LX + N$ with $L : E \to F$ being a bounded operator and N being an F-valued Gaussian random variable independent of X.

Lemma 3.1.3. *Let Assumption A be satisfied and $L : E \to F$ be a bounded operator. Moreover, let*

$$Y := LX + N,$$

with $N : \Omega \to F$ being a Gaussian random variable independent of X. Then we have

$$L_W = L|_{W_X}.$$

Proof. Let us fix a $g \in G_X$, since g is centered and N is independent of X, we then have $\int_\Omega gN \,\mathrm{d}\mu = 0$, and hence we obtain

$$L_{|W_X}\left(\int_\Omega gX \,\mathrm{d}\mu\right) = \int_\Omega gLX \,\mathrm{d}\mu = \int_\Omega g(LX+N)\,\mathrm{d}\mu = \int_\Omega gY \,\mathrm{d}\mu = \hat{V}_Y g$$
$$= L_W\left(\int_\Omega gX \,\mathrm{d}\mu\right),$$

where in the last line we used the definition of L_W as in (2.6), i.e. $L_W = \hat{V}_Y\hat{V}_X^*$. □

3.2 Conditional Expectation

In this section, we state some general results about conditional expectation that are needed to prove the main theorems of this chapter. These results also account for null sets: the conditional expectation of a random variable remains unchanged if the variables one conditions on are modified only on a null set. This technical detail is often overlooked in the literature.

Lemma 3.2.1. *Let Assumption A be satisfied and $\mathcal{C} \subseteq \mathcal{A}$ be a sub-σ-algebra. Then for all $g \in L^2(\mathcal{C})$ the equality*

$$\int_\Omega g\mathbb{E}(X|\mathcal{C})\,\mathrm{d}\mu = \int_\Omega gX \,\mathrm{d}\mu \tag{3.3}$$

is true. Moreover for all $w \in W_Y$ we have $L_W^* w \in W_Z$, where $L_W = \hat{V}_Y\hat{V}_X^*$ as in (2.6).

Proof. Equation (3.3) follows by using [34, Proposition 2.6.31] with

$$\beta : \mathbb{R} \times E \to E$$
$$(a, e) \mapsto ae.$$

To prove the second assertion we fix a $w \in W_Y$ and set $\mathcal{C} := \sigma(Y)$. For $g := V_Y^* w \in G_Y$, we then have $w = \int_\Omega gY \,\mathrm{d}\mu$, which in turn leads to

$$L_W^* w = \hat{V}_X\hat{V}_Y^* w = \int_\Omega gX \,\mathrm{d}\mu = \int_\Omega gZ \,\mathrm{d}\mu \in W_Z\,,$$

where in the last equation we used (3.3). □

The next Lemma can be found for example in [34, Proposition 2.6.31].

Lemma 3.2.2. *Let E be a separable Banach space, $X \in L^1(\mathcal{A}, E)$, and $\mathcal{B} \subseteq \mathcal{A}$ be a σ-algebra. Then for all $e' \in E'$ there exists an $N_{e'} \in \mathcal{A}$ with $\mu(N_{e'}) = 0$ such that*

$$e'\left(\mathbb{E}(X|\mathcal{B})\right)(\omega) = \mathbb{E}(e'(X)|\mathcal{B})(\omega) \tag{3.4}$$

holds true for all $\omega \in \Omega \setminus N_{e'}$.

Lemma 3.2.3. *Under Assumption A the following equality holds true for all $e' \in E'$*

$$\langle \mathrm{cov}(X|Y)e', e'\rangle_{E,E'} = \mathrm{cov}(e'(X)|Y).$$

Proof. First we set $X_0 := X - \mathbb{E}(X|Y)$. We apply Lemma 3.2.2 thrice to obtain

$$\begin{aligned}
\langle \mathrm{cov}(X|Y)e', e'\rangle_{E,E'} &= \langle e'\left(\mathbb{E}((X - \mathbb{E}(X|Y)) \otimes (X - \mathbb{E}(X|Y))|Y)\right), e'\rangle_{E,E'} \\
&= \langle e'\left(\mathbb{E}(X_0 \otimes X_0|Y)\right), e'\rangle_{E,E'} \\
&= \langle \mathbb{E}(e'(X_0)X_0|Y), e'\rangle_{E,E'} \\
&= e'\left(\mathbb{E}(e'(X_0)X_0)|Y)\right) \\
&= \mathbb{E}(e'(X_0)e'(X_0)|Y) \\
&= \mathbb{E}((e'(X) - \mathbb{E}(e'(X)|Y))^2|Y) \\
&= \mathrm{cov}(e'(X)|Y).
\end{aligned}$$

□

The next theorem can be found in [34, Theorem 3.3.2]. Moreover, the upcoming theorem is also essential for the proofs of [26].

Theorem 3.2.4. *Let $(\mathcal{A}_n)_{n\in\mathbb{N}}$ be a filtration in the probability space $(\Omega, \mathcal{A}, \mu)$ and we define $\mathcal{A}_\infty := \sigma\left(\cup \mathcal{A}_n \,:\, n \geq 1\right)$. Then for all $p \in [1, \infty)$. and all $X \in L^p(\mathcal{A}, E)$, we have*

$$\lim_{n\to\infty} \mathbb{E}(X|\mathcal{A}_n) = \mathbb{E}(X|\mathcal{A}_\infty)\,,$$

where the convergence is almost everywhere and in the norm of $L^p(\mathcal{A}, E)$.

The next lemma shows that the Gaussian random variable Y induces the same σ-algebra as the pre-Gaussian Hilbert space $\{f'(Y) \,|\, f' \in F'\}$:

Lemma 3.2.5. *Let Assumption A be satisfied and $\mathcal{G} := \{f'(Y) \,|\, f' \in F'\}$. Then we have $\sigma(Y) = \sigma(\mathcal{G})$.*

Proof. For $f' \in F'$, we have

$$Y^{-1}(\sigma(f')) = Y^{-1}(f'^{-1}(\mathcal{B})) = (f' \circ Y)^{-1}(\mathcal{B}) = \sigma(f'(Y))\,,$$

where $\mathcal{B}$ is the Borel-σ-algebra of $\mathbb{R}$. This leads to

$$\begin{aligned}
\sigma(Y) = Y^{-1}(\sigma(F')) = Y^{-1}\left(\sigma\left(\bigcup_{f'\in F'} \sigma(f')\right)\right) &= \sigma\left(Y^{-1}\left(\bigcup_{f'\in F'} \sigma(f')\right)\right) \\
&= \sigma\left(\bigcup_{f'\in F'} Y^{-1}(\sigma(f'))\right) \\
&= \sigma\left(\bigcup_{f'\in F'} \sigma(f'(Y))\right) \\
&= \sigma\left(\bigcup_{g\in\mathcal{G}} \sigma(g)\right) \\
&= \sigma(\mathcal{G})\,,
\end{aligned}$$

where in the third step we used [46, Theorem 1.81]. □

Lemma 3.2.6. *Let $(\Omega, \mathcal{A}, \mu)$ be a probability space, E be a separable Banach space, and $X, Y : \Omega \to E$ random variables. Moreover, assume that for all $e' \in E'$ there exists a null-set $N_{e'} \in \mathcal{A}$ such that*

$$e'(X(\omega)) = e'(Y(\omega)), \qquad \omega \in \Omega \setminus N_{e'} .$$

Then $X = Y$ almost everywhere.

Proof. Let $\mathcal{D} \subset B_{E'}$ be a countable weak-$*$-dense subset, according to Lemma B.1.2. We set

$$N_{\mathcal{D}} := \cup_{e' \in \mathcal{D}} N_{e'} .$$

Then $N_{\mathcal{D}}$ is a union of countable null-sets, thus again a null-set. Moreover, our construction shows

$$e'(X(\omega)) = e'(Y(\omega)), \qquad \omega \in \Omega \setminus N_{\mathcal{D}},\, e' \in \mathcal{D}.$$

Applying Lemma B.1.1 leads to $X(\omega) = Y(\omega)$ for all $\omega \in \Omega \setminus N_{\mathcal{D}}$. □

In the following, we like to introduce conditional expectations $\mathbb{E}(X|G_Y)$ with respect to a given Gaussian Hilbert space G_Y. Here we note that formally G_Y is a subset of $L^2(\mathcal{A})$ that is, a collection of *μ-equivalence classes*. Consequently, a definition like

$$\mathbb{E}(X|G_Y) := \mathbb{E}\big(X \,\big|\, \sigma\left(\{g : g \in G_Y\}\right)\big)$$

is *not* well defined. To address this problem we need a couple of auxilliary results. We begin with the following characterization of conditional expectations, where in its proof we need the indicator function $\chi_A : \Omega \to \mathbb{R}$ of $A \subseteq \Omega$, that is, the function defined by

$$\chi_A(\omega) := \begin{cases} 1, & \text{if } \omega \in A \\ 0, & \text{else.} \end{cases}$$

Lemma 3.2.7. *Let $(\Omega, \mathcal{A}, \mu)$ be a probability space, $X \in L^1(\mathcal{A}, E)$, and $\mathcal{B} \subset \mathcal{A}$ be a sub-σ-algebra. Furthermore let $\mathcal{E} \subset \mathcal{A}$ be an $\cap$-stable system with $\Omega \in \mathcal{E}$ and $\sigma(\mathcal{E}) = \mathcal{B}$. If $Z : \Omega \to E$ is $\mathcal{B}$-measurable, μ-integrable, and satisfies*

$$\int_B Z \,\mathrm{d}\mu = \int_B X \,\mathrm{d}\mu \tag{3.5}$$

for all $B \in \mathcal{E}$, then we have $Z \in \mathbb{E}(X|B)$.

Proof. We consider the Banach space valued mappings $Q, P : \mathcal{B} \to E$ given by

$$Q(B) := \int_B X \,\mathrm{d}\mu$$
$$P(B) := \int_B Z \,\mathrm{d}\mu.$$

Moreover we define

$$\mathcal{D} := \{B \in \mathcal{B} \,|\, Q(B) = P(B)\}.$$

Note that $\Omega \in \mathcal{E}$ by assumption, and (3.5) implies $\mathcal{E} \subseteq \mathcal{D}$. Let us now show that $\mathcal{D}$ is a Dynkin system in the sense of [33, Chapter 1.6]. Obviously we have $\Omega \in \mathcal{E} \subseteq \mathcal{D}$. Moreover, if $A, B \in \mathcal{D}$ with $A \subseteq B$ then

$$\int_{B\setminus A} Z \,\mathrm{d}\mu = \int_B Z \,\mathrm{d}\mu - \int_A Z \,\mathrm{d}\mu = \int_B X \,\mathrm{d}\mu - \int_A X \,\mathrm{d}\mu = \int_{B\setminus A} X \,\mathrm{d}\mu$$

and therefore $B \setminus A \in \mathcal{D}$. Finally if we have an increasing sequence $(A_n) \subseteq \mathcal{D}$ then for $A := \cup_{n\in\mathbb{N}} A_n$ we have

$$\int_A Z \,\mathrm{d}\mu = \int_\Omega \lim_{n\to\infty} \chi_{A_n} Z \,\mathrm{d}\mu = \lim_{n\to\infty} \int_{A_n} Z \,\mathrm{d}\mu = \lim_{n\to\infty} \int_{A_n} X \,\mathrm{d}\mu = \int_A X \,\mathrm{d}\mu,$$

where we used in the second and last step the dominated convergence theorem for Bochner integrals [34, Proposition 1.2.5]. Therefore $\mathcal{D}$ is indeed Dynkin class. By [33, Theorem 1.6.2] we find that $\sigma(\mathcal{E}) = d(\mathcal{E}) \subseteq d(\mathcal{D}) = \mathcal{D}$, where $d(\mathcal{E})$ and $d(\mathcal{D})$ denote the smallest Dynkin system containing $\mathcal{E}$ and $\mathcal{D}$, respectively. Since $\sigma(\mathcal{E}) = \mathcal{B}$ we find the assertion. □

Let us now consider the σ-algebra $\mathcal{T}$ of μ-trivial events, that is

$$\mathcal{T} := \{T \in \mathcal{A} \,|\, \mu(T) \in \{0, 1\}\}.$$

Given a σ-algebra $\mathcal{B} \subseteq \mathcal{A}$ we further define

$$\mathcal{B}_\cap := \{B \cap T \,|\, B \in \mathcal{B},\, T \in \mathcal{T}\},$$
$$\mathcal{B}_\cup := \{B \cup T \,|\, B \in \mathcal{B},\, T \in \mathcal{T}\}.$$

The next lemma collects some simple but yet important properties of these sets.

Lemma 3.2.8. *For all σ-algebras $\mathcal{B} \subseteq \mathcal{A}$ the following statements hold true:*

i) The set $\mathcal{B}_\cap$ is $\cap$-stable and $\Omega \in \mathcal{B}_\cap$.

ii) We have $\sigma(\mathcal{B}_\cap) = \sigma(\mathcal{B}_\cup) =: \hat{\mathcal{B}}$.

iii) It holds that $\mathcal{B} \subseteq \hat{\mathcal{B}}$.

iv) We have $\hat{\mathcal{B}} = \hat{\hat{\mathcal{B}}}$.

v) Given a σ-algebra $\mathcal{C} \subseteq \hat{\mathcal{B}}$ we have $\hat{\mathcal{C}} \subseteq \hat{\mathcal{B}}$.

Proof. *i*). The first statement follows by the fact that intersections are commutative and associative, and the second statement is obvious.

ii). We first show that $\mathcal{B}_\cap \subseteq \sigma(B_\cup)$. To this end let $B \in \mathcal{B}$ and $T \in \mathcal{T}$. Then we have

$$B \cap T = \Omega \setminus (\Omega \setminus (B \cap T)) = \Omega \setminus ((\Omega \setminus B) \cup (\Omega \setminus T)) \in \sigma(\mathcal{B}_\cup).$$

To prove $\mathcal{B}_\cup \subseteq \sigma(\mathcal{B}_\cap)$ we fix again some $B \in \mathcal{B}$ and $T \in \mathcal{T}$. Then we have

$$B \cup T = \Omega \setminus (\Omega \setminus (B \cup T)) = \Omega \setminus ((\Omega \setminus B) \cap (\Omega \setminus T))) \in \sigma(\mathcal{B}_\cap).$$

In summary we thus find $\sigma(\mathcal{B}_\cap) \subseteq \sigma(\sigma(\mathcal{B}_\cup)) = \sigma(\mathcal{B}_\cup) \subseteq \sigma(\sigma(\mathcal{B}_\cap)) = \sigma(\mathcal{B}_\cap)$.

iii). This follows from $\mathcal{B} \subseteq \mathcal{B}_\cap$.

iv). For the first statement we note that $\hat{\mathcal{B}} \subseteq \hat{\hat{\mathcal{B}}}$ is obvious. For the converse inclusion we note that $\hat{\mathcal{B}}_\cap \subseteq \hat{\mathcal{B}}$ holds true. Since $\hat{\mathcal{B}}$ is a σ-algebra this statement follows.

v). This follows by $\hat{\mathcal{C}} \subseteq \hat{\hat{\mathcal{B}}} = \hat{\mathcal{B}}$. □

Lemma 3.2.9. *Given σ-algebras $\mathcal{B} \subseteq \mathcal{C} \subseteq \hat{\mathcal{B}} \subseteq \mathcal{A}$, we have $\hat{\mathcal{C}} = \hat{\mathcal{B}}$.*

Proof. We first show $\hat{\mathcal{C}} \subseteq \hat{\mathcal{B}}$. To this end let $C \in \mathcal{C}$ and $T \in \mathcal{T}$. By assumption we find $C \in \hat{\mathcal{B}}$ and $T \in \mathcal{T} \subseteq \hat{\mathcal{B}}$, and thus $C \cap T \in \hat{\mathcal{B}}$. This shows $\hat{\mathcal{C}} = \sigma(\mathcal{C}_\cap) \subseteq \hat{\mathcal{B}}$.

For the converse conclusion we note that $\mathcal{B} \subseteq \mathcal{C}$ implies $\mathcal{B}_\cap \subseteq \mathcal{C}_\cap$. By Lemma 3.2.8 we conclude $\hat{\mathcal{B}} = \sigma(\mathcal{B}_\cap) \subseteq \sigma(\mathcal{C}_\cap) = \hat{\mathcal{C}}$. □

Lemma 3.2.10. *Let $X \in L^1(\mathcal{A}, E)$. Then for all σ-algebras $\mathcal{B} \subseteq \mathcal{C} \subseteq \hat{\mathcal{B}} \subseteq \mathcal{A}$, we have*

$$\mathbb{E}(X|\mathcal{B}) \subseteq \mathbb{E}(X|\mathcal{C}) \subseteq \mathbb{E}(X|\hat{\mathcal{B}}).$$

Furthermore, for all $Z \in \mathbb{E}(X|\mathcal{B})$, $Z_\mathcal{C} \in \mathbb{E}(X|\mathcal{C})$, and $\hat{Z} \in \mathbb{E}(X|\hat{\mathcal{B}})$, we have

$$\mu(Z \neq Z_\mathcal{C}) = \mu(Z \neq \hat{Z}) = \mu(Z_\mathcal{C} \neq \hat{Z}) = 0. \tag{3.6}$$

Proof. Let us fix a $Z \in \mathbb{E}(X|\mathcal{B})$. Our first goal is to show $Z \in \mathbb{E}(X|\hat{\mathcal{B}})$. Then Z is $\hat{\mathcal{B}}$-measurable. Moreover, for $B \in \mathcal{B}$ and $T \in \mathcal{T}$ with $\mu(T) = 1$, we have

$$\int_{B\cap T} Z \,\mathrm{d}\mu = \int_B Z \,\mathrm{d}\mu = \int_B X \,\mathrm{d}\mu = \int_{B\cap T} X \,\mathrm{d}\mu, \tag{3.7}$$

where in the first and last step we used $B = (B\cap T)\cup(B\cap T^c)$ and $\mu(B\cap T^c) \leq \mu(T^c) = 0$. Furthermore, for $B \in \mathcal{B}$ and $T \in \mathcal{T}$ with $\mu(T) = 0$, Equation (3.7) is obviously satisfied. Consequently, we have

$$\int_B Z \,\mathrm{d}\mu = \int_B X \,\mathrm{d}\mu \tag{3.8}$$

for all $B \in \mathcal{B}_\cap$. With the help of Lemma 3.2.8 and Lemma 3.2.7, we conclude that (3.8) holds for all $B \in \hat{\mathcal{B}}$. This finishes the proof of $Z \in \mathbb{E}(X|\hat{\mathcal{B}})$.

Next, we note that for $\hat{Z} \in \mathbb{E}(X|\hat{\mathcal{B}})$ we have $\mu(Z \neq \hat{Z}) = 0$ by the almost sure uniqueness of the conditional expectation of X given $\hat{\mathcal{B}}$.

Our next goal is to show that $Z \in \mathbb{E}(X|\mathcal{C})$. Since $\mathcal{B} \subseteq \mathcal{C}$, it follows that Z is $\mathcal{C}$-measurable. Since we have already seen that (3.8) holds for all $B \in \hat{\mathcal{B}}$ it also holds for all $B \in \mathcal{C}$. This shows $Z \in \mathbb{E}(X|\mathcal{C})$.

Again, we note that for $Z_\mathcal{C} \in \mathbb{E}(X|\mathcal{C})$ we have $\mu(Z \neq Z_\mathcal{C}) = 0$ by the almost sure uniqueness of the conditional expectation of X given $\mathcal{C}$.

Our last goal is to establish the inclusion $\mathbb{E}(X|\mathcal{C}) \subseteq \mathbb{E}(X|\hat{\mathcal{B}})$ together with the last identity of (3.6). To this end we note that we already know that $\mathbb{E}(X|\mathcal{C}) \subseteq \mathbb{E}(X|\hat{\mathcal{C}})$ and

$\mu(Z_{\mathcal{C}} \neq \hat{Z}) = 0$ for all $Z_{\mathcal{C}} \in \mathbb{E}(X|\mathcal{C})$ and $\hat{Z} \in \mathbb{E}(X|\hat{\mathcal{C}})$. Therefore it suffices to show that $\hat{\mathcal{C}} = \hat{\mathcal{B}}$. For the proof of the inclusion $\hat{\mathcal{C}} \subseteq \hat{\mathcal{B}}$. We fix a $C \in \mathcal{C}$ and a $T \in \mathcal{T}$. This gives $C \in \hat{\mathcal{B}}$ by assumption and $T \in \hat{\mathcal{B}}$ by the definition of $\hat{\mathcal{B}}$, and therefore $C \cap T \in \hat{\mathcal{B}}$. In other words we have $\mathcal{C}_\cap \subseteq \hat{\mathcal{B}}$ and by Lemma 3.2.8, we conclude $\hat{\mathcal{C}} = \sigma(\mathcal{C}_\cap) \subseteq \hat{\mathcal{B}}$. The converse inclusion follows by $\mathcal{B} \subseteq \mathcal{C}$. □

In the following we need the space

$$\mathcal{L}^2(\mathcal{A}) := \left\{ g : \Omega \to \mathbb{R} \,\middle|\, g \text{ is measurable, } \int_\Omega g^2 \,\mathrm{d}\mu < \infty \right\}$$

equipped with the usual $\|\cdot\|_{\mathcal{L}^2(\mathcal{A})}$ semi-norm. Moreover, for $f \in \mathcal{L}^2(\mathcal{A})$ we denote its μ-equivalence class by $[f]$. Note that we have $[f] \in L^2(\mathcal{A})$.

Lemma 3.2.11. *Let $G \subseteq \mathcal{L}^2(\mathcal{A})$ be non empty and*

$$\hat{G} := \left\{ g \in \mathcal{L}^2(\mathcal{A}) \,\middle|\, \exists (g_n) \subseteq G : [g_n] \to [g] \text{ in } L^2(\mathcal{A}) \right\}. \tag{3.9}$$

Then for $\mathcal{B} := \sigma(G)$ we have $\mathcal{B} \subseteq \sigma(\hat{G}) \subseteq \hat{\mathcal{B}}$.

Proof. The statement $\mathcal{B} \subseteq \sigma(\hat{G})$ follows directly from the fact that $G \subseteq \hat{G}$.

Next, we prove the statement $\sigma(\hat{G}) \subseteq \hat{\mathcal{B}}$. To this end it suffices to show that all $g \in \hat{G}$ are $\hat{\mathcal{B}}$-measurable. Let us therefore fix a $g \in \hat{G}$. Then there exists a sequence $(g_n) \subseteq G$ such that $[g_n] \to [g]$ in $L^2(\mathcal{A})$. By the properties of convergence in $L^2(\mathcal{A})$, there exists a subsequence (g_{n_k}) such that $g_{n_k} \to g$ μ-almost surely. Consequently the set $T := \{\omega \in \Omega \mid g_{n_k}(\omega) \to g(\omega)\}$ is $\mathcal{A}$-measurable with $\mu(T) = 1$, i.e., we have $T \in \mathcal{T}$.

Let us now define $\hat{g}_{n_k} := \chi_T g_{n_k}$ and $\hat{g} := \chi_T g$. Obviously we have $\hat{g}_{n_k} \to \hat{g}$ pointwise. Moreover, the definition of $\hat{G}$ ensures $\hat{g}_{n_k}, \hat{g} \in \hat{G}$. Furthermore, $g_{n_k} \in G$ shows that g_{n_k} is $\mathcal{B}$-measurable. Since χ_T is $\mathcal{T}$-measurable and $\mathcal{T} \subseteq \hat{\mathcal{B}}$, it follows that $\hat{g}_{n_k} = \chi_T g_{n_k}$ is $\hat{\mathcal{B}}$-measurable. Thus, the limit function $\hat{g}$ is also $\hat{\mathcal{B}}$-measurable.

With these preparations we now show that $g^{-1}(A) \in \hat{\mathcal{B}}$ for all measurable $A \subseteq \mathbb{R}$. To this end we note that

$$\begin{aligned} g^{-1}(A) &= \{\omega \in \Omega \,|\, g(\omega) \in A\} \\ &= \big(\{\omega \in \Omega \,|\, g(\omega) \in A\} \cap T\big) \cup \big(\{\omega \in \Omega \,|\, g(\omega) \in A\} \setminus T\big) \\ &= \big(\{\omega \in \Omega \,|\, \hat{g}(\omega) \in A\} \cap T\big) \cup \big(\{\omega \in \Omega \,|\, g(\omega) \in A\} \setminus T\big). \end{aligned}$$

Since $\{\omega \in \Omega \mid \hat{g}(\omega) \in A\} \in \hat{\mathcal{B}}$ and $T \in \hat{\mathcal{B}}$, it follows that $\{\omega \in \Omega \mid \hat{g}(\omega) \in A\} \cap T \in \hat{\mathcal{B}}$. Furthermore, we have

$$\begin{aligned} \mu\big(\{\omega \in \Omega \,|\, g(\omega) \in A\} \setminus T\big) &= \mu\big(\{\omega \in \Omega \,|\, g(\omega) \in A\} \cap (\Omega \setminus T)\big) \\ &\leq \mu(\Omega \setminus T) = 0. \end{aligned}$$

Therefore, we have $\{\omega \in \Omega \mid g(\omega) \in A\} \setminus T \in \hat{\mathcal{B}}$ and by combining this with our previous consideration we conclude $g^{-1}(A) \in \hat{\mathcal{B}}$. □

With these preparations and under Assumption A we now define

$$\mathbb{E}(X|G_Y) := \mathbb{E}(X|\sigma(\hat{G})),$$

where $G := \{f'(Y) \,|\, f' \in F'\}$ and $\hat{G}$ is defined by (3.9). The following theorem describes $\mathbb{E}(X|G_Y)$.

Theorem 3.2.12. *Let Assumption A be satisfied. Then, we have*

$$\mathbb{E}(X|\sigma(Y)) \subseteq \mathbb{E}(X|G_Y) \subseteq \mathbb{E}(X|\widehat{\sigma(Y)}).$$

Furthermore, for all $Z \in \mathbb{E}(X|\sigma(Y))$, $Z_G \in \mathbb{E}(X|G_Y)$, *and* $\hat{Z} \in \mathbb{E}(X|\widehat{\sigma(Y)})$, *we have*

$$\mu(Z \neq Z_G) = \mu(Z \neq \hat{Z}) = \mu(Z_G \neq \hat{Z}) = 0.$$

Proof. We set $\mathcal{B} := \sigma(G)$ and obtain $\mathcal{B} \subseteq \sigma(\hat{G}) \subseteq \hat{\mathcal{B}}$ by Lemma 3.2.11. For $\mathcal{C} := \sigma(\hat{G})$ Lemma 3.2.10 thus shows

$$\mathbb{E}(X|\mathcal{B}) \subseteq \mathbb{E}(X|\sigma(\hat{G})) \subseteq \mathbb{E}(X|\hat{\mathcal{B}})$$

and (3.6). Finally, Lemma 3.2.5 shows $\sigma(Y) = \sigma(G) = \mathcal{B}$, thus the assertion follows. □

Lemma 3.2.13. *Let Assumption A be satisfied. Then for all* $g \in L^2(\mathcal{A})$ *we have*

$$\int_\Omega gX \,\mathrm{d}\mu = \int_\Omega \Pi_{G_X} gX \,\mathrm{d}\mu\,,$$

where $\Pi_{G_X} : L^2(\mathcal{A}) \to L^2(\mathcal{A})$ *is the orthogonal projection onto* G_X.

Proof. Since G_X is a closed subspace of $L^2(\mathcal{A})$, we decompose g into $g = \Pi_{G_X} g + \Pi_{G_X^\perp} g$, with $\Pi_{G_X^\perp}$ being the orthogonal projection onto the orthogonal complement of G_X. For simplicity we set $g_X := \Pi_{G_X} g$ and $g_X^\perp = \Pi_{G_X^\perp} g$. Given $e' \in E'$, we then have

$$\begin{aligned} e'\left(\int_\Omega gX \,\mathrm{d}\mu\right) = \int_\Omega g e'(X) \,\mathrm{d}\mu = \int_\Omega (g_X + g_X^\perp) e'(X) \,\mathrm{d}\mu = \int_\Omega g_X e'(X) \,\mathrm{d}\mu \\ = e'\left(\int_\Omega g_X X \,\mathrm{d}\mu\right), \end{aligned}$$

where we used $e'(X) \in G_X$ and $g_X^\perp \in G_X^\perp$. Since this holds true for all $e' \in E'$ we find the assertion by the Hahn-Banach theorem. □

Lemma 3.2.14. *Under Assumption A it holds* $G_Z \subseteq G_Y = G_{(Z,Y)}$ *and if* $G_Y \subseteq G_X$, *we additionally have* $G_Y = G_Z$.

Proof. For $G_Z \subseteq G_Y$, we refer to [26, Theorem 3.3 vii)]. We conclude $G_Y = G_{(Z,Y)}$ by

$$G_Y \subseteq G_{(Z,Y)} = \overline{G_Z + G_Y} \subseteq \overline{G_Y + G_Y} = G_Y\,,$$

where in the first and second step we used (2.2).

To prove the second assertion, we fix a $g \in G_Y$. Since $G_Y \subseteq G_X$, there then exists a sequence (e'_n) in E' such that $e'_n(X) \to g$ in $L^2(\mathcal{A})$. Now Lemma 3.2.1 with $\mathcal{C} := \sigma(Y)$ implies for all $g_Y \in G_Y \subseteq L^2(\mathcal{C})$

$$\int_\Omega g_Y e'_n(Z) \,\mathrm{d}\mu = \int_\Omega g_Y e'_n(X) \,\mathrm{d}\mu \to \int_\Omega g_Y g \,\mathrm{d}\mu\,.$$

We conclude that the sequence $(e'_n(Z))$ converges $L^2(\mathcal{C})$-weakly against g. Moreover, using [26, Theorem 3.3 vii)], we find

$$\|e'_n(Z) - e'_m(Z)\|_{L^2(\mathcal{C})} = \|\Pi_{G_Y}(e'_n(X) - e'_m(X))\|_{L^2(\mathcal{C})} \leq \|e'_n(X) - e'_m(X)\|_{L^2(\mathcal{A})} .$$

Since $(e'_n(X))$ is an $L^2(\mathcal{A})$-Cauchy sequence, we conclude that $(e'_n(Z))$ is an $L^2(\mathcal{C})$-Cauchy sequence. Combining the latter with the weak convergence, we conclude that $e'_n(Z) \to g$ in $L^2(\mathcal{C})$. In other words we have found $g \in G_Z$. □

Lemma 3.2.15. *Under Assumption A, it holds $W_Z \subseteq W_X$ and for all $w \in W_Z$, we have $\|w\|_{W_X} \leq \|w\|_{W_Z}$.*

Proof. By Lemma 3.2.1 with $\mathcal{C} := \sigma(Y)$ and Lemma 3.1.1 we find

$$\begin{aligned} W_Z &= \left\{ f \in E \,\middle|\, \exists g \in G_Z, f = \int_\Omega gX \,\mathrm{d}\mu \right\} \\ &\subseteq \left\{ f \in E \,\middle|\, \exists g \in L^2(\mathcal{A}), f = \int_\Omega gX \,\mathrm{d}\mu \right\} = W_X . \end{aligned}$$

For the proof of the second assertion we fix a $w \in W_Z$. We then obtain

$$\|w\|_{W_Z} = \inf_{g \in G_Z \, w = V_Z g} \|g\|_{L^2(\mathcal{A})} = \inf_{g \in G_Z \, w = \hat{V}_X g} \|g\|_{L^2(\mathcal{A})} \geq \inf_{g \in L^2(\mathcal{A}) \, w = \hat{V}_X g} \|g\|_{L^2(\mathcal{A})} = \|w\|_{W_X} ,$$

where in the second step we used Lemma 3.2.1 and in the last step we used Lemma 3.1.1. □

Lemma 3.2.16. *Let Assumption A be satisfied, $\mathcal{C} \subseteq \mathcal{A}$ be a sub-σ-algebra, and $X_\mathcal{C} := \mathbb{E}(X|\mathcal{C})$. Then for all $e'_1, e'_2 \in E'$ the following identities hold true*

i) $\operatorname{cov}(X, X_\mathcal{C}) = \operatorname{cov}(X_\mathcal{C}, X) = \operatorname{cov}(X_\mathcal{C})$,

ii) $\operatorname{cov}(X - X_\mathcal{C}) = \operatorname{cov}(X) - \operatorname{cov}(X_\mathcal{C})$.

Proof. We start with *i*). Taking $e'_1, e'_2 \in E'$, we obtain with Lemma 3.2.1

$$\langle \operatorname{cov}(X_\mathcal{C}, X) e'_1, e'_2 \rangle_{E,E'} = \int_\Omega e'_1(X_\mathcal{C}) e'_2(X) \,\mathrm{d}\mu = \int_\Omega e'_1(X_\mathcal{C}) e'_2(X_\mathcal{C}) \,\mathrm{d}\mu = \langle \operatorname{cov}(X_\mathcal{C}) e'_1, e'_2 \rangle_{E,E'} .$$

The identity $\operatorname{cov}(X, X_\mathcal{C}) = \operatorname{cov}(X_\mathcal{C})$ can be shown analogously.

The second assertion follows from the bilinearity of the cross covariance operator, namely

$$\begin{aligned} \operatorname{cov}(X - X_\mathcal{C}) &= \operatorname{cov}(X - X_\mathcal{C}, X - X_\mathcal{C}) \\ &= \operatorname{cov}(X, X) - \operatorname{cov}(X, X_\mathcal{C}) - \operatorname{cov}(X_\mathcal{C}, X) + \operatorname{cov}(X_\mathcal{C}, X_\mathcal{C}) \\ &= \operatorname{cov}(X, X) - \operatorname{cov}(X_\mathcal{C}, X_\mathcal{C}), \end{aligned}$$

where in the last step we used the first assertion. □

Lemma 3.2.17. *Let Assumption A be satisfied and W_Y be dense in F. Then the following statements are equivalent:*

i) The operator $M_W : W_Y \to W_Z$ *is an isometric isomorphism.*

ii) The equality $G_Y = G_Z$ *holds true.*

Note that i) and ii) are equivalent even if W_Y is not dense in F.

Proof. ii) $\Rightarrow$ *i).* As discussed around (2.7), the operator M_W is surjective. Furthermore, for $w \in W_Y$ we find a $g \in G_Y = G_Z$ such that $w = V_Y g$. This gives

$$\|M_W w\|_{W_Z} = \|\hat{V}_Z \hat{V}_Y^* V_Y g\|_{W_Z} = \|\hat{V}_Z g\|_{W_Z} = \|g\|_{G_Z} = \|g\|_{G_Y} = \|V_Y g\|_{W_Y} = \|w\|_{W_Y}.$$

Therefore, $M_W : W_Y \to W_Z$ is an isometric isomorphism.

i) $\Rightarrow$ *ii).* By Lemma 3.2.14 we know that $G_Z \subseteq G_Y$. To prove the converse inclusion we consider the orthogonal complement $G_Z^\perp$ of G_Z in G_Y. Let $g \in G_Z^\perp$. Then by the definition of M_W and Lemma 3.1.1 we have

$$M_W \left(\int_\Omega gY \, \mathrm{d}\mu \right) = \hat{V}_Z g = 0.$$

By assumption the latter implies $V_Y g = 0$ and since V_Y is injective we find $g = 0$. Therefore, we have $G_Z^\perp = \{0\}$. □

3.3 Convergence Analysis

In this section we present the main results of this chapter that investigate conditioning of Gaussian random variables. We start with a technical yet crucial lemma that we need to prove an even more technical theorem that allows us to essentially view the process of conditioning as orthogonal projections in the spaces $G_{(X,Y)}, W_{(X,Y)}$ and $H_{(X,Y)}$.

Lemma 3.3.1. *Under Assumption A, let* G *be a Gaussian Hilbert space, such that* $G_X \subseteq G$*. Additionally, let* $\mathcal{G} \subseteq G$ *be a closed subspace and* $(r_j)_{j\in\mathbb{N}}$ *be an ONB of* $\mathcal{G}$*. Then the series*

$$\Pi_{\mathcal{G}} X := \sum_{j=1}^{\infty} r_j \int_\Omega r_j X \, \mathrm{d}\mu$$

converges almost everywhere as well as in $L^p(\mathcal{A}, E)$ *for all* $p \geq 1$ *and its limit* $\Pi_{\mathcal{G}} X$ *is jointly Gaussian with* X*. Moreover the limit is independent of the choice of the ONB, that is if* $(\tilde{r}_j)_{j\in\mathbb{N}}$ *is another ONB of* $\mathcal{G}$*, then*

$$\sum_{j=1}^{\infty} r_j \int_\Omega r_j X \, \mathrm{d}\mu = \sum_{j=1}^{\infty} \tilde{r}_j \int_\Omega \tilde{r}_j X \, \mathrm{d}\mu$$

almost everywhere. Finally, if $\Pi_{\mathcal{G}} : G \to G$ *denotes the orthogonal projection onto* $\mathcal{G}$*, then for all* $e' \in E'$ *we have*

$$e'(\Pi_{\mathcal{G}} X) = \Pi_{\mathcal{G}} e'(X) \,. \tag{3.10}$$

Proof. In order to apply Theorem 3.2.4, we define the filtration given by the σ-algebras

$$\mathcal{A}_n := \sigma(r_1, \dots, r_n) .$$

For $e' \in E'$, Lemma 3.2.2 and (2.3) applied to $V := \operatorname{span}\{r_1, \dots, r_n\}$ give

$$\begin{aligned} e'\left(\mathbb{E}(X|\mathcal{A}_n)(\omega)\right) = \mathbb{E}(e'(X)|\mathcal{A}_n)(\omega) &= \sum_{j=1}^{n} r_j(\omega) \int_\Omega r_j e'(X)\,\mathrm{d}\mu \\ &= e'\left(\sum_{j=1}^{n} r_j(\omega) \int_\Omega r_j X\,\mathrm{d}\mu\right) \end{aligned}$$

for all $\omega \in \Omega \setminus N_{e'}$. Consequently, Lemma 3.2.6 shows

$$\mathbb{E}(X|\mathcal{A}_n) = \sum_{j=1}^{n} r_j \int_\Omega r_j X\,\mathrm{d}\mu.$$

Moreover, using Theorem 3.2.4, we obtain the convergence

$$\mathbb{E}(X|\mathcal{A}_n) \to \mathbb{E}(X|\mathcal{A}_\infty),$$

which occurs almost everywhere and in $L^p(\mathcal{A}, E)$. This implies that the series

$$\sum_{j=1}^{\infty} r_j \int_\Omega r_j X \mathrm{d}\mu = \mathbb{E}(X|\mathcal{A}_\infty)$$

is convergent in the same sense. In order to show that the resulting random variable is Gaussian we consider the sequence

$$Z_n := \mathbb{E}(X|\mathcal{A}_n) = \sum_{j=1}^{n} r_j \int_\Omega r_j X\,\mathrm{d}\mu.$$

Each Z_n is an E-valued Gaussian random variable and, as noted above, the sequence (Z_n) converges to the random variable $\Pi_{\mathcal{G}} X = \mathbb{E}(X|\mathcal{A}_\infty)$ pointwise on the set $\Omega \setminus N$, where N is a suitable null-set. Consequently, we have $e'(Z_n(\omega)) \to e'(\Pi_{\mathcal{G}}(\omega))$ for all $e' \in E'$ and $\omega \in \Omega \setminus N$.

We note for $e_1', e_2' \in E'$ we have $e_1'(\Pi_{\mathcal{G}} X) \in \mathcal{G} \subseteq G$ and $e_2'(X) \in G_X \subseteq G$, and thus $e_1'(\Pi_{\mathcal{G}} X) + e_2'(X)$ is as an element of G Gaussian. We conclude $\Pi_{\mathcal{G}} X$ is jointly Gaussian with X.

Moreover, since each $e'(Z_n)$ is an $\mathbb{R}$-valued Gaussian random variable, so is the limit $e'(Z_\infty)$, see e.g. [26, Theorem C.4].

To verify the last assertion we first note that applying (2.3) to $V := \mathcal{G}$ and $g := e'(X)$ yields

$$\sum_{j=1}^{\infty} r_j \int_\Omega r_j e'(X)\,\mathrm{d}\mu = \Pi_{\mathcal{G}} e'(X)\,, \tag{3.11}$$

where both sides are $L^2(\mathcal{A})$ μ-equivalences classes. Morevover, the series

$$\sum_{j=1}^{\infty} r_j \int_\Omega r_j X \,\mathrm{d}\mu$$

converges pointwise up to some null set N. Consequently, we find

$$e'(\Pi_{\mathcal{G}} X(\omega)) = e'\left(\sum_{j=1}^{\infty} r_j(\omega) \int_\Omega r_j X \,\mathrm{d}\mu\right) = \sum_{j=1}^{\infty} r_j(\omega) \int_\Omega r_j e'(X) \,\mathrm{d}\mu \tag{3.12}$$

for all $\omega \in \Omega \setminus N$. Combining this with (3.11) yields (3.10).

Finally, let $(\tilde{r}_j)_{j\in\mathbb{N}}$ be another ONB of $\mathcal{G}$. The already proven parts then show that

$$\tilde{\Pi}_{\mathcal{G}} X := \sum_{j=1}^{\infty} \tilde{r}_j \int_\Omega \tilde{r}_j X \,\mathrm{d}\mu$$

converges almost everywhere and for all $e' \in E'$ we have $e'(\tilde{\Pi}_{\mathcal{G}} X) = \Pi_{\mathcal{G}} e'(X) = e'(\Pi_{\mathcal{G}} X)$ as $L^2(\mathcal{A})$ μ-equivalences classes. Now Lemma 3.3.1 leads to the desired result. □

Theorem 3.3.2. *Under Assumption A and given an ONB $(r_j)_{j\in J}$ of G_Y the equality*

$$\mathbb{E}(X|Y) = \sum_{j\in J} r_j \int_\Omega r_j X \,\mathrm{d}\mu$$

holds true, where the convergence of the series is almost everywhere and in $L^p(\mathcal{A}, E)$ for all $p \in [1, \infty)$. Additionally, the following statements hold true:

i) The orthogonal projection $\Pi_G : G_{(X,Y)} \to G_{(X,Y)}$ onto $G_{(Z,Y)} = G_Y$ is given by

$$\Pi_G g = \mathbb{E}(g|Y).$$

ii) The orthogonal projection $\Pi_W : W_{(X,Y)} \to W_{(X,Y)}$ onto $W_{(Z,Y)}$ is given by

$$\Pi_W(w_x, w_y) = (L_w^* w_y, w_y).$$

iii) The orthogonal projection $\Pi_H : H_{(X,Y)} \to H_{(X,Y)}$ onto $H_{(Z,Y)}$ is given by

$$\Pi_H(h_x, h_y) = (\iota_X L_w^* \iota_Y^{-1} h_y, h_y).$$

Finally, the orthogonal projections are related via the following relation

$$\Pi_G = V_{(X,Y)}^* \Pi_W V_{(X,Y)} = U_{(X,Y)} \Pi_H U_{(X,Y)}^*.$$

Proof. Applying Lemma 3.2.2, Lemma 3.2.5, Identity (2.3), and Lemma 3.3.1 leads to

$$e'\left(\mathbb{E}(X|Y)\right) = \mathbb{E}(e'(X)|Y) = \mathbb{E}(e'(X)|G_Y) = \Pi_{G_Y} e'(X) = e'\left(\Pi_{G_Y} X\right),$$

where Π_{G_Y} denotes the orthogonal projection $\Pi_{G_Y} : G_{(X,Y)} \to G_{(X,Y)}$ onto G_Y. Using Lemma 3.2.6 and Lemma 3.3.1, we obtain

$$\mathbb{E}(X|Y) = \Pi_{G_Y} X = \sum_{j\in J} r_j \int_\Omega r_j X \,\mathrm{d}\mu, \tag{3.13}$$

where the series converges almost everywhere and in $L^p(\mathcal{A}, E)$.

i). Since $\mathbb{E}(\,\cdot\,|Y) : G_{(X,Y)} \to G_{(X,Y)}$ is the orthogonal projection onto G_Y, see [26, Theorem 3.3 vii)], we conclude $\Pi_{G_Y} = \mathbb{E}(\,\cdot\,|Y)$. We have $G_Y = G_{(Z,Y)}$ with Lemma 3.2.14. We obtain $\Pi_G = \Pi_{G_Y} = \mathbb{E}(\,\cdot\,|Y)$.

ii). We consider the map $\Pi_W : W_{(X,Y)} \to W_{(X,Y)}$ that is defined by

$$\Pi_W(w_x, w_y) = (L_W^* w_y, w_y),$$

and show that it is the orthogonal projection onto $W_{(Z,Y)}$. To this end, let $g \in G_{(X,Y)}$ and $g_Y := \Pi_G g$. Then we find

$$\begin{aligned} V_{(X,Y)}^* \Pi_W V_{(X,Y)} g &= V_{(X,Y)}^* \Pi_W \int_\Omega g\cdot (X,Y)\,\mathrm{d}\mu \\ &= V_{(X,Y)}^* \left(L_W^* \int_\Omega gY \,\mathrm{d}\mu, \int_\Omega gY\,\mathrm{d}\mu \right) \\ &= V_{(X,Y)}^* \left(L_W^* \int_\Omega g_Y Y \,\mathrm{d}\mu, \int_\Omega g_Y Y\,\mathrm{d}\mu \right) \\ &= V_{(X,Y)}^* \left(\int_\Omega g_Y X \,\mathrm{d}\mu, \int_\Omega g_Y Y\,\mathrm{d}\mu \right) \\ &= g_Y\,, \end{aligned}$$

where the third identity is a consequence of Lemma 3.2.13 and the fourth identity follows by the definition of L_W, see (2.6). We use Lemma B.1.3 to conclude that Π_W is an orthogonal projection onto the space

$$\begin{aligned} V_{(X,Y)}\mathrm{ran}(\Pi_G) = V_{(X,Y)} G_Y &= \left\{ \int_\Omega g\cdot (X,Y)\,\mathrm{d}\mu \,\middle|\, g\in G_Y \right\} \\ &= \left\{ \int_\Omega g\cdot (Z,Y)\,\mathrm{d}\mu \,\middle|\, g\in G_Y \right\} \\ &= \left\{ \int_\Omega g\cdot (Z,Y)\,\mathrm{d}\mu \,\middle|\, g\in G_{(Z,Y)} \right\} \\ &= W_{(Z,Y)}\,. \end{aligned}$$

where in the third identity we used Lemma 3.2.1, with $\mathcal{C} = \sigma(Y)$ and in the fourth identity we used Lemma 3.2.14.

iii). We define the mapping $\Pi_H : H_{(X,Y)} \to H_{(X,Y)}$ by

$$\Pi_H(h_x, h_y) := (\iota_X L_W^* \iota_Y^{-1} h_y, h_y)$$

and show that this is the orthogonal projection onto $H_{(Z,Y)}$. To this end we first note that for $(h_X, h_Y) \in H_{(X,Y)}$ and $(w_X, w_Y) := (\iota_X^{-1} h_X, \iota_Y^{-1} h_Y) \in W_{(X,Y)}$, we have

$$\iota_{(X,Y)} \Pi_W \iota_{(X,Y)}^{-1}(h_X, h_Y) = \iota_{(X,Y)}\Pi_W(w_X, w_Y) = \iota_{(X,Y)}(L_W^* w_Y, w_Y) = (\iota_X L_W^* \iota_Y^{-1} h_Y, h_Y)\,.$$

Moreover recall from (2.8) that we have

$$U_{(X,Y)} = V_{(X,Y)}^{-1}\iota_{(X,Y)}^{*}.$$

Combining both identities leads to

$$\begin{aligned} U_{(X,Y)}\Pi_H U_{(X,Y)}^{*} &= V_{(X,Y)}^{*}\iota_{(X,Y)}^{-1} \circ \Pi_H \circ \iota_{(X,Y)}V_{(X,Y)} \\ &= V_{(X,Y)}^{*}\iota_{(X,Y)}^{-1} \circ \iota_{(X,Y)} \circ \Pi_W \circ \iota_{(X,Y)}^{-1} \circ \iota_{(X,Y)}V_{(X,Y)} \\ &= V_{(X,Y)}^{*}\Pi_W V_{(X,Y)}\,. \end{aligned}$$

By Lemma B.1.3 we conclude that Π_H is an orthogonal projection with

$$\operatorname{ran}(\Pi_H) = \iota_{(X,Y)}\operatorname{ran}(\Pi_W) = \left\{\iota_{(X,Y)}\int_\Omega g\cdot(Z,Y)\,\mathrm{d}\mu \,\middle|\, g\in G_{(Z,Y)}\right\} = H_{(Z,Y)}\,,$$

where in the last equality we used

$$\iota_X\left(\int_\Omega gZ\,\mathrm{d}\mu\right) = \left(e'\mapsto \int_\Omega ge'(Z)\,\mathrm{d}\mu\right)\Big|_{B_{E'}} = \iota_Z\left(\int_\Omega gZ\,\mathrm{d}\mu\right)\,.$$

□

Theorem 3.3.2 establishes that conditioning corresponds to orthogonal projections in the spaces $G_{(X,Y)}$, $W_{(X,Y)}$, and $H_{(X,Y)}$. Moreover, it demonstrates that the operator L_W^*, and consequently M_W, are intrinsically linked to the conditioning process.

The next theorem uses the orthogonal projection statement of Theorem 3.3.2 by showing that the kernel of the conditioned random variable is given by a projection of the initial kernel.

Lemma 3.3.3. *Let $X:\Omega\to E$ be a Gaussian random variable and $\mathcal{E}\subset E'$. Then there exists a Gaussian random variable $Y:\Omega\to\ell^2(\mathbb{N})$ such that for all*

$$Z_{\mathcal{E}}\in\mathbb{E}\left(X|\sigma(\{e'(X)\,:\,e'\in\mathcal{E}\})\right) \text{ and all } Z\in\mathbb{E}(X|Y),$$

we have $\mu(Z_{\mathcal{E}}\neq Z)=0$. Moreover we have

$$G_Y = G_{\mathcal{E}} := \overline{\operatorname{span}\{e'(X)\,|\,e'\in\mathcal{E}\}}^{\|\cdot\|_{L^2(\mathcal{A})}}.$$

Proof. For simplicity we prove the statement only for $\dim G_{\mathcal{E}}=\infty$. The finite dimensional cases are analogous yet simpler.

Let us fix an ONB $([r_j])$ of $G_{\mathcal{E}}$ such that $r_j\in\operatorname{span}(\{e'(X)\,|\,e'\in\mathcal{E}\})$ for all $j\geq 1$. Here we note that this is possible by the Gram-Schmidt construction if in each step we pick a vector in $\operatorname{span}(\{e'(X)\,|\,e'\in\mathcal{E}\})$ that is linearly independent of the previously constructed members of the ONB.

We further note that $r_j\sim\mathcal{N}(0,1)$. Moreover, for each $(r_{j_1},\dots,r_{j_m})$ we know that for linear combinations we have $\sum_{i=1}^m a_i r_{j_i}\in\operatorname{span}(\mathcal{E}\circ X)$ and hence $(r_{j_1},\dots,r_{j_m})$ is a Gaussian vector. Since it components are pairwise uncorrelated we conclude that $(r_{j_1},\dots,r_{j_m})$ are independent. In other words, (r_j) is an i.i.d. sequence.

Let (e_j) be the standard ONB of $\ell^2(\mathbb{N})$ and

$$\alpha_j := \left(\sqrt{\ln(j\ln^2(j+1))}\right)^{-1}$$
$$Y_n := \sum_{j=1}^{n} \alpha_j r_j \frac{e_j}{j}.$$

We note that by Lemma A.1.1 we μ-almost surely have

$$\left\| \left(\alpha_j r_j \frac{1}{j}\right) \right\|_{\ell^2(\mathbb{N})}^2 \leq \|(\alpha_j r_j)\|_{\ell^\infty(\mathbb{N})}^2 \cdot \left\| \left(\frac{1}{j}\right) \right\|_{\ell^2(\mathbb{N})}^2 < \infty,$$

and thus

$$\sum_{j=1}^{\infty} \alpha_j r_j \frac{e_j}{j}$$

converges almost everywhere in $\ell^2(\mathbb{N})$. We denote the set where the series does converge by T and we set

$$Y(\omega) := \begin{cases} \sum_{j=1}^{\infty} \alpha_j r_j(\omega) \frac{e_j}{j}, & \omega \in T, \\ 0, & \text{else.} \end{cases}$$

Next we prove that $Y : \Omega \to \ell^2(\mathbb{N})$ is a Gaussian random variable. To this end let us fix an $a' \in \ell^2(\mathbb{N})$. Then, for $\omega \in T$, we have

$$a'(Y - Y_n) = a'\left(\sum_{j=n+1}^{\infty} \alpha_j r_j(\omega) \frac{e_j}{j}\right) = \sum_{j=n+1}^{\infty} \alpha_j \frac{a'(e_j)}{j} r_j(\omega).$$

By Parseval's identity and the fact that $([r_j])$ is an orthonormal system in $L^2(\mathcal{A})$ we conclude

$$\int_\Omega a'(Y - Y_n)^2 \, \mathrm{d}\mu = \sum_{j=n+1}^{\infty} \left(\alpha_j \frac{a'(e_j)}{j}\right)^2 \leq \|a'\|_{\ell^2(\mathbb{N})}^2 \sum_{j=n+1}^{\infty} \frac{\alpha_j^2}{j^2} \to 0. \tag{3.14}$$

Due to the fact that each $a'(Y_n)$ is a Gaussian random variable and $a'(Y_n) \to a'(Y)$ in $L^2(\mathcal{A})$ it follows that Y is indeed a Gaussian random variable.

Our next goal is to establish the inclusion

$$\sigma(\mathcal{E} \circ X) \subseteq \widehat{\sigma(Y)} \subseteq \sigma(\widehat{\mathcal{E} \circ X}). \tag{3.15}$$

To this end we fix a $g \in \mathcal{E} \circ X$. Then there exists a sequence $(b_j) \in \ell^2(\mathbb{N})$ such that $g = \sum_{j=1}^{\infty} b_j r_j$ in $\mathcal{L}^2(\mathcal{A})$. By

$$e_j(Y)(\omega) = \begin{cases} \alpha_j r_j(\omega) \frac{1}{j}, & \text{if } \omega \in T \\ 0, & \text{else} \end{cases} \tag{3.16}$$

we conclude that

$$\chi_T r_j = \frac{j}{\alpha_j} e_j(Y).$$

Therefore, $\chi_T r_j$ is $\sigma(Y)$-measurable and thus also $\widehat{\sigma(Y)}$-measurable. Using $r_j = \chi_T r_j + \chi_{\Omega \setminus T} r_j$ we conclude that r_j is $\widehat{\sigma(Y)}$-measurable. Consequently, each finite sum $\sum_{j=1}^{m} b_j r_j$ is $\widehat{\sigma(Y)}$-measurable. Since a subsequence of these finite sums converge almost surely to g we conclude that g itself is $\widehat{\sigma(Y)}$-measurable. This shows the first inclusion $\sigma(\mathcal{E} \circ X) \subseteq \widehat{\sigma(Y)}$.

For the second inclusion $\widehat{\sigma(Y)} \subseteq \sigma\widehat{(\mathcal{E} \circ X)}$ we first prove $\sigma(Y) \subseteq \sigma\widehat{(\mathcal{E} \circ X)}$. To this end, we note that for $a' \in \ell^2(\mathbb{N})$ we have

$$a'(Y) = \sum_{j=1}^{\infty} \chi_T \alpha_j r_j \frac{a'(e_j)}{j} \tag{3.17}$$

with pointwise convergence. We further note that χ_T is $\mathcal{T}$-measurable and that all r_j are $\sigma(\mathcal{E} \circ X)$-measurable since by our construction they are finite linear combinations of elements in $\mathcal{E} \circ X$. Therefore, $a'(Y)$ is $\sigma\widehat{(\mathcal{E} \circ X)}$-measurable and by Lemma 3.2.5 we find $\sigma(Y) \subseteq \sigma\widehat{(\mathcal{E} \circ X)}$. By Lemma 3.2.8 we obtain $\widehat{\sigma(Y)} \subseteq \sigma\widehat{(\mathcal{E} \circ X)}$.

Combining (3.15) with Theorem 3.2.12 we obtain the first assertion.

For the proof of $G_{\mathcal{E}} \subseteq G_Y$ we first note that $[r_j] \in G_Y$ by (3.16). Since $([r_j])$ is an ONB of $G_{\mathcal{E}}$ and G_Y is a closed subspace of $L^2(\mathcal{A})$ we then find the desired inclusion.

To prove the other inclusion $G_Y \subseteq G_{\mathcal{E}}$, we fix an $a' \in \ell^2(\mathbb{N})$. Analogously to (3.14) we have

$$\left(\alpha_j \frac{a'(e_j)}{j} \right) \in \ell^2(\mathbb{N}).$$

By (3.17) we conclude that $a'(Y) \in G_{\mathcal{E}}$, and since $G_{\mathcal{E}}$ is a closed space we then find $G_Y \subseteq G_{\mathcal{E}}$. □

Theorem 3.3.4. *Under Assumption A and given $\mathcal{E} \subset B_{E'}$, we set*

$$H(\mathcal{E}) := \overline{\operatorname{span}\{k_X(\,\cdot\,, e') \,|\, e' \in \mathcal{E}\}}^{\|\cdot\|_{H_X}}. \tag{3.18}$$

Then $X_{\mathcal{E}} := \mathbb{E}(X|\,\sigma\{e'(X) \,:\, e' \in \mathcal{E}\})$ is a Gaussian random variable, whose kernel is given by

$$k_{\mathcal{E}}(\,\cdot\,, e') = \Pi_{H(\mathcal{E})} k_X(\,\cdot\,, e'), \quad \text{for all } e' \in B_{E'}.$$

Also, the kernel $k_{X-X_{\mathcal{E}}} : B_{E'} \times B_{E'} \to \mathbb{R}$ of the Gaussian random variable $X - X_{\mathcal{E}}$ is given by

$$k_{X-X_{\mathcal{E}}}(e_1', e_2') = k_X(e_1', e_2') - k_{\mathcal{E}}(e_1', e_2'), \quad \text{for all } e_1', e_2' \in B_{E'}.$$

Additionally, the following equality holds true

$$\sup_{e' \in B_{E'}} \|k_X(\,\cdot\,, e') - k_{\mathcal{E}}(\,\cdot\,, e')\|_{H_X}^2 = \|\operatorname{cov}(X - X_{\mathcal{E}})\|_{E' \to E} = \|\operatorname{cov}(X) - \operatorname{cov}(X_{\mathcal{E}})\|_{E' \to E}.$$

Proof. We set $G_{\mathcal{E}} := \overline{\operatorname{span}\{e'(X) \,|\, e' \in \mathcal{E}\}}^{\|\cdot\|_{L^2(\mathcal{A})}}$ and use Lemma 3.2.2 to obtain

$$e'(X_{\mathcal{E}}) = e'\left(\mathbb{E}(X|\sigma(\{e'(X) \,|\, e' \in \mathcal{E}\})\right) = \mathbb{E}(e'(X)|\sigma(\{e'(X) \,|\, e' \in \mathcal{E}\}))$$

for all $e' \in E'$. We now fix a Gaussian random variable Y according to Lemma 3.3.3. This leads to

$$\mathbb{E}(e'(X)|\sigma(\{e'(X) \,|\, e' \in \mathcal{E}\})) = \mathbb{E}(e'(X)|Y).$$

By [26, Theorem 3.3 vii)] and Lemma 3.3.3 we also have

$$\mathbb{E}(e'(X)|Y) = \Pi_{G_Y} e'(X) = \Pi_{G_{\mathcal{E}}} e'(X) = e'\left(\Pi_{G_{\mathcal{E}}} X\right)$$

where in the last step we used Lemma 3.3.1. In summary we have

$$e'(X_{\mathcal{E}}) = e'(\Pi_{G_{\mathcal{E}}} X) \tag{3.19}$$

almost surely. We conclude by Lemma 3.2.6, that $\Pi_{G_{\mathcal{E}}} X = X_{\mathcal{E}}$ and by Lemma 3.3.1, that $X_{\mathcal{E}}$ is a Gaussian random variable.

Next we prove that the kernel $k_{\mathcal{E}}$ is given by

$$k_{\mathcal{E}}(\,\cdot\,, e') = \Pi_{H(\mathcal{E})} k_X(\,\cdot\,, e') \quad \text{ with } e' \in B_{E'}\,.$$

Recall that the operator U_X from (2.8) is an isometry and for $e' \in \mathcal{E}$ we have

$$U_X k_X(\,\cdot\,, e') = V_X^{-1}\left(\int_\Omega e'(X) X \,\mathrm{d}\mu\right) = e'(X). \tag{3.20}$$

By taking the span and the closure we conclude $U_X H(\mathcal{E}) = G_{\mathcal{E}}$. Using the reproducing property in H_X, Equation (3.20), Lemma B.1.3, and Lemma 3.3.1 in combination with Equation (3.19) we find

$$\begin{aligned}
\left(\Pi_{H(\mathcal{E})} k_X(\,\cdot\,, e'_1)\right)(e'_2) &= \langle \Pi_{H(\mathcal{E})} k_X(\,\cdot\,, e'_1), k_X(\,\cdot\,, e'_2)\rangle_{H_X} \\
&= \langle U_X \Pi_{H(\mathcal{E})} U_X^* e'_1(X), U_X \Pi_{H(\mathcal{E})} U_X^* e'_2(X)\rangle_{G_X} \\
&= \langle \Pi_{G_{\mathcal{E}}} e'_1(X), \Pi_{G_{\mathcal{E}}} e'_2(X)\rangle_{G_X} \\
&= \int_\Omega e'_1(X_{\mathcal{E}}) e'_2(X_{\mathcal{E}}) \,\mathrm{d}\mu \\
&= \langle \operatorname{cov}(X_{\mathcal{E}}) e'_1, e'_2\rangle_{E,E'} = k_{\mathcal{E}}(e'_1, e'_2)
\end{aligned}$$

for all $e'_1, e'_2 \in B_{E'}$.

Next we note that $X - X_{\mathcal{E}}$ is Gaussian since $X, X_{\mathcal{E}}$ are jointly Gaussian by Lemma 3.3.1. Moreover, for $e'_1, e'_2 \in B_{E'}$, we obtain by Lemma 3.2.16

$$\begin{aligned}
k_{X - X_{\mathcal{E}}}(e'_1, e'_2) &= \langle \operatorname{cov}(X - X_{\mathcal{E}}) e'_1, e'_2\rangle_{E,E'} \\
&= \langle \operatorname{cov}(X) e'_1, e'_2\rangle_{E,E'} - \langle \operatorname{cov}(X_{\mathcal{E}}) e'_1, e'_2\rangle_{E,E'}\,.
\end{aligned}$$

Finally we establish the last assertion. With what we have already proven we conclude

$$k_{X - X_{\mathcal{E}}}(e', e') = \|k_X(\cdot, e') - \Pi_{H(\mathcal{E})} k_X(\cdot, e')\|^2_{H_X} \qquad \text{for } e' \in B_{E'}\,.$$

With the operator norm definition this leads to

$$\|\operatorname{cov}(X - X_{\mathcal{E}})\|_{E' \to E} = \sup_{e' \in B_{E'}} \|k_X(\cdot, e') - k_{\mathcal{E}}(\cdot, e')\|^2_{H_X}.$$

The second identity follows by Lemma 3.2.16. □

For a short example of Theorem 3.3.4, let $S \subseteq T \subseteq \mathbb{R}^d$ be compact, $X : \Omega \to C(T)$ be a Gaussian random variable, and $\mathcal{E} := \{\delta_s \mid s \in S\}$ with δ_s being the point evaluation in the point s. Then to obtain $X_{\mathcal{E}}$ we only have to calculate $\Pi_{H(\mathcal{E})} k_X(\,\cdot\,, \delta_t)$ for $t \in T$. Here, we note that $k_X(\delta_s, \delta_t) = k(s,t)$ holds true, where k is the covariance function. We write H for its RKHS. This gives $H = W_X$ and thus $\iota_X H = H_X$. Lastly we have

$$H(\mathcal{E}) = \iota H(S) := \iota_X \overline{\operatorname{span}\{k(\,\cdot\,, s) \mid s \in S\}}^H.$$

Corollary 3.3.5. *Under Assumption A and given a bounded operator $L : E \to F$ such that*

$$Y = LX,$$

one obtains $G_Z = G_Y \subseteq G_X$ and the kernel to the space H_Z is given by

$$k_Z(\,\cdot\,, e') = \Pi_{H(\mathcal{E})} k_X(\,\cdot\,, e')$$

for all $e' \in B_{E'}$ and $H(\mathcal{E}) := \overline{\operatorname{span}\{k_X(\,\cdot\,, L'f') \mid f' \in F' \wedge L'f' \in B_{E'}\}}^{\|\cdot\|_{H_X}}$.

Proof. We apply Theorem 3.3.4 with the set

$$\mathcal{E} := \{0\} \cup \left\{ \frac{L'f'}{\|L'f'\|_{E'}} \,\middle|\, f' \in F' \text{ with } L'f' \neq 0 \right\},$$

where we note that $Z = X_{\mathcal{E}}$ since $(L'f')(X) = f'(LX) = f'(Y)$ for all $f' \in F'$. Therefore it remains to prove that $G_Z = G_Y$ holds true. To this end, we note that

$$G_Y = \overline{\operatorname{span}\{f'(LX) \mid f' \in F'\}}^{\|\cdot\|} = \overline{\operatorname{span}\{L'f' \circ X \mid f' \in F'\}}^{\|\cdot\|} \subseteq G_X$$

holds true. Now Lemma 3.2.14 implies $G_Y = G_Z$. □

Our next goal is to calculate the conditional expectation $Z = \mathbb{E}(X|Y)$. To this end we recall the mapping $M_W : W_Y \to W_Z$

$$M_W w_y = \hat{V}_Z \hat{V}_Y^* w_y$$

from (2.7). Moreover, recall from Theorem 3.3.2 that given an ONB $(r_j)_{j\in J} \subseteq G_Y$ we find

$$Y = \sum_{j\in J} r_j \int_\Omega r_j Y \,\mathrm{d}\mu.$$

If we could apply M_W to Y, we would obtain

$$M_W Y = \sum_{j\in J} r_j M_W \left(\int_\Omega r_j Y \,\mathrm{d}\mu \right) = \sum_{j\in J} r_j \int_\Omega r_j Z \,\mathrm{d}\mu = Z, \tag{3.21}$$

thus solving the conditioning problem. Unfortunately, however M_W is only defined on W_Y, therefore Equation (3.21) is not valid and should only be seen as a heuristic.

In the following we investigate whether we can turn this heuristic into a rigorous argument. To this end we explore under which conditions we can continuously extend the operator M_W. We begin with the following lemma.

Lemma 3.3.6. *Let Assumption A be satisfied and W_Y be dense in F. If there exists a $C > 0$ with*

$$\|M_W w\|_E \leq C\|w\|_F \tag{3.22}$$

for all $w \in W_Y$, then there exists a bounded linear operator $M : F \to E$ such that $MY = Z$ almost everywhere and $M|_{W_Y} = M_W$.

Proof. By the B.L.T. Theorem, see [47, Theorem 1.7] there exists a bounded linear extension $M : F \to E$ of $M_W : W_Y \to W_Z$. Repeating (3.21) we then obtain

$$MY = M \sum_{j\in J} r_j \int_\Omega r_j Y \,\mathrm{d}\mu = \sum_{j\in J} r_j M_W \left(\int_\Omega r_j Y \,\mathrm{d}\mu \right) = \sum_{j\in J} r_j \int_\Omega r_j Z \,\mathrm{d}\mu = Z.$$

□

An example where the inequality in (3.22) occurs, is given by partial differential equations with well-posedness inequalities, e.g. [48, Chapter 2.5]

If M_W does not satisfy (3.22), we can change one of the spaces E or F or both to obtain that M_W can be extended, as we exemplify by the following theorem.

Theorem 3.3.7. *Under Assumption A, assuming $G_Y = G_Z$, and W_Y is dense in F, there always exists a Banach space $\tilde{E}$ such that $W_Z \subseteq \tilde{E}$ and there exists an extension $\hat{M} : F \to \tilde{E}$ of M_W which is an isometric isomorphism and satisfies $\hat{M}Y = Z$.*

Proof. We first note that by Lemma 3.2.17, we have that M_W is an isometry. We now define a norm on W_Z by

$$\|w\|_{\tilde{E}_0} := \|M_W^{-1} w\|_F$$

for all $w \in W_Z$. Since M_W is invertible as an isometric isomorphism, this norm is well-defined. We define $\tilde{E}$ to be the completion of the space $(W_Z, \|\cdot\|_{\tilde{E}_0})$.

Our first goal is to show that $\tilde{E}$ is separable. To this end we note that $W_Y \subseteq F$ is separable with respect to the $\|\cdot\|_F$-norm since F is separable. Consequently there exists a countable and $\|\cdot\|_F$-dense $D \subseteq W_Y$. Clearly it suffices to show that $\hat{D} := M_W D \subseteq W_Z$ is $\|\cdot\|_{\tilde{E}_0}$-dense in W_Z. To this end we fix a $w \in W_Z$. Then there exists a $v \in W_Y$ such that $M_W v = w$, and additionally a sequence $(v_n) \subset D$ such that $v_n \to v$. For $w_n := M_W v_n \in \hat{D}$, we then find

$$\|w_n - w\|_{\tilde{E}_0} = \|M_W^{-1}(w_n - w)\|_F = \|v_n - v\|_F \to 0.$$

Let us now consider the operator $\tilde{M}_W : W_Y \to \tilde{E}$ that is given by

$$\tilde{M}_W w = M_W w.$$

Our next goal is to show that this operator is continuous with respect to the norms $\|\cdot\|_F$ and $\|\cdot\|_{\tilde{E}_0}$. To this end we fix $w \in W_Y$, then we have $\tilde{M}_W w \in W_Z$ and thus we find

$$\|\tilde{M}_W w\|_{\tilde{E}_0} = \|M_W^{-1}\tilde{M}_W w\|_F = \|w\|_F.$$

By [49, Theorem 1.9.1] there exists a unique bounded and linear extension $\hat{M} : F \to \tilde{E}$ of $\tilde{M}_W$. Clearly it is even an isometry.

Lastly, we show that $\hat{M}Y = Z$ holds true. To this end, let $(r_j)_{j\in J} \subset G_Y$ be an ONB. By Theorem 3.3.2 we then find

$$Y = \mathbb{E}(Y|Y) = \sum_{j\in J} r_j \int_\Omega r_j Y \,\mathrm{d}\mu,$$

where the convergence is pointwise almost sure in F. Applying $\hat{M}$ leads to

$$\begin{aligned}\hat{M}Y = \hat{M}\left(\sum_{j\in J} r_j \int_\Omega r_j Y \,\mathrm{d}\mu\right) = \sum_{j\in J} r_j \hat{M}\left(\int_\Omega r_j Y \,\mathrm{d}\mu\right) &= \sum_{j\in J} r_j M_W \left(\int_\Omega r_j Y \,\mathrm{d}\mu\right)\\ &= \sum_{j\in J} r_j \int_\Omega r_j Z \,\mathrm{d}\mu\\ &= \mathbb{E}(Z|Y)\\ &= Z,\end{aligned}$$

where in the fourth step we used the definition of M_W. □

We note that one could instead of changing the norm on E change the norm on F to obtain a continuous extension of M_W into the space E.

Having shown criteria for extending the operator M_W to an operator $M : F \to E$, we make the following assumption:

Assumption M. *Under Assumption A, we assume that there exists a bounded linear operator $M : F \to E$ such that $MY = Z$ almost everywhere holds true.*

The operator M allows us to obtain an easy formula for the conditional variance.

Remark 3.3.8. *Assume now that W_X and W_Y are RKHSs on some domains T_X and T_Y, respectively. For simplicity, we write k_X and k_Y for their corresponding kernels. Then, for all $t_X \in T_X$ and $t_Y \in T_Y$, we have*

$$\begin{aligned}(L_W k_X(\,\cdot\,, t_X))\,(t_Y) &= \langle L_W k_X(\,\cdot\,, t_X), , k_Y(\,\cdot\,, t_Y)\rangle_{W_Y},\\ \langle k_X(\,\cdot\,, t_X), L_W^* k_Y(\,\cdot\,, t_Y)\rangle_{W_X} &= (L_W^* k_Y(\,\cdot\,, t_Y))\,(t_X).\end{aligned}$$

Since $\mathrm{Id}_{W_Z\to W_X} M_W = L_W^*$, it follows that

$$(M_W k_Y(\,\cdot\,, t_Y))\,(t_X) = (L_W k_X(\,\cdot\,, t_X))\,(t_Y),$$

so M_W can be explicitly evaluated on kernel translates.

This is particularly relevant when conditioning on finitely many points $(t_{Y,j}) \in (T_Y)^n$, where the conditional expectation is given by

$$s_n = \sum_{j=1}^n a_j k_Y(\,\cdot\,, t_{Y,j}),$$

for some coefficients $(a_j) \in \mathbb{R}^n$*. We can then compute*

$$\begin{aligned}(M_W s_n)(t_X) &= \sum_{j=1}^{n} a_j \left(M_W k_Y(\,\cdot\,, t_{Y,j})\right)(t_X) \\ &= \sum_{j=1}^{n} a_j \left(L_W k_X(\,\cdot\,, t_X)\right)(t_{Y,j}),\end{aligned}$$

thus expressing $M_W s_n$ *in terms of* L_W *and the kernel* k_X*.*

The next theorem covers a simple formula, which allows us to easily work with the conditional variance of X given Y.

Theorem 3.3.9. *Under Assumption A the conditional variance of* X *is given via*

$$\operatorname{cov}(X|Y) = \operatorname{cov}(X) - \operatorname{cov}(Z)$$

with $Z := \mathbb{E}(X|Y)$*. Additionally, under Assumption M the conditional variance is also given via*

$$\operatorname{cov}(X|Y) = \operatorname{cov}(X) - M\operatorname{cov}(Y)M'.$$

Proof. The first assertion can be found for example in [26, Theorem 3.3 vi).], but here we give an independent proof. To this end we define $X_u := X - Z$, this gives

$$Z = \mathbb{E}(X|Y) = \mathbb{E}(X_u + Z|Y) = \mathbb{E}(X_u|Y) + \mathbb{E}(Z|Y).$$

Using the definition of the conditional expectation we conclude $\mathbb{E}(Z|Y) = Z$, which implies $\mathbb{E}(X_u|Y) = 0$. We conclude that

$$\langle g_u, g_y\rangle_{L^2(\mathcal{A})} = 0 \tag{3.23}$$

for all $g_u \in G_{X_u}$ and $g_y \in G_Y$. Additionally, since $\mathbb{E}(X) = \mathbb{E}(Z) = 0$ we have

$$\mathbb{E}(X_u) = 0.$$

Utilizing Lemma 3.2.16 leads to

$$\begin{aligned}\langle \operatorname{cov}(X|Y)e', e'\rangle_{E,E'} = \langle \operatorname{cov}(X_u + Z|Y)e', e'\rangle_{E,E'} &= \mathbb{E}(\left(e'(X_u + Z - \mathbb{E}(X_u + Z|Y))\right)^2 |Y) \\ &= \mathbb{E}(\left(e'(X_u + Z - Z)\right)^2 |Y) \\ &= \mathbb{E}(\left(e'(X_u)\right)^2 |Y) \\ &= \langle \operatorname{cov}(X_u)e', e'\rangle_{E,E'} \\ &= \langle \operatorname{cov}(X - Z)e', e'\rangle_{E,E'} \\ &= \langle \operatorname{cov}(X)e', e'\rangle_{E,E'} - \langle \operatorname{cov}(Z)e', e'\rangle_{E,E'}\,,\end{aligned}$$

where in the fifth step we used that $e'(X_u)$ is independent of Y, since $\mathbb{E}(e'(X_u)|Y) = 0$. Where we used the well known fact that it suffices to calculate the diagonal of the covariance operator, see for example [26, Lemma B.2.].

For the second assertion we use $MY = Z$ and some well known formulas for cross covariance operators, see e.g. [26, Equation (B.3)]:

$$\operatorname{cov}(X|Y) = \operatorname{cov}(X) - \operatorname{cov}(Z) = \operatorname{cov}(X) - M\operatorname{cov}(Y)M'\,.$$

□

We use Theorem 3.3.9 to derive convergence rates for the conditional variance leading to the following theorem.

Theorem 3.3.10. *Under Assumption A, Assumption M, and given a subset $F'_n \subset F'$. We define*

$$Y_n := \mathbb{E}(Y|\sigma\{f'(Y) \,:\, f' \in F'_n\})$$

We obtain

$$\|\mathrm{cov}(X|Y) - \mathrm{cov}(X|Y_n)\|_{E'\to E} \leq \|M\|^2_{F\to E}\|\mathrm{cov}(Y) - \mathrm{cov}(Y_n)\|_{F'\to F}.$$

Proof. We set

$$\begin{aligned} Z_n &:= \mathbb{E}(X|\sigma\{f'(Y) \,:\, f' \in F'_n\}) \\ Z &:= \mathbb{E}(X|Y). \end{aligned}$$

By applying Theorem 3.3.9 and Lemma 3.2.16 we obtain

$$\begin{aligned} \mathrm{cov}(X|Y_n) - \mathrm{cov}(X|Y) = \mathrm{cov}(X) - \mathrm{cov}(Z_n) - \mathrm{cov}(X) - \mathrm{cov}(Z) &= \mathrm{cov}(Z) - \mathrm{cov}(Z_n) \\ &= \mathrm{cov}(Z - Z_n). \end{aligned}$$

Using Assumption M, Theorem 3.3.2 and given an ONB $(r_j)_{j\in J} \subseteq G_{Y_n}$ we obtain

$$MY_n = M\sum_{j\in J} r_j \int_\Omega r_j Y \,\mathrm{d}\mu = \sum_{j\in J} r_j M \int_\Omega r_j Y \,\mathrm{d}\mu = \sum_{j\in J} r_j \int_\Omega r_j Z \,\mathrm{d}\mu = Z_n.$$

Now we prove the inequality via

$$\begin{aligned} \|\mathrm{cov}(X|Y) - \mathrm{cov}(X|Y_n)\|_{E'\to E} &= \|\mathrm{cov}(Z) - \mathrm{cov}(Z_n)\|_{E'\to E} \\ &= \|\mathrm{cov}(MY) - \mathrm{cov}(MY_n)\|_{E'\to E} \\ &\leq \|M\|^2_{F\to E}\|\mathrm{cov}(Y) - \mathrm{cov}(Y_n)\|_{F'\to F}. \end{aligned}$$

□

In summary, Theorem 3.3.10 states that if $M : F \to E$ is bounded then the convergence rates of $\mathrm{cov}(Y) - \mathrm{cov}(Y_n)$ translate into convergence rates of $\mathrm{cov}(X|Y) - \mathrm{cov}(X|Y_n)$. The following corollary demonstrates how this can be done by applying the P-greedy algorithm on Y.

Corollary 3.3.11. *Assume that we are in the setting of Theorem 3.3.10 with Assumption M and that the statement $\|\mathrm{cov}(Y)\|_{F'\to F} \leq 1$ holds true. If we obtain the set F'_n by applying the weak-P-greedy method on k_Y. We obtain the inequality*

$$\|\mathrm{cov}(X|Y) - \mathrm{cov}(X|Y_n)\|_{E'\to E} \leq 2\|M\|^2_{F\to E} \min_{1\leq m<n} \gamma^{-2} d_m^{\frac{2(n-m)}{n}}(\mathcal{F}), \tag{3.24}$$

with $\mathcal{F} = \{k_Y(\,\cdot\,, f') \mid f' \in B_{F'}\}$. *Lastly assuming there exist* $C, \alpha > 0$ *such that for each* $n \in \mathbb{N}$ *there exists a set* $\mathfrak{F}'_n \subset F'$ *with* $|\mathfrak{F}'_n| = n$ *and* $Y_n^\star := \mathbb{E}(Y|\sigma\{f'(Y) \,:\, f' \in \mathfrak{F}'_n\})$

$$\|\mathrm{cov}(Y) - \mathrm{cov}(Y_n^\star)\|_{F'\to F} \le Cn^{-\alpha} \tag{3.25}$$

then the inequality

$$\|\mathrm{cov}(X|Y) - \mathrm{cov}(X|Y_n)\|_{E'\to E} \le \|M\|^2_{F\to E} 2^{5\alpha+1}\gamma^{-2}Cn^{-\alpha}.$$

holds true.

Proof. We apply Theorem 3.3.10, meaning we only have to show that

$$\|\mathrm{cov}(Y) - \mathrm{cov}(Y_n)\|_{F'\to F} \le 2 \min_{1\le m<n} \gamma^{-2} d_m^{\frac{2(n-m)}{n}}(\mathcal{F})$$

holds true. Using Corollary 2.2.5, we obtain

$$\sup_{f\in B_{H_Y}} \|f - \Pi_{F'_n} f\|_{C(B_{F'})} \le \sqrt{2} \min_{1\le m<n} \gamma^{-1} d_m^{\frac{(n-m)}{n}}(\mathcal{F}),$$

with $\Pi_{F'_n}$ being the orthogonal projection in H_Y onto the subspace $\mathrm{span}\{k_Y(\,\cdot\,, f'_n) \mid f'_n \in F'_n\}$. Utilizing Theorem 3.3.4 and [40, Lemma 2.3] we end up with

$$\begin{aligned}
\sup_{f\in B_{H_Y}} \|f - \Pi_{F'_n} f\|^2_{C(B_{F'})} &= \sup_{f\in B_{H_Y}} \sup_{f'\in B_{F'}} |(f - \Pi_{F'_n} f)(f')|^2 \\
&= \sup_{f'\in B_{F'}} \|k_Y(\,\cdot\,, f') - \Pi_{F'_n} k_Y(\,\cdot\,, f')\|^2_{H_Y} \\
&= \sup_{f'\in F'} \|\mathrm{cov}(Y) - \mathrm{cov}(Y_n)\|_{F'\to F}.
\end{aligned}$$

Again using Theorem 3.3.4 and [40, Lemma 2.3] we obtain

$$\begin{aligned}
d_n(\mathcal{F})^2 &\le \sup_{f\in B_{H_Y}} \|f - \Pi_{\mathfrak{F}'_n} f\|^2_{C(B_{F'})} \\
&= \|\mathrm{cov}(Y) - \mathrm{cov}(Y_n^\star)\|_{F'\to F} \le Cn^{-\alpha}.
\end{aligned}$$

Using [39, Corollary 3.3 (ii)], we end up with

$$\|\mathrm{cov}(Y) - \mathrm{cov}(Y_n)\|_{F'\to F} \le 2^{5\alpha+1}\gamma^{-2}Cn^{-\alpha}.$$

□

Corollary 3.3.11 shows that we obtain at least the optimal convergence rate of $\mathrm{cov}(Y)$ for $\mathrm{cov}(X|Y)$ by applying the P-greedy algorithm on Y. A short example when the Inequality (3.25) is satisfied, is given by $T \subseteq \mathbb{R}^d$ compact, $F = C(T)$ and W_Y is equal to a Sobolev space on T with regularity $s > d/2$. Then the sampling inequalities in [50] give (3.25) with $\alpha = 2s/d - 1$.

Remark 3.3.12. *If one replaces the Banach space* E *by a Banach space* $\tilde{E}$*, as mentioned in Theorem 3.3.7 the convergence rate results from Theorem 3.3.10 still hold. However* $\mathrm{cov}(X|Y)$ *is not necessarly a mapping from* $\tilde{E}'$ *to* $\tilde{E}$ *but the difference* $\mathrm{cov}(X|Y) - \mathrm{cov}(X|Y_n)$ *is a mapping from* $\tilde{E}'$ *to* $\tilde{E}$.

3.4 Examples

In this section we give a few examples that show how to calculate M_W and when M_W can be extended and when not. We begin with a positive example in which the spaces W_X and W_Y are well known and understood.

Example 3.4.1. *Let $E := C([0,1])$, X be the Brownian motion, and $L : C([0,1]) \to C([1/2,1])$ be the restriction operator on the intervall $[1/2,1]$, that is*

$$Lf := f|_{[1/2,1]}.$$

Let us further consider $Y := LX$. Then W_X is an RKHS with kernel

$$k_{W_X}(t,s) = \min(s,t)$$

and its scalar product is given by

$$\langle u, v\rangle_{W_X} = \int_0^1 u'(t)v'(t)\,\mathrm{d}t,$$

see e.g. [19, Subsection 8.1.2].

We note that by $Y = LX$ we have that $W_Y = LW_X$ is an RKHS and its kernel is given by

$$k_{W_Y}(s,t) = \min(s,t)$$

for all $s,t \in [a,b]$ with $a := 1/2$ and $b := 1$. By Lemma A.1.2 we thus have

$$\langle u, v\rangle_{W_Y} = 2 \cdot u(1/2)v(1/2) + \int_{1/2}^1 u'(t)v'(t)\,\mathrm{d}t$$

for all for $u, v \in W_Y$. Let us now calculate the adjoint of $L_W : W_X \to W_Y$, which is given by

$$L_W w = w|_{[1/2,1]},$$

see Lemma 3.1.3. To this end, we first note that for $u \in W_X$ and $v \in W_Y$ we have

$$2 \cdot u(1/2)v(1/2) + \int_{1/2}^1 u'(t)v'(t)\,\mathrm{d}t = \langle L_W u, v\rangle_{W_Y} = \langle u, L_W^* v\rangle_{W_X} = \int_0^1 u'(t)(L_W^* v)'(t)\,\mathrm{d}t.$$

In particular, for $u \in W_X$ with $\mathrm{supp}(u) \subset [1/2,1]$ we can conclude that

$$\int_{1/2}^1 u'(t)v'(t)\,\mathrm{d}t = \int_{1/2}^1 u'(t)(L_W^* v)'(t)\,\mathrm{d}t.$$

Recognizing that for all $\tilde{u} \in L^2([1/2,1])$ we find an $u \in W_X$ with $\mathrm{supp}(u) \subset [1/2,1]$ and $u' = \tilde{u}$, we conclude $v'(s) = (L_W^* v)'(s)$ for almost all $s \in [1/2,1]$. Combining the calculations leads to

$$2 \cdot u(1/2)v(1/2) = \int_0^{1/2} u'(t)(L_W^* v)'(t)\,\mathrm{d}t \tag{3.26}$$

for all $u \in W_X$ *and* $v \in W_Y$*. Using the fundamental theorem of calculus and* $u(0) = 0$ *we conclude*

$$u(1/2) = \int_0^{1/2} u'(t)\,\mathrm{d}t.$$

Combining this with (3.26) *leads to*

$$2\int_0^{1/2} u'(t)v(1/2)\,\mathrm{d}t = \int_0^{1/2} u'(t)(L_W^* v)'(t)\,\mathrm{d}t.$$

Again we can conclude that for almost all $s \in [0, 1/2]$ *we have* $(L_W^* v)'(s) = 2 \cdot v(1/2)$*. Putting these results together and taking into consideration that* $L_W^* v$ *has to be continuous with* $L_W^* v(0) = 0$*, we end up with*

$$(L_W^* v)(s) = \begin{cases} 2v(1/2)s, & \text{for } s \leq 1/2 \\ v(s), & \text{for } s > 1/2. \end{cases} \tag{3.27}$$

We note for $v \in W_Y$ *we have* $\|L_w^* v\|_\infty \leq \|v\|_\infty$*, and therefore* L_W^* *can be uniquely extended to a bounded linear operator* $M : C([1/2, 1]) \to C([0, 1])$*. Finally, it is not hard to see that* Mv *can be calculated as in* (3.27)*.*

In the following example, we add some noise to the Gaussian random variable of Example 3.4.1.

Example 3.4.2. *Let* X *and* L *be as in Example 3.4.1 and*

$$Y := LX + N,$$

where N *is a Gaussian random variable independent of* X*, whose kernel is given by*

$$k_{W_N}(s, t) = \sigma^2.$$

This changes the kernel of W_Y *to*

$$k_{W_Y}(s, t) = \sigma^2 + \min(s, t),$$

and by Lemma A.1.2 the scalar product is therefore given by

$$\langle u, v \rangle_{W_Y} = \frac{u(1/2)v(1/2)}{1/2 + \sigma^2} + \int_{1/2}^1 u'(t)v'(t)\,\mathrm{d}t.$$

Repeating the steps of Example 3.4.1 up to (3.26)*, we end up with* $v'(s) = (L_W^* v)'(s)$ *for almost all* $s \in [1/2, 1]$ *as well as*

$$\frac{u(1/2)v(1/2)}{1/2 + \sigma^2} = \int_0^{1/2} u'(t)(L_W^* v)'(t)\,\mathrm{d}t.$$

Again we use the fundamental theorem of calculus to conclude

$$\frac{1}{1/2 + \sigma^2}\int_0^{1/2} u'(t)v(1/2)\,\mathrm{d}t = \int_0^{1/2} u'(t)(L_W^* v)'(t)\,\mathrm{d}t.$$

Thus for almost all $s \in [0, 1/2]$ we have

$$(L_W^* v)'(s) = \frac{v(1/2)}{1/2 + \sigma^2}.$$

Combining these considerations and respecting the continuity of $L_W^* v$ and $(L_W^* v)(0) = 0$ we end up with

$$(L_W^* v)(s) = \begin{cases} \frac{v(1/2)}{1/2+\sigma^2} \cdot s, & \text{for } s \leq 1/2 \\ v(s) - \frac{\sigma^2}{1/2+\sigma^2} v(1/2), & \text{for } s > 1/2. \end{cases}$$

We note that L_W^* can also be extended onto the set of continuous functions as in Example 3.4.1.

The last example provides a situation in which the operator M_W *cannot* be extended.

Example 3.4.3. *Let $E := C^1([0, 1])$ be the space of once continuously differentiable functions and $F := C([0, 1])$ be the space of continuous functions. Moreover, let X be an E-valued Gaussian random variable such that W_X is dense in E and*

$$Y := \mathrm{Id}(X),$$

where $\mathrm{Id} : E \to F$ is the embedding. Note that we have $\mathrm{Id}(W_X) = W_Y$, and therefore we can define $\mathrm{Id}_W : W_X \to W_Y$ as $\mathrm{Id}_W(w) = \mathrm{Id}(w)$ for all $w \in W_X$. For $w \in W_X$ and $g := V_X^* w$ we thus find

$$L_W w = \hat{V}_Y \hat{V}_X^* w = \hat{V}_Y g = \int_\Omega gY \, \mathrm{d}\mu = \int_\Omega g\mathrm{Id}(X) \, \mathrm{d}\mu = \mathrm{Id}_W \left(\int_\Omega gX \, \mathrm{d}\mu \right) = \mathrm{Id}_W(w).$$

Moreover, setting $v := \mathrm{Id}_W(w) \in W_Y$ and $h := V_Y^* v$, we find

$$V_Y h = V_Y V_Y^* \mathrm{Id}_W(w) = \mathrm{Id}_W(w) = \hat{V}_Y g,$$

where in the last step we re-used parts of the previous calculation. Applying V_Y^* on both sides leads to $h = g$. Consequently we obtain

$$\|\mathrm{Id}_W(w)\|_{W_Y} = \|v\|_{W_Y} = \|h\|_{G_Y} = \|g\|_{G_X} = \|w\|_{W_X}.$$

In other words, Id_W is an isometric isomorphism and thus $M_W = L_W^* = \mathrm{Id}_W^* = \mathrm{Id}_W^{-1}$. Now observe that $\mathrm{Id}_W^{-1} : W_Y \to W_X$ is given by $\mathrm{Id}_W^{-1} w = w$, where both side are viewed as functions. Moreover, since W_X is dense in E and E is dense in F we quickly see that W_Y is dense in F. Consequently, the mapping Id_W^{-1} is not continuous with respect to the norms $\|\cdot\|_E$ and $\|\cdot\|_F$ and therefore M_W cannot be extended to a desired M.

In view of Theorem 3.3.7 we note that for $\tilde{E} := F$ we obtain $Mw = w$ as a continuous linear mapping, see Example 3.4.4.

Lastly we give some final example on how to deal with a general operator $L : E \to F$ and some noisy observation such that $Y = LX + N$ with N being a Gaussian random variable independent of X.

Example 3.4.4. *We assume that W_Y is dense in F and that we have a continuous invertible operator $L : E \to F$ such that $Y = LX$. A simple calculation shows*

$$Z = \mathbb{E}(X|Y) = \mathbb{E}(X|LX) = \mathbb{E}(X|X) = X = L^{-1}Y,$$

and this suggests $M = L^{-1}$. Let us now verify this. We first observe that since L is invertible the equality

$$G_X = G_Y$$

holds true. For $g \in G_X$ the norm of $w := V_X g \in W_X$ is given by

$$\|L_W w\|_{W_Y} = \left\| \int_\Omega gY \,\mathrm{d}\mu \right\|_{W_Y} = \|g\|_{G_Y} = \|g\|_{G_X} = \left\| \int_\Omega gX \,\mathrm{d}\mu \right\|_{W_X} = \|w\|_{W_X},$$

where we used the definition $L_W = \hat{V}_Y \hat{V}_X^*$, see (2.6). Consequently, L_W is isometric, and since it is also surjective it is an isometric isomorphism. Here we used that L_W^{-1} exists because L is invertible. By the definition of M_W we conclude that $M_W = L_W^* = L_W^{-1}$. Since we further know that $L|_{W_X} = L_W$, see Lemma 3.1.3, we obtain $M_W = L_W^{-1} = L^{-1}|_{W_Y}$. Consequently L^{-1} is indeed a continous extension of M_W and since W_Y is dense in F it is the only one, in other words we have

$$M = L^{-1}.$$

The next example generalizes the previous Example 3.4.4 to L that are not invertible.

Example 3.4.5. *Let $L : E \to F$ be a bounded operator and $Y = LX$. Then we have $M_W = L_W^\dagger$, where*

$$L_W^\dagger w := \operatorname{argmin} \left\{ \|w_x\|_{W_X} \,|\, L_W w_x = w \right\}, \qquad w \in W_Y \tag{3.28}$$

is the Moore-Penrose inverse of L_W, see [51]. To verify this we first note that $L_W^\dagger$ does exist since $\operatorname{ran}(L_W) = W_Y$ is obviously closed in W_Y. Moreover, we find $G_Y \subseteq G_X$ by $Y = LX$, and obviously G_Y is closed in G_X.

We define the space

$$W_X(Y) := \left\{ \int_\Omega gX \,\mathrm{d}\mu \,\middle|\, g \in G_Y \right\}.$$

Note that the space $W_X(Y)$ is a closed subspace of W_X because $G_Y \subseteq G_X$ is closed and $W_X(Y) = V_X G_Y$. Now let $w \in W_Y$. Then there exists a $g \in G_Y$ such that $V_Y g = w$. We calculate

$$\begin{aligned}\|w\|_{W_Y} = \left\| \int_\Omega gY \,\mathrm{d}\mu \right\|_{W_Y} = \left\| \int_\Omega \Pi_{G_Y} gY \,\mathrm{d}\mu \right\|_{W_Y} &= \|\Pi_{G_Y} g\|_{G_Y} = \|\Pi_{G_Y} g\|_{G_X} \\ &= \left\| \int_\Omega \Pi_{G_Y} gX \,\mathrm{d}\mu \right\|_{W_X},\end{aligned}$$

where in the second step we used $g \in G_Y$ and thus $\Pi_{G_Y} g = g$. Moreover, since for $U := V_X^*$ we have $\Pi_{G_X} U = U \Pi_{W_X(Y)}$, see Lemma B.1.3, we find

$$\begin{aligned}\left\| \int_\Omega \Pi_{G_Y} gX \,\mathrm{d}\mu \right\|_{W_X} &= \left\| \Pi_{W_X(Y)} \left(\int_\Omega gX \,\mathrm{d}\mu \right) \right\|_{W_X} \\ &= \min \left\{ \|\tilde{w}\|_{W_X} \,\middle|\, \tilde{w} - \int_\Omega gX \,\mathrm{d}\mu \in W_X(Y)^\perp ,\, \tilde{w} \in W_X \right\}.\end{aligned}$$

Now observe that $G_Y \subseteq G_X$ implies $L_W = \hat{V}_Y \hat{V}_X^* = V_Y V_X^*$. In addition, for $\tilde{w} \in W_X$ we have

$$\begin{aligned}\tilde{w} - \int_\Omega gX \,\mathrm{d}\mu \in W_X(Y)^\perp &\Leftrightarrow V_X^* \left(\tilde{w} - \int_\Omega gX \,\mathrm{d}\mu \right) \in G_Y^\perp \\ &\Leftrightarrow \quad V_Y V_X^* \left(\tilde{w} - \int_\Omega gX \,\mathrm{d}\mu \right) = 0 \\ &\Leftrightarrow \quad L_W \tilde{w} = \int_\Omega gLX \,\mathrm{d}\mu.\end{aligned}$$

In summary we thus find

$$\begin{aligned}\|w\|_{W_Y} = \left\| \int_\Omega \Pi_{G_Y} gX \,\mathrm{d}\mu \right\|_{W_X} &= \min \left\{ \|\tilde{w}\|_{W_X} \,\middle|\, \tilde{w} - \int_\Omega gX \,\mathrm{d}\mu \in W_X(Y)^\perp ,\, \tilde{w} \in W_X \right\} \\ &= \min \left\{ \|\tilde{w}\|_{W_X} \,\middle|\, L_W \tilde{w} = \int_\Omega gLX \,\mathrm{d}\mu ,\, \tilde{w} \in W_X \right\} \\ &= \left\| L_W^\dagger \left(\int_\Omega gY \,\mathrm{d}\mu \right) \right\|_{W_X} \\ &= \|L_W^\dagger w\|_{W_X}.\end{aligned}$$

Now calculating the adjoint given $u \in W_X$ and $v \in W_Y$ we obtain

$$\langle L_W u, v \rangle_{W_Y} = \langle L_W^\dagger L_W u, L_W^\dagger v \rangle_{W_X} = \langle u, (L_W^\dagger L_W)^* L_W^\dagger v \rangle_{W_X} = \langle u, L_W^\dagger v \rangle_{W_X},$$

where in the last step we used one of the defining properties of the Moore-Penrose inverse. In other words we have $M_W = L_W^\dagger$. Note that depending on the situation we sometimes can extend M_W, see Examples 3.4.3 and 3.4.1.

The next example extends the previous example to noisy observations of LX.

Example 3.4.6. *Let $L : E \to F$ be a bounded operator and $N : \Omega \to F$ be a Gaussian random variable independent of X. For*

$$Y := LX + N,$$

we then have

$$M_W = L_W^\dagger \mathrm{Id}_L^* \tag{3.29}$$

with $L_W^\dagger$ *as in* (3.28) *and* Id_L^* *being the adjoint of the embedding* $\mathrm{Id}_L : W_{LX} \to W_Y$*, where we note that this embedding is well-defined since the independence of* N *and* X *implies* $G_Y = G_{LX} \oplus G_N$*. To verify* (3.29) *we note that for* $g \in G_X$

$$L_w \left(\int_\Omega gX \, \mathrm{d}\mu \right) = \int_\Omega gY \, \mathrm{d}\mu = \int_\Omega g(LX + N) \, \mathrm{d}\mu = \int_\Omega gLX \, \mathrm{d}\mu,$$

where we used that N *and* X *are independent, thus* $\int_\Omega gN \, \mathrm{d}\mu = 0$*. In other words, we have* $L_W = \mathrm{Id}_L \hat{L}_W$ *with*

$$\begin{aligned} \hat{L}_W &: W_X \to W_{LX} \\ \int_\Omega gX \, \mathrm{d}\mu &\mapsto \int_\Omega gLX \, \mathrm{d}\mu. \end{aligned}$$

We already showed in Example 3.4.5 that the adjoint of $\hat{L}_W$ *is* $L_W^\dagger$*. In summary, the adjoint of* L_W *is*

$$L_W^* = (\mathrm{Id}_L \hat{L}_W)^* = \hat{L}_W^* \mathrm{Id}_L^* = L_W^\dagger \mathrm{Id}_L^*.$$

Example 3.4.6 shows that one can factorize the conditioning problem into solving the minimization problem for $L_W^\dagger$ in (3.28) and calculating the adjoint of an embedding. We further note that for invertible L, we have $L_W^\dagger = L^{-1}|_{W_Y}$, which makes the computation of $M_W = L_W^*$ in Example 3.4.6 easier.

The next example investigates a variant of the noise model in Example 3.4.6.

Example 3.4.7. *Let* $L : E \to F$ *be a bounded operator,* $N : \Omega \to E$ *be a Gaussian random variable independent of* X*, and*

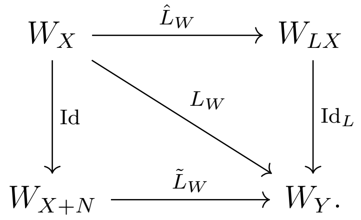
$$Y := L(X + N) .$$

Clearly this is a special case of Example 3.4.6 and in the following we provide an alternative way to compute M_W*.*

To this end, we define the following operators

$$\begin{aligned} \hat{L}_W &: W_X \to W_{LX}, & \tilde{L}_W &: W_{X+N} \to W_Y, \\ w &\mapsto Lw & w &\mapsto Lw \\ \mathrm{Id} &: W_X \to W_{X+N}, & \mathrm{Id}_L &: W_{LX} \to W_Y. \\ w &\mapsto w & w &\mapsto w \end{aligned}$$

Repeating the calculations of Example 3.4.6 we obtain the following commutative diagram



In other words we have $L_W = \mathrm{Id}_L \hat{L}_W = \tilde{L}_W \mathrm{Id}$*. Calculating the adjoint as in Example 3.4.6 we obtain*

$$M_W = \hat{L}_W^\dagger \mathrm{Id}_L^* = \mathrm{Id}^* \tilde{L}_W^\dagger .$$

We note that the Moore-Penrose inverses of $\hat{L}_W$ *and* $\tilde{L}_W$ *can differ.*

In our final example, we investigate Hilbert space valued Gaussian random variables for which our observational Y is based upon a subset of the eigenvectors of the covariance operator of X.

Example 3.4.8. *Let* H *be a separable Hilbert space and* $X : \Omega \to H$ *be a Gaussian random variable such that* W_X *is dense in* H*. Then the covariance operator can be viewed as a symmetric and positive operator* $\mathrm{cov}(X) : H \to H$*. Clearly,* $\mathrm{cov}(X)$ *is also compact and our denseness assumption ensures that it is injective. Consequently, all its, at most countably many, eigenvalues* $(\lambda_i)_{i\in I}$ *are greater than zero. The corresponding eigenvectors, denoted by* $(e_i)_{i\in I}$*, form an ONB of* H *and one can show that the sequence* $(\sqrt{\lambda_i} e_i)_{i\in I}$ *is an ONB of* W_X *with*

$$\|w\|_{W_X}^2 = \sum_{i\in I} \langle w, e_i\rangle_{W_X}^2 = \sum_{i\in I} \frac{\langle w, \mathrm{cov}(X) e_i\rangle_{W_X}^2}{\lambda_i} = \sum_{i\in I} \frac{\langle w, e_i\rangle_H^2}{\lambda_i} \tag{3.30}$$

for all $w \in W_X$*. Now, given a non-empty* $J \subseteq I$ *we consider the map* $L : H \to \ell^2(J)$ *given by*

$$Lf := \left(\langle f, e_j\rangle_H\right)_{j\in J} .$$

Lastly, we set $Y := LX$*. By Example 3.4.5 we know that* $M_W : W_Y \to W_X$ *is given by*

$$M_W w = \mathrm{argmin}\left\{\|w_x\|_{W_X} \mid Lw_x = w\right\}$$

for all $w \in W_Y$*. To solve this optimization problem, we fix a* $w := (w_j)_{j\in J} \in W_Y \subseteq \ell^2(J)$*. For* $w_x \in W_X$ *with* $Lw_x = w$ *we then know* $w_x = \sum_{i\in I} \alpha_i \sqrt{\lambda_i} e_i$ *for some* $(\alpha_i) \in \ell^2(I)$*, and therefore* $Lw_x = w$ *implies*

$$(\alpha_j \sqrt{\lambda_j})_{j\in J} = L_W w_x = (w_j)_{j\in J} .$$

In other words, we have $\alpha_j = \lambda_j^{-1/2} w_j$ *for all* $j \in J$*. In view of* (3.30) *we conclude that the sought minimizer satisfies* $\alpha_i = 0$ *for all* $i \in I \setminus J$ *and therefore we find*

$$M_W w = \sum_{j\in J} \alpha_j \sqrt{\lambda_j} e_j = \sum_{j\in J} w_j e_j$$

with convergence in W_X*. Since this shows* $\|M_W w\|_H \le \|w\|_{\ell^2(J)}$ *for all* $w \in \ell^2(J)$ *we conclude that* M_W *can be uniquely extended to a bounded linear operator* $M : \ell^2(J) \to H$*, which is given by*

$$Mw = \sum_{j\in J} w_j e_j .$$

We note that Example 3.4.8 covers the case of rotational invariant Gaussian random variables on a sphere $\mathbb{S}^n$, with $H = L^2(\mathbb{S}^n, \lambda)$ and λ being a rotational invariant measure on $\mathbb{S}^n$.

Chapter 4

Escaping the Native Space

In contrast to Chapters 3 and 5, this chapter focuses primarily on kernels and RKHSs rather than on Gaussian random variables. A standard assumption in kernel approximation is that the target function f belongs to the RKHS associated with the kernel.

This assumption is typically not satisfied for Gaussian random variables. As is well-known in the literature [29, 30, 42, 52], we have $\mu(\iota X|_{B(E')} \in H_X) = 0$ when H_X is infinite-dimensional. This phenomenon precludes the direct application of classical kernel approximation results [31] and necessitates addressing the so-called escaping-the-native-space problem [32].

In simple terms, the escaping-the-native-space question is as follows: assume we are given an RKHS H, and Banach spaces B_1, B_2 such that $H \subseteq B_1 \subseteq B_2$ and $f \in B_1$. Moreover, given a set of functionals $\Lambda_n := \{\lambda_1, \dots, \lambda_n\} \subset B_1'$. Can we derive a convergence rate $c_n > 0$ such that

$$\|f - s_{f,n}\|_{B_2} \le c_n \|f\|_{B_1}, \qquad \text{for all } f \in B_1,$$

holds true, with $s_{f,n}$ being the generalized kernel interpolant of f?

For Sobolev RKHSs H on some domain T and Sobolev spaces B_1 on the same domain with lower smoothness than H, this question has been answered with $B_2 = C(T)$, see [32]. The precise convergence rate depends on both the smoothness of the spaces H and B_1 and the dimensionality of the domain [53].

It is also possible to generalize the escaping-the-native-space question to the setting of generalized interpolation. To illustrate this, consider the case where λ is a measure on T such that $\int_T k(t,t)\,\mathrm{d}\lambda < \infty$. Under mild technical assumptions, there exists a sequence $(e_j) \subseteq L^2(\lambda)$ such that (e_j) forms an ONB in $L^2(\lambda)$, and a monotonically decreasing sequence (λ_j) such that $(\sqrt{\lambda_j} e_j) \subseteq H$ is an ONB in H, see [43].

For $f \in L^2(\lambda)$, point evaluations are in general not well defined, so we cannot compute standard kernel interpolants. Instead, we define the approximation $s_n(f) := \sum_{j=1}^n \langle f, e_j \rangle_{L^2(\lambda)} e_j$. We then derived in (2.1)

$$\|f - s_n(f)\|_{L^2(\lambda)}^2 \le \lambda_{n+1}^\theta \|f\|_{H^\theta}^2,$$

for all $f \in H^\theta$, where H^θ denotes the power space with $\theta \in [0,1]$. This answers the question of escaping-the-native-space, in the context of orthogonal projection onto span$\{(e_j)_{j=1}^n\}$ and in the $L^2(\lambda)$-norm. However, for general RKHSs and interpolation, the problem remains open. This is the setting we address.

To approach this problem, we first need to classify the spaces into which we "escape" into. In the Sobolev and $L^2(\lambda)$ cases, the target spaces were interpolation spaces. In our setting, we escape into scalable RKHSs. In Section 4.1, we directly provide the proof of the escaping-the-native-space theorem, and in Section 4.2, we compare this result with the convergence rate in (2.1).

4.1 Scalable RKHS

Let H be an RKHS. For $j \in \mathbb{N}$, let $d_j \in \mathbb{N}$ and $1 \le m \le d_j$, and let $(v_{j,m}) \subseteq H$ be an ONB. Then, for all $t_1, t_2 \in T$, we have $k(t_1, t_2) = \sum_{j=1}^{\infty} \sum_{m=1}^{d_j} v_{j,m}(t_1) v_{j,m}(t_2)$, see [38, Theorem 4.20].

The use of two indices for the ONB $(v_{j,m})$ is motivated by Example 5.3.4. There, we consider a stationary Gaussian random variable on a product of spheres and, instead of conditioning on point evaluations, we condition on harmonic polynomials these are the eigenfunctions of the Laplacian on the sphere. Since the eigenspaces E_j of the Laplacian have dimension greater than one, it is more natural to condition on all harmonic polynomials within a given eigenspace rather than on individual functions.

The following lemma is essential for answering the escaping-the-native-space question.

Lemma 4.1.1. *Let $(v_{j,m}) \subseteq H$ be an ONB, and let $a_j > 0$ be a monotonically increasing sequence, and let $c_j > 0$ be a monotonically decreasing sequence, such that*

$$\sup_{t\in T} \sum_{j=n+1}^{\infty} \sum_{m=1}^{d_j} v_{j,m}^2(t) \le c_n, \tag{4.1}$$

$a_j c_j \to 0$, and $(a_j \cdot (c_{j-1} - c_j)) \in \ell^1(\mathbb{N})$. We then have for all $t \in T$

$$\sum_{j=n+1}^{\infty} a_j \sum_{m=1}^{d_j} v_{j,m}(t)^2 \le \sum_{j=n+1}^{\infty} a_j (c_{j-1} - c_j) < \infty.$$

Proof. Set $\overline{P}_n^2(t) := \sum_{j=n+1}^{\infty} \sum_{m=1}^{d_j} v_{j,m}^2(t)$. We then have

$$\sum_{j=n+1}^{\infty} a_j \sum_{m=1}^{d_j} v_{j,m}^2(t) = \sum_{j=n+1}^{\infty} a_j \cdot \left(\overline{P}_{j-1}^2(t) - \overline{P}_j^2(t)\right).$$

For a finite sum with $l > n+1$, we compute

$$\begin{aligned}
\sum_{j=n+1}^{l} a_j \cdot \left(\overline{P}_{j-1}^2(t) - \overline{P}_j^2(t)\right) &= \sum_{j=n+1}^{l} a_j \overline{P}_{j-1}^2(t) - \sum_{j=n+1}^{l} a_j \overline{P}_j^2(t) \\
&= a_{n+1}\overline{P}_n^2(t) - a_l \overline{P}_l^2(t) + \sum_{j=n+2}^{l} a_j \overline{P}_{j-1}^2(t) - \sum_{j=n+1}^{l-1} a_j \overline{P}_j^2(t) \\
&= a_{n+1}\overline{P}_n^2(t) - a_l \overline{P}_l^2(t) + \sum_{j=n+1}^{l-1} a_{j+1} \overline{P}_j^2(t) - \sum_{j=n+1}^{l-1} a_j \overline{P}_j^2(t) \\
&= a_{n+1}\overline{P}_n^2(t) - a_l \overline{P}_l^2(t) + \sum_{j=n+1}^{l-1} \overline{P}_j^2(t)\left(a_{j+1} - a_j\right) \\
&\leq a_{n+1}c_n - a_l \overline{P}_l^2(t) + \sum_{j=n+1}^{l-1} c_j (a_{j+1} - a_j),
\end{aligned}$$

where in the last step we used the monotonicity of (a_j). By applying summation of parts on the right hand side we obtain

$$\begin{aligned}
a_{n+1}c_n - a_l \overline{P}_l^2(t) + \sum_{j=n+1}^{l-1} c_j (a_{j+1} - a_j) &= a_{n+1}c_n - a_l \overline{P}_l^2(t) + \sum_{j=n+1}^{l-1} c_j a_{j+1} - \sum_{j=n+1}^{l-1} c_j a_j \\
&= a_{n+1}c_n - a_l \overline{P}_l^2(t) + \sum_{j=n+2}^{l} c_{j-1} a_j - \sum_{j=n+1}^{l-1} c_j a_j \\
&= c_l a_l - c_l a_l - a_l \overline{P}_l^2(t) + \sum_{j=n+1}^{l} c_{j-1} a_j - \sum_{j=n+1}^{l-1} c_j a_j \\
&= c_l a_l - a_l \overline{P}_l^2(t) + \sum_{j=n+1}^{l} c_{j-1} a_j - \sum_{j=n+1}^{l} c_j a_j \\
&= c_l a_l - a_l \overline{P}_l^2(t) + \sum_{j=n+1}^{l} a_j \left(c_{j-1} - c_j\right).
\end{aligned}$$

Since $0 \leq a_l \overline{P}_l^2(t) \leq a_l c_l$ and by assumption $a_l c_l \to 0$ we have $a_l \overline{P}_l^2(t) \to 0$. Taking the limit $l \to \infty$ gives

$$\sum_{j=n+1}^{\infty} a_j \cdot \left(\overline{P}_{j-1}^2(t) - \overline{P}_j^2(t)\right) \leq \sum_{j=n+1}^{\infty} a_j \left(c_{j-1} - c_j\right).$$

□

Note that the key assumption in Lemma 4.1.1 is (4.1), which, as we will see later, implies a convergence rate of the posterior variance. Moreover, (4.1) with the conditions that there exist a monotonically increasing sequence $a_j > 0$ with $(a_j \cdot (c_{j-1} - c_j)) \in \ell^1(\mathbb{N})$

and $a_j c_j \to 0$ allows us to define a kernel $k_{a,V,d} : T \times T \to \mathbb{R}$ by

$$k_{a,V,d}(t_1, t_2) := \sum_{j=1}^{\infty} a_j \sum_{m=1}^{d_j} v_{j,m}(t_1) v_{j,m}(t_2),$$

see [38, Lemma 4.2]. We denote the associated RKHS by $H_{a,V,d}$. Furthermore, we recall $H \subseteq H_{a,V,d}$, see [42, Proposition 3.3], and that $(\sqrt{a_j} v_{j,m}) \subset H_{a,V,d}$ is an ONB, see [42, Proposition 3.3]

Let $V_n := \operatorname{span}\{v_{j,m} \,|\, 1 \leq j \leq n,\ 1 \leq m \leq d_j\}$ and $\Pi_n : H \to H$ denote the orthogonal projection onto V_n in H. Similarly, let $\Pi_{a,n,d} : H_{a,V,d} \to H_{a,V,d}$ be the orthogonal projection onto V_n in $H_{a,V,d}$. It follows that these two projections coincide on V_n, as the following lemma shows.

Lemma 4.1.2. *Let $\Pi_n : H \to H$ be the orthogonal projection in H onto*

$$V_n := \operatorname{span}\{v_{j,m} \,|\, 1 \leq j \leq n,\ 1 \leq m \leq d_j\}$$

and $\Pi_{a,n,d} : H_{a,V,d} \to H_{a,V,d}$ be the orthogonal projection in $H_{a,V,d}$ onto the same V_n. Then, for all $f \in H_X \subseteq H_{a,V,d}$, we have

$$\Pi_n f = \Pi_{a,n,d} f.$$

Proof. For $f \in H \subseteq H_{a,V,d}$ there exists a sequence $(b_{j,m})$ with $\sum_{j=1}^{\infty} \sum_{m=1}^{d_j} b_{j,m}^2 < \infty$ such that $f = \sum_{j=1}^{\infty} \sum_{m=1}^{d_j} b_{j,m} v_{j,m}$. Since $(\sqrt{a_j} v_{j,m})$ is an ONB of $H_{a,V,d}$. We conclude

$$\Pi_n f = \sum_{j=1}^{n} \sum_{m=1}^{d_j} b_{j,m} v_{j,m} = \sum_{j=1}^{n} \sum_{m=1}^{d_j} \frac{b_{j,m}}{\sqrt{a_j}} (\sqrt{a_j} v_{j,m}) = \Pi_{a,n,d} f.$$

□

It is essential that the orthogonal projections Π_n and $\Pi_{a,n,d}$ coincide. In the context of interpolation, computing Π_n involves inverting the kernel matrix. Since the projections are the same, we can use the same kernel matrix to compute $\Pi_{a,n,d}$. Consequently, the precise values of $k_{a,V,d}$ are never needed.

We can now prove the escaping-the-native-space Theorem.

Theorem 4.1.3. *Let $(v_{j,m}) \subseteq H$ be an ONB, and let $a_j > 0$ be a monotonically increasing sequence. Moreover, let $c_j > 0$ be a monotonically decreasing sequence, such that (4.1) is satisfied and we have both $a_j c_j \to 0$ and $(a_j \cdot (c_{j-1} - c_j)) \in \ell^1(\mathbb{N})$. Then, for all $f \in H_{a,V,d}$ the following inequality holds*

$$\|f - \Pi_{a,n,d} f\|_{C(T)}^2 \leq \sum_{j=n+1}^{\infty} a_j (c_{j-1} - c_j) \cdot \|f - \Pi_{a,n,d} f\|_{H_{a,V,d}}^2,$$

where $\Pi_{a,n,d}$ denotes the orthogonal projection onto $V_n := \operatorname{span}\{v_{j,m} \,|\, 1 \leq j \leq n,\ 1 \leq m \leq d_j\}$ in $H_{a,V,d}$.

Proof. Let $\mathrm{Id} : H_{a,V,d} \to H_{a,V,d}$ denote the identity mapping. By the reproducing property of $H_{a,V,d}$ and the fact that $\Pi_{a,n,d}$ is an orthogonal projection, which we use in the second and third equation, we have

$$\begin{aligned}(f - \Pi_{a,n,d} f)(t) &= \langle k_{a,V,d}(\,\cdot\,, t), f - \Pi_{a,n,d} f\rangle_{H_{a,V,d}} \\ &= \langle k_{a,V,d}(\,\cdot\,, t), (\mathrm{Id} - \Pi_{a,n,d})^2 f\rangle_{H_{a,V,d}} \\ &= \langle k_{a,V,d}(\,\cdot\,, t) - \Pi_n k_{a,V,d}(\,\cdot\,, t), f - \Pi_{a,n,d} f\rangle_{H_{a,V,d}}.\end{aligned}$$

Applying the Cauchy-Schwarz inequality gives

$$|(f - \Pi_{a,n,d} f)(t)| \leq \|k_{a,V,d}(\,\cdot\,, t) - \Pi_{a,n,d} k_{a,V,d}(\,\cdot\,, t)\|_{H_{a,V,d}} \|f - \Pi_{a,n,d} f\|_{H_{a,V,d}}.$$

Set $\overline{P}_n^2(t) := \sum_{j=n+1}^{\infty} \sum_{m=1}^{d_j} v_{j,m}^2(t)$. By applying [40, Lemma 2.3] to the kernel $k_{a,V,d}$, we have

$$\|k_{a,V,d}(\,\cdot\,, t) - \Pi_n k_{a,V,d}(\,\cdot\,, t)\|_{H_{a,V,d}}^2 = \sum_{j=n+1}^{\infty} a_j \sum_{m=1}^{d_j} v_{j,m}^2(t).$$

By Lemma 4.1.1 we obtain

$$\sum_{j=n+1}^{\infty} a_j \sum_{m=1}^{d_j} v_{j,m}(t)^2 \leq \sum_{j=n+1}^{\infty} a_j (c_{j-1} - c_j).$$

Finally, putting everything together and taking $\sup_{t\in T}$ leads to the assertion. □

4.2 Example

One of the issues with scalable RKHSs is that they highly depend on the considered ONB and can in general not easily be identified with well-known spaces such as Sobolev spaces. Therefore, comparing the convergence result obtained in Theorem 4.1.3 with results from the literature is difficult. However, as shown in Section 2.2.3, scalable RKHSs generalize power spaces. In the upcoming example we compare the convergence rate (2.1) for functions that lie in power spaces with the convergence rate in Theorem 4.1.3.

Example 4.2.1. *Let $(\tilde{T}, \lambda)$ be a probability space, moreover let $k : \tilde{T} \times \tilde{T} \to \mathbb{R}$ be a kernel such that there exists an ONB (e_j) in $L^2(\lambda)$ and a positive monotonically decreasing sequence $(\lambda_j) \in \ell^1$ such that $k(t, t') = \sum_{j=1}^{\infty} \lambda_j e_j(t) e_j(t')$ and $(\sqrt{\lambda_j} e_j)$ is an ONB in the associated RKHS H. Moreover, let $1 \geq \beta > 0$ such that $\sum_{j=1}^{\infty} \lambda_j^{\beta} e_j^2(t) < \infty$ and consider the orthogonal projection $\Pi_n : L^2(\lambda) \to L^2(\lambda)$ onto $V_n := \operatorname{span}\{e_j \mid 1 \leq j \leq n\}$. We recall that the convergence rate in* (2.1)*,*

$$\|f - \Pi_n f\|_{L^2(\lambda)} \leq \lambda_{n+1}^{\beta/2} \|f\|_{H^{\beta}},$$

holds true. We note that this setting falls under generalized interpolation. We take $T = B_{L^2(\lambda)}$ and the functionals are given by $\langle e_j, \cdot\,\rangle_{L^2(\lambda)} \in L^2(\lambda)'$.

We now reconstruct the power space within the scalable RKHS framework. Set $d_j = 1$*, we then omit the second* m *index in* $v_{j,m}$*, and define* $(v_j) := (\sqrt{\lambda_j} e_j)$ *as an ONB in* H*. Define* $a_j := \lambda_j^{\beta-1}$*, then, the resulting scaled kernel matches the power kernel. Moreover,* (2.1) *holds for* $\beta = 1$*, so we set* $c_j = \lambda_{j+1}$*. Assuming that* $\lambda_j^{\beta-1} \cdot \lambda_{j+1} \to 0$ *and* $(\lambda_j^{\beta-1} \cdot (\lambda_j - \lambda_{j+1})) \in \ell^1(\mathbb{N})$*, allows us to apply Theorem 4.1.3, which yields*

$$\|f - \Pi_n f\|_{L^2(\lambda)} \leq \sqrt{\sum_{j=n+1}^{\infty} \lambda_j^{\beta-1}(\lambda_j - \lambda_{j+1})} \cdot \|f\|_{H^\beta}. \tag{4.2}$$

The rate in (4.2) *is not yet explicit. To clarify the behavior, we further specify the example by assuming that for some* $\alpha > 1$*, we have* $\lambda_j = j^{-\alpha}$*. It follows from Lemma A.2.1 that the sequences* (a_j) *and* (c_j) *satisfy the conditions of Theorem 4.1.3. Consequently,* (4.2) *becomes*

$$\|f - \Pi_n f\|_{L^2(\lambda)} \leq \sqrt{\sum_{j=n+1}^{\infty} j^{\alpha(\beta-1)}(j^{-\alpha} - (j+1)^{-\alpha})} \cdot \|f\|_{H^\beta}.$$

Applying Lemma A.2.1, we obtain

$$\|f - \Pi_n f\|_{L^2(\lambda)} \leq \frac{1}{\sqrt{\beta}} n^{-\frac{\alpha \cdot \beta}{2}} \cdot \|f\|_{H^\beta}.$$

This matches the convergence rate obtained from (2.1) *in this setting*

$$\|f - \Pi_n f\|_{L^2(\lambda)} \leq n^{-\frac{\alpha \cdot \beta}{2}} \cdot \|f\|_{H^\beta},$$

which implies that, in this case, Theorem 4.1.3 yields the optimal convergence rate up to the constant factor $1/\sqrt{\beta}$*.*

Although only one example is given for Theorem 4.1.3, the class of examples it applies to is quite rich. The next chapter can also be viewed as an example of Theorem 4.1.3.

Chapter 5

Convergence Rates for Gaussian Random Variables

In this chapter, we first derive a simple concentration inequality and then relate it to the results from Chapter 4. We show that the escaping-the-native-space results can be used to obtain improved convergence rates for realizations of Gaussian random variables compared to those already established in the literature.

To place our result in context, we briefly compare it with existing concentration inequalities for Gaussian processes. The concentration inequality in [54] is based on a maximal inequality for Gaussian processes [55, Theorems 1.3.3 and 2.1.1], whereas the inequality we derive is based on a concentration inequality for chi-squared random variables. Additionally, the inequality in [54] applies in the misspecification setting but requires stronger assumptions on the Gaussian random variable than our approach. We discuss these differences in more detail later in this chapter.

Moreover, the concentration inequality that we derive is expressed in terms of the posterior variance. This is relevant in the context of Chapter 3, where we explicitly derived convergence rates of the posterior variance in Corollary 3.3.11.

Furthermore, since we established in Chapter 3 that we are working in a setting where an operator $M : F \to E$ exists with $MY = Z$, we restrict our attention to the case where X is conditioned on observations of X itself. This simplifies the notation without any loss of generality.

In Section 5.1, we derive a concentration inequality that allows us to bound the norm of the Gaussian random variable X in the scaled RKHS. In Section 5.2, we use the results from Chapter 4 in conjunction with the established concentration inequality to derive a concentration inequality for the Gaussian random variable. Finally, we provide examples in Section 5.3.

5.1 Concentration Inequality

The following lemma states a concentration inequality for generalized-chi-squared distributed random variables. Such a lemma has already been proven for the sum of *finitely* many squared independent one-dimensional Gaussian random variables, see [56, Theorem 4.2.3], but this is not enough for our context since we require the statement for the sum of *infinitely* many squared independent one dimensional Gaussian random variables. Let us briefly clarify the reason, let H be a Hilbert space, (v_j) an ONB of H, $r_j \sim \mathcal{N}(0,1)$ i.i.d., and $(a_j) \in \ell^1(\mathbb{N})$ i.i.d. assume we can write

$$X = \sum_{j=1}^{\infty} r_j \sqrt{a_j} v_j.$$

Then it can be shown that $\|X\|_H = \sum_{j=1}^{n} a_j r_j^2 < \infty$ almost everywhere holds true. Thus, the following concentration inequality bounds the norm of the Gaussian random variable.

Lemma 5.1.1. *Let $(r_j) \sim \mathcal{N}(0,1)$ be i.i.d. and let $(b_j) \in \ell^1(\mathbb{N})$ with $b_j \geq 0$ for all $j \geq 1$. Then the series $\sum_{j=1}^{\infty} b_j r_j^2$ converges almost surely, and for $Z := \sum_{j=1}^{\infty} b_j(r_j^2 - 1)$ we have, for all $\tau > 0$*

$$\mu(Z \geq 2\|(b_j)\|_{\ell^2}\sqrt{\tau} + 2\|(b_j)\|_{\ell^\infty}\tau) \leq \mathrm{e}^{-\tau}.$$

Proof. First, we show that $\sum_{j=1}^{\infty} b_j r_j^2$ converges almost surely. Since $b_j r_j^2 \geq 0$, we can apply Fubini's theorem to calculate the expectation

$$\mathbb{E}\sum_{j=1}^{\infty} b_j r_j^2 = \sum_{j=1}^{\infty} b_j \mathbb{E} r_j^2 = \sum_{j=1}^{\infty} b_j = \|b_j\|_{\ell^1} < \infty.$$

A series of nonnegative random variables with finite expectation converges almost surely, so the series converges. For the concentration inequality, we follow the argument from [57, Lemma 1]. For $x < 1/2$, [56, Theorem 4.2.3] gives

$$\psi(x) := \ln\left(\mathbb{E}\mathrm{e}^{x(r_j^2-1)}\right) = \ln\left(\mathrm{e}^{-x}\mathbb{E}\mathrm{e}^{r_j^2}\right) = -x - \frac{1}{2}\ln(1-2x).$$

Using Lemma A.3.3, for $0 < x < 1/2$ we have

$$0 < \psi(x) \leq \frac{x^2}{1-2x}.$$

Thus, for $0 < x < 1/(2\|(b_j)\|_{\ell^\infty})$, the independence of (r_j) yields

$$\ln\left(\mathbb{E}\mathrm{e}^{xZ}\right) = \sum_{j=1}^{\infty} \ln\left(\mathbb{E}\mathrm{e}^{b_j x(r_j^2-1)}\right) \leq \sum_{j=1}^{\infty} \frac{b_j^2 x^2}{1-2b_j x} \leq \frac{x^2\|(b_j)\|_{\ell^2}^2}{1-2x\|(b_j)\|_{\ell^\infty}} < \infty.$$

All series converge since the terms are nonnegative. Finally, applying Lemma A.3.4 with $v = 2\|(b_j)\|_{\ell^2}^2$ and $c = 2\|(b_j)\|_{\ell^\infty}$ gives the stated bound. □

5.2 Convergence Rates for Gaussian Random Variables

In this section, we combine the concentration inequality from Corollary 5.2.3 with Theorem 4.1.3 to derive a convergence rate for the concentration of Gaussian random variables.

The key quantity controlling this rate is the decay of the posterior variance. This connection should be understood in the light of Chapter 3, where we analyzed how the posterior variance decreases with the number of observations.

As already mentioned in Chapter 4, we will condition X on more than one functional at a time in Example 5.3.4. For this reason, we need to adapt our notation.

In the case where we condition in each step on $d_j \in \mathbb{N}$ many functionals $(e'_{j,m}) \in E'$ simultaneously, we write

$$X^{[n]} := \mathbb{E}\left(X \,|\, \sigma\left(\left(\left(e'_{j,m}(X)\right)_{m=1}^{d_j}\right)_{j=1}^{n}\right)\right) \tag{5.1}$$

and if we only condition on one functional in each step, meaning $d_j = 1$ we write X_n.

We note that in (5.1) we have a sequence $(e'_{j,m}) \subset E'$, and by definition of H_X we have

$$h_{j,m} := \int_\Omega e'_{j,m}(X)\iota_{B_{E''}} X \,\mathrm{d}\mu \in H_X, \tag{5.2}$$

where $\iota_{B_{E''}} : E \to E''|_{B'_E}$ denotes the canonical embedding into the bidual space, which is then restricted to the unit ball. We use the sequence $(e'_{j,m})$ to construct an orthonormal system in H_X. This leads to the following assumption.

Assumption V. *Given a sequence $(e'_{j,m}) \subseteq E'$, we define $h_{j,m}$ by (5.2). In addition let $(v_{j,m}) \subseteq H_X$ be an ONB of $\overline{\operatorname{span}\{h_{j,m}\}} \subseteq H_X$ satisfying*

$$\operatorname{span}\{v_{j,m} \,|\, 1 \le j \le n,\, 1 \le m \le d_j\} = \operatorname{span}\{h_{j,m} \,|\, 1 \le j \le n,\, 1 \le m \le d_j\}$$

for all n.

Remark 5.2.1. *Under Assumption V, if $\operatorname{span}\{e'_{j,m} \,|\, j \in \mathbb{N},\, 1 \le m \le d_j\}$ is weak*-dense in E', then by [19, Lemma 8.2.3], the family $(v_{j,m})$ forms an ONB of H_X.*

Lemma 5.2.2. *If Assumption V is satisfied, then there exist independent $\mathcal{N}(0,1)$ distributed random variables $r_{j,m} : \Omega \to \mathbb{R}$ such that $\int_\Omega r_{j,m}\iota_{B_{E''}} X \,\mathrm{d}\mu = v_{j,m}$ and*

$$\iota_{B_{E''}} X^{[n]} = \sum_{j=1}^{n}\sum_{m=1}^{d_j} r_{j,m} v_{j,m}.$$

Moreover, if $\|\operatorname{cov}(X) - \operatorname{cov}(X^{[n]})\|_{E' \to E} \to 0$, we also have

$$\iota_{B_{E''}} X = \sum_{j=1}^{\infty}\sum_{m=1}^{d_j} r_{j,m} v_{j,m}.$$

Proof. Our first goal is to construct the random variables $r_{j,m}$. To this end let $(v_{j,m})$ be an ONB of $V_n := \operatorname{span}\{\int_\Omega e'_{j,m}(X)\iota_{B_{E''}} X \,\mathrm{d}\mu \,|\, 1 \le j \le n, 1 \le m \le d_j\}$ and $(v_j^\perp)$ be an ONB of $V^\perp := (\operatorname{span}_{n\in\mathbb{N}} V_n)^\perp$. We note that since H_X is separable $(v_{j,m})$ and $(v_j^\perp)$ are both countable. Furthermore, $\{v_{j,m}\} \cup \{v_j^\perp\}$ is an ONB of H_X and for all $e' \in B_{E'}$ we have

$$\operatorname{cov}(X)e' = k(e', \cdot\,) = \sum_{j=1}^{\infty}\sum_{m=1}^{d_j} v_{j,m}(e')v_{j,m}(\,\cdot\,) + \sum_{j=1}^{\infty} v_j^\perp(e')v_j^\perp(\,\cdot\,). \tag{5.3}$$

By Theorem 3.3.4, for all $e' \in B_{E'}$ we also have

$$\operatorname{cov}(X^{[n]})e' = \sum_{j=1}^{n} \sum_{m=1}^{d_j} v_{j,m}(e')v_{j,m}(\,\cdot\,). \tag{5.4}$$

By [58, Theorem 3.1], we have that $(v_{j,m}) \subset H_{X^{[n]}}$ is a Parseval frame, that is

$$\sum_{j=1}^{n} \sum_{m=1}^{d_j} |\langle f, v_{j,m}\rangle_{H_{X^{[n]}}}|^2 = \|f\|^2_{H_{X^{[n]}}}$$

for all $f \in H_{X^{[n]}}$. Consequently, we have $H_{X^{[n]}} = \operatorname{span}\{v_{j,m} \,|\, 1 \le j \le m, 1 \le m \le d_j\}$ by (5.4) and [38, Theorem 4.21] and since the vectors $(v_{j,m})$ are linearly independent, they are a basis of $H_{X^{[n]}}$. Moreover, by [59, Proposition 1.11], we have, for all $i, l \in \mathbb{N}$,

$$v_{i,l} = \sum_{j=1}^{n} \sum_{m=1}^{d_j} \langle v_{i,l}, v_{j,m}\rangle_{H_{X^{[n]}}} v_{j,m}.$$

Since $(v_{j,m})$ is a basis, the coefficients are unique. We conclude $\langle v_{i,l}, v_{j,m}\rangle_{H_{X^{[n]}}} = 1$ if $(j,m) = (i,l)$ and 0 else. In summary, $(v_{j,m})$ is an ONB of $H_{X^{[n]}}$.

By Lemma 3.1.1 the mapping $V_{X^{[n]}} : G_{X^{[n]}} \to W_{X^{[n]}}$ defined by

$$V_{X^{[n]}} g := \int_\Omega g X^{[n]} \,\mathrm{d}\mu$$

is an isometry. Furthermore, the mapping $\iota_{B_{E''}} : W_{X^{[n]}} \to H_{X^{[n]}}$ is an isometry by Lemma 3.1.2. Thus we define $r_{j,m} := V^{-1}_{X^{[n]}} \iota^{-1} v_{j,m}$. Furthermore, we have $r_{j,m} \in G_{X^{[n]}}$ and by definition they are normalized and orthogonal, thus $r_{j,m} \sim \mathcal{N}(0,1)$ i.i.d.. By Theorem 3.3.2, we have

$$X^{[n]} = \sum_{j=1}^{n} \sum_{m=1}^{d_j} r_{j,m} v_{j,m}.$$

Let us now assume that $\operatorname{cov}(X^{[n]}) \to \operatorname{cov}(X)$ in the operator norm. Our first goal is to show

$$\overline{\operatorname{span}\left\{\int_\Omega e'_{j,m}(X)\iota X \,\mathrm{d}\mu\right\}}^{\|\cdot\|_{H_X}} = H_X. \tag{5.5}$$

It suffices to show $V^\perp = \{0\}$. By assumption, $\operatorname{cov}(X_n)e' \to \operatorname{cov}(X)e'$, so (5.3) and (5.4) give

$$\operatorname{cov}(X)e' - \operatorname{cov}(X^{[n]})e' = \sum_{j=1}^{\infty} v_j^\perp(e')v_j^\perp(\,\cdot\,) + \sum_{j=n+1}^{\infty} \sum_{m=1}^{d_j} v_{j,m}(e')v_{j,m}(\,\cdot\,) \to 0.$$

This implies $\sum_{j=1}^\infty v_j^\perp(e')v_j^\perp(e') + \sum_{j=n+1}^\infty \sum_{m=1}^{d_j} v_{j,m}(e')v_{j,m}(e') \to 0$ and since the first sum is independent of n, we conclude that $\sum_{j=1}^\infty v_j^\perp(e')v_j^\perp(e') = 0$ for all $e' \in E'$. This implies $v_j^\perp = 0$ for all $j \in \mathbb{N}$, so $V^\perp = \{0\}$. Therefore, (5.5) holds true.

By [34, Theorem 3.3.2] we have

$$\sum_{j=1}^{n}\sum_{m=1}^{d_j} r_{j,m}v_{j,m} \to \sum_{j=1}^{\infty}\sum_{m=1}^{d_j} r_{j,m}v_{j,m},$$

where the convergence is almost surely. Finally, $X - X^{[n]}$ is Gaussian with mean zero and covariance $\operatorname{cov}(X - X^{[n]})$. By Lemma 3.2.16 we have

$$\operatorname{cov}(X - X^{[n]}) = \operatorname{cov}(X) - \operatorname{cov}(X^{[n]}),$$

which converges to 0 by assumption, thus the limit of $X^{[n]}$ is truly X and we conclude

$$\iota X = \sum_{j=1}^{\infty}\sum_{m=1}^{d_j} r_{j,m}v_{j,m}.$$

□

Corollary 5.2.3. *Under Assumption V, let (a_j) be a sequence with $a_j > 0$ for all $j \in \mathbb{N}$ and $(d_j/a_j) \in \ell^1(\mathbb{N})$. Then we have $\iota_{B_{E''}}X, \iota_{B_{E''}}X^{[n]} \in H_{a,V}$ almost surely, and*

$$\mu\left(\|\iota_{B_{E''}}X - \iota_{B_{E''}}X^{[n]}\|^2_{H_{a,V,d}} \leq 5\max\{1,\tau\}\sum_{j=n+1}^{\infty}\frac{d_j}{a_j}\right) \geq 1 - \mathrm{e}^{-\tau}.$$

Proof. By Lemma 5.2.2, we can write $\iota_{B_{E''}}X = \sum_{j=1}^{\infty}\sum_{m=1}^{d_j} r_{j,m}v_{j,m}$. Applying Lemma 5.1.1, we get

$$\sum_{j=1}^{\infty}\frac{1}{a_j}\sum_{m=1}^{d_j} r_{j,m}^2 < \infty$$

almost surely. Since $(\sqrt{a_j}v_{j,m})$ is an ONB of $H_{a,V,d}$ we conclude $\iota_{B_{E''}}X \in H_{a,V,d}$. Moreover, by Lemma 5.2.2, we have $\iota_{B_{E''}}X - \iota_{B_{E''}}X^{[n]} = \sum_{j=n+1}^{\infty}\sum_{m=1}^{d_j} r_{j,m}v_{j,m}$. Applying the concentration inequality from Lemma 5.1.1, we obtain for any $\tau > 0$

$$\mu\left(\|\iota_{B_{E''}}X - \iota_{B_{E''}}X_n\|^2_{H_{a,V}} \geq \sum_{j=n+1}^{\infty} d_j/a_j + 2\sqrt{\tau}\sqrt{\sum_{j=n+1}^{\infty} d_j^2/a_j^2} + 2\tau\sup_{j\geq n+1} d_j/a_j\right) \leq \mathrm{e}^{-\tau}.$$

Furthermore, note that we have $\|\,(d_{j+n}/a_{j+n})\,\|_{\ell^\infty} \leq \|\,(d_{j+n}/a_{j+n})\,\|_{\ell^2} \leq \|\,(d_{j+n}/a_{j+n})\,\|_{\ell^1}$ and thus we find

$$\begin{aligned}&\|(d_{j+n}/a_{j+n})\|_{\ell^1(\mathbb{N})} + 2\sqrt{\tau}\,\|(d_{j+n}/a_{j+n})\|_{\ell^2(\mathbb{N})} + 2\tau\,\|(d_{j+n}/a_{j+n})\|_{\ell^\infty(\mathbb{N})}\\ \leq& 5\max\{1,\tau\}\,\|(d_{j+n}/a_{j+n})\|_{\ell^1(\mathbb{N})}\,.\end{aligned}$$

Putting everything together leads to the assertion. □

Our main result shows that convergence rates for the posterior variance also imply convergence rates for the realizations. We then derive our other results from this theorem. In particalur we treat the cases where the posterior variance converges at a polynomial rate or at an exponential rate.

Theorem 5.2.4. *Let X be a Gaussian random variable and $(e'_{j,m})_{m=1}^{d_j} \subset E'$. Suppose there exists a monotonically decreasing sequence $(c_j) \in \ell^1(\mathbb{N})$ such that*

$$\|\operatorname{cov}(X) - \operatorname{cov}(X^{[n]})\|_{E' \to E} \leq c_n$$

for all $n \in \mathbb{N}$. Then, for every monotonically increasing sequence (a_j) satisfying $(d_j/a_j) \in \ell^1(\mathbb{N})$, $(a_j(c_{j-1} - c_j)) \in \ell^1(\mathbb{N})$, $a_j c_j \to 0$, and for all $n \in \mathbb{N}$, $\tau > 0$ we have

$$\mu\left(\|X - X^{[n]}\|_E \leq \sqrt{5\max\{1,\tau\}\left(\sum_{j=n+1}^{\infty} a_j(c_{j-1} - c_j)\right)\left(\sum_{j=n+1}^{\infty} \frac{d_j}{a_j}\right)}\right) > 1 - \mathrm{e}^{-\tau}.$$

Proof. Let $\left((v_{j,m})_{m=1}^{d_j}\right)_{j=1}^{n} \subseteq H_{X^{[n]}}$ be an ONB of $H_{X^{[n]}}$. We then have

$$\sup_{e' \in B_{E'}} \sum_{j=n+1}^{\infty} \sum_{m=1}^{d_j} v_{j,m}^2(e') = \|\operatorname{cov}(X) - \operatorname{cov}(X^{[n]})\|_{E' \to E} \leq c_n.$$

Moreover, we have $\|\iota X - \iota X^{[n]}\|_{E''} = \|X - X^{[n]}\|_E$. By Lemma 5.2.2 we have $X^{[n]} = \Pi_n X$. Applying Lemma 4.1.3, we obtain

$$\|X - X^{[n]}\|_E = \|\iota X - \iota X^{[n]}\|_{E''} \leq \sqrt{\sum_{j=n+1}^{\infty} a_j(c_{j-1} - c_j) \cdot \|\iota X - \iota X^{[n]}\|_{H_{a,V,d}}^2}.$$

Finally, using Corollary 5.2.3 leads to the assertion. □

In the case where the posterior variance converges at a polynomial rate, we obtain the following corollary.

Corollary 5.2.5. *Let X be a Gaussian random variable and let $(e'_{j,m}) \subseteq E'$. Suppose there exist constants $C, C_d > 0$, $\alpha > 1 + \beta \geq 1$, such that*

$$\|\operatorname{cov}(X) - \operatorname{cov}(X^{[n]})\|_{E' \to E} \leq C(n+1)^{-\alpha}, \qquad \text{and} \qquad d_n \leq C_d(n+1)^{\beta}$$

for all $n \in \mathbb{N}$. Then, for $n \geq 1$ and $\tau > 0$, we have

$$\mu\left(\|X - X^{[n]}\|_E \leq \frac{\sqrt{20 \cdot 2^{\beta} C C_d \alpha \max\{1,\tau\}}}{\alpha - \beta - 1} n^{\frac{-\alpha+\beta+1}{2}}\right) > 1 - \mathrm{e}^{-\tau}.$$

Proof. We will apply Theorem 5.2.4 with $c_n := C(n+1)^{-\alpha}$ and $a_j := j^{\gamma}$ for $\beta+1 < \gamma < \alpha$. The condition $\beta + 1 < \gamma$ implies $(d_j/a_j) \in \ell^1(\mathbb{N})$. We note that (a_j) is monotonically increasing. By Lemma A.2.1 we have that the assumptions of Theorem 5.2.4 are satisfied and

$$\sum_{j=n+1}^{\infty} a_j(c_{j-1} - c_j) \leq C\alpha \frac{n^{\gamma-\alpha}}{\alpha - \gamma},$$

$$\sum_{j=n+1}^{\infty} \frac{d_j}{a_j} \leq C_d 2^{\beta} \frac{n^{\beta+1-\gamma}}{\gamma - \beta - 1}.$$

Multiplying the two terms gives

$$\left(\sum_{j=n+1}^{\infty} a_j(c_{j-1}-c_j)\right)\left(\sum_{j=n+1}^{\infty} \frac{d_j}{a_j}\right) \le \alpha C C_d 2^{\beta} \frac{n^{\gamma-\alpha}}{\alpha-\gamma}\frac{n^{\beta+1-\gamma}}{\gamma-\beta-1}$$
$$= \alpha C C_d 2^{\beta} \frac{n^{-\alpha+\beta+1}}{(\alpha-\gamma)(\gamma-\beta-1)}.$$

By setting $\gamma := \frac{\alpha+\beta+1}{2}$ and putting everything together we obtain by Theorem 5.2.4

$$\mu\left(\|X-X^{[n]}\|_{E'\to E} \le \frac{\sqrt{20\alpha C C_d 2^{\beta}\max\{1,\tau\}}}{\alpha-\beta-1} n^{\frac{-\alpha+\beta+1}{2}}\right) > 1-\mathrm{e}^{-\tau}.$$

□

The next theorem should be viewed as a simplified version of our main result, in which only one functional is added at each conditioning step.

Theorem 5.2.6. *Let X be a Gaussian random variable and $(e'_j) \subset E'$ a sequence of functionals for which there exists a positive, monotonically decreasing sequence $(c_j) \in \ell^1(\mathbb{N})$ such that*

$$\|\mathrm{cov}(X)-\mathrm{cov}(X_n)\|_{E'\to E} \le c_n, \tag{5.6}$$

for all $n \in \mathbb{N}$. Assume further that $(c_{j-1}-c_j)$ is monotonically decreasing, $(\sqrt{c_{j-1}-c_j}) \in \ell^1(\mathbb{N})$, and $c_j/\sqrt{c_{j-1}-c_j} \to 0$. Then, for all $n \in \mathbb{N}$ and $\tau > 0$ we have

$$\mu\left(\|X-X_n\|_E \le \sqrt{5\max\{1,\tau\}} \sum_{j=n+1}^{\infty} \sqrt{c_{j-1}-c_j}\right) > 1-\mathrm{e}^{-\tau}.$$

Proof. The result follows directly from Theorem 5.2.4 by setting $d_j := 1$ and $a_j := 1/\sqrt{c_{j-1}-c_j}$ for all $j \in \mathbb{N}$. □

The following corollary specializes Corollary 5.2.5 to the case in which only one functional is added at each conditioning step.

Corollary 5.2.7. *Let X be a Gaussian random variable and $(e'_j) \subset E'$ be a sequence for which there exist $C > 0$ and $\alpha > 1$ such that*

$$\|\mathrm{cov}(X)-\mathrm{cov}(X_n)\|_{E'\to E} \le C(n+1)^{-\alpha}$$

for all $n \in \mathbb{N}$. Then, for $n \ge 1$ and $\tau > 0$, we have

$$\mu\left(\|X-X_n\|_E \le \frac{\sqrt{20C\alpha\max\{1,\tau\}}}{(\alpha-1)} n^{\frac{1-\alpha}{2}}\right) > 1-\mathrm{e}^{-\tau}.$$

Proof. This follows directly by Corollary 5.2.5 with $d_j = 1$. □

Similarly, if the posterior variance converges at an exponential rate, we obtain the following bound.

Corollary 5.2.8. *Let X be a Gaussian random variable and $(e'_j) \subset E'$ be a sequence for which there exist $C_1, C_2 > 0$ and $\alpha \geq 1$ such that*

$$\|\mathrm{cov}(X) - \mathrm{cov}(X_n)\|_{E' \to E} \leq C_1 \mathrm{e}^{-C_2 n^{1/\alpha}}$$

for all $n \in \mathbb{N}$. Then, for $\alpha > 1$, $n > \left(\frac{11}{C_2}(\alpha - 1)\right)^{\alpha}$, and $\tau > 0$, we have

$$\mu\left(\|X - X_n\|_E \leq \sqrt{\frac{121 C_1 C_2 \alpha \max\{1, \tau\}}{20}\left(\frac{C_2}{2}\right)^{\alpha-2}} (n-1)^{\frac{\alpha-1}{2\alpha}} \mathrm{e}^{-\frac{C_2}{2}(n-1)^{1/\alpha}}\right) > 1 - \mathrm{e}^{-\tau}.$$

For $\alpha = 1$ we obtain for all $\tau > 0$ and $n \in \mathbb{N}$

$$\mu\left(\|X - X_n\|_E \leq \sqrt{5 \max\{1, \tau\} C_1 \cdot (\mathrm{e}^{C_2} - 1)} \cdot \frac{\mathrm{e}^{C_2}}{\sqrt{\mathrm{e}^{C_2} - 1}} \mathrm{e}^{-\frac{C_2}{2} n}\right) > 1 - \mathrm{e}^{-\tau}.$$

Proof. We will apply Theorem 5.2.6 with $c_n := C_1 \mathrm{e}^{-C_2 n^{1/\alpha}}$. First, we note that $(c_n) \in \ell^1(\mathbb{N})$ and (c_n) is positive monotonically decreasing. To show that $(c_{j-1} - c_j)$ is monotonically decreasing, observe that this is equivalent to $c_{j-1} - c_j - (c_j - c_{j+1}) \geq 0$. Applying the mean value theorem twice to the function $f : [0, \infty) \to \mathbb{R}$, $f(t) := C_1 \mathrm{e}^{-C_2 t^{1/\alpha}}$, there exist $\xi \in [j-1, j]$,$\zeta \in [j, j+1]$ such that

$$c_{j-1} - c_j - (c_j - c_{j+1}) = -f'(\xi) + f'(\zeta).$$

Since f is convex and $\zeta > \xi$, we conclude $f'(\zeta) - f'(\xi) \geq 0$.

By the mean value theorem, there also exist $\xi_j \in [j-1, j]$ such that

$$\begin{aligned}\sum_{j=1}^{\infty} \left(\sqrt{c_{j-1} - c_j}\right) = \sum_{j=1}^{\infty} \sqrt{-f'(\xi_j)} &= \sum_{j=1}^{\infty} \sqrt{\frac{C_1 C_2}{\alpha} \xi_j^{\frac{1}{2\alpha} - \frac{1}{2}} \mathrm{e}^{-\frac{C_2}{2} \xi_j^{1/\alpha}}} \\ &\leq \sum_{j=1}^{\infty} \sqrt{\frac{C_1 C_2}{\alpha} j^{\frac{1}{2\alpha} - \frac{1}{2}} \mathrm{e}^{-\frac{C_2}{2}(j-1)^{1/\alpha}}}.\end{aligned}$$

The final sum is finite because of the exponential decay, hence $(\sqrt{c_{j-1} - c_j}) \in \ell^1(\mathbb{N})$.

Finally, we show that $c_j / \sqrt{c_{j-1} - c_j} \to 0$. Again, using the mean value theorem, for $j > 1$:

$$\begin{aligned}\frac{c_j}{\sqrt{c_{j-1} - c_j}} = \frac{C_1 \mathrm{e}^{-C_2 j^{1/\alpha}}}{\sqrt{\frac{C_1 C_2}{\alpha} \xi_j^{\frac{1}{2\alpha} - \frac{1}{2}} \mathrm{e}^{-\frac{C_2}{2} \xi_j^{1/\alpha}}}} &\leq \frac{\sqrt{\alpha C_1}}{\sqrt{C_2} (j-1)^{\frac{1}{2\alpha} - \frac{1}{2}}} \cdot \frac{\mathrm{e}^{-C_2 j^{1/\alpha}}}{\mathrm{e}^{-\frac{C_2}{2} \xi_j^{1/\alpha}}} \\ &= \frac{\sqrt{\alpha C_1}}{\sqrt{C_2} (j-1)^{\frac{1}{2\alpha} - \frac{1}{2}}} \cdot \mathrm{e}^{-\frac{C_2}{2} j^{1/\alpha}} \to 0.\end{aligned}$$

Thus all conditions of Theorem 5.2.6 are satisfied. For $\alpha > 1$ applying Corollary A.3.2 gives the desired assertion.

For the case $\alpha = 1$ we calculate

$$\begin{aligned}\sum_{j=n+1}^{\infty} \sqrt{c_{j-1} - c_j} &= \sum_{j=n+1}^{\infty} \sqrt{C_1 \cdot (\mathrm{e}^{C_2})^{-(j-1)}(1 - \mathrm{e}^{-C_2})} \\ &= \sqrt{C_1 \cdot (1 - \mathrm{e}^{-C_2})\mathrm{e}^{C_2}} \sum_{j=n+1}^{\infty} (\sqrt{\mathrm{e}^{C_2}})^{-j} \\ &= \sqrt{C_1 \cdot (\mathrm{e}^{C_2} - 1)}(\mathrm{e}^{C_2})^{-\frac{n+1}{2}} \sum_{j=0}^{\infty} (\sqrt{\mathrm{e}^{C_2}})^{-j} \\ &= \sqrt{C_1 \cdot (\mathrm{e}^{C_2} - 1)} \cdot \frac{\mathrm{e}^{C_2}}{\sqrt{\mathrm{e}^{C_2}} - 1} \mathrm{e}^{-\frac{C_2}{2}n}.\end{aligned}$$

With Theorem 5.2.6 we obtain the assertion. □

Remark 5.2.9. *Assume that $E = C(T)$ and that the covariance kernel of X can be written as $k_X(t,s) = \kappa(t-s)$ for all $t, s \in T$, where $\kappa : \mathbb{R}^d \to \mathbb{R}$. In addition assume that there exists an $0 < \alpha \leq 1$ such that*

$$\int_{\mathbb{R}^d} \|\omega\|^{\alpha} \hat{\kappa}(\omega) \, \mathrm{d}\mu < \infty, \tag{5.7}$$

where $\hat{\kappa}$ denotes the Fourier transform of κ. If we consider conditioning with respect to point evaluations and assume that (5.6) holds then the results in [54] lead to the concentration inequality

$$\mu\left(\|X - X_n\|_{C(T)} \leq C\left(\sqrt{c_n} \ln\left(\frac{1}{c_n}\right) + \sqrt{\tau c_n}\right)\right) > 1 - \mathrm{e}^{-\tau},$$

for all $n \in \mathbb{N}$ and all $\tau > 0$, where $C > 0$ is a suitable constant. This is comparable to Theorem 5.2.6. We will discuss in the next Section 5.3, when which result is better.

5.3 Examples

We focus our examples on Gaussian random variables taking their values in the Banach space $E = C([0,1]^d)$. For simplicity, we assume that $\mathbb{E}(X) = 0$ and describe the Gaussian random variable X using the covariance operator defined on point evaluations $\delta_t(X)(\omega) := X_t(\omega) := X(t, \omega)$ for $\omega \in \Omega$ and $t \in [0,1]^d$.

5.3.1 Sobolev Covariance

Let $X : \Omega \to C([0,1]^d)$ be a centered Gaussian random variable with kernel function given by

$$k_X(t,s) = \frac{2^{1-(s-d/2)}}{\Gamma(s - d/2)} \|r - t\|^{s-d/2} K_{s-d/2}(\|r - t\|),$$

where $s > d$, $K_{s-d/2}$ is the modified Bessel function of the second kind, and Γ is the Gamma function. This kernel is known as the Matérn kernel, whose RKHS is the Sobolev space $H^s([0,1]^d)$, see [60].

Assume that we have chosen points $(t_n) \in [0,1]^d$, such that for some constant $C > 0$, we have

$$\sup_{t\in[0,1]^d} |k_X(t,t) - k_{X_n}(t,t)| \le C(n+1)^{-\frac{2s}{d}+1}, \quad \text{for all } n \in \mathbb{N}.$$

Such points can be constructed by algorithms like P-greedy, see [40, Corollary 2.2]. Applying Lemma A.3.5 and Corollary 5.2.7 leads to

$$\mu\left(\|X - X_n\|_{C([0,1]^d)} \le \frac{\sqrt{20C\left(\frac{2s}{d}-1\right)\max\{1,\tau\}}}{\frac{2s}{d}-2} n^{-\frac{s}{d}+1} \right) > 1 - \mathrm{e}^{-\tau}$$

for $n \ge 1$ and $\tau > 0$. We note that this converges to zero only if $s > d$, recall that this condition also ensures that almost all realizations of X are contained in the Sobolev space $H^r([0,1]^d)$ for all $r \in (d/2, s-d/2)$, see [61, Example 5.6].

We note that the convergence rates provided in [54] leads to the following concentration inequality

$$\mu\left(\|X - X_n\|_{C([0,1]^d)} \le C\left(n^{-\frac{s}{d}+\frac{1}{2}}\sqrt{\ln(n)} + \sqrt{\tau} n^{-\frac{s}{d}+\frac{1}{2}} \right) \right) > 1 - \mathrm{e}^{-\tau}, \tag{5.8}$$

for some $C > 0$. The rate in (5.8) is faster than the one provided in our setting, however it only holds for stationary processes satisfying (5.7).

5.3.2 Gaussian Covariance

Let $X : \Omega \to C([0,1]^d)$ be a centered Gaussian random variable with the kernel function

$$k_X(t,s) = \mathrm{e}^{-\|t-s\|_2^2}, \quad t,s \in [0,1]^d.$$

We identify points t_n with the functionals δ_{t_n}, meaning that the conditional expectation X_n is given by

$$X_n = \mathbb{E}(X \mid \sigma((X_{t_j})_{j=1}^n)).$$

Assume that the points $(t_n) \in [0,1]^d$ are chosen such that there exist constants $C_1, C_2 > 0$ with

$$\sup_{t\in[0,1]^d} |k_X(t,t) - k_{X_n}(t,t)| \le C_1 \mathrm{e}^{-C_2 n^{1/d}}, \quad \text{for all } n \in \mathbb{N}.$$

Such points do exist and can be constructed, for instance using P-greedy algorithms, see [40, Corollary 2.2]. We then apply Lemma A.3.5 and Corollary 5.2.8 to obtain

$$\mu\left(\|X - X_n\|_{C([0,1]^d)} \le \sqrt{\frac{121 C_1 C_2 d \max\{1,\tau\}}{20}} \left(\frac{C_2}{2}\right)^{d-2} (n-1)^{\frac{d-1}{2d}} \mathrm{e}^{-\frac{C_2}{2}(n-1)^{1/d}} \right) > 1 - \mathrm{e}^{-\tau},$$

for $d > 1$, $n > \left(\frac{11}{C_2}(d-1)\right)^d + 1$, and $\tau > 0$.

For $d = 1$, $n \in \mathbb{N}$, and $\tau > 0$ we obtain

$$\mu\left(\|X - X_n\|_E \le \sqrt{5\max\{1,\tau\}C_1 \cdot (\mathrm{e}^{C_2} - 1)} \cdot \frac{\mathrm{e}^{C_2}}{\sqrt{\mathrm{e}^{C_2} - 1}}\mathrm{e}^{-\frac{C_2}{2}n}\right) > 1 - \mathrm{e}^{-\tau}.$$

We note that the convergence rates provided in [54] leads to the following concentration inequality

$$\mu\left(\|X - X_n\|_{C([0,1]^d)} \le C\left(\mathrm{e}^{-\frac{C_2}{2}n^{1/d}} \cdot n^{\frac{1}{2d}} + \sqrt{\tau}\mathrm{e}^{-\frac{C_2}{2}n^{1/d}}\right)\right) > 1 - \mathrm{e}^{-\tau} \tag{5.9}$$

for some $C > 0$. The rate in (5.9) is faster if $d \ge 3$ and for $d = 2$ the rates coincide. For the case $d = 1$ our setting provides an improvement in the realistic case $\tau \le n^{1/d}$. Note, however that these improvements are only obtained for stationary processes satisfying (5.7).

5.3.3 Conditioning on Eigenfunctions

Let (T, λ) be a probability space such that $L^2(\lambda)$ is separable. In the case of $E = L^2(\lambda)$, it is often more natural to condition on the eigenfunctions of the covariance operator rather than on point evaluations, which do not exist. Recall that for a separable Banach space, the covariance operator is nuclear, symmetric and positive, see [62]. Let (λ_j) denote the monotone decreasing eigenvalues and (e_j) denote the eigenfunctions. By conditioning on $e'_j := \langle e_j, \cdot\rangle_{L^2(\lambda)} \in L^2(\lambda)'$, we have by Lemma A.3.6

$$\|\mathrm{cov}(X) - \mathrm{cov}(X_n)\|_{L^2(T)\to L^2(T)} = \lambda_{n+1}.$$

If we further assume that there exist constants $C > 0$ and $\alpha > 1$ such that $\lambda_{n+1} \le C(n+1)^{-\alpha}$, as it is the case for Sobolev spaces, then Corollary 5.2.7 yields

$$\mu\left(\|X - X_n\|_E \le \frac{\sqrt{20C\alpha\max\{1,\tau\}}}{\alpha - 1}n^{(1-\alpha)/2}\right) > 1 - \mathrm{e}^{-\tau}.$$

5.3.4 Gaussian Random Variables on the Product of Spheres

We now make the previous eigenfunction-based example more concrete by considering a product of spheres $T = \mathbb{S}^{\mathbf{d}_1} \times \mathbb{S}^{\mathbf{d}_2}$, following the setup in [63]. We note, $\mathbb{S}^d \subset \mathbb{R}^{d+1}$ denotes the sphere and $d, \mathbf{d}_1, \mathbf{d}_2 \in \mathbb{N}$. To avoid repeating the full harmonic analysis framework, we refer the reader to [63], but recall the key elements needed for our discussion.

Let Δ denote the Laplace-Beltrami operator on $\mathbb{S}^d$ with eigenvalues μ_j and corresponding eigenspaces $H_j(d)$. The dimension of $H_j(d)$ is denoted by $D_j(d)$, with $D_j(1) = 2$ and, for $d \ge 2$,

$$D_j(d) = \dim(H_j(d)) = (2j + d - 1)\frac{j + d - 2}{j!(d-1)!} \le c_d(j+1)^{d-1}$$

for some $c_d \ge 1$, see [63, Equation (5.2)]. We denote by $S^d_{j,l}$ the classical spherical harmonics on $\mathbb{S}^d$, see [64], which form an ONB of $H_j(d)$.

For $\mathbf{j} = (j_1, j_2) \in \mathbb{N}^2$ and $\mathbf{m} = (m_1, m_2) \in \mathbb{N}^2$, let $r_{\mathbf{j},\mathbf{m}} \sim \mathcal{N}(0,1)$ be i.i.d., and let $B_{\mathbf{j}} > 0$ satisfy

$$\sum_{\mathbf{j}\in\mathbb{N}_0^2} B_{\mathbf{j}} D_{j_1}(\mathbf{d}_1) D_{j_2}(\mathbf{d}_2) < \infty. \tag{5.10}$$

We define the Gaussian random variable

$$X := \sum_{\mathbf{j}\in\mathbb{N}_0^2} \sqrt{B_{\mathbf{j}}} \sum_{m_1=1}^{D_{k_1}(\mathbf{d}_1)} \sum_{m_2=1}^{D_{k_2}(\mathbf{d}_2)} r_{\mathbf{j},\mathbf{m}} S^{\mathbf{d}_1}_{j_1,m_1} \cdot S^{\mathbf{d}_2}_{j_2,m_2}.$$

By construction, the covariance operator of X has eigenvectors $S^{\mathbf{d}_1}_{j_1,m_1}, S^{\mathbf{d}_2}_{j_2,m_2}$ with eigenvalues $B_{\mathbf{j}}$, so that

$$\|\mathrm{cov}(X)\|_{L^2(\lambda)\to L^2(\lambda)} = \max_{\mathbf{j}} |B_{\mathbf{j}}|.$$

By conditioning X on $\left(S^{\mathbf{d}_1}_{j_1,m_1} \cdot S^{\mathbf{d}_2}_{j_2,m_2}\right)_{\mathbf{m}=(1,1)}^{(D_{j,1}(\mathbf{d}_1), D_{j_2}(\mathbf{d}_2))}$ for $|\mathbf{j}| \leq n$, we have

$$d_n = \sum_{l=0}^{n} D_{n-l}(\mathbf{d}_1) D_l(\mathbf{d}_2).$$

Furthermore, by Lemma A.3.6 we obtain

$$X^{[n]} = \sum_{|\mathbf{j}|\leq n} \sqrt{B_{\mathbf{j}}} \sum_{m_1=1}^{D_{j_1}(\mathbf{d}_1)} \sum_{m_2=1}^{D_{j_2}(\mathbf{d}_2)} r_{\mathbf{j},\mathbf{m}} S^{\mathbf{d}_1}_{j_1,m_1} \cdot S^{\mathbf{d}_2}_{j_2,m_2}.$$

For $\alpha, C > 0$, let the coefficients satisfy

$$B_{\mathbf{j}} \leq C(1 + |\mathbf{j}|)^{-2\alpha - \mathbf{d}_1 - \mathbf{d}_2}, \tag{5.11}$$

where we note that (5.11) implies (5.10). By Lemma A.3.6 we have

$$\|\mathrm{cov}(X) - \mathrm{cov}(X^{[n]})\|_{L^2(\lambda)\to L^2(\lambda)} \leq C(1+n)^{-2\alpha-\mathbf{d}_1-\mathbf{d}_2}.$$

Counting the functionals one adds in each step we estimate d_n by

$$d_n = \sum_{l=0}^{n} D_{n-l}(\mathbf{d}_1) D_l(\mathbf{d}_2) \leq \sum_{l=0}^{n} c_{\mathbf{d}_1} c_{\mathbf{d}_2} (n-l+1)^{\mathbf{d}_1-1} (l+1)^{\mathbf{d}_2-1} \leq c_{\mathbf{d}_1} c_{\mathbf{d}_2} (1+n)^{\mathbf{d}_1+\mathbf{d}_2-1}.$$

Applying Corollary 5.2.5, we obtain the concentration inequality

$$\mu\left(\|X - X^{[n]}\|_{L^2(\lambda)} \leq \frac{1}{2\sqrt{\alpha}} \sqrt{20 \cdot 2^{\mathbf{d}_1+\mathbf{d}_2-1} C c_{\mathbf{d}_1} c_{\mathbf{d}_2} \max\{1,\tau\} n^{-\alpha}} \right) > 1 - \mathrm{e}^{-\tau}.$$

In [63, Theorem 5.1] the authors have shown an asymptotic almost surely bound $\|X - X^{[n]}\|_{L^2(\lambda)} \leq n^{-\gamma}$ for all $0 < \gamma < \alpha$. In contrast, we obtain the 'limiting' $n^{-\alpha}$ as convergence rate but instead of holding almost surely, our bound is formulated in terms of a concentration inequality with explicit constants.

Chapter 6

Conclusion

In this chapter, I present my concluding perspective. Section 6.1 provides a brief example showing how all the results obtained previously can be combined to derive a convergence rate for a Dirichlet boundary value problem. In Section 6.2, the main results of the thesis are summarized. Finally, in Section 6.3, we address some related questions that remain unsolved.

6.1 Assembling the Parts

As mentioned in Section 1.1, Chapters 3, 4, and 5 cover the results of [1] and [2]. The attentive reader may have noticed that the results of [3] have not yet been included. I will use the setting of [3] to connect the statements of Chapter 3 and Chapter 5.

Ref. [3] considers a Dirichlet boundary value problem. In the context of Gaussian random variables, this corresponds to the equation $LX = Y$, where L is the Dirichlet-Laplace operator. Moreover, Ref. [3] applies the P-greedy onto the right-hand side of the Dirichlet boundary value problem. This leads to a convergence rate of the diagonal of the kernel of the right-hand-side, which as we have established in Lemma A.3.5 can be interpreted as convergence rate of the posterior variance of the right-hand-side (Y). The Dirichlet boundary value problem in [3] is well-posed, which implies that the operator $M : F \to E$ exists and is bounded. By Theorem 3.3.10 we obtain a convergence rate for the posterior variance of the solution. This convergence rate of the posterior variance for the solution implies, by Theorem 5.2.4, a convergence rate for the realizations of the solution. This leads to the following example.

Example 6.1.1. *Let us consider the setting where $T = B_{\mathbb{R}^d}$, $E = C^2(T) \cap C(\overline{T})$, $F = C(T) \times C(\partial T)$, and $L : E \to F$ is defined by $Lu := (\Delta u, u|_{\partial T})$, where Δ denotes the Laplace operator.*

Moreover, let X be a Gaussian random variable with $H_X = H^\alpha(T)$, where $H^\alpha(T)$ is the Sobolev space of order $\alpha > 2 + d/2$. For $Y = LX$, it follows from [3, Theorem 3.6] that there exist functionals $(f_j^*)_{j=1}^n \subset F'$ such that

$$\|\mathrm{cov}(Y) - \mathrm{cov}(Y|(f_j^*(Y))_{j=1}^n)\|_{F' \to F} \leq C n^{1-\frac{2\tau-4}{d}}$$

holds true for some constant $C > 0$ and all $n \in \mathbb{N}$. Furthermore, since L corresponds to the Dirichlet boundary value problem, it is invertible by the maximum principle. Thus, by Example 3.4.4, the operator M exists and is given by L^{-1}, and we have $Z = X$.

Corollary 3.3.11 then implies that for functionals $(f_j')_{j=1}^n \subset F'$ selected via the P-greedy algorithm, we have

$$\|\mathrm{cov}(X) - \mathrm{cov}(X_n)\|_{E' \to E} = \|\mathrm{cov}(X|Y) - \mathrm{cov}(X|Y_n)\|_{E' \to E'} \leq C' n^{1 - \frac{2\tau - 4}{d}},$$

holds true for some $C' > 0$ and all $n \in \mathbb{N}$, where $X_n := \mathbb{E}(X \mid (f_j'(Y))_{j=1}^n)$ and $Y_n := \mathbb{E}(Y \mid (f_j'(Y))_{j=1}^n)$. In the first equality, we used Theorem 3.3.9 and the identity $X = Z$.

Thus, we are in a position to apply Corollary 5.2.7 to obtain concentration inequalities for realizations of X, meaning that

$$\mu\left(\|X - X_n\|_{C(T)} \leq C'' n^{-\frac{\tau - 2}{d} + 1}\right) > 1 - \mathrm{e}^{-\tau}$$

for some $C'' > 0$, all $n \in \mathbb{N}$ and all $\tau > 0$ holds true.

The results of [3] provide a convergence rate for functions within the RKHS, while Example 6.1.1 establishes a convergence rate for a class of functions that lies outside the RKHS. This falls under the category of the escaping-the-native-space problem, see Chapter 4.

6.2 Summary

As mentioned in Section 1.2, Gaussian random variables have a long-standing history and are well established within the mathematical community. In this work, we focus on the conditioning of Gaussian random variables with infinite-dimensional information, which plays a key role in applications such as Bayesian inverse problems and machine learning. Unlike usual settings that rely heavily on finite-dimensional approximations, our focus is on the structural properties of Gaussian random variables in infinite-dimensional spaces. In particular, we aim to understand how conditioning interacts with the underlying functional analytic framework and how this perspective can aid both theory and computation.

Following a brief exposition of the necessary background in Chapter 2, the first part of our main results in Chapter 3 is of a theoretical nature. We focus on the connection between the different Hilbert spaces (Gaussian Hilbert space, Abstract Wiener space, and Reproducing Covariance space) associated with Gaussian random variables and analyzed how the process of conditioning acts on each of these three spaces. This provides valuable insights, which ultimately led us to interpret the conditioning process as a mapping defined by a linear operator M. Our main result in this chapter, Theorem 3.3.10, uses the assumption of boundedness on M to derive convergence rates for the posterior variance of the conditioned random variable. Moreover, we formulated the result in a general way so that it applies not only to the P-greedy algorithm but also to any suitable functional selection algorithm. Additionally, we ensured that our results remain valid for arbitrary separable Banach spaces.

Chapter 4 did not focus on Gaussian random variables but instead on kernels, by addressing the escaping-the-native-space question. The main result of this chapter, Theorem 4.1.3, provides a positive result for scalable RKHSs. It states, in simple terms, that even when the target function lies outside the RKHS, we can still expect a meaningful convergence rate. This is notable, because the assumption that the target function lies in the RKHS is often unverifiable or not satisfied in practice, thereby offering an explanation for the observed robustness of kernel methods, although we only treated power spaces as a specific example in that chapter.

Lastly and building upon the results of Chapter 4, in Chapter 5, we derived a concentration inequality that allows us to bound the norm of the Gaussian random variable X in scalable RKHSs. We combined this insight with Theorem 4.1.3 to obtain our main result in this chapter, Theorem 5.2.4. In essence, it establishes that the convergence rate of the posterior variance of a conditioned Gaussian random variable implies a corresponding concentration inequality for its realizations. We then compared our result to the main results of [54] and demonstrated that there are cases where Theorem 5.2.4 yields a sharper concentration inequality than the one established in [54]. Moreover, we formulated our result so that the concentration inequality is expressed in terms of the posterior variance, thus the results of Chapter 3 directly imply a concentration inequality for the realizations of Gaussian random variables.

6.3 Outlook

In mathematics, it is quite common that answering one question leads to several new ones. In this section, we discuss the questions that arose during the development of this thesis and that remain open.

The results in Chapter 3 relied heavily on the assumption that there exists a bounded operator $M : F \to E$ such that $MY = Z$. A natural question is: Under which conditions does such an operator M exist? More precisely, consider the setting where $Y = LX + N$ for some bounded operator $L : E \to F$ and a Gaussian random variable N that is independent of X. Can we classify the conditions on X, N, L, E, and F under which M is a bounded operator?

One possible approach to tackle this question is to adapt techniques from partial differential equation (PDE) theory, as the question is similar to the classical well-posedness problem in that field. To demonstrate the well-posedness of a PDE, several key theorems such as the Lax-Milgram Theorem or the Fredholm alternative are typically used [65, Chapter 6]. These theorems then need to be adapted to the setting of Gaussian random variables, translating assumptions like ellipticity and weak solutions in Sobolev spaces into conditions on X, L, N, E, and F.

In Chapter 4, we answered the escaping-the-native-space question for elements of the scaled RKHS. However, these scaled RKHSs are not well understood in comparison to power spaces or Sobolev spaces. It is unclear how much larger the scaled RKHS is compared to the original RKHS. Moreover, the scaled RKHS depends heavily on the choice of the ONB, depending on this choice, the scaled RKHS can vary significantly. While we demonstrated in Chapter 5 that realizations of Gaussian random variables can lie in scaled RKHSs, it remains unclear whether there are larger classes of functions that are contained in the scaled RKHSs. Conversely, one can also ask whether there are smaller classes of functions that are included in the scaled RKHSs.

To resolve this issue, one can analyze the ONB for specific properties and, depending on which properties are satisfied, guarantee that the scaled RKHS is contained in a larger space. For the reverse, direction, checking whether a space $\tilde{H}$ is contained in the scaled RKHS, assume additionally that the ONB can be identified with a set of functionals (λ_j) that can be applied to a larger space. Then, for all $f \in \tilde{H}$, one can consider the sequence $(\lambda_j(f))$. Depending on specific properties of this sequence, we can determine whether $\tilde{H}$ is contained in the scaled RKHS.

The results in [54] also apply in the misspecified case, meaning when the covariance of X is unknown. For Chapter 5, it is unclear whether the results remain valid in such a setting. One possible approach to address this is to investigate under which conditions $X \in H_{a,V,d}$ holds when H is not the RKHS associated with X, and whether the corresponding norm can still be computed explicitly.

Bibliography

[1] Daniel Winkle, Ingo Steinwart, and Bernard Haasdonk. Convergence analysis of a greedy algorithm for conditioning Gaussian random variables. *arXiv:2502.10772*, 2025.

[2] Daniel Winkle, Ingo Steinwart, and Bernard Haasdonk. Convergence rates for realizations of Gaussian random variables. *arXiv:2508.13940*, 2025.

[3] Tizian Wenzel, Daniel Winkle, Gabriele Santin, and Bernard Haasdonk. Adaptive meshfree approximation for linear elliptic partial differential equations with PDE-greedy kernel methods. *BIT Numerical Mathematics*, 65, 2025.

[4] Stephen M. Stigler. *The history of statistics*. The Belknap Press of Harvard University Press, 1990.

[5] Norbert Wiener. Differential-space. *J. Math. and Phys.*, 2:131–174, 1923.

[6] A. N. Kolmogorov. *Foundations of the theory of probability: Second English Edition*. Courier Dover Publications, 2018.

[7] Thomas J. Santner, Brian J. Williams, and William I. Notz. *The design and analysis of computer experiments*. Springer, 2003.

[8] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian processes for machine learning*. MIT Press, 2006.

[9] Michael L. Stein. *Interpolation of spatial data*. Springer, 1999.

[10] Zoubin Ghahramani. Bayesian non-parametrics and the probabilistic approach to modelling. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 2013.

[11] Marc C. Kennedy and Anthony O'Hagan. Bayesian calibration of computer models. *J. R. Stat. Soc. Ser. B*, 2001.

[12] Havard Rue and Leonhard Held. *Gaussian Markov random fields: theory and applications*. Chapman and Hall/CRC, 2005.

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

[14] Elias M. Stein. *Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals*. Princeton University Press, 1993.

[15] Motonobu Kanagawa, Philipp Hennig, Dino Sejdinovic, and Bharath K Sriperumbudur. Gaussian processes and kernel methods: A review on connections and equivalences. *arXiv preprint arXiv:1807.02582*, 2018.

[16] William Fulton and Joe Harris. *Representation theory*. Springer-Verlag, 1991.

[17] John B Conway. *A course in functional analysis*, volume 96. Springer, 2019.

[18] Svante Janson. *Gaussian Hilbert spaces*. Cambridge University Press, 1997.

[19] Daniel W. Stroock. *Probability theory: An Analytic View*. Cambridge University Press, 2010.

[20] Mikhail Sh. Birman and Mikhail Z. Solomjak. *Spectral theory of selfadjoint operators in Hilbert space*. D. Reidel Publishing Co., 1987.

[21] Yifan Chen, Bamdad Hosseini, Houman Owhadi, and Andrew M. Stuart. Solving and learning nonlinear PDEs with Gaussian processes. *J. Comput. Phys.*, 2021.

[22] Huiyan Sang and Jianhua Z. Huang. A full scale approximation of covariance functions for large spatial data sets. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 2012.

[23] Toni Karvonen and Chris J Oates. Maximum likelihood estimation in Gaussian process regression is ill-posed. *JMLR*, 24, 2023.

[24] Houman Owhadi and Clint Scovel. Conditioning Gaussian measure on Hilbert space. *arXiv preprint arXiv:1506.04208*, 2015.

[25] Cédric Travelletti and David Ginsbourger. Disintegration of Gaussian measures for sequential assimilation of linear operator data. *Electron. J. Stat.*, 2024.

[26] Ingo Steinwart. Conditioning of Banach Space Valued gaussian random variables: An approximation approach based on martingales. *arXiv:2404.03453*, 2024.

[27] Marvin Pförtner, Jonathan Wenger, Jon Cockayne, and Philipp Hennig. Computation-Aware Kalman filtering and smoothing. *arXiv:2405.08971*, 2024.

[28] Helmut Harbrecht, Michael Multerer, Olaf Schenk, and Christoph Schwab. Multiresolution kernel matrix algebra. *Numer. Math.*, 156:1085–1114, 2024.

[29] Michael F. Driscoll. The reproducing kernel Hilbert space structure of the sample paths of a Gaussian process. *Z. Wahrsch. verw. Geb.*, 26:309–316, 1973.

[30] Emanuel Parzen. Probability density functionals and reproducing kernel Hilbert spaces. *Proc. Symp. Time Ser. Anal.*, pages 155–169, 1963.

[31] Holger Wendland. *Scattered data approximation*. Cambridge University Press, 2004.

[32] Francis J. Narcowich, Joseph D. Ward, and Holger Wendland. Sobolev error estimates and a Bernstein inequality for scattered data interpolation via radial basis functions. *Constr. Approx.*, 24:175–186, 2006.

[33] Donald L Cohn. *Measure Theory.* Springer, 2013.

[34] Mark Veraar Tuomas Hytönen, Jan Van Neerven and Lutz Weis. *Analysis in Banach Spaces: Volume I: Martingales and Littlewood-Paley Theory.* Springer, 2016.

[35] Nicholas Vakhania, Vazha Tarieladze, and Sergei Chobanyan. *Probability distributions on Banach Spaces.* D. Reidel Publishing Co., 1987.

[36] Svante Janson and Sten Kaijser. *Higher moments of Banach space valued random variables*, volume 238. American Mathematical Society, 2015.

[37] Aris Spanos. *Probability theory and statistical inference: Empirical modeling with observational data.* Cambridge University Press, 2019.

[38] Ingo Steinwart and Andreas Christmann. *Support Vector Machines.* Springer, 2008.

[39] Ronald DeVore, Guergana Petrova, and Przemyslaw Wojtaszczyk. Greedy algorithms for reduced bases in Banach spaces. *Constr. Approx.*, 37(3):455–466, 2013.

[40] Gabriele Santin and Bernard Haasdonk. Convergence rate of the data-independent P-greedy algorithm in kernel-based approximation. *Dolomites Res. Notes Approx.*, 10:68–78, 2017.

[41] Tizian Wenzel, Gabriele Santin, and Bernard Haasdonk. Analysis of target data-dependent greedy kernel algorithms: convergence rates for f-, $f \cdot P$-, and f/P-greedy. *Constr. Approx.*, 2023.

[42] Toni Karvonen. Small sample spaces for Gaussian processes. *Bernoulli*, 29:875–900, 2023.

[43] Ingo Steinwart and Clint Scovel. Mercer's theorem on general domains: on the interaction between measures, kernels, and RKHSs. *Constr. Approx.*, 35:363–417, 2012.

[44] Francis Bach. Information theory with kernel methods. *IEEE Transactions on Information Theory*, 69(2):752–775, 2022.

[45] Aad W. Van Der Vaart and Harry Van Zanten. Information rates of nonparametric Gaussian process methods. *J. Mach. Learn. Res.*, 12:2095–2119, 2011.

[46] Achim Klenke. *Probability theory: a comprehensive course.* Springer, 2008.

[47] Michael Reed, Barry Simon, Barry Simon, and Barry Simon. *Methods of modern mathematical physics*, volume 1. Elsevier, 1972.

[48] Jacques Louis Lions and Enrico Magenes. *Non-homogeneous boundary value problems and applications: Volume 1.* Springer, 2012.

[49] Robert E Megginson. *An introduction to Banach space theory*, volume 183. Springer Science & Business Media, 2012.

[50] Christian Rieger and Barbara Zwicknagl. Sampling inequalities for infinitely smooth functions, with applications to interpolation and machine learning. *Advances in Computational Mathematics*, 2010.

[51] Jiu Ding and Liang Jiao Huang. On the perturbation of the least squares solutions in hilbert spaces. *Linear Algebra and its Applications*, 212:487–500, 1994.

[52] Milan N. Lukić and Jay H. Beder. Stochastic processes with sample paths in reproducing kernel Hilbert spaces. *Trans. Amer. Math. Soc.*, 2001.

[53] Toni Karvonen, George Wynne, Filip Tronarp, Chris Oates, and Simo Sarkka. Maximum likelihood estimation and uncertainty quantification for Gaussian process approximation of deterministic functions. *SIAM/ASA J. on Uncertain. Quantif.*, 8:926–958, 2020.

[54] Wenjia Wang, Rui Tuo, and C. F. Jeff Wu. On prediction properties of Kriging: uniform error bounds and robustness. *J. Amer. Statist. Assoc.*, 115:920–930, 2020.

[55] Robert J. Adler and Jonathan E. Taylor. *Random fields and geometry.* Springer, 2007.

[56] Kandethody M. Ramachandran and Chris P. Tsokos. *Mathematical statistics with applications in R.* Academic Press, 2020.

[57] Beatrice Laurent and Pascal Massart. Adaptive estimation of a quadratic functional by model selection. *Ann. Stat.*, 28:1302–1338, 2000.

[58] Roland Opfer. Tight frame expansions of multiscale reproducing kernels in Sobolev spaces. *Appl. Comput. Harmon. Anal.*, 20:357–374, 2006.

[59] Peter G. Casazza and Gitta Kutyniok. *Finite frames.* Birkhäuser, 2013.

[60] Ari Solin and Samppa Särkkä. Practical Hilbert space approximate Bayesian Gaussian processes for probabilistic programming. *Stat. Comput.*, 30:411–426, 2020.

[61] Ingo Steinwart. When does a Gaussian process have its paths in a reproducing kernel Hilbert space? *arXiv:2407.11898*, 2024.

[62] Werner Linde, Vazha I. Tarieladze, and Sergei A. Chobanyan. Characterization of certain classes of Banach spaces by properties of Gaussian measures. *Theory Probab. Appl.*, 25:159–164, 1980.

[63] Alfredo Alegría, Galatia Cleanthous, Athanasios G. Georgiadis, Emilio Porcu, and Philip A. White. Gaussian random fields on the product of spheres: Theory and applications. *Electron. J. Stat.*, 18:1394–1435, 2024.

[64] George E. Andrews, Richard Askey, and Ranjan Roy. *Special functions.* Cambridge University Press, 1999.

[65] Lawrence C. Evans. *Partial differential equations.* American Mathematical Society, Providence, 2010.

[66] Pierpaolo Natalini and Biagio Palumbo. Inequalities for the incomplete gamma function. *Math. Inequal. Appl.*, 3:69–77, 2000.

[67] Lucien Birgé and Pascal Massart. Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli*, 4:329–375, 1998.

[68] Walter Rudin. *Functional Analysis*. McGraw-Hill, 2nd edition, 1991.

[69] Friedrich Wille. Galerkins Lösungsnäherungen bei monotonen Abbildungen. *Mathematische Zeitschrift*, 127:10–16, 1972.

Appendices

Appendix A

Calculations

A.1 Statements for Chapter 3

Lemma A.1.1. *Let* $(r_j) \sim \mathcal{N}(0,1)$ *be a sequence of i.i.d. random variables, then*

$$r_j \cdot \left(\sqrt{\ln(j \ln^2(j+1))}\right)^{-1}$$

is a bounded sequence μ*-almost everywhere.*

Proof. We define

$$E_j := \left\{ \frac{|r_j|}{\sqrt{\ln(j \ln^2(j+1))}} > \sqrt{2} \right\}$$

and $E := \limsup_{j\to\infty} E_j$. By the Borel-Cantelli lemma, we have

$$P(E) = 0 \quad \text{if} \quad \sum_{j=1}^{\infty} P(E_j) < \infty.$$

Using a well known tail bound for one dimensional standard normal random variables we find $P(|r_j| > t) \leq 2\mathrm{e}^{-t^2/2}$ for all $t > 0$. We thus obtain

$$\begin{aligned} \sum_{j=1}^{\infty} P(E_j) = \sum_{j=1}^{\infty} P\left(|r_j| > \sqrt{2} \cdot \sqrt{\ln(j \ln^2(j+1))}\right) &\leq \sum_{j=1}^{\infty} 2\mathrm{e}^{-\left(2 \cdot \sqrt{\ln(j \ln^2(j+1))}^2 /2\right)} \\ &= \sum_{j=1}^{\infty} 2\mathrm{e}^{-\left(2 \cdot \ln(j \ln^2(j+1))/2\right)} \\ &= \sum_{j=1}^{\infty} \frac{2}{j \ln^2(j+1)} . \end{aligned}$$

The latter series converges, thus $P(E) = 0$. □

Lemma A.1.2. *Let $a, b, c \in \mathbb{R}$ such that $a < b$ and $c^2 + a > 0$. Moreover let $k : [a,b] \times [a,b] \to \mathbb{R}$ be the function given by*

$$k(t,s) := c^2 + \min(t,s).$$

Then k is a kernel and the scalar product of the associated RKHS H is given by

$$\langle f, g \rangle_H = \frac{f(a)g(a)}{c^2 + a} + \int_a^b f'(t)g'(t)\,\mathrm{d}t \tag{A.1}$$

for all $f, g \in H$.

Proof. Clearly k is a kernel. Let us consider the set $H = H^1([a,b])$, equipped with the scalar product from (A.1). Because $\|\cdot\|_H$ is induced by a scalarproduct and $H \subseteq C([a,b])$ we have that $(H, \|\cdot\|_H)$ is an RKHS. Consequently, it suffices to show that k is the reproducing kernel of H. To this end, we pick an $f \in H$. For $s \in [a,b]$, we then have

$$\langle f, c^2 + \min(s, \cdot\,) \rangle_H = \frac{f(a)(c^2+a)}{c^2+a} + \int_a^s f'(t)\,\mathrm{d}t = f(a) + f(s) - f(a) = f(s),$$

we conclude $k(s, \cdot\,) \in H$ for all $s \in [a,b]$. □

A.2 Statements for Chapter 4

We note that the following result is also needed for Chapter 5.

Lemma A.2.1. *Given constants $C, C_d > 0$, $\alpha > \gamma > 1 + \beta \geq 1$ and sequences $a_j = j^\gamma$, $c_j = C(j+1)^{-\alpha}$ and $d_j = C_d(j+1)^\beta$, we have*

$$\sum_{j=n+1}^{\infty} a_j(c_{j-1} - c_j) \leq C\alpha \frac{n^{\gamma-\alpha}}{\alpha - \gamma}$$

and

$$\sum_{j=n+1}^{\infty} \frac{d_j}{a_j} \leq C_d 2^\beta \frac{n^{\beta+1-\gamma}}{\gamma - \beta - 1}$$

for all $n \in \mathbb{N}$. Additionally, we have $a_j \cdot c_j \to 0$.

Proof. The condition $\beta + 1 < \gamma$ implies $(d_j/a_j) \in \ell^1(\mathbb{N})$. We note (a_j) is monotonically increasing. Furthermore, $\gamma < \alpha$ implies $a_j c_j \to 0$.

Next, we show that $(Cj^{-\gamma}(j^{-\alpha} - (j+1)^{-\alpha})) \in \ell^1(\mathbb{N})$. To this end we define $f : (0,\infty) \to \mathbb{R}$ by $f(t) := C(t+1)^{-\alpha}$. By the mean value theorem there exist $\xi_j \in [j-1, j]$ such that $c_{j-1} - c_j = -f'(\xi_j) = C\alpha(\xi_j+1)^{-\alpha-1} \leq C\alpha j^{-\alpha-1}$. Thus $(a_j(c_{j-1} - c_j)) \in \ell^1(\mathbb{N})$ is equivalent to $\gamma - \alpha - 1 < -1$ meaning $\gamma < \alpha$. In other words the assumptions of Theorem 4.1.3 are satisfied.

Now we estimate

$$\sum_{j=n+1}^{\infty} a_j(c_{j-1} - c_j) = \sum_{j=n+1}^{\infty} -a_j f'(\xi_j) = \sum_{j=n+1}^{\infty} Cj^\gamma \alpha(\xi_j+1)^{-\alpha-1} \leq C\alpha \sum_{j=n+1}^{\infty} j^{\gamma-\alpha-1}.$$

Using that $j^{\gamma-\alpha-1}$ is decreasing, we bound the sum by an integral

$$\sum_{j=n+1}^{\infty} j^{\gamma-\alpha-1} \leq \int_n^{\infty} t^{\gamma-\alpha-1}\mathrm{d}t = \left[\frac{t^{\gamma-\alpha}}{\gamma-\alpha}\right]_{t=n}^{\infty} = \frac{n^{\gamma-\alpha}}{\alpha-\gamma}.$$

Note that $d_j = C_d(j+1)^\beta \leq C_d 2^\beta j^\beta$. Using this estimate, we obtain

$$\sum_{j=n+1}^{\infty} \frac{d_j}{a_j} \leq C_d 2^\beta \sum_{j=n+1}^{\infty} j^{\beta-\gamma} \leq C_d 2^\beta \int_n^{\infty} t^{\beta-\gamma}\mathrm{d}t = C_d 2^\beta \left[\frac{t^{\beta+1-\gamma}}{\beta+1-\gamma}\right]_{t=n}^{\infty} = C_d 2^\beta \frac{n^{\beta+1-\gamma}}{\gamma-\beta-1}.$$

□

A.3 Statements for Chapter 5

Lemma A.3.1. *Given a decreasing sequence (c_j) with $c_j > 0$ and $f : [0,\infty) \to [0,\infty)$ a twice differentiable, convex, and monotonically decreasing function, with $f(j) = c_j$ for all $j \in \mathbb{N}$, we have*

$$\sum_{j=n+1}^{\infty} \sqrt{c_{j-1} - c_j} \leq \int_{n-1}^{\infty} \sqrt{|f'(t)|}\,\mathrm{d}t \quad \text{for all } n \geq 1.$$

Proof. This is a consequence of the mean value theorem, we note that we have by the property of f being monotonically decreasing and convex

$$c_{j-1} - c_j = f(j-1) - f(j) \leq \sup_{\xi\in[j-1,j]} -f'(\xi) \leq -f'(j-1) \leq -f'(t)$$

for all $t \in [j-2, j-1]$, thus

$$\sum_{j=n+1}^{\infty} \sqrt{c_{j-1} - c_j} \leq \int_{n-1}^{\infty} \sqrt{|f'(t)|}\,\mathrm{d}t.$$

□

Corollary A.3.2. *Let $C_1, C_2 > 0$, and $d > 1$. For $n > \left(\frac{11}{C_2}(d-1)\right)^d + 1$ we then have*

$$\sum_{j=n+1}^{\infty} \sqrt{C_1 \mathrm{e}^{-C_2(j-1)^{1/d}} - C_1 \mathrm{e}^{-C_2 j^{1/d}}} \leq \frac{11\sqrt{C_1C_2d}}{10}\left(\frac{C_2}{2}\right)^{d-2}(n-1)^{\frac{d-1}{2d}}\mathrm{e}^{-\frac{C_2}{2}(n-1)^{1/d}}.$$

Proof. We apply Lemma A.3.1 with $f(t) := C_1\mathrm{e}^{-C_2 t^{1/d}}$. The derivative of f is given by $f'(t) = -C_2C_1\frac{t^{1/d-1}}{d}\mathrm{e}^{-C_2t^{1/d}}$ and we obtain

$$\sum_{j=n+1}^{\infty} \sqrt{C_1 \mathrm{e}^{-C_2(j-1)^{1/d}} - C_1 \mathrm{e}^{-C_2 j^{1/d}}} \leq \sqrt{C_2C_1}\int_{n-1}^{\infty} \sqrt{\frac{t^{1/d-1}}{d}}\mathrm{e}^{-\frac{C_2}{2}t^{1/d}}\,\mathrm{d}t.$$

Substituting $u = t^{1/d}$ we obtain

$$\int_{n-1}^{\infty} \sqrt{\frac{t^{1/d-1}}{d}} \mathrm{e}^{-\frac{C_2}{2}t^{1/d}} \,\mathrm{d}t = \int_{(n-1)^{1/d}}^{\infty} \sqrt{\frac{u^{1-d}}{d}} \mathrm{e}^{-\frac{C_2}{2}u} du^{d-1} \,\mathrm{d}u = \int_{(n-1)^{1/d}}^{\infty} \sqrt{du^{d-1}} \mathrm{e}^{-\frac{C_2}{2}u} \,\mathrm{d}u.$$

Now substituting $\frac{C_2}{2}u = x$ leads to

$$\int_{(n-1)^{1/d}}^{\infty} \sqrt{du^{d-1}} \mathrm{e}^{-\frac{C_2}{2}u} \,\mathrm{d}u = \sqrt{d} \int_{\frac{C_2}{2}(n-1)^{1/d}}^{\infty} \left(\frac{C_2}{2}x\right)^{\frac{d-1}{2}} \mathrm{e}^{-x} \frac{2}{C_2} \,\mathrm{d}x.$$

This integral matches the incomplete gamma function. Applying [66, Inequality (3.5)], with $B = \frac{11}{10}$, we obtain the following inequality for $C_2(n-1)^{1/d} > 11(d-1)$

$$\sqrt{d\left(\frac{C_2}{2}\right)^{d-3}} \int_{\frac{C_2}{2}(n-1)^{1/d}}^{\infty} x^{\frac{d-1}{2}} \mathrm{e}^{-x} \,\mathrm{d}x \leq \frac{11\sqrt{d}}{10} \left(\frac{C_2}{2}\right)^{d-2} (n-1)^{\frac{d-1}{2d}} \mathrm{e}^{-\frac{C_2}{2}(n-1)^{1/d}},$$

thus the assertion follows. □

Lemma A.3.3. *For* $0 \leq x < 1/2$ *we have*

$$0 \leq -x - 1/2\ln(1-2x) \leq \frac{x^2}{1-2x}.$$

Proof. We start by showing that $f(x) := -x - 1/2\ln(1-2x) \geq 0$. We note that $f(0) = 0$ and $f'(x) = -1 + \frac{1}{1-2x} = \frac{2x}{1-2x} \geq 0$ for $x \in [0, 1/2)$, thus we conclude $f(x) \geq 0$.

Next we prove $g(x) := \frac{x^2}{1-2x} + x + 1/2\ln(1-2x) \geq 0$. We note that $g(0) = 0$ and $g'(x) = \frac{2x^2}{(1-2x)^2} \geq 0$ for $x \in [0, 1/2)$, thus the assertion follows. □

Lemma A.3.4. *Let* $Z : \Omega \to \mathbb{R}$ *be a random variable with* $\mathbb{E}(Z) = 0$*. If for fixed* $v, c > 0$*, and all* $0 < x < 1/c$ *we have*

$$\ln\left(\mathbb{E}\mathrm{e}^{xZ}\right) \leq \frac{vx^2}{2(1-cx)},$$

then for any $\tau > 0$*, we have*

$$\mu(Z \geq c\tau + \sqrt{2v\tau}) \leq \mathrm{e}^{-\tau}.$$

Proof. This proof is taken from [67, Lemma 8], we also add all the necessary technical details. We use the well known Markov inequality and obtain with the assumptions for all $x \geq 0$

$$\mu\left(Z \geq \varepsilon\right) = \mu\left(\mathrm{e}^{xZ} \geq \mathrm{e}^{x\varepsilon}\right) \leq \mathbb{E}\mathrm{e}^{xZ}\mathrm{e}^{-x\varepsilon} = \mathrm{e}^{-x\varepsilon + \ln\left(\mathbb{E}\mathrm{e}^{xZ}\right)} \leq \mathrm{e}^{-x\varepsilon + \frac{vx^2}{2(1-cx)}}$$

for all $\varepsilon > 0$. We define $h_\varepsilon : (0, \infty) \to \mathbb{R}$ by

$$h_\varepsilon(x) := \left(x\varepsilon - \frac{vx^2}{2(1-cx)}\right).$$

Next we show that the function $h(x)$ is maximized at $x^* := c^{-1}\left(1 - \sqrt{v/(2\varepsilon c + v)}\right) < 1/c$. To this end we note that

$$h'_\varepsilon(x) = \varepsilon - \frac{vx(2-cx)}{2(1-cx)^2}.$$

Moreover, the denominator at x^* is

$$2(1-cx^*)^2 = \frac{2v}{2\varepsilon c + v},$$

while the enumerator is

$$vx^*(2-cx^*) = \frac{v}{c}\left(1 - \sqrt{\frac{v}{2\varepsilon c + v}}\right)\left(1 + \sqrt{\frac{v}{2\varepsilon c + v}}\right) = \frac{2\varepsilon v}{2\varepsilon c + v}.$$

Inserting both equations in h'_ε gives $h'_\varepsilon(x^*) = \varepsilon - \varepsilon = 0$. To verify that $h''_\varepsilon(x^*) < 0$, we note that the second derivative is given by

$$h''_\varepsilon(x) = \frac{-v}{(1-cx)^3}.$$

Inserting x^* gives

$$h''_\varepsilon(x^*) = -v\left(\frac{v}{2\varepsilon c + v}\right)^{-3/2} < 0.$$

Thus x^* maximizes h_ε.

In the next step we calculate $h_\varepsilon(x^*)$. To this end we set

$$t := \sqrt{\frac{v}{2\varepsilon c + v}}$$

and note that $x^* = \frac{1-t}{c}$ and $1 - cx^* = t$ holds true. We thus obtain

$$h_\varepsilon(x^*) = \varepsilon\frac{1-t}{c} - \frac{v}{2}\frac{(1-t)^2}{tc^2} = \frac{2c\varepsilon(1-t)t}{2c^2t} - \frac{v(1-t)^2}{2c^2t} = \frac{(1-t)(2\varepsilon ct - v(1-t))}{2c^2t}.$$

Since we have $(2\varepsilon c + v)t = v/t$ we conclude

$$2\varepsilon ct - v(1-t) = (2\varepsilon c + v)t - v = v\frac{1-t}{t}.$$

It follows that

$$h_\varepsilon(x^*) = \frac{(1-t)(v(1-t)/t)}{2c^2t} = \frac{v(1-t)^2}{2c^2t^2}.$$

For $u := 1/t = \sqrt{1 + \frac{2\varepsilon c}{v}}$ we now have

$$\frac{(1-t)^2}{t^2} = \left(\frac{1}{t} - 1\right)^2 = (u-1)^2,$$

and with $u^2+1=\frac{2v+2\varepsilon c}{v}$

$$h_\varepsilon(x^*)=\frac{v}{2c^2}(u-1)^2=\frac{v(u^2-2u+1)}{2c^2}=\frac{2(\varepsilon c+v)-2vu}{2c^2}=\frac{\varepsilon c+v-vu}{c^2}.$$

Multiplying numerator and denominator by $\varepsilon c+v+vu$ gives

$$h_\varepsilon(x^*)=\frac{(\varepsilon c+v-vu)(\varepsilon c+v+vu)}{c^2(\varepsilon c+v+vu)}=\frac{(\varepsilon c+v)^2-(vu)^2}{c^2(\varepsilon c+v+vu)}.$$

Moreover, we have

$$(\varepsilon c+v)^2-(vu)^2=\varepsilon^2c^2+2\varepsilon cv+v^2-v^2\left(1+\frac{2\varepsilon c}{v}\right)=\varepsilon^2c^2$$

and hence

$$h_\varepsilon(x^*)=\frac{\varepsilon^2c^2}{c^2(\varepsilon c+v+vu)}=\frac{\varepsilon^2}{\varepsilon c+v+v\sqrt{1+\frac{2\varepsilon c}{v}}}.$$

We now show that for $\varepsilon:=c\tau+\sqrt{2v\tau}$ we have $h_{c\tau+\sqrt{2v\tau}}(x^*)=\tau$. To this end we note that

$$\varepsilon^2=\tau(\tau c^2+2c\sqrt{2v\tau}+2v)\quad\text{and}\quad c\varepsilon+v=\tau c^2+c\sqrt{2v\tau}+v.$$

This gives

$$v\sqrt{1+\frac{2\varepsilon c}{v}}=\sqrt{v^2+2\varepsilon cv}=\sqrt{c^2v\tau+2cv\sqrt{2v\tau}+v^2}=\sqrt{(v+c\sqrt{2v\tau})^2}=v+c\sqrt{2v\tau},$$

and thus we conclude

$$h_{c\tau+\sqrt{2v\tau}}(x^*)=\frac{\tau(\tau c^2+2c\sqrt{2v\tau}+2v)}{\tau c^2+c\sqrt{2v\tau}+v+v+c\sqrt{2v\tau}}=\frac{\tau(\tau c^2+2c\sqrt{2v\tau}+2v)}{\tau c^2+2c\sqrt{2v\tau}+2v}=\tau.$$

In summary we obtain

$$\mu(Z\geq c\tau+\sqrt{2v\tau})\leq \mathrm{e}^{-\tau}.$$

□

Lemma A.3.5. *If $E=C(T)$ and X is an E-valued Gaussian random variable, we have*

$$\sup_{t\in T}|k_X(t,t)-k_{X^{[n]}}(t,t)|=\|\mathrm{cov}(X)-\mathrm{cov}(X^{[n]})\|_{C(T)'\to C(T)}.$$

Proof. We set $Z=X-X^{[n]}$ and show that

$$\sup_{t\in T}|k_Z(t,t)|=\|\mathrm{cov}(Z)\|_{C(T)'\to C(T)}$$

holds true for all E-valued Gaussian random variables Z. First, recall that

$$\sup_{t\in T}|k_Z(t,t)| = \sup_{t,s\in T}|k_Z(t,s)| = \sup_{t,s\in T}\langle \operatorname{cov}(Z)\delta_t,\delta_s\rangle_{C(T),C(T)'},$$

where $\delta_t \in C(T)'$ denotes the point evaluation at t. Furthermore, by the Krein-Milman theorem we obtain $\overline{\operatorname{aco}\{\delta_t\,|\,t\in T\}}^{w^*} = B_{C(T)'}$, see for instance [68, Theorem 3.23]. Moreover, by [26, Theorem B.3] the kernel k_Z is w^*-continuous and thus we find

$$\sup_{t,s\in T}\langle \operatorname{cov}(Z)\delta_t,\delta_s\rangle_{E,E'} = \sup_{e_1',e_2'\in B_{C(T)'}}\langle \operatorname{cov}(Z)e_1',e_2'\rangle_{E,E'}.$$

In summary we obtain

$$\sup_{t\in T}|k_Z(t,t)| = \|\operatorname{cov}(Z)\|_{C(T)'\to C(T)}.$$

By Lemma 3.2.16, we have

$$k_Z = \operatorname{cov}(X - X^{[n]}) = \operatorname{cov}(X) - \operatorname{cov}(X^{[n]}) = k_X - k_{X^{[n]}}.$$

□

Lemma A.3.6. *Let X be a Gaussian random variable, $E = L^2(\lambda)$ and denote by λ_j the monotone decreasing eigenvalues of $\operatorname{cov}(X): L^2(\lambda)\to L^2(\lambda)$. Furthermore, we denote by H_j the eigenspace associated to λ_j, we write $d_j := \dim(H_j)$, and let $(e_{j,m})_{m=1}^{d_j} \subseteq H_j$ be an ONB of H_j. For $(e_{j,m}') = \langle e_{j,m},\,\cdot\,\rangle \in L^2(\lambda)'$ we then have*

$$X^{[n]} = \sum_{j=1}^{n}\sqrt{\lambda_j}\sum_{m=1}^{d_j} r_{j,m}e_{j,m} \quad \text{and} \quad \|\operatorname{cov}(X)-\operatorname{cov}(X^{[n]})\|_{L^2(\lambda)\to L^2(\lambda)} = \lambda_{n+1}.$$

Proof. Because $\operatorname{cov}(X)$ is a symmetric, nuclear operator, see [62], we can write

$$\operatorname{cov}(X)f = \sum_{j=1}^{\infty}\lambda_j\sum_{m=1}^{d_j}\langle f, e_{j,m}\rangle_{L^2(\lambda)}e_{j,m}, \quad \text{for } f\in L^2(\lambda).$$

By Example 3.4.8 we know that $(\sqrt{\lambda_j}e_{j,m}) \subset H_X$ is an ONB and

$$h_{j,m} := \int_\Omega e_{j,m}'(X)\iota_{B_{E''}}X\,\mathrm{d}\mu = \operatorname{cov}(X)e_{j,m} = \lambda_j e_{j,m}.$$

Consequently, Assumption V is satisfied and Lemma 5.2.2 thus shows

$$X^{[n]} = \sum_{j=1}^{n}\sqrt{\lambda_j}\sum_{m=1}^{d_j} r_{j,m}e_{j,m}.$$

We conclude

$$\begin{aligned}\sup_{\|f\|_{L^2(\lambda)}\le 1}\|\operatorname{cov}(X)-\operatorname{cov}(X^{[n]})\|_{L^2(\lambda)} &= \sup_{\|f\|_{L^2(\lambda)}\le 1}\left\|\sum_{j=n+1}^{\infty}\lambda_j\sum_{m=1}^{d_j}\langle f, e_{j,m}\rangle_{L^2(\lambda)}e_{j,m}\right\|_{L^2(\lambda)}\\ &= \max_{j\ge n+1}\lambda_j.\end{aligned}$$

Because $\lambda_j > \lambda_{j+1} > 0$, we conclude the assertion. □

Appendix B

Auxiliary Theorems

B.1 Theorems for Chapter 3

Lemma B.1.1. *Let E be a Banach space and $\mathcal{D} \subset B_{E'}$ a weak-$*$-dense subset. Then for all $e_1, e_2 \in E$ with $e'(e_1) = e'(e_2)$ for all $e' \in \mathcal{D}$, we have $e_1 = e_2$.*

Proof. Let $e' \in B_{E'}$ then there exists a net $e'_\alpha \subseteq \mathcal{D}$ with $\langle e'_\alpha, e_j \rangle_{E',E} \to \langle e', e_j \rangle_{E',E}$ for $j = 1, 2$. This shows by our assumption that $e'(e_1) = e'(e_2)$. Now a simple application of Hahn-Banach Theorem gives the assertion. □

The next Lemma can be found in [49, Theorem 2.6.18 and 2.6.23].

Lemma B.1.2. *Let E be a separable Banach space. Then $B_{E'}$ endowed with the weak-$*$-topology is a compact metrizable space and there exists a countable, weak-$*$-dense subset $\mathcal{D} \subseteq B_{E'}$.*

Lemma B.1.3. *Let H_1, H_2 be Hilbert spaces, $\Pi_1 : H_1 \to H_1$ be an orthogonal projection, and $U : H_1 \to H_2$ be an isometric isomorphism. Then $\Pi_2 := U\Pi_1 U^*$ is an orthogonal projection in H_2 with $U(\mathrm{ran}(\Pi_1)) = \mathrm{ran}(\Pi_2)$.*

Proof. Using [17, Proposition II.3.3], we note that we only have to prove that Π_2 is idempotent and self adjoint. We first note that Π_2 is idempotent since

$$\Pi_2^2 = U\Pi_1 U^* U\Pi_1 U^* = U\Pi_1^2 U^* = U\Pi_1 U^* = \Pi_2 \,.$$

Moreover, we also have

$$\Pi_2^* = (U\Pi_1 U^*)^* = ((U^*)^* \Pi_1^* U^*) = U\Pi_1 U^* = \Pi_2 \,,$$

i.e. Π_2 is self-adjoint. The last assertion is trivial. □

Declaration

I, *Daniel Winkle*, declare herewith, that this work with the title

Conditioning of
Gaussian Random Variables

is my own work. I declare that I am the sole author of this dissertation and that I have not reproduced, without acknowledgement, the work of another. I also declare that the electronic copy of this thesis coincides with the physical ones.

Place, Date

Daniel Winkle

Danksagung

Hallo, mein Name ist Daniel Winkle und Sie lesen meine Danksagung.

In diesem Abschnitt, werde ich mich nun bei all denen bedanken, die zu dieser Dissertation direkt oder indirekt beigetragen haben.

Ich bedanke mich bei Ingo Steinwart für seine Geduld bei den wiederholten Iterationen beim Schreiben der Paper, da ich vermutlich öfter als einmal die gleichen Fehler gemacht habe. Insbesondere danke ich ihm für seine handschriftlichen Korrekturen, auch wenn diese eher altägyptischen Hieroglyphen gleichen als dem lateinischen Alphabet.

Mein Dank gilt auch Bernard Haasdonk, dafür, dass er mir die Freiheit gegeben hat, die Mathematik so anzugehen, wie ich es für richtig gehalten habe, auch wenn ich in die ein oder andere mathematische Sackgasse geraten bin.

Bei meinen beiden Betreuern hat es mich besonders gefreut, dass sie meine mathematischen Fähigkeiten anerkennen und wertschätzen, danke sehr.

Weiterhin bedanke ich mich bei Elke Maurer für die netten Klatsch-und-Tratsch-Stunden im Sekretariat und für die Hilfe im Kampf gegen die Dienstreiseanträge, ich sehe diesbezüglich noch keinen Sieg. Insbesondere danke ich auch Elke Gangl, die mir besonders in den letzten zwei Wochen vor der Abgabe geholfen hat, alle bürokratischen Hürden zu bewältigen.

Da ich mehr als nur ein Computerproblem hatte, bedanke ich mich außerdem bei Jörg Hörner, der immer schnell erreichbar war und stets eine Lösung parat hatte.

Ich bedanke mich bei den Korrekturlesern dieser Arbeit. Besonders hervorheben möchte ich meinen guten Freund Jann van der Meer, der mir bereits während meiner Bachelor- und Masterstudentenzeit geholfen hat und sich auch dieses Mal die Zeit genommen hat, mir bei präziseren Formulierungen in der Dissertation zu helfen, trotz der Tatsache, dass er sich momentan auf der anderen Seite der Welt befindet.

Ein weiterer Korrekturleser, den ich namentlich erwähnen und bei dem ich mich bedanken möchte, ist mein ISA-Kollege Michael Bitzer, der sein Büro zum inoffiziellen Pausenraum erklärt hat. Dort fanden einige spannende mathematische Unterhaltungen statt, die wesentlich zu diversen Theoremen beigetragen haben. Unter anderem mit meinen weiteren ISA-Kollegen Max Schölpple und Aleksandar Arsenijevic. Besonders bei Aleks bedanke ich mich dafür, dass er mir sehr viel Arbeit in der Lehre abgenommen hat, sodass ich mich auf meine Dissertation konzentrieren konnte.

Ein weiterer "Pausenraum", bei dem ich mich bedanken möchte, ist die Fachgruppe Mathematik, wo ich oft Zeit verbracht habe, wenn ich bei einem Beweis nicht weiterkam.

Ich bedanke mich bei Tizian Wenzel dafür, dass er die gleiche Leidenschaft für Kernmethoden teilt wie ich und für die zahlreichen Diskussionen über Kernapproximation zu jeder Nachtzeit. Des Weiteren hat Tizian mich auf die Stelle in Finnland aufmerksam gemacht, die ich demnächst antreten werde, und mich zu zwei Konferenzen mitgeschleppt, die ich beide als meine bisher interessantesten Konferenzen bezeichne.

Weiterhin bedanke ich mich bei meiner Familie dafür, dass sie mir Kost und Logis zur Verfügung gestellt hat. Dadurch konnte ich mich voll und ganz auf mein Mathematikstudium konzentrieren. Besonders habe ich das leckere Essen meines Vaters Dietmar Winkle geschätzt, und es hat mich sehr gefreut, dass mein Bruder die samstags um 12:00 Uhr Spaghetti Tradition aufrechterhält. Außerdem danke ich Kevin Winkle dafür, dass er mir einen Computer zur Verfügung gestellt hat, nachdem bei meinem die CPU den Geist aufgegeben hatte.

Darüber hinaus war die Promotionszeit nicht immer nur mit Arbeit verbunden. Ich habe währenddessen auch viele schöne Freizeitstunden mit meiner DSA-Truppe verbracht. Besonders hat mich gefreut, dass wir es trotz großer Distanzen geschafft haben, uns regelmäßig zum Spielen zu treffen. Mein Dank gilt außerdem meinen Computerspielfreunden für einige Stunden der Ablenkung, wenn ich einmal wieder ausschließlich an meine Dissertation gedacht habe, ohne diese Pausen wäre ich vermutlich irgendwann verrückt geworden.

Abschließend bedanke ich mich bei Jens Wirth für seine Unterstützung während meiner Zeit im Bachelor und im Master. Er hat mich insbesondere auf folgendes Zitat aufmerksam gemacht, das meinen derzeitigen Gemütszustand treffend beschreibt und als Abschluss dieser Arbeit dienen soll

“Damit ist unser Satz bewiesen, das wollen wir nun froh begießen.” [69]