THE GAUSSIAN RADON TRANSFORM FOR BANACH SPACES

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor in Philosophy

in

The Department of Mathematics

by
Irina Holmes
B. S., Louisiana State University, 2008
M. S., Louisiana State University, 2010
August 2014

For Iuky, who I miss every day

Acknowledgments

I would like to thank my loving husband Taylor for being the foundation of my American Dream, my wonderful parents Dan and Ileana Craciun who instilled in me the love for the written word, for the intangible and for hard work, my terrific parents-in-law Keith and Gail Holmes who made Louisiana my home by opening their hearts to me, and my siblings-in-law Emily, Amy, Daniel and Claire for many years of laughter and togetherness.

I would like to thank the entire LSU Mathematics Department for 10 beautiful years of learning; I would especially like to thank Professor Leonard Richardson for being incredibly supportive of me from my very first day at LSU and for always having a kind word or useful advice, Professor Boris Rubin for teaching me functional analysis the right way, Professor William Adkins for all his help and support, Professor Neal Stoltzfus, whose topology lectures I've always enjoyed tremendously, Professor Hui-Hsiung Kuo for sharing many interesting stories, Professor Lawrence Smolinsky for also being supportive of me throughout my time at LSU, Professor Robert Perlis, whose lectures made me love algebra, and Professor James Oxley for our many conversations and for all his useful advice through the years. I am also grateful to Charlotte Senn, Karen Cooper, Kathy Ulkins and Jeff Sheldon for all their help, hard work, and kindness. I would also like to thank LSU's Center for Computation and Technology, and especially Dr. Gabrielle Allen , Dr. Ed Seidel and Professor Horst Beyer, for introducing me to research as an undergraduate and for offering me the opportunity and support to study in the United States.

I am very grateful for the Louisiana State University Graduate School Dissertation Year Fellowship and the US Department of Education GAANN grant P200A100080, which supported this research. Also, this work is part of a research project covered by NSA grant H98230-13-1-0210.

Finally, I would like to express my deepest gratitude to my adviser, Professor Ambar N. Sengupta, for offering me the best graduate school experience I could have had. I will never be able to thank him enough for how much I've enjoyed our research, for his unconventional lectures which always led me to see things in a light I could not have seen by myself, for teaching me how to think differently and creatively, and for his particular way of never giving me the entire picture, instead subtly guiding me to discover new things on my own. The only disadvantage to having a fantastic adviser is that it makes it very hard to leave graduate school.

Contents

Acknowledgments iv

List of Figures vi

Abstract vii

1 Introduction 1
- 1.1 Gaussian Measures 2
- 1.2 Abstract Wiener Spaces 6

2 The Gaussian Radon Transform 10
- 2.1 Measurably Adapted Sequences 10
- 2.2 Definition of the Gaussian Radon Transform 13
- 2.3 Closed Affine Subspaces 17

3 Properties of the Gaussian Radon Transform 22
- 3.1 A Disintegration of Wiener Measure 29
- 3.2 The Gaussian Radon Transform and Conditional Expectation 34
- 3.3 The Gaussian Radon Transform and the Classical Wiener Space . . . 37
- 3.4 An Inversion Procedure for the Gaussian Radon Transform 47

4 The Gaussian Radon Transform and Machine Learning 50
- 4.1 Ridge Regression 51
- 4.2 Probabilistic Interpretations 54
- 4.3 Direct Sums of Abstract Wiener Spaces 63

Bibliography 67

Appendix A Category Theory and the Kolmogorov Existence Theorem 71
- A.1 Kolmogorov's Existence Theorem 71
- A.2 Category Theory 73
- A.3 A Categorical Interpretation of Kolmogorov's Existence Theorem . . 77

Appendix B Permission for Use 84

Vita 85

List of Figures

1.1 The classical Radon transform. 1

3.1 Disintegration theorem 30

4.1 A geometric interpretation of Theorem 4.2. 53

4.2 The point on $(T^*)^{-1}(y)$ closest to the origin. 62

A.1 Axioms of category theory. 73

A.2 An indexed diagram and an indexed cone. 76

A.3 A morphism of indexed cones. 76

A.4 A possible cone based on a Kolmogorov diagram. 78

Abstract

The classical Radon transform can be thought of as a way to obtain the density of an n-dimensional object from its $(n-1)$-dimensional sections in different directions. A generalization of this transform to infinite-dimensional spaces has the potential to allow one to obtain a function defined on an infinite-dimensional space from its conditional expectations. We work within a standard framework in infinite-dimensional analysis, that of abstract Wiener spaces, developed by L. Gross. An abstract Wiener space is a triple (H, B, μ) where H is a real separable Hilbert space, B is the Banach space obtained by completing H with respect to a measurable norm, and μ is Wiener measure on B.

The main obstacle in infinite-dimensional analysis is the absence of a useful version of Lebesgue measure. To overcome this, we construct Gaussian measures μ_{M_p} on B, concentrated on closed affine subspaces $p + \overline{M_0}$ of B, where M_0 is any closed subspace of H, and then define the Gaussian Radon transform Gf of a bounded Borel function f on B using these measures. We investigate the relationship between the closed subspaces of finite codimension in B and those in H, and also investigate properties of the Gaussian Radon transform. Among these, we prove a disintegration theorem and express Gf as a conditional expectation. We provide an inversion procedure for the Gaussian Radon transform which uses the Segal-Bargmann transform. Finally, we present some possible applications of the Gaussian Radon transform to machine learning, by showing that Gf provides a stochastic interpretation of the ridge regression problem.

Chapter 1
Introduction

The classical Radon transform, first developed by Johann Radon in 1917, is defined for a function $f : \mathbb{R}^n \to \mathbb{R}$ as:

$$Rf(P) = \int_P f \, dl_P, \tag{1.1}$$

for all hyperplanes P in $\mathbb{R}^n$, where for every P integration is with respect to Lebesgue measure l_P on P. One may think of the hyperplane P as a "ray" shooting through the support of f, as pictured in Figure 1.1, and integrating f over P may be viewed as measuring the changes in the "density" of f as the ray passes through it. In this sense, Rf provides a way to reconstruct the density of an n-dimensional object from its known $(n-1)$-dimensional cross sections. It is through this line of thinking that the Radon transform became the mathematical backbone of medical CT scans, tomography and other image reconstruction applications.

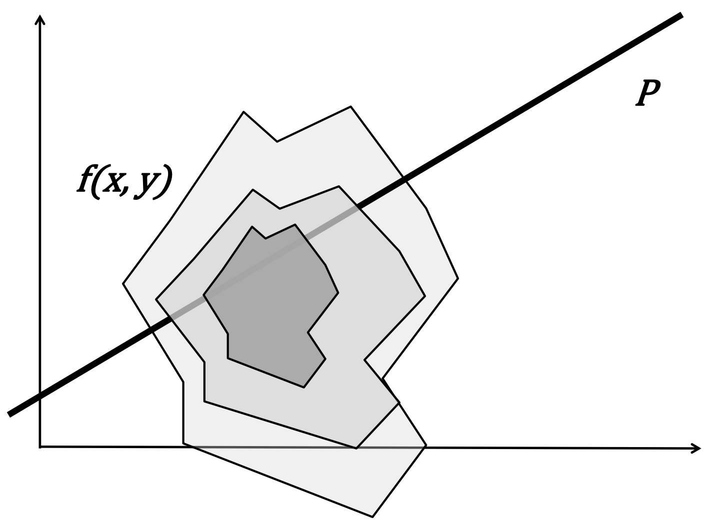


Figure 1.1: The classical Radon transform.

The goal of this work is to develop an infinite-dimensional version of the Radon transform. The main problem in infinite-dimensional analysis is the absence of a useful version of Lebesgue measure. However, Gaussian measures are known to be well-behaved in infinite-dimensional settings; with this in mind, we will be taking a probabilistic approach to this problem. Therefore the previously mentioned property of the classical Radon transform, that of recovering n-dimensional objects from their

$(n-1)$-dimensional sections, will become in our setting the ability to recover information about a function defined on an infinite-dimensional space from its conditional expectations.

Of the two standard frameworks in infinite-dimensional analysis, nuclear spaces and abstract Wiener spaces, we work within the latter. Abstract Wiener spaces were first developed by Leonard Gross in the celebrated work [13]. We continue the present chapter with some background material. In Section 1.1 we present the basic definitions and some results about Gaussian measures, as well as some of the reasons behind the popularity of Gaussian measures in infinite-dimensional analysis. Section 1.2 introduces the concept of measurable norm, necessary to then define abstract Wiener spaces, and also presents some of the basic properties of abstract Wiener spaces.

Before we proceed, we present a short outline of this work. Our first goal in developing a Radon transform for an infinite-dimensional Banach space B was to construct an appropriate measure on every hyperplane of B, which would correspond to the measures l_P in (1.1). In Chapter 2 we more generally construct probability measures μ_{M_p} on an infinite-dimensional Banach space B which are concentrated on closed affine subspaces. In this chapter we also explore the relationship between the closed subspace of finite codimension in B and those in an underlying dense Hilbert space H. Once the measures μ_{M_p} have been constructed, we use them to define the Gaussian Radon transform. Chapter 3 then explores the properties of the Gaussian Radon transform, including a disintegration formula, an expression of the Gaussian Radon transform as a conditional expectation, and an inversion procedure. Finally, Chapter 4 will explore some possible applications of the Gaussian Radon transform to the field of machine learning.

1.1 Gaussian Measures

Lebesgue measure l on $\mathbb{R}^n$ is uniquely determined (up to a constant) by the following three conditions:

i. l assigns finite values to bounded Borel sets.

ii. l assigns positive numbers to non-empty open sets.

iii. l is translation-invariant.

For this reason, we say that a Borel measure μ on a real Hilbert space H is a "Lebesgue measure" on H if it satisfies the three conditions above. By "the" Lebesgue measure, one would mean a particular choice of Lebesgue measure that has been specified (for instance by requiring that a given set, such as the unit cube, have measure 1).

Suppose V is a real finite-dimensional Hilbert space. Lebesgue measure l_V on V is given by:

$$l_V(E) := l\left[\pi_V^{-1}(E)\right] \text{ for all } E \in \mathcal{B}(V), \tag{1.2}$$

where for any topological vector space X, $\mathcal{B}(X)$ denotes the Borel σ-algebra of X, and:

$$\pi_V : \mathbb{R}^n \to V \tag{1.3}$$

is the Hilbert space isomorphism given by:

$$\pi_V(x) := x_1 e_1 + x_2 e_2 + \ldots + x_n e_n,$$

for all $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$, where $\{e_1, e_2, \ldots, e_n\}$ is an orthonormal basis for V. The measure l_V is independent of the choice of orthonormal basis.

Unfortunately, this does not work in infinite dimensions: suppose H is a real separable infinite-dimensional Hilbert space and that μ is a Lebesgue measure on H. Let $\{e_n\}_{n\in\mathbb{N}}$ be an orthonormal basis for H and for every integer $n \geq 1$ let B_n denote the open ball of radius $1/2$ centered at e_n. Since each B_n is a translate of the open ball of radius $1/2$ centered at 0, a non-empty open, bounded set, we have:

$$\mu(B_n) = a \text{ for all } n \in \mathbb{N},$$

where $0 < a < \infty$. Moreover, the sets B_n are mutually disjoint and are all contained in B, the open ball of radius 2 centered at the origin. But then:

$$\mu(B) \geq \sum_{n=1}^{\infty} \mu(B_n) = \sum_{n=1}^{\infty} a = \infty,$$

which contradicts the fact that μ assigns finite values to bounded sets.

The absence of a useful version of Lebesgue measure on infinite-dimensional spaces is one of the major obstacles of infinite-dimensional analysis. However, it was observed that Gaussian measures behave relatively well in this setting, and they have become key tools in infinite-dimensional analysis. We begin our review of Gaussian measures with the simple Euclidean case: for $n \in \mathbb{N}$, $m \in \mathbb{R}^n$ and $\sigma > 0$, the Gaussian measure on $\mathbb{R}^n$ with mean m and variance σ^2 is the Borel probability measure $\gamma_{m,\sigma}$ given by:

$$\gamma_{m,\sigma}(E) = \frac{1}{\left(\sigma\sqrt{2\pi}\right)^n} \int_E e^{-\frac{1}{2\sigma^2}\|x-m\|^2}\, dx, \tag{1.4}$$

for all $E \in \mathcal{B}(\mathbb{R}^n)$, where $\|\cdot\|$ denotes the usual Euclidean norm on $\mathbb{R}^n$. A Gaussian measure with mean 0 is said to be *centered* and the Gaussian measure $\gamma_{0,1}$ with mean 0 and variance 1 is known as the *standard Gaussian measure* on $\mathbb{R}^n$. Recall that for $m \in \mathbb{R}^n$, *Dirac measure* δ_m on $\mathbb{R}^n$ is the probability measure concentrated at the point m:

$$\delta_m(E) = \begin{cases} 1 & \text{, if } m \in E \\ 0 & \text{, if } m \notin E, \end{cases}$$

for all Borel subsets $E \subset \mathbb{R}^n$. In fact, δ_m is the weak limit as $\sigma \to 0$ of the probability measures $\gamma_{m,\sigma}$. For this reason, δ_m is considered a *degenerate* Gaussian measure, with mean m and variance 0. Gaussian measures with positive variance are said to be *non-degenerate*.

Now suppose $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space. A measurable function $X : \Omega \to \mathbb{R}$ is said to be a *Gaussian* (or *normal*) random variable provided that the distribution measure $\mathbb{P} \circ X^{-1}$ of X is a Gaussian measure on $\mathbb{R}$. Specifically, we say that X is Gaussian with mean m and variance σ^2, denoted:

$$X \sim \mathcal{N}(m, \sigma^2),$$

if the density function $\rho_X : \mathbb{R} \to \mathbb{R}$ is given by:

$$\rho_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-m)^2}, \text{ for all } x \in \mathbb{R}.$$

An important property of Gaussian random variables X is that:

$$\mathbb{E}\left[e^{tX}\right] = e^{t\mathbb{E}[X] + \frac{t^2}{2}\mathrm{Var}(X)}, \text{ for all } t \in \mathbb{C}. \tag{1.5}$$

In particular, the *characteristic function* of a Gaussian random variable X is given by:

$$\Phi_X(t) = \mathbb{E}\left[e^{itX}\right] = e^{it\mathbb{E}[x] - \frac{t^2}{2}\mathrm{Var}(X)}, \text{ for all } t \in \mathbb{C}. \tag{1.6}$$

Conversely, if $X : \Omega \to \mathbb{R}$ is a random variable with characteristic function Φ_X given by:

$$\Phi_X(t) = e^{itm - \frac{t^2\sigma^2}{2}}, \text{ for all } t \in \mathbb{R},$$

for some $m \in \mathbb{R}$ and $\sigma \geq 0$, then X is Gaussian with mean m and variance σ^2.

More generally, Gaussian measures may be defined on topological vector spaces.

Definition 1.1. Let X be a real locally convex topological vector space. A Borel probability measure μ on X is said to be a *Gaussian measure* provided that every continuous linear functional $f \in X^*$, viewed as a random variable on (X, μ), is Gaussian. In this case, the *mean* of μ is the linear function $m_\mu : X^* \to \mathbb{R}$ given by:

$$m_\mu(f) \stackrel{\text{def}}{=} \int_X f \, d\mu, \text{ for all } f \in X^*, \tag{1.7}$$

and the *covariance operator* $R_\mu : X^* \times X^* \to \mathbb{R}$ is given by:

$$R_\mu(f, g) \stackrel{\text{def}}{=} \int_X [f - m_\mu(f)]\,[g - m_\mu(g)]\, d\mu, \text{ for all } f, g \in X^*. \tag{1.8}$$

Moreover, μ is said to be *centered* if $m_\mu(f) = 0$ for all $f \in X^*$, and said to be *non-degenerate* if $\mu \circ f^{-1}$ is a non-degenerate measure on $\mathbb{R}^n$ for all non-zero $f \in X^*$.

We make a few remarks about this definition.

i. It is easily seen that R_μ is a positive definite, symmetric bilinear form. Moreover, μ is non-degenerate if and only if R_μ is *strictly* positive definite, that is if:

$$R_\mu(f, f) = \mathrm{Var}(f) > 0, \text{ for all non-zero } f \in X^*.$$

If μ is centered and non-degenerate, then $R_\mu(f, g)$ is simply the inner product of f and g in $L^2(X, \mu)$.

ii. Since every $f \in X^*$ is Gaussian with mean $m_\mu(f)$ and variance $R_\mu(f, f)$, we see from (1.6) that the *characteristic functional* of a Gaussian measure μ on X is:

$$\widehat{\mu}(f) \stackrel{\text{def}}{=} \int_X e^{if}\, d\mu = e^{im_\mu(f) - \frac{1}{2}R_\mu(f,f)}, \text{ for all } f \in X^*. \tag{1.9}$$

Conversely, if μ is a probability measure on X with characteristic functional:

$$\widehat{\mu}(f) = e^{iL(f) - \frac{1}{2}K(f,f)}, \text{ for all } f \in X^* \tag{1.10}$$

where $L : X^* \to \mathbb{R}$ is linear and $K : X^* \times X^* \to \mathbb{R}$ is a positive definite symmetric bilinear form, then μ is the Gaussian measure with mean L and covariance operator K.

Now suppose V is a real finite-dimensional Hilbert space and $\{e_1, \ldots, e_n\}$ is an orthonormal basis for V. Moreover, let $Z_1, \ldots, Z_n$ be independent standard Gaussian random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and consider the V-valued random variable on Ω:

$$Z := Z_1 e_1 + \ldots + Z_n e_n.$$

The Borel distribution measure γ_V induced by Z on V:

$$\gamma_V(E) := \mathbb{P}[Z \in E] = \mathbb{P}\left[(Z_1, \ldots, Z_n) \in \pi_V^{-1}(E)\right], \tag{1.11}$$

for all Borel subsets E of V, is called *standard Gaussian measure* on V, where $\pi_V : \mathbb{R}^n \to V$ is the isomorphism in (1.3). Remark that, regardless of the choice of Ω, Z_k or e_k, the random vector $(Z_1, \ldots, Z_n)$ induces standard Gaussian measure γ_n on $\mathbb{R}^n$, so for any Borel subset E of V:

$$\begin{aligned} \gamma_V(E) &= \gamma_n(\pi_V^{-1}(E)) \\ &= \int_{\pi_V^{-1}(E)} \frac{1}{\sqrt{2\pi}^n} e^{-\frac{1}{2}\|x\|^2_{\mathbb{R}^n}}\, dx \\ &= \int_E \frac{1}{\sqrt{2\pi}^n} e^{-\frac{1}{2}\|v\|^2_V}\, dl_V(v), \end{aligned} \tag{1.12}$$

where l_V is Lebesgue measure on V. The characteristic functional of γ_V is then:

$$\int_V e^{i\langle \cdot, h\rangle}\, d\gamma_V = e^{-\frac{1}{2}\|h\|^2}, \tag{1.13}$$

for all $h \in V$.

We can already see that the expression in (1.12) makes little sense in infinite dimensions: $(\sqrt{2\pi})^{-n} \to 0$ as $n \to \infty$ and there is no Lebesgue measure on infinite-dimensional spaces. Nonetheless, let us try to repeat the process above for a real separable infinite-dimensional Hilbert space H: let $\{e_k\}_{k\in\mathbb{N}}$ be an orthonormal basis for H and $\{Z_k\}_{k\in\mathbb{N}}$ be an independent sequence of standard Gaussian random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Next, we would like to define an H-valued random variable Z on Ω as:

$$Z = \sum_{k=1}^{\infty} Z_k e_k. \tag{1.14}$$

A problem immediately arises: this series does *not* converge almost surely. For any $\omega \in \Omega$:

$$Z(\omega) = \sum_{k=1}^{\infty} Z_k(\omega) e_k$$

belongs to H if and only if:

$$\sum_{k=1}^{\infty} Z_k(\omega)^2 < \infty. \tag{1.15}$$

Clearly (1.15) cannot hold almost everywhere, since the random variables Z_k are independent standard Gaussian.

The central idea behind abstract Wiener spaces is that of "measurable norm", introduced by Gross in [13]. The inspiration behind the definition of a measurable norm came from attempting to "force" the series in (1.14) to converge; we know this series does not converge with respect to the original Hilbert norm $\|\cdot\|$, so instead we consider a weaker norm $|\cdot|$ on H, complete H with respect to this norm to obtain a Banach space B, and see if (1.14) converges almost everywhere as a B*-valued* random variable. Of course, this new norm $|\cdot|$ must have certain properties in order to lead to this desired convergence, and we define these next.

1.2 Abstract Wiener Spaces

Definition 1.2. Let H be a real separable Hilbert space with Hilbert norm $\|\cdot\|$ and let $\mathcal{J}(H)$ be the collection of all finite-dimensional subspaces of H. We say that a norm $|\cdot|$ on H is *measurable* if for every $\epsilon > 0$ there is $F_\epsilon \in \mathcal{J}(H)$ such that:

$$\gamma_F \left[x \in F : |x| > \epsilon \right] < \epsilon, \tag{1.16}$$

for all $F \in \mathcal{J}(H)$ with $F \perp F_\epsilon$, where γ_F denotes standard Gaussian measure on $F \in \mathcal{J}(H)$.

Remark that, since all norms are equivalent on finite-dimensional spaces, the set $[x \in F : |x| > \epsilon]$ is Borel in $(F, \|\cdot\|)$, so (1.16) makes sense. In fact, much more is true - as the next result shows, every measurable norm $|\cdot|$ on H is *weaker* than the original Hilbert norm.

Theorem 1.1. *Let H be a real separable Hilbert space with Hilbert norm $\|\cdot\|$ and let $|\cdot|$ be a measurable norm on H. Then there is $c > 0$ such that:*

$$|x| \leq c\|x\|, \text{ for all } x \in H. \tag{1.17}$$

Moreover, if H is infinite-dimensional, then the original Hilbert norm $\|\cdot\|$ on H is not *a measurable norm.*

For a proof, see [21]. As a consequence of Theorem 1.1, if $|\cdot|$ is a measurable norm on a real separable infinite-dimensional Hilbert space H, then H is not complete with respect to $|\cdot|$. If it were complete, then $(H, |\cdot|)$ would be a Banach space, and since $|\cdot|$ is weaker than $\|\cdot\|$, the identity map:

$$id : (H, |\cdot|) \to (H, \|\cdot\|),$$

would be continuous. By the open mapping theorem, id would then be an open map, and then $|\cdot|$ and $\|\cdot\|$ would be equivalent. But then $\|\cdot\|$ would be a measurable norm, which is a contradiction.

Let B be the Banach space obtained by completing H with respect to a measurable norm $|\cdot|$. Then every $x^* \in B^*$ is continuous on H with respect to the Hilbert norm $\|\cdot\|$. To see this, note that since x^* is $|\cdot|$-continuous, there is $K > 0$ such that:

$$|(x, x^*)| \leq K|x|, \text{ for all } x \in B,$$

where (x, x^*) denotes the usual pairing $B - B^*$ for all $x \in B$ and $x^* \in B^*$. But from Theorem 1.1, there is $c > 0$ such that:

$$|(x, x^*)| \leq K|x| \leq Kc\|x\|, \text{ for all } x \in H.$$

By the Riesz representation theorem, we associate to every x^* a unique $h_{x^*} \in H$ given by:

$$\langle h, h_{x^*} \rangle = (h, x^*), \text{ for all } h \in H. \tag{1.18}$$

Moreover, the map $B^* \to H$; $x^* \mapsto h_{x^*}$ is linear and continuous. Clearly:

$$h_{x^*+\alpha y^*} = h_{x^*} + \alpha h_{y^*},$$

for all $x^*, y^* \in B^*$ and $\alpha \in \mathbb{R}$, and for every $x^* \in B^*$:

$$\begin{aligned} \|h_{x^*}\| = \sup_{\substack{h\in H \\ h\neq 0}} \frac{|\langle h, h_{x^*}\rangle|}{\|h\|} &\leq \sup_{\substack{h\in H \\ h\neq 0}} \frac{|\langle h, h_{x^*}\rangle|}{\frac{1}{c}|h|} \\ &\leq \sup_{\substack{x\in B \\ x\neq 0}} \frac{c|(x, x^*)|}{|x|} = c|x^*|_*, \end{aligned}$$

where $|\cdot|_*$ is the usual norm on B^* and c is a positive real number such that $|h| \leq c\|h\|$ for all $h \in H$.

We let H_{B^*} denote the image of this map in H:

$$H_{B^*} := \{h_{x^*} \in H : x^* \in B^*\}. \tag{1.19}$$

Then H_{B^*} is clearly a subspace of H. Now suppose that $h \in H_{B^*}^{\perp}$. Then $\langle h, h_{x^*}\rangle = 0$ for all $x^* \in B^*$, or $(h, x^*) = 0$ for all $x^* \in B^*$. But the only element of B that is mapped to 0 by all continuous linear functionals on B is 0, so $H_{B^*}^{\perp} = \{0\}$. In this manner, B^* is continuously embedded as the *dense* subspace H_{B^*} of H.

Theorem 1.2. *Let H be a real separable Hilbert space and B be the Banach space obtained by completing H with respect to a measurable norm $|\cdot|$. Then there is a unique Borel probability measure μ on B, called* Wiener measure*, such that:*

$$\int_B e^{i(x,x^*)}\, d\mu(x) = e^{-\frac{1}{2}\|h_{x^*}\|^2}, \tag{1.20}$$

for all $x^* \in B^*$.

This theorem may be proved from the perspective of cylindrical measures, which we do not focus on here, or from a more probabilistic point of view, using Lemma 2.3. One may find both of these proofs in [21]. Note however that Theorem 2.4, which we prove in Chapter 2, will directly imply the result above. We now have all the ingredients to define abstract Wiener spaces.

Definition 1.3. An *abstract Wiener space* is a triple:

$$(H, B, \mu),$$

where H is a real separable Hilbert space with Hilbert norm $\|\cdot\| = \sqrt{\langle\cdot,\cdot\rangle}$, B is the Banach space obtained by completing H with respect to a measurable norm $|\cdot|$, and μ is Wiener measure on B.

Let (H, B, μ) be an abstract Wiener space. From (1.20), every $x^* \in B^*$, as a random variable on (B, μ), is centered Gaussian with variance $\|h_{x^*}\|^2$. Then:

$$\|x^*\|^2_{L^2(B,\mu)} = \int_B |x^*|^2 \, d\mu = \|h_{x^*}\|^2.$$

So the map $H_{B^*} \to L^2(B, \mu)$; $h_{x^*} \mapsto x^*$ is continuous with respect to $\|\cdot\|$, and since H_{B^*} is a dense subspace of H, this map has a unique extension to H. We denote this extension by:

$$I : H \to L^2(B, \mu); h \mapsto Ih. \tag{1.21}$$

Specifically, for $h \in H$ we let $\{h_{x_k^*}\}_{k\in\mathbb{N}}$ be a sequence in H_{B^*} that converges to h in H. Then the sequence $\{x_k^*\}_{k\in\mathbb{N}}$ is Cauchy in $L^2(B, \mu)$, and thus converges to a limit $Ih \in L^2(B, \mu)$; this limit does not depend on the choice of $\{h_{x_k^*}\}$.

The map I was first introduced by Gross in [13] and is sometimes referred to as the *Paley-Wiener map*. As the L^2-limit of a sequence of centered Gaussians with variances $\|h_{x_k^*}\|^2 \to \|h\|^2$, Ih is centered Gaussian with variance $\|h\|^2$. Moreover, I is an *i*sometry: let $h, h' \in H$ and $\{h_{x_k^*}\}_k$, $\{h_{y_k^*}\}_k$ be sequences in H_{B^*} converging in H to h, h', respectively. Then:

$$\begin{aligned}
\langle Ih, Ih' \rangle_{L^2(B,\mu)} &= \int_B (Ih)(Ih') \, d\mu \\
&= \lim_{k\to\infty} \int_B x_k^* y_k^* \, d\mu \\
&= \lim_{k\to\infty} \langle h_{x_k^*}, h_{y_k^*} \rangle \\
&= \langle h, h' \rangle.
\end{aligned}$$

Remark 1.1. Many authors simply denote this map by $h \mapsto (\cdot, h)$ and think of it as a sort of extension of the inner-product map from H to B, but some measure-theoretic technicalities arising in Chapter 3 require us to be careful about the fact that Ih is really an equivalence class of functions defined almost everywhere.

Example 1.1. Let $\mathcal{H}$ be the space of all *absolutely continuous* functions $h : [0,1] \to \mathbb{R}$ such that $h(0) = 0$ and $h' \in L^2[0,1]$. Then $\mathcal{H}$ is a real separable infinite-dimensional Hilbert space with the inner product:

$$\langle h_1, h_2 \rangle := \int_0^1 h_1'(x) h_2'(x) dx,$$

for all $h_1, h_2 \in \mathcal{H}$. The supremum norm:

$$\|h\|_\infty := \sup_{0 \le x \le 1} |h(x)|,$$

for all $h \in \mathcal{H}$, is a *measurable norm* on $\mathcal{H}$, and the completion of $\mathcal{H}$ with respect to $\|\cdot\|_\infty$ is the space $\mathcal{C}$ of continuous functions $f : [0,1] \to \mathbb{R}$ with $f(0) = 0$. The resulting triple:

$$(\mathcal{H}, \mathcal{C}, \mu),$$

is known as the *classical Wiener space*, and the resulting measure μ induced by $\mathcal{H}$ on $\mathcal{C}$ is known as *classical Wiener measure*. This space will be explored further in Section 3.3.

Example 1.2. If $T : H \to H$ is an injective Hilbert-Schmidt operator on a real separable infinite-dimensional Hilbert space, then:

$$|h| := \|Th\|, \text{ for all } h \in H,$$

defines a measurable norm on H.

So far we have presented the original approach by Gross, that of starting with H and constructing B and μ by completing H with respect to a measurable norm. However, one can start with any real separable Banach space and turn it into an abstract Wiener space. Specifically, let B be a real separable Banach space with norm $|\cdot|$, μ be a centered, non-degenerate Gaussian measure on B, and $R_\mu : B^* \times B^* \to \mathbb{R}$ be the covariance operator:

$$R_\mu(x^*, y^*) = \int_B x^* y^* \, d\mu, \text{ for all } x^*, y^* \in B^*.$$

The *Cameron-Martin space* of (B, μ) is the subspace $H \subset B$ defined as:

$$H := \left\{ x \in B : \|x\| := \sup_{0 \ne x^* \in B^*} \frac{|(x, x^*)|}{\sqrt{R_\mu(x^*, x^*)}} < \infty \right\}. \tag{1.22}$$

Then the norm $\|\cdot\|$ defined above is a complete inner-product norm on H, it is stronger than the Banach norm $|\cdot|$, and H is dense in B. Moreover, the Banach norm $|\cdot|$ is a measurable norm on H (for an ingenious proof, due to Stroock, of this fact, see Section VIII of Bruce Driver's notes [10]). So (H, B, μ) is an abstract Wiener space, and note that H is uniquely determined by (B, μ). Conversely, if (H, B, μ) is an abstract Wiener space, then H is the Cameron-Martin space of (B, μ).

Chapter 2
The Gaussian Radon Transform

The focus of this chapter will be to define the Gaussian Radon transform. Our first goal will be to construct the appropriate measures on B, which will be probability measures concentrated on closed affine subspaces of B. Before we proceed, we introduce the notion of measurably adapted sequence.

2.1 Measurably Adapted Sequences

The next definition and Lemma 2.2 following it are instrumental in obtaining the main result of this chapter.

Definition 2.1. Let $|\cdot|$ be a measurable norm on a real separable infinite-dimensional Hilbert space H. We say that a sequence $\{F_n\}_{n\in\mathbb{N}}$ of closed subspaces of H is *measurably adapted* provided that it satisfies the following conditions:

(1). The sequence is strictly increasing: $F_1 \subset F_2 \subset \ldots \subset H$, with $F_n \neq F_{n+1}$, and F_n has *finite codimension* in F_{n+1} for all $n \in \mathbb{N}$:

$$1 \leq \dim\left(F_{n+1} \cap F_n^{\perp}\right) < \infty, \text{ for all } n \in \mathbb{N}. \tag{2.1}$$

(2). The union $\bigcup_{n=1}^{\infty} F_n$ is dense in H.

(3). For every $n \in \mathbb{N}$:

$$\gamma_{Q_n}\left[x \in Q_n : |x| > \frac{1}{2^n}\right] < \frac{1}{2^n}, \tag{2.2}$$

where γ_{Q_n} denotes standard Gaussian measure on $Q_n = F_{n+1} \cap F_n^{\perp}$.

Before we proceed, we make a few simple but useful observations about increasing sequences of closed subspaces of a real separable Hilbert space.

Proposition 2.1. *Let H be a real separable Hilbert space.*

(i). If K and L are closed subspaces of H:

$$(L + K^{\perp}) \cap K = P_K(L), \tag{2.3}$$

where P_K denotes the orthogonal projection of H onto K.

(ii). If $F_1 \subset F_2 \subset \ldots \subset H$ is an increasing sequence of closed subspaces of H:

$$F_{m+1} = F_n \oplus (F_{n+1} \cap F_n^{\perp}) \oplus \ldots \oplus (F_{m+1} \cap F_m^{\perp}), \tag{2.4}$$

for all integers $m \geq n > 0$. If, in addition, $\bigcup_{n=1}^{\infty} F_n$ is dense in H, then:

$$F_n^{\perp} = \oplus_{j=n}^{\infty}(F_{j+1} \cap F_j^{\perp}), \tag{2.5}$$

for all $n \in \mathbb{N}$.

Proof. (i). Let $k = l + k' \in (L + K^\perp) \cap K$, where $l \in L$ and $k' \in K^\perp$. Then since $k \in K$:

$$k = P_K k = P_K(l + k') = P_K l \in P_K(L),$$

so $(L + K^\perp) \cap K \subset P_K(L)$. Conversely, suppose $l \in L$. Then:

$$P_K l = P_K l + P_{K^\perp} l - P_{K^\perp} l = l - P_{K^\perp} l \in (L + K^\perp) \cap K,$$

which proves (2.3).

(ii). Let $n \leq m$ be positive integers. Since $F_m \subset F_{m+1}$, we may express F_{m+1} as:

$$F_{m+1} = F_m \oplus (F_{m+1} \cap F_m^\perp).$$

Similarly, we may express F_m as $F_m = F_{m-1} \oplus (F_m \cap F_{m-1}^\perp)$, so:

$$F_{m+1} = F_{m-1} \oplus (F_m \cap F_{m-1}^\perp) \oplus (F_{m+1} \cap F_m^\perp).$$

Continuing in this manner, (2.4) follows inductively.

Now suppose that $\bigcup_{m=1}^\infty F_m$ is dense in H. Since the sequence $\{F_m\}_{m\in\mathbb{N}}$ is increasing, for $n \in \mathbb{N}$:

$$F_n^\perp \supset \oplus_{j=n}^\infty (F_{j+1} \cap F_j^\perp).$$

Let $h \in F_n^\perp$ be such that $h \perp (F_{j+1} \cap F_j^\perp)$ for all $j \geq n$. By (2.4), $h \perp F_{m+1}$ for all $m \geq n$, so then $h \perp F_m$ for all $m \in \mathbb{N}$. But since $\bigcup_{m=1}^\infty F_m$ is dense in H, it follows that $h = 0$, and then (2.5) holds. □

The following result proves the existence of a measurably adapted sequence "starting" at the orthogonal complement of a given closed subspace of H.

Lemma 2.2. *Let $|\cdot|$ be a measurable norm on a real separable infinite-dimensional Hilbert space H and $M_0 \subset H$ be an infinite-dimensional closed subspace. Then there is a measurably adapted sequence $\{F_n\}_{n\in\mathbb{N}}$ of closed subspaces of H such that:*

$$F_1 \supsetneq F_0 := M_0^\perp, \tag{2.6}$$

and:

$$dim(F_1 \cap M_0) < \infty. \tag{2.7}$$

Moreover, the linear span of the subspaces $F_n \cap F_{n-1}^\perp$ for $n \in \mathbb{N}$ is dense in M_0.

Proof. Since M_0 is a closed subspace of a separable Hilbert space, it is also separable. So let $D = \{d_n\}_{n\in\mathbb{N}}$ be a countable dense subset of M_0, with $d_n \neq 0$ and $d_n \neq d_m$ for all $n, m \in \mathbb{N}$. Since $|\cdot|$ is a measurable norm, there is for every $n \in \mathbb{N}$ a finite-dimensional subspace E_n of H such that:

$$\gamma_E \left[x \in E : |x| > \frac{1}{2^n} \right] < \frac{1}{2^n}, \tag{2.8}$$

for every $E \in \mathcal{J}(H)$ with $E \perp E_n$, where $\mathcal{J}(H)$ denotes the collection of all finite-dimensional subspaces of H and γ_E denotes standard Gaussian measure on E.

Let:

$$F_1 := M_0^\perp + E_1 + \mathbb{R}d_1, \tag{2.9}$$

where $\mathbb{R}h$ denotes the span $\{\alpha h : \alpha \in \mathbb{R}\}$ of the vector h for all $h \in H$. Note that $F_1 \supset M_0^\perp$ and the inclusion is strict because $d_1 \notin M_0^\perp$. Also, by (2.3):

$$F_1 \cap M_0 = P_{M_0}(E_1 + \mathbb{R}d_1),$$

so $F_1 \cap M_0$ is the image of a finite-dimensional subspace under a continuous map and thus $\dim(F_1 \cap M_0) < \infty$.

Now F_1 is a closed subspace, as the sum of two closed subspaces, one of which is finite-dimensional. So $M_0 \setminus F_1 = M_0 \cap F_1^{\mathrm{C}}$ is an open subset of M_0 and it is non-empty: if $M_0 \setminus F_1 = \emptyset$, then $M_0 \cap F_1 = M_0$, a contradiction because M_0 is infinite-dimensional and $F_1 \cap M_0$ is not. Thus $M_0 \setminus F_1$ is a non-empty open subset of M_0 which does not contain d_1, so there is $n_1 > 1$ such that:

$$d_{n_1} \in M_0 \setminus F_1.$$

Consider now:

$$F_2 := F_1 + E_2 + \mathbb{R}d_2 + \ldots + \mathbb{R}d_{n_1}. \tag{2.10}$$

As before, the inclusion $F_2 \supset F_1$ is strict, since $d_{n_1} \notin F_1$, and $F_2 \cap F_1^\perp$ is a non-empty finite-dimensional subspace that is orthogonal to F_1, and thus also to E_1. By (2.8):

$$\gamma_{Q_1}\left[x \in Q_1 : |x| > \frac{1}{2}\right] < \frac{1}{2},$$

where $Q_1 = F_2 \cap F_1^\perp$. By the same reasoning as above, $M_0 \setminus F_2$ is a non-empty open subset of M_0 that does not contain $d_1, d_2, \ldots, d_{n_1}$, so there is $n_2 > n_1$ such that:

$$d_{n_2} \in M_0 \setminus F_2.$$

Then let:

$$F_3 := F_2 + E_3 + \mathbb{R}d_{n_1+1} + \ldots + \mathbb{R}d_{n_2}. \tag{2.11}$$

As above, it follows that:

$$\gamma_{Q_2}\left[x \in Q_2 : |x| > \frac{1}{2^2}\right] < \frac{1}{2^2},$$

where $Q_2 = F_3 \cap F_2^\perp$.

Continuing this process inductively, we obtain a sequence $F_1 \subset F_2 \subset \ldots \subset H$ that satisfies (2.6) and (2.7) such that $1 \leq \dim(F_{n+1} \cap F_n^\perp) < \infty$ and:

$$\gamma_{Q_n}\left[x \in Q_n : |x| > \frac{1}{2^n}\right] < \frac{1}{2^n},$$

for all $n \in \mathbb{N}$, where $Q_n = F_{n+1} \cap F_n^\perp$. Since $F_0 = M_0^\perp \subset F_1 \subset \ldots \subset H$, by (2.4):

$$F_k = M_0^\perp \oplus (F_1 \cap M_0) \oplus (F_2 \cap F_1^\perp) \oplus \ldots \oplus (F_k \cap F_{k-1}^\perp),$$

for all $k \in \mathbb{N}$, and since $M_0^\perp \subset F_k$:

$$F_k = M_0^\perp \oplus (F_k \cap M_0),$$

for all $k \in \mathbb{N}$. Consequently:

$$F_k \cap M_0 = (F_1 \cap M_0) \oplus (F_2 \cap F_1^\perp) \oplus \ldots \oplus (F_k \cap F_{k-1}^\perp),$$

for all $k \in \mathbb{N}$, and since (by construction) $F_k \cap M_0$ contains $\{d_1, \ldots, d_{n_k}\}$, we conclude that the closed linear span of $F_j \cap F_{j-1}^\perp$ for $j \in \mathbb{N}$ is M_0. Finally, the closed linear span of $\{F_n\}_{n\in\mathbb{N}}$ is then H, so $\{F_n\}_{n\in\mathbb{N}}$ is a measurably adapted sequence. $\square$

2.2 Definition of the Gaussian Radon Transform

Our next goal is to construct measures concentrated on closed affine subspaces of B, measures which will lead to the definition of the Gaussian Radon transform. We will need Lemma 2.2 in conjunction with the following result, which is a standard trick in probability to test the convergence of a series of random variables.

Lemma 2.3. *Let B be a real separable Banach space with norm $|\cdot|$ and $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of B-valued random variables:*

$$X_n : \Omega \to B,$$

on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Suppose that:

$$\sum_{n=1}^{\infty} \mathbb{P}\left[|X_n| > \frac{1}{2^n}\right] < \infty. \tag{2.12}$$

Then $\sum_{n=1}^{\infty} X_n$ is almost surely absolutely convergent.

Proof. By the first Borel-Cantelli Lemma, (2.12) yields:

$$\mathbb{P}\left(\limsup_{n\to\infty}\left[|X_n| > \frac{1}{2^n}\right]\right) = 0. \tag{2.13}$$

Now suppose $\omega \in \Omega'$, where:

$$\Omega' := \left(\limsup_{n\to\infty}\left[|X_n| > \frac{1}{2^n}\right]\right)^{\mathrm{C}}.$$

Then there is $N \in \mathbb{N}$ such that:

$$|X_n(\omega)| \leq \frac{1}{2^n}, \text{ for all } n \geq N.$$

Then:

$$\begin{aligned}\sum_{n=1}^{\infty}|X_n(\omega)| &= \sum_{n=1}^{N-1}|X_n(\omega)| + \sum_{n=N}^{\infty}|X_n(\omega)| \\ &\leq \sum_{n=1}^{N-1}|X_n(\omega)| + \sum_{n=N}^{\infty}\frac{1}{2^n} \\ &< \infty.\end{aligned}$$

So $\sum_{n=1}^{\infty} X_n(\omega)$ is absolutely convergent for all $\omega \in \Omega'$, and $\mathbb{P}(\Omega') = 1$ by (2.13), which proves our claim. $\square$

Finally, we may construct the measures needed for the Gaussian Radon transform.

Theorem 2.4. *Let (H, B, μ) be an abstract Wiener space and M_0 be a closed subspace of H. For every $p \in M_0^{\perp}$ there exists a unique Borel measure μ_{M_p} on B such that:*

$$\int_B e^{ix^*}\, d\mu_{M_p} = e^{i(p,x^*) - \frac{1}{2}\|P_{M_0}h_{x^*}\|^2}, \tag{2.14}$$

for all $x^* \in B^*$, where P_{M_0} denotes the orthogonal projection of H onto M_0. Moreover, the measure μ_{M_p} is concentrated on the closure $\overline{M_p}$ of $M_p = p + M_0$ in B:

$$\mu_{M_p}(\overline{M_p}) = 1. \tag{2.15}$$

Proof. Suppose first that $\dim(M_0) = \infty$. By Lemma 2.2, there is a measurably adapted sequence $\{F_n\}_{n\in\mathbb{N}}$ of closed subspaces of H with $F_1 \supsetneq F_0 := M_0^{\perp}$ and $\dim(F_1 \cap M_0) < \infty$ such that the linear span of $F_n \cap F_{n-1}^{\perp}$ for $n \in \mathbb{N}$ is dense in M_0.

Let $\{e_1, \ldots, e_{k_1}\}$ be an orthonormal basis for $F_1 \cap M_0$, which we extend inductively to an orthonormal sequence $\{e_k\}_{k\in\mathbb{N}}$ with $\{e_{k_{n-1}+1}, \ldots, e_{k_n}\}$ forming an orthonormal basis for $F_n \cap F_{n-1}^{\perp}$ for all $n \in \mathbb{N}$ and $k_0 := 0 < k_1 < k_2 < \ldots$. Then $\{e_1, \ldots, e_{k_n}\}$ is an orthonormal basis for the subspace:

$$(F_1 \cap M_0) \oplus (F_2 \cap F_1^{\perp}) \oplus \ldots \oplus (F_n \cap F_{n-1}^{\perp}) = F_n \cap M_0.$$

Therefore $\{e_k\}_{k\in\mathbb{N}}$ is an orthonormal basis for M_0.

Now let $\{Z_k\}_{k\in\mathbb{N}}$ be an independent sequence of standard Gaussian random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Since $\{e_{k_{n-1}+1}, \ldots, e_{k_n}\}$ is an orthonormal basis for $Q_{n-1} := F_n \cap F_{n-1}^{\perp}$, the standard Gaussian measure $\gamma_{Q_{n-1}}$ on Q_{n-1} is the distribution measure of the random variable:

$$X_{n-1} := Z_{k_{n-1}+1}e_{k_{n-1}+1} + \ldots + Z_{k_n}e_{k_n}, \tag{2.16}$$

for all $n \in \mathbb{N}$.

By the measurably adapted property in (2.2):

$$\gamma_{Q_{n-1}}\left[x \in Q_{n-1} : |x| > \frac{1}{2^{n-1}}\right] < \frac{1}{2^{n-1}}.$$

This becomes:

$$\mathbb{P}\left[|X_{n-1}| > \frac{1}{2^{n-1}}\right] < \frac{1}{2^{n-1}}.$$

By Lemma 2.3, the (appropriately grouped) series:

$$\begin{aligned} Z_{M_0} &:= \sum_{n=0}^{\infty} X_n \qquad (2.17)\\ &= (Z_1 e_1 + \ldots + Z_{k_1} e_{k_1}) + \\ &\quad +(Z_{k_1+1} e_{k_1+1} + \ldots + Z_{k_2} e_{k_2}) + \\ &\quad \vdots \end{aligned}$$

is $\mathbb{P}$-a.s. absolutely convergent (with respect to the norm $|\cdot|$ on B). Moreover, since $\{e_k\}_{k\in\mathbb{N}}$ is an orthonormal basis for M_0, the random variable Z_{M_0} takes values in $\overline{M_0}$, the closure of M_0 in B.

Now for every $x^* \in B^*$, by continuity of x^*:

$$\begin{aligned} (Z_{M_0}, x^*) &= \sum_{n=0}^{\infty} (X_n, x^*) \text{ a.s.} \\ &= \sum_{n=1}^{\infty} \left(Z_{k_{n-1}+1} e_{k_{n-1}+1} + \ldots + Z_{k_n} e_{k_n}, x^*\right) \\ &= \sum_{n=1}^{\infty} \left(Z_{k_{n-1}+1} \langle e_{k_{n-1}+1}, h_{x^*}\rangle + \ldots + Z_{k_n} \langle e_{k_n}, h_{x^*}\rangle\right) \\ &= \sum_{k=1}^{\infty} Z_k \langle e_k, h_{x^*}\rangle. \end{aligned}$$

Then:

$$\begin{aligned} \int_\Omega e^{i(Z_{M_0},x^*)}\, d\mathbb{P} &= \int_\Omega e^{i\sum_{k=1}^{\infty} Z_k\langle e_k, h_{x^*}\rangle}\, d\mathbb{P} \\ &= \lim_{N\to\infty} \int_\Omega e^{i\sum_{k=1}^{N} Z_k\langle e_k, h_{x^*}\rangle}\, d\mathbb{P} \qquad \text{(Dominated Convergence Theorem)} \\ &= \lim_{N\to\infty} \int_\Omega \prod_{k=1}^{N} e^{iZ_k\langle e_k, h_{x^*}\rangle}\, d\mathbb{P} \\ &= \lim_{N\to\infty} \prod_{k=1}^{N} \int_\Omega e^{iZ_k\langle e_k, h_{x^*}\rangle}\, d\mathbb{P} \qquad \text{(by independence of the } Z_k\text{'s)} \\ &= \lim_{N\to\infty} \prod_{k=1}^{N} e^{-\frac{1}{2}\langle e_k, h_{x^*}\rangle^2} \qquad \text{(because } Z_k \text{ is standard Gaussian)} \\ &= e^{-\frac{1}{2}\sum_{k=1}^{\infty}\langle e_k, h_{x^*}\rangle^2} \\ &= e^{-\frac{1}{2}\|P_{M_0} h_{x^*}\|^2} \qquad \text{(because } \{e_k\}_{k\in\mathbb{N}} \text{ is an o.n.b. for } M_0\text{).} \end{aligned}$$

So (Z_{M_0}, x^*) is a centered Gaussian random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ with variance $\|P_{M_0} h_{x^*}\|^2$. Let μ_{M_0} be the distribution measure Z_{M_0} induces on B:

$$\mu_{M_0}(E) := \mathbb{P}\left[Z_{M_0} \in E\right], \tag{2.18}$$

for all Borel subsets E of B. Then $x^* \in B^*$, as a random variable on (B, μ_{M_0}), is centered Gaussian with variance $\|P_{M_0} h_{x^*}\|^2$:

$$\int_B e^{ix^*}\, d\mu_{M_0} = \int_\Omega e^{i(Z_{M_0}, x^*)}\, d\mathbb{P} = e^{-\frac{1}{2}\|P_{M_0} h_{x^*}\|^2}.$$

Moreover, since $Z_{M_0}(\omega) \in \overline{M_0}$ for $\mathbb{P}$-almost all $\omega \in \Omega$:

$$1 = \mathbb{P}[Z_{M_0} \in \overline{M_0}] = \mu_{M_0}(\overline{M_0}).$$

Now let $p \in M_0^\perp$ and let μ_{M_p} be the measure specified by:

$$\mu_{M_p}(E) := \mu_{M_0}(E - p), \tag{2.19}$$

for all Borel subsets E of B. Equivalently, μ_{M_p} is the distribution measure of the random variable $Z_{M_p} := p + Z_{M_0}$:

$$\mu_{M_p}(E) = \mathbb{P}\left[p + Z_{M_0} \in E\right]. \tag{2.20}$$

In this case:

$$\int_B f(x)\, d\mu_{M_p}(x) = \int_B f(x + p)\, d\mu_{M_0}(x), \tag{2.21}$$

whenever either side exists (this reduces to (2.19) for $f = 1_E$ and the general case for a Borel function f follows as usual). Then:

$$\int_B e^{ix^*}\, d\mu_{M_p} = \int_B e^{i(x+p, x^*)}\, d\mu_{M_0}(x) = e^{i(p,x^*) - \frac{1}{2}\|P_{M_0} h_{x^*}\|^2}.$$

Moreover:

$$\mu_{M_p}(\overline{M_p}) = \mathbb{P}\left[p + Z_{M_0} \in p + \overline{M_0}\right] = \mu_{M_0}(\overline{M_0}) = 1.$$

Finally, if M_0 is finite-dimensional, we can simply take:

$$Z_{M_0} := Z_1 e_1 + \ldots + Z_n e_n,$$

where $\{e_1, \ldots e_n\}$ is an orthonormal basis for M_0 and $Z_1, \ldots, Z_n$ are independent standard Gaussians on $(\Omega, \mathcal{F}, \mathbb{P})$. We then define μ_{M_0} and μ_{M_p} the same as in (2.18) and (2.19), and (2.14) follows. Note that, in this case, μ_{M_0} is simply standard Gaussian measure on M_0.

Uniqueness of the measure μ_{M_p} follows from uniqueness of characteristic functions for probability measures.

□

We are now ready to define the Gaussian Radon transform.

Definition 2.2. Let (H, B, μ) be an abstract Wiener space and f be a Borel function on B. For every closed subspace M_0 of H and $p \in M_0^{\perp}$, the Gaussian Radon transform Gf of f is defined by:

$$Gf(p + M_0) := \int_B f \, d\mu_{M_p}, \tag{2.22}$$

where μ_{M_p} is the measure concentrated on $\overline{M_p} = p + \overline{M_0}$ constructed in Theorem 2.4.

Note that for a generic Borel function f, the Gaussian Radon transform Gf does not necessarily exist, although Gf does exist if f is bounded or non-negative. In our initial paper [16] we followed the classical Radon transform approach and defined Gf on the set of all hyperplanes in H (because, as will be discussed next in Section 2.3, there are in a sense "more" hyperplanes in H than in B). However, broader results can be obtained if one works with general closed affine subspaces and the notation in this more general case becomes less cumbersome if we define Gf as in (2.22).

2.3 Closed Affine Subspaces

The classical Radon transform is defined on the set of all hyperplanes in $\mathbb{R}^n$, which naturally led us to study hyperplanes in B and their relationship to hyperplanes in H. Eventually, we obtained the complete relationship between the closed subspaces of finite codimension in B and those in H. This result is Theorem 2.6, which relies on the following lemma.

Lemma 2.5. *Let (H, B, μ) be an abstract Wiener space and $\{u_1, \ldots, u_n\} \subset H_{B^*}$ be an orthonormal set, where $u_k = h_{y_k^*}$ for some $y_k^* \in B^*$ for all $1 \leq k \leq n$. Then:*

$$\bigcap_{k=1}^{n} (y_k^*)^{-1}(p_k) = p_1 u_1 + \ldots + p_n y_n + \overline{V^{\perp}}, \tag{2.23}$$

where V =span$\{u_1, \ldots, u_n\} \subset H_{B^*}$, $p_1, \ldots, p_n \in \mathbb{R}$, and $\overline{V^{\perp}}$ is the closure of $V^{\perp}$ in B.

Proof. Let:

$$L_1 := \bigcap_{k=1}^{n} (y_k^*)^{-1}(p_k) \text{ and } L_1' := p_1 u_1 + \ldots + p_n u_n + V^{\perp} \subset H.$$

Then for every $v \in V^{\perp}$ and $1 \leq k \leq n$:

$$(p_1 u_1 + \ldots + p_n u_n + v, y_k^*) = \langle p_1 u_1 + \ldots + p_n u_n + v, u_k \rangle = p_k,$$

so $L_1' \subset L_1$. Since L_1 is a closed subspace of B:

$$L_1 \supset \overline{L_1'} = p_1 u_1 + \ldots + p_n u_n + \overline{V^{\perp}}.$$

Now fix an element $x \in L_1$ and consider:

$$x' := x - p_1 u_1 - \ldots - p_n u_n.$$

Since $(x, y_k^*) = p_k$, we have $(x', y_k^*) = 0$ for all $1 \leq k \leq n$, so:

$$x' \in \bigcap_{k=1}^{n} \mathrm{Ker}(y_k^*).$$

We show that there is a $|\cdot|$-Cauchy sequence $\{h_j'\}_{j\in\mathbb{N}} \subset V^\perp$ such that:

$$h_j' \xrightarrow{j\to\infty} x' \text{ in } B. \tag{2.24}$$

In turn, this will give us that the sequence $\{h_j\}_{j\in\mathbb{N}} \subset L_1'$ given by:

$$h_j := h_j' + p_1 u_1 + \ldots + p_n u_n,$$

converges to x in B. Since this holds for all $x \in L_1$, we will have $L_1 \subset \overline{L_1'}$.

To prove the claim in (2.24), note that since H is dense in B there is a sequence $\{g_j\}_{j\in\mathbb{N}} \subset H$ such that $g_j \to x'$ in B. By $|\cdot|$-continuity of y_k^*, $(g_j, y_k^*) \xrightarrow{j\to\infty} (x', y_k^*)$ for all k, so:

$$\lim_{j\to\infty} \langle g_j, u_k\rangle = 0, \text{ for all } 1 \leq k \leq n. \tag{2.25}$$

For every $1 \leq k \leq n$, consider the sequence $\{g_j^{(k)}\}_{j\in\mathbb{N}}$ given by:

$$g_j^{(k)} := g_j - \langle g_j, u_k\rangle u_k, \text{ for all } j \in \mathbb{N}.$$

By (2.25):

$$\lim_{j\to\infty} g_j^{(k)} = x' \text{ in } B, \text{ for all } 1 \leq k \leq n. \tag{2.26}$$

For every k:

$$\langle g_j^{(k)}, u_k\rangle = \langle g_j, u_k\rangle - \langle g_j, u_k\rangle\langle u_k u_k\rangle = 0, \text{ for all } j \in \mathbb{N},$$

so:

$$\{g_j^{(k)}\}_{j\in\mathbb{N}} \subset u_k^\perp, \text{ for all } 1 \leq k \leq n. \tag{2.27}$$

Finally, consider the sequence:

$$h_j' := \frac{1}{n}\left(\sum_{k=1}^{n} g_j^{(k)} - \sum_{k=1}^{n} \sum_{\substack{1\leq i\leq n \\ i\neq k}} \langle g_j^{(i)}, u_k\rangle u_k\right).$$

By (2.26) and $|\cdot|$-continuity of each y_k^*:

$$\lim_{j\to\infty} h_j' = \frac{1}{n}(nx') = x' \text{ in } B.$$

Moreover, $\{h'_j\}_{j\in\mathbb{N}} \subset V^\perp$ as desired; for every $1 \le m \le n$:

$$\begin{aligned}
\langle h'_j, u_m\rangle &= \frac{1}{n}\left(\sum_{k=1}^{n}\langle g_j^{(k)}, u_m\rangle - \sum_{k=1}^{n}\sum_{\substack{1\le i\le n\\ i\neq k}}\langle g_j^{(i)}, u_k\rangle\langle u_k, u_m\rangle\right)\\
&= \frac{1}{n}\left(\sum_{k=1}^{n}\langle g_j^{(k)}, u_m\rangle - \sum_{\substack{1\le i\le n\\ i\neq m}}\langle g_j^{(i)}, u_m\rangle\right)\\
&= \frac{1}{n}\langle g_j^{(m)}, u_m\rangle\\
&= 0,
\end{aligned}$$

where the last equality follows from (2.27). □

Theorem 2.6. *Let (H, B, μ) be an abstract Wiener space. Then:*

(i). If L is a closed subspace of finite codimension n in B then there is a unique closed subspace M of codimension n in H such that $L = \overline{M}$, where we are taking closures in B. Specifically, $M = L \cap H$.

(ii). Let M be a closed subspace of finite codimension in H. Then the closure $\overline{M}$ of M in B is a closed subspace of codimension k in B, where:

$$0 \le k := dim(M^\perp \cap H_{B^*}) \le n. \tag{2.28}$$

In particular, if $M^\perp \cap H_{B^*} = \{0\}$, then M is dense in B.

Proof. (i). Let L be a closed subspace of codimension n in B. Then there is a linearly independent set $\{x_1^*, \ldots, x_n^*\}$ of non-zero elements in B^*, such that:

$$L = \bigcap_{k=1}^{n}\mathrm{Ker}(x_k^*).$$

Consider a translate L_1 of L:

$$L_1 = \bigcap_{k=1}^{n}(x_k^*)^{-1}(t_k),$$

for some $t_1, \ldots, t_n \in \mathbb{R}$. Applying the Gram-Schmidt orthonormalization process, we obtain an orthonormal basis $\{u_1, \ldots, u_n\}$ of $V :=\text{span}\{h_{x_1^*}, \ldots, h_{x_n^*}\} \subset H_{B^*}$, where $u_k = h_{y_k^*}$ for some $y_k^* \in B^*$ for all $1 \le k \le n$. Then:

$$\begin{aligned}
L_1 &= \bigcap_{k=1}^{n}\{x \in B : (x, x_k^*) = t_k\}\\
&= \bigcap_{k=1}^{n}\{x \in B : \sum_{i=1}^{n}\alpha_{k,i}(x, y_k^*) = t_k\}, \text{ where } \alpha_{k,i} = \langle h_{x_k^*}, u_i\rangle \text{ for all } k, i,\\
&= \{x \in B : A[(x, y_1^*)\ldots(x, y_n^*)]^T = [t_1 \ldots t_n]^T\},
\end{aligned}$$

where A is the $n \times n$ matrix given by $A_{k,i} = \alpha_{k,i}$ for all k, i. Let $p = (p_1, \ldots, p_n)^T$ be the unique element of $\mathbb{R}^n$ such that $Ap = t$. Then:

$$\begin{aligned} L_1 &= \bigcap_{k=1}^{n} (y_k^*)^{-1}(p_k) \\ &= p_1 u_1 + \ldots + p_n u_n + \overline{V^\perp}, \end{aligned}$$

where the last equality follows from (2.5).

If we let $t_1 = \ldots = t_n = 0$, we see that:

$$L = \overline{M} = \bigcap_{k=1}^{n} \operatorname{Ker}(y_k^*),$$

where $M = V^\perp$ is a closed subspace of codimension n in H. Moreover, $L \cap H = M$. To see that M is unique, suppose N is a subspace of codimension n in H such that $\overline{N} = L$ in B. Then $N \subset L \cap H = M$, so $N^\perp \supset M^\perp$ and since both $N^\perp$ and $M^\perp$ have dimension n, $N^\perp = M^\perp$ and then $N = M$.

(ii). Let M be a subspace of finite codimension n in H. Then $H = M \oplus M^\perp$ and $\dim(M^\perp) = n$, so $M^\perp \cap H_{B^*}$ is a subspace of H with dimension at most n. Suppose first that $\dim(M^\perp \cap H_{B^*}) = n$. Then $M^\perp \subset H_{B^*}$, so there is an orthonormal basis $\{h_{y_1^*}, \ldots, h_{y_n^*}\} \subset H_{B^*}$ for $M^\perp$, where $y_1^*, \ldots, y_n^* \in B^*$. Then by Lemma 2.5:

$$\overline{M} = \bigcap_{k=1}^{n} (y_k^*)^{-1}(0) = \bigcap_{k=1}^{n} \operatorname{Ker}(y_k^*).$$

So $\overline{M}$ is a closed subspace of codimension n in B.

Next, suppose $1 \leq k :=\dim(M^\perp \cap H_{B^*}) < n$. Then:

$$M \subset N := (M^\perp \cap H_{B^*})^\perp.$$

By our discussion above, $\overline{N}$ is a closed subspace of codimension $k \geq 1$ in B, therefore $\overline{M} \subset \overline{N}$ are both proper subspaces. Let $x \in B \setminus \overline{M}$. By the Hahn-Banach theorem, there is a non-zero $x^* \in B^*$ such that $(x, x^*) \neq 0$ and $x^*|_M = 0$. But then:

$$h_{x^*} \in M^\perp \cap H_{B^*} = N^\perp,$$

so $x^*|_N = 0$. Since $(x, x^*) \neq 0$, the Hahn-Banach theorem gives us that $x \in B \setminus \overline{N}$, so $\overline{M} \supset \overline{N}$. Then $\overline{M} = \overline{N}$, so $\overline{M}$ is a closed subspace of codimension k in B.

Finally, suppose $M^\perp \cap H_{B^*} = \{0\}$ and assume that $\overline{M}$ is a proper subspace of B. By the Hahn-Banach theorem, there is a non-zero $x^* \in B^*$ such that $x^*|_M = 0$. Then $h_{x^*} \in M^\perp \cap H_{B^*}$, but then $h_{x^*} = 0$ and $x^* = 0$, a contradiction. So $\overline{M} = B$. $\square$

Remark that in both Lemma 2.5 and Theorem 2.6, the discussion was purely topological and the measure μ played no role. Therefore, both results are valid in the more general setting of a real separable Hilbert space H with norm $\|\cdot\|$ and the Banach space B obtained by completing H with respect to a weaker norm $|\cdot|$.

Every hyperplane in H is of the form:

$$\xi_{p,u} = pu + u^{\perp},$$

where $u \in H$ is a unit vector (uniquely determined as the unit vector normal to $\xi_{p,u}$) and $p \geq 0$ is a non-negative real number (uniquely determined as the distance from $\xi_{p,u}$ to the origin). The result in Theorem 2.6, applied to closed subspaces of codimension 1, shows that every hyperplane in B is the B-closure of a hyperplane in H, that is every hyperplane in B is of the form:

$$\overline{\xi_{p,u}} = pu + \overline{u^{\perp}},$$

where $u \in H_{B^*}$. However, this relationship is not one-to-one and, in a sense, there are "more" hyperplanes in H than in B; specifically, if $\xi_{p,u}$ is a hyperplane in H, then:

- If $u \in H_{B^*}$ then the closure $\overline{\xi_{p,u}}$ is a hyperplane in B.
- If $u \notin H_{B^*}$ then $\xi_{p,u}$ is dense in B, that is the closure $\overline{\xi_{p,u}}$ is all of B.

Chapter 3
Properties of the Gaussian Radon Transform

Let us first explore some of the properties of the measures μ_{M_p} constructed in Theorem 2.4. The equation in (2.14) shows that every $x^* \in B^*$ is, with respect to μ_{M_p}, a (possibly degenerate) Gaussian random variable with mean $(p, x^*) = \langle p, h_{x^*} \rangle$ and variance $\|P_{M_0} h_{x^*}\|^2$. The degenerate case occurs when $x^* \in B^*$ is such that $x^*|_{M_0} = 0$. In this case, $h_{x^*} \in H_{B^*} \cap M_0^{\perp}$ and the distribution of x^* is the Dirac distribution $\delta_{(p,x^*)}$:

$$\int_B e^{ix^*} \, d\mu_{M_p} = e^{i(p,x^*)}.$$

So:

$$\text{If } x^* \in B^* \text{ satisfies } x^*|_{M_0} = 0, \text{ then } (x, x^*) = (p, x^*) \text{ for } \mu_{M_p}\text{-almost all } x \in B. \tag{3.1}$$

Let us now compute the covariance operator of the Gaussian measure μ_{M_p}; let $x^*, y^* \in B^*$ and apply (2.14) to $x^* + y^*$:

$$\begin{aligned} \mathbb{E}\left[(x^* + y^*)^2\right] &= (p, x^* + y^*)^2 + \|P_{M_0} h_{x^*+y^*}\|^2 \\ &= (p, x^*)^2 + (p, y^*)^2 + 2(p, x^*)(p, y^*) + \\ &\quad + \|P_{M_0} h_{x^*}\|^2 + \|P_{M_0} h_{y^*}\|^2 + 2\langle P_{M_0} h_{x^*}, P_{M_0} h_{y^*} \rangle, \end{aligned}$$

where all expectations are with respect to μ_{M_p}. In the same time:

$$\begin{aligned} \mathbb{E}\left[(x^* + y^*)^2\right] &= \mathbb{E}[(x^*)^2] + \mathbb{E}[(y^*)^2] + 2\mathbb{E}[x^* y^*] \\ &= (p, x^*)^2 + \|P_{M_0} h_{x^*}\|^2 + (p, y^*)^2 + \|P_{M_0} h_{y^*}\|^2 + 2\mathbb{E}[x^* y^*]. \end{aligned}$$

Then:

$$\mathbb{E}[x^* y^*] = (p, x^*)(p, y^*) + \langle P_{M_0} h_{x^*}, P_{M_0} h_{y^*} \rangle,$$

so the covariance of x^*, y^* with respect to μ_{M_p} is given by:

$$\mathrm{Cov}_{\mu_{M_p}}(x^*, y^*) = \langle P_{M_0} h_{x^*}, P_{M_0} h_{y^*} \rangle. \tag{3.2}$$

The measures μ_{M_p} in Theorem 2.4, while concentrated on closed affine subspaces of B, are all Borel probability measures on the same space B, which facilitates computations involving more than one of these measures. Next, we explore the relationship between the measures μ_{M_p} and Wiener measure μ. First of all, note that if we let $M_0 = H$ and $p = 0$ in Theorem 2.4, we obtain exactly Wiener measure $\mu_{M_0} = \mu$:

$$\int_B e^{ix^*} \, d\mu_{M_0} = e^{-\frac{1}{2}\|P_H h_{x^*}\|^2} = e^{-\frac{1}{2}\|h_{x^*}\|^2},$$

for all $x^* \in B^*$. Moreover, μ_{M_0} is concentrated on $\overline{H} = B$. The following result shows that even when $M_0 \subsetneqq H$ we obtain an abstract Wiener space.

Proposition 3.1. *Let (H, B, μ) be an abstract Wiener space and M_0 be a closed subspace of H. Then $(M_0, \overline{M_0}, \mu_{M_0})$ is an abstract Wiener space, where $\overline{M_0}$ is the closure of M_0 in B and μ_{M_0} is the measure constructed in Theorem 2.4, considered on the Borel σ-algebra of $\overline{M_0}$.*

Proof. First remark that:

$$\overline{M_0}^* = \{x^*|_{\overline{M_0}} : x^* \in B^*\}.$$

For every $x^* \in B^*$, the restriction $x^*|_{\overline{M_0}}$ is continuous on M_0 with respect to the Hilbert norm $\|\cdot\|$ and corresponds to $P_{M_0} h_{x^*}$ in M_0^*. To see this, note that:

$$(h, x^*|_{\overline{M_0}}) = (h, x^*) = \langle h, h_{x^*} \rangle = \langle h, P_{M_0} h_{x^*} \rangle,$$

for all $h \in M_0$. Moreover, the set $P_{M_0}(H_{B^*}) := \{P_{M_0} h_{x^*} : x^* \in B^*\}$ is dense in $(M_0, \|\cdot\|)$: if $h \in M_0$ is such that $h \perp P_{M_0}(H_{B^*})$, then $(h, x^*|_{\overline{M_0}}) = 0$ for all $x^* \in B^*$, so $h = 0$.

Now μ_{M_0} is a centered, non-degenerate Gaussian measure on $\overline{M_0}$ and the Cameron-Martin space H_0 of $(\overline{M_0}, \mu_{M_0})$ is given by:

$$H_0 = \{x \in \overline{M_0} : x^*|_{\overline{M_0}} \mapsto (x, x^*|_{\overline{M_0}}) \text{ is continuous on } \overline{M_0}^* \text{ with respect to } q\},$$

where q is the inner product on $\overline{M_0}^*$ induced by the covariance operator of μ_{M_p}:

$$q(x^*|_{\overline{M_0}}, y^*|_{\overline{M_0}}) = \langle P_{M_0} h_{x^*}, P_{M_0} h_{y^*} \rangle.$$

Clearly $M_0 \subset H_0$. Now let $x \in H_0$. Then $x \in \overline{M_0}$ and there is $c > 0$ such that:

$$|(x, x^*|_{\overline{M_0}})| \leq c \|P_{M_0} h_{x^*}\|, \text{ for all } x^* \in B^*.$$

Therefore the linear map:

$$P_{M_0}(H_{B^*}) \to \mathbb{R};\ P_{M_0} h_{x^*} \mapsto (x, x^*|_{\overline{M_0}})$$

is continuous on $P_{M_0}(H_{B^*})$ with respect to $\|\cdot\|$, so it extends uniquely to M_0. Then there is $h \in M_0$ such that:

$$(x, x^*|_{\overline{M_0}}) = \langle h, P_{M_0} h_{x^*} \rangle, \text{ for all } x^* \in B^*.$$

But $\langle h, P_{M_0} h_{x^*} \rangle = (h, x^*|_{\overline{M_0}})$, so:

$$(x, x^*|_{\overline{M_0}}) = (h, x^*|_{\overline{M_0}}), \text{ for all } x^* \in B^*.$$

Therefore $x = h \in M_0$, so $H_0 \subset M_0$. This shows that M_0 is indeed the Cameron-Martin space of $(\overline{M_0}, \mu_{M_0})$. □

Note that if $\overline{M_0}$ is a proper subspace of B then μ_{M_0} is a centered degenerate Gaussian measure on B. However, if $\overline{M_0} = B$ then (M_0, B, μ_{M_0}) is an abstract Wiener space. The distinction between these two cases will be relevant for the proof of our next result, which shows that μ_{M_0} and Wiener measure μ are equivalent (in fact, identical) if and only if $M_0 = H$, and are orthogonal otherwise.

Theorem 3.2. *Let (H, B, μ) be an abstract Wiener space, M_0 be a closed proper subspace of H, and $p \in M_0^\perp$. Then the measure μ_{M_p} is orthogonal to Wiener measure μ.*

Proof. Consider first the case when $\overline{M_0}$ is a proper subspace of B. By the Hahn-Banach theorem, there is a non-zero $x^* \in B^*$ such that $x^*|_{M_0} = 0$. From (3.1), μ_{M_p} assigns full measure 1 to the set $\{x \in B : (x, x^*) = (p, x^*)\}$, which has μ-measure 0, so $\mu_{M_p} \perp \mu$.

Now suppose $\overline{M_0} = B$. By Proposition 3.1, (M_0, B, μ_{M_0}) is an abstract Wiener space. If ν_1, ν_2 are centered, non-degenerate Gaussian measures on a real separable Banach space, then ν_1 and ν_2 are either equivalent or orthogonal; moreover, if $\nu_1 \sim \nu_2$ then ν_1 and ν_2 have identical Cameron-Martin spaces (see Theorem 2.7.2 and Proposition 2.7.3 in [7] for proofs). Since M_0 and H are the Cameron-Martin spaces of (B, μ_{M_0}) and (B, μ), respectively, and M_0 is a proper subspace of H, it follows that $\mu_{M_0} \perp \mu$.

Finally, if $0 \neq p \in M_0^\perp$ then $\mu_{M_p} \perp \mu_{M_0}$ by the Cameron-Martin theorem, so also $\mu_{M_p} \perp \mu$. □

Next, we remark that, as in the classical case, the map:

$$H_{B^*} \to L^2(B, \mu_{M_p});\ h_{x^*} \mapsto x^*,$$

is continuous with respect to the Hilbert norm $\|\cdot\|$ for every $p \in M_0^\perp$:

$$\begin{aligned} \|x^*\|^2_{L^2(B,\mu_{M_p})} &= \langle p, h_{x^*}\rangle^2 + \|P_{M_0} h_{x^*}\|^2 \\ &\leq \|p\|^2 \|h_{x^*}\|^2 + \|h_{x^*}\|^2 \\ &= (\|p\|^2 + 1)\|h_{x^*}\|^2. \end{aligned}$$

For every $p \in M_0^\perp$, we denote the extension of this map to H by:

$$I_{M_p} : H \to L^2(B, \mu_{M_p});\ h \mapsto I_{M_p} h. \tag{3.3}$$

Then every $I_{M_p} h$ is, with respect to μ_{M_p}, Gaussian with mean $\langle p, h\rangle$ and variance $\|P_{M_0} h\|^2$:

$$\int_B e^{i I_{M_p} h}\, d\mu_{M_p} = e^{i\langle p,h\rangle - \frac{1}{2}\|P_{M_0} h\|^2}. \tag{3.4}$$

To be more specific, for every $h \in H$ let $\{x_n^*\}_{n\in\mathbb{N}} \subset B^*$ be a sequence such that $h_{x_n^*}$ converges to h in H. Then the sequence $\{x_n^*\}_{n\in\mathbb{N}}$ is Cauchy in $L^2(B, \mu_{M_p})$ and converges to $I_{M_p} h$.

However, unlike the classical case, $I_{M_p} h$ is *not* an isometry:

$$\begin{aligned} \langle I_{M_p} h, I_{M_p} k\rangle_{L^2(B,\mu_{M_p})} &= \int_B (I_{M_p} h)(I_{M_p} k)\, d\mu_{M_p} \\ &= \lim_{n\to\infty} \int_B x_n^* y_n^*\, d\mu_{M_p} \\ &= \lim_{n\to\infty} \langle P_{M_0} h_{x_n^*}, P_{M_0} h_{y_n^*}\rangle \\ &= \langle P_{M_0} h, P_{M_0} k\rangle, \end{aligned}$$

where $\{x_n^*\}_{n\in\mathbb{N}}$ and $\{y_n^*\}_{n\in\mathbb{N}}$ are sequences in B^* such that $h_{x_n^*}$ and $h_{y_n^*}$ converge to h and k in H, respectively.

Proposition 3.3. *Let (H, B, μ) be an abstract Wiener space and $F_1 \subset F_2 \subset \dots$ be a measurably adapted sequence of closed subspaces of H. Then:*

$$\lim_{n\to\infty} \mu_{F_n^\perp}[x \in B : |x| > R] = 0, \tag{3.5}$$

for any $R > 0$.

Proof. As in the proof of Theorem 2.4, let $\{e_k\}_{k\in\mathbb{N}}$ be an orthonormal sequence in H such that $\{e_1, \dots, e_{k_1}\}$ is an orthonormal basis for $F_2 \cap F_1^\perp$ and $\{e_{k_n+1}, \dots, e_{k_{n+1}}\}$ is an orthonormal basis for $F_{n+2} \cap F_{n+1}^\perp$, for all $n \geq 1$, where $k_1 < k_2 < \dots$ is an increasing sequence of positive integers.

As proved in Theorem 2.4, the measure $\mu_{F_n^\perp}$ is the distribution of the B-valued random variable:

$$Z_{F_n^\perp} := \sum_{j=n-1}^{\infty} \left(\sum_{l=k_j+1}^{k_{j+1}} Z_l e_l \right),$$

where $\{Z_k\}_{k\in\mathbb{N}}$ is an independent sequence of standard Gaussian random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For each j, the term:

$$S_j = \sum_{l=k_j+1}^{k_{j+1}} Z_l e_l,$$

takes values in $Q_{j+1} = F_{j+2} \cap F_{j+1}^\perp$ and its distribution measure is exactly standard Gaussian measure $\gamma_{Q_{j+1}}$ on this space. By the measurably adapted property in (2.2):

$$\mathbb{P}\left[|S_j| > \frac{1}{2^{j+1}}\right] < \frac{1}{2^{j+1}}, \tag{3.6}$$

for all $j \in \mathbb{N}$.

Let $R > 0$ and choose $N \in \mathbb{N}$ large enough such that:

$$R > \frac{1}{2^{N-1}}.$$

Now if:

$$\omega \in \bigcap_{j=N-1}^{\infty} \left[|S_j| \leq \frac{1}{2^{j+1}}\right],$$

then:

$$\left|Z_{F_n^\perp}(\omega)\right| \leq \sum_{j=N-1}^{\infty} |S_j| \leq \sum_{j=N-1}^{\infty} \frac{1}{2^{j+1}} = \frac{1}{2^{N-1}}.$$

So:

$$\left[|Z_{F_n^\perp}| > \frac{1}{2^{N-1}}\right] \subseteq \bigcup_{j=N-1}^{\infty} \left[|S_j| > \frac{1}{2^{j+1}}\right],$$

therefore:

$$\begin{aligned}
\mathbb{P}\left[|Z_{F_n^\perp}|>R\right] &\le \mathbb{P}\left[|Z_{F_n^\perp}|>\frac{1}{2^{N-1}}\right]\\
&\le \sum_{j=N-1}^{\infty}\mathbb{P}\left[|S_j|>\frac{1}{2^{j+1}}\right]\\
&< \sum_{j=N-1}^{\infty}\frac{1}{2^{j+1}} && \text{(by (3.6))}\\
&= \frac{1}{2^{N-1}}.
\end{aligned}$$

We showed that for all $N\in\mathbb{N}$ such that $\frac{1}{2^{N-1}}<R$:

$$\mathbb{P}\left[|Z_{F_n^\perp}|>R\right]<\frac{1}{2^{N-1}}\xrightarrow{N\to\infty}0,$$

which translates to (3.5). $\square$

Corollary 3.4. *Let (H,B,μ) be an abstract Wiener space and f be a bounded Borel function on B. If $F_1\subset F_2\subset\dots$ is a measurably adapted sequence of subspaces of H, then:*

$$\begin{aligned}
f(p) &= \lim_{n\to\infty} Gf(p+F_n^\perp)\\
&= \lim_{n\to\infty}\int_B f\,d\mu_{p+F_n^\perp}, \qquad (3.7)
\end{aligned}$$

for all $p\in B$ such that f is continuous at p.

Proof. Using the translation property in (2.21):

$$\int_B f\,d\mu_{p+F_n^\perp}-f(p)=\int_B\left(f(x+p)-f(p)\right)\,d\mu_{F_n^\perp}(x),$$

for all n. Let $\epsilon>0$. Since f is continuous at p, there is $\delta>0$ such that $|f(x+p)-f(p)|<\epsilon$ for all $x\in B$ with $|x|\le\delta$. Then:

$$\begin{aligned}
\left|\int_B f\,d\mu_{p+F_n^\perp}-f(p)\right| &\le \int_B|f(x+p)-f(p)|\,d\mu_{F_n^\perp}(x)\\
&= \int_{[|x|\le\delta]}|f(x+p)-f(p)|\,d\mu_{F_n^\perp}(x)+\\
&\quad+\int_{[|x|>\delta]}|f(x+p)-f(p)|\,d\mu_{F_n^\perp}(x)\\
&< \epsilon+\int_{[|x|>\delta]}\left(|f(x+p)|+|f(p)|\right)\,d\mu_{F_n^\perp}(x)\\
&\le \epsilon+2\|f\|_\infty\mu_{F_n^\perp}[x\in B:|x|>\delta],
\end{aligned}$$

where $\|f\|_\infty := \sup_{x\in B}|f(x)|$. Since this holds for all $\epsilon > 0$ and, by Proposition 3.3:

$$\lim_{n\to\infty} \mu_{F_n^\perp}[x \in B : |x| > \delta] = 0,$$

we obtain (3.7). □

Next, we look at some inequalities. The following result is the celebrated Fernique Theorem. For a proof, see III, Theorem 3.1 in [21].

Theorem 3.5. *Let (H, B, μ) be an abstract Wiener space. Then there is $\alpha > 0$ such that:*

$$\int_B e^{\alpha|x|^2}\, d\mu(x) < \infty,$$

where $|\cdot|$ is the norm on B.

As a consequence:

$$\int_B |x|^t\, d\mu(x) < \infty,$$

for all $t > 0$. Now suppose M_0 is a closed subspace of H. As noted in Proposition 3.1, $(M_0, \overline{M_0}, \mu_{M_0})$ is an abstract Wiener space. By Fernique's Theorem:

$$\int_B e^{\alpha|x|^2}\, d\mu_{M_0}(x) = \int_{\overline{M_0}} e^{\alpha|x|^2}\, d\mu_{M_0}(x) < \infty,$$

and consequently:

$$\int_B |x|^t\, d\mu_{M_0}(x) < \infty,$$

for all $t > 0$.

Let B_1^* denote the closed unit ball in B^*:

$$B_1^* := \{x^* \in B^* : |x^*|_* \le 1\},$$

where $|\cdot|_*$ is the usual operator norm on B^*. Then:

$$\{x^*\}_{x^*\in B_1^*} \tag{3.8}$$

is a centered Gaussian process on (B, μ_{M_0}). Moreover, this process is bounded (that is, all sample paths are bounded): for a fixed $x \in B$,

$$|(x, x^*)| \le |x||x^*|_* \le |x|, \text{ for all } x^* \in B_1^*.$$

Recall that for every closed subspace M_0 of H we constructed in Theorem 2.4 a random variable:

$$Z_{M_0} : \Omega \to \overline{M_0} \subset B,$$

on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with values in the closure $\overline{M_0}$ of M_0 in B, and μ_{M_0} is then defined by:

$$\mu_{M_0}(E) := \mathbb{P}[Z_{M_0} \in E],$$

for all Borel subsets $E \subset B$. Then:

$$\int_B f \, d\mu_{M_0} = \int_\Omega f(Z_{M_0}) \, d\mathbb{P},$$

whenever either side exists. Therefore:

$$\{x^*(Z_{M_0})\}_{x^* \in B_1^*}$$

is a bounded centered Gaussian process on $(\Omega, \mathcal{F}, \mathbb{P})$ and:

$$\int_\Omega (x^*(Z_{M_0}))^2 \, d\mathbb{P} = \int_B (x^*)^2 \, d\mu_{M_0} = \|P_{M_0} h_{x^*}\|^2,$$

for all $x^* \in B_1^*$. This allows us to compare properties of the Gaussian process in (3.8) considered on (B, μ_{L_0}) and (B, μ_{M_0}) for different closed subspaces L_0 and M_0 of H and employ another famous result, the Sudakov-Fernique inequality:

Theorem 3.6. *Let $\{X_t\}_{t \in T}$ and $\{Y_t\}_{t \in T}$ be almost surely bounded centered Gaussian processes such that:*

$$\mathbb{E}[(X_t - X_s)^2] \leq \mathbb{E}[(Y_t - Y_s)^2],$$

for all $s, t \in T$. Then:

$$\mathbb{E}\left[\sup_{t \in T} X_t\right] \leq \mathbb{E}\left[\sup_{t \in T} Y_t\right].$$

See [1] for a detailed proof of this result, which we use to obtain the next theorem.

Theorem 3.7. *Let (H, B, μ) be an abstract Wiener space and $L_0 \subset M_0$ be closed subspaces of H. Then:*

$$\int_B |x| \, d\mu_{L_0}(x) \leq \int_B |x| \, d\mu_{M_0}(x). \tag{3.9}$$

Proof. Recall that:

$$|x| = \sup_{x^* \in B_1^*} |(x, x^*)|,$$

for all $x \in B$. So if we consider the Gaussian process (3.8) on (B, μ_{M_0}) for any closed subspace M_0 of H, then:

$$\left(\sup_{x^* \in B_1^*} x^*\right)(x) = \sup_{x^* \in B_1^*} (x, x^*) = \sup_{x^* \in B_1^*} |(x, x^*)| = |x|.$$

Now consider:

$$\{x^*(Z_{L_0})\}_{x^* \in B_1^*} \text{ and } \{x^*(Z_{M_0})\}_{x^* \in B_1^*},$$

both bounded, centered Gaussian processes on $(\Omega, \mathcal{F}, \mathbb{P})$. For every $x^*, y^* \in B_1^*$:

$$\begin{aligned}
\mathbb{E}\left[(x^*(Z_{L_0}) - y^*(Z_{L_0}))^2\right] &= \|P_{L_0}(h_{x^*} - h_{y^*})\|^2 \\
&\leq \|P_{M_0}(h_{x^*} - h_{y^*})\|^2 = \mathbb{E}\left[(x^*(Z_{M_0}) - y^*(Z_{M_0}))^2\right].
\end{aligned}$$

By Sudakov-Fernique:

$$\mathbb{E}\left[\sup_{x^*\in B_1^*} x^*(Z_{L_0})\right] \leq \mathbb{E}\left[\sup_{x^*\in B_1^*} x^*(Z_{M_0})\right],$$

so:

$$\int_B \left(\sup_{x^*\in B_1^*} x^*\right) d\mu_{L_0} \leq \int_B \left(\sup_{x^*\in B_1^*} x^*\right) d\mu_{M_0},$$

which is exactly (3.9). $\square$

3.1 A Disintegration of Wiener Measure

In this section we focus on the measures μ_{M_p} in the case when we are dealing with subspaces of finite codimension. First, we provide a disintegration of μ_{M_0} through the measures μ_{L_p}, where L_0 is a subspace of M_0 having finite codimension. As a particular case, we will have a disintegration of Wiener measure through the measures μ_{Q_p}, where Q_0 is a subspace of finite codimension in H.

Theorem 3.8. *Let (H, B, μ) be an abstract Wiener space and $L_0 \subset M_0$ be closed subspaces of H such that L_0 has finite codimension in M_0, that is the subspace $K_0 := L_0^\perp \cap M_0$ is finite-dimensional. For every $p \in K_0$, consider the translate $L_p = p + L_0$ of L_0. Then the map:*

$$K_0 \ni p \mapsto Gf(L_p) = \int_B f\, d\mu_{L_p}, \tag{3.10}$$

is Borel measurable on K_0 for all non-negative Borel functions f on B. Moreover:

$$\int_B f\, d\mu_{M_0} = \int_{K_0} \left(\int_B f\, d\mu_{L_p}\right) d\gamma_{K_0}(p), \tag{3.11}$$

for all Borel functions for which the left hand side exists, where γ_{K_0} denotes standard Gaussian measure on K_0.

In particular, if Q_0 is a closed subspace of finite codimension in H:

$$\int_B f\, d\mu = \int_{Q_0^\perp} \left(\int_B f\, d\mu_{Q_p}\right) d\gamma_{Q_0^\perp}, \tag{3.12}$$

whenever the left hand side exists, where μ is Wiener measure.

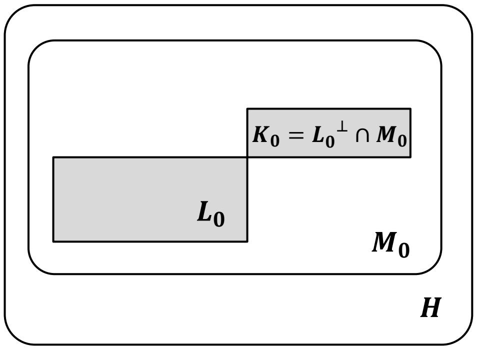
These results are pictured below in Figure 3.1.

Proof. Let f be a non-negative Borel function on B. To prove measurability of the map in (3.10), consider the map $g : (B, \mu_{L_0}) \times (K_0, \gamma_{K_0}) \to \mathbb{R}_+$ given by:

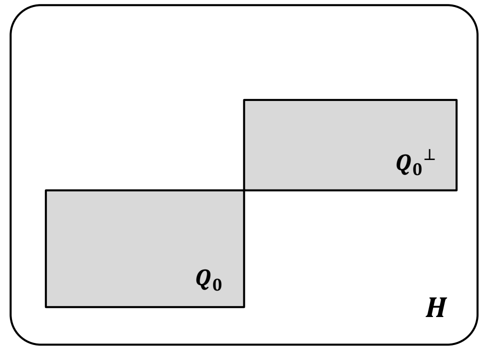
$$g(x, p) = f(x + p), \text{ for all } x \in B, p \in K_0.$$

By Fubini's theorem, the map:

$$p \mapsto \int_B g(x, p)\, d\mu_{L_0}(x) = \int_B f(x + p)\, d\mu_{L_0}(x) = \int_B f\, d\mu_{L_p} = Gf(L_p),$$



(a) Closed subspaces $L_0 \subset M_0 \subset H$ where $K_0 = L_0^\perp \cap M_0$ is finite-dimensional.



(b) Closed subspace $Q_0 \subset H$ with finite codimension.

Figure 3.1: Disintegration theorem

is Borel measurable, where the second equality follows from (2.21).

To prove (3.11), it suffices to show that the characteristic functional of the Borel probability measure μ'_{M_0} on B specified by:

$$\int_B b\, d\mu'_{M_0} = \int_{K_0} Gb(L_p)\, d\gamma_{K_0}(p),$$

for all bounded Borel functions b, coincides with that of μ_{M_0}. To see this, note that for all $x^* \in B^*$:

$$\begin{aligned}
\int_B e^{i(x,x^*)}\, d\mu'_{M_0}(x) &= \int_{K_0} \left(\int_B e^{i(x,x^*)}\, d\mu_{L_p}(x) \right) d\gamma_{K_0}(p) \\
&= \int_{K_0} e^{i(p,x^*) - \frac{1}{2}\|P_{L_0} h_{x^*}\|^2}\, d\gamma_{K_0}(p) && \text{(by (2.14))} \\
&= e^{-\frac{1}{2}\|P_{L_0} h_{x^*}\|^2} \int_{K_0} e^{i\langle p, P_{K_0} h_{x^*}\rangle}\, d\gamma_{K_0} \\
&= e^{-\frac{1}{2}\|P_{L_0} h_{x^*}\|^2} e^{-\frac{1}{2}\|P_{K_0} h_{x^*}\|^2} \\
&= e^{-\frac{1}{2}\|P_{M_0} h_{x^*}\|^2},
\end{aligned}$$

which proves our claim. Finally, (3.12) follows by taking $M_0 = H$ and $L_0 = Q_0$ in (3.11). $\square$

Next, we explore some of the consequences of this result.

Corollary 3.9. *Let (H, B, μ) be an abstract Wiener space and Q_0 be a closed subspace of finite codimension in H. Then for every measurable $f : B \to \mathbb{R}$ with $\int_B |f|^r\, d\mu < \infty$, where $1 \leq r < \infty$, we have:*

$$\|f\|_{L^r(B,\mu_{Q_p})} < \infty \text{ for } \gamma_{Q_0^\perp}\text{-a.a. } p \in Q_0^\perp, \tag{3.13}$$

and:
$$p \mapsto Gf(Q_p) \in L^r\left(Q_0^\perp, \gamma_{Q_0^\perp}\right). \tag{3.14}$$
In particular, if $f = 0$ μ-a.e. for some measurable function f on B, then $Gf(Q_p) = 0$ for $\gamma_{Q_0^\perp}$-a.a $p \in Q_0^\perp$.

Proof. Using $|f|^r$ in place of f in (3.11) we obtain:
$$\int_B |f|^r \, d\mu = \int_{Q_0^\perp} \left(\int_B |f(x)|^r \, d\mu_{Q_p}(x) \right) d\gamma_{Q_0^\perp}(p).$$
So:
$$\|f\|^r_{L^r(B,\mu)} = \int_{Q_0^\perp} \|f\|^r_{L^2(B,\mu_{Q_p})} \, d\gamma_{Q_0^\perp}(p) < \infty. \tag{3.15}$$
Consequently, the map:
$$p \mapsto \|f\|^r_{L^r(B,\mu_{Q_p})},$$
is $\gamma_{Q_0^\perp}$-a.e. finite. Moreover:
$$\begin{aligned}
\int_{Q_0^\perp} |Gf(Q_p)|^r \, d\gamma_{Q_0^\perp}(p) &= \int_{Q_0^\perp} \left| \int_B f \, d\mu_{Q_p} \right|^r d\gamma_{Q_0^\perp}(p) \\
&\leq \int_{Q_0^\perp} \left(\int_B |f|^r \, d\mu_{Q_p} \right) d\gamma_{Q_0^\perp}(p) \\
&= \int_B |f|^r \, d\mu < \infty,
\end{aligned}$$
which proves (3.14). We proved above that:
$$\|Gf(Q_p)\|_{L^r(Q_0^\perp, \gamma_{Q_0^\perp})} \leq \|f\|_{L^r(B,\mu)}. \tag{3.16}$$
Now the last statement in the theorem follows readily from (3.16): if $f = 0$ μ-a.e. then $\|f\|_{L^r(B,\mu)} = 0$, so:
$$\|Gf(Q_p)\|_{L^r(Q_0^\perp, \gamma_{Q_0^\perp})} = 0,$$
and then $Gf(Q_p) = 0$ for $\gamma_{Q_0^\perp}$-a.a. $p \in Q_0^\perp$. $\square$

Note that this result also implies that if $E \subset B$ is a Borel subset with $\mu(E) = 0$, then $\mu_{Q_p}(E) = 0$ for $\gamma_{Q_0^\perp}$-a.a. $p \in Q_0^\perp$. To see this, remark that $\mathbb{1}_E = 0$ μ-a.e. so $G\mathbb{1}_E(Q_p) = \mu_{Q_p}(E) = 0$ for almost all p.

Corollary 3.10. *Let (H, B, μ) be an abstract Wiener space, Q_0 be a closed subspace of finite codimension in H and $h \in H$. If $\tilde{h}$ is any representative of Ih in $L^2(B, \mu)$, then $\tilde{h}$ is a representative of $I_{Q_p}h$ in $L^2(B, \mu_{Q_p})$ for $\gamma_{Q_0^\perp}$-a.a. $p \in Q_0^\perp$. Moreover:*
$$\text{for } \gamma_{Q_0^\perp}\text{-a.a. } p \in Q_0^\perp\text{: } \tilde{h}(x) = \langle p, h \rangle \text{ for } \gamma_{Q_0^\perp}\text{-a.a. } x \in B, \tag{3.17}$$
holds whenever $h \in Q_0^\perp$.

Proof. For every $p \in Q_0^\perp$, let $\tilde{h}_{Q_p}$ be a representative of $I_{Q_p}h$ in $L^2(B, \mu_{Q_p})$. Let $\{h_{x_k^*}\}_{k\geq 1}$ be a sequence in H_{B^*} converging to h in H. Then:

$$\lim_{k\to\infty} \|\tilde{h} - x_k^*\|_{L^2(B,\mu)} = 0, \tag{3.18}$$

and:

$$\lim_{k\to\infty} \|\tilde{h}_{Q_p} - x_k^*\|_{L^2(B,\mu_{Q_p})} = 0, \tag{3.19}$$

for all $p \in Q_0^\perp$.

Now note that if $\tilde{h}$ is defined μ-a.e. on B, Corollary 3.9 shows that for $\gamma_{Q_0^\perp}$-a.a. $p \in Q_0^\perp$, $\tilde{h}$ is defined μ_{Q_p}-a.e. on B and $\tilde{h} \in L^2(B, \mu_{Q_p})$. From the disintegration formula:

$$\int_B |\tilde{h} - x_k^*|^2 \, d\mu = \int_{Q_0^\perp} \left(\int_B |\tilde{h} - x_k^*|^2 \, d\mu_{Q_p} \right) d\gamma_{Q_0^\perp}(p),$$

so:

$$\|\tilde{h} - x_k^*\|^2_{L^2(B,\mu)} = \int_{Q_0^\perp} \|\tilde{h} - x_k^*\|^2_{L^2(B,\mu_{Q_p})} \, d\gamma_{Q_0^\perp}(p) = \|g_k\|^2_{L^2(Q_0^\perp, \gamma_{Q_0^\perp})},$$

where

$$g_k(p) = \|\tilde{h} - x_k^*\|_{L^2(B,\mu_{Q_p})} \text{ for all } p \in Q_0^\perp,$$

is defined $\gamma_{Q_0^\perp}$-a.e. From (3.18):

$$\lim_{k\to\infty} g_k = 0 \text{ in } L^2(Q_0^\perp, \gamma_{Q_0^\perp}). \tag{3.20}$$

Now since $x_k^* \to \tilde{h}_{Q_p}$ in $L^2(B, \mu_{Q_p})$, $(\tilde{h} - x_k^*)$ converges to $(\tilde{h} - \tilde{h}_{Q_p})$ in $L^2(B, \mu_{Q_p})$ for $\gamma_{Q_0^\perp}$-a.a. $p \in Q_0^\perp$ (namely for all p such that $\tilde{h}$ is defined μ_{Q_p}-a.e.). Therefore:

$$\lim_{k\to\infty} g_k(p) = \|\tilde{h} - \tilde{h}_{Q_p}\|_{L^2(B,\mu_{Q_p})} \text{ for } \gamma_{Q_0^\perp}\text{-a.a. } p \in Q_0^\perp. \tag{3.21}$$

From (3.20) and (3.21), since mean-square limits and pointwise-a.e. limits agree, we have:

$$\|\tilde{h} - \tilde{h}_{Q_p}\|_{L^2(B,\mu_{Q_p})} = 0 \text{ for } \gamma_{Q_0^\perp}\text{-a.a. } p \in Q_0^\perp.$$

Therefore $\tilde{h}$ is a representative of $I_{Q_p}h$ for $\gamma_{Q_0^\perp}$-a.a. $p \in Q_0^\perp$.

Finally, recall that any representative of $I_{Q_p}h$ is, with respect to μ_{Q_p}, Gaussian with mean $\langle p, h\rangle$ and variance $\|P_{Q_0}h\|^2$. Therefore, if $h \in Q_0^\perp$ then $\|P_{Q_0}h\| = 0$ and then any representative of $I_{Q_p}h$ is μ_{Q_p}-a.e. equal to $\langle p, h\rangle$. Since $\tilde{h}$ is a representative of $I_{Q_p}h$ for $\gamma_{Q_0^\perp}$-a.a. $p \in Q_0^\perp$, (3.17) follows.

□

Remark 3.1. In light of these calculations, it is very tempting to say something like

"$Ih = I_{Q_p}h$ for almost all p."

However, Ih and $I_{Q_p}h$ are elements of $L^2(B,\mu)$ and $L^2(B,\mu_{Q_p})$, respectively, and these are spaces whose elements are functions defined almost everywhere with respect to *different* measures - so the statement above makes little sense. For the sake of accuracy, we remained sensitive to the true quotient-space structure of L^2-spaces and stated the result as in Corollary 3.10.

We can also use the disintegration theorem to give an alternate proof of the inequality in Theorem 3.7, and in fact strengthen it. Instead of using the Sudakov-Fernique inequality approach, we will use Anderson's inequality, which we state next and for which a proof can be found in [7].

Theorem 3.11. *Let μ be a centered Gaussian measure on a real separable Banach space B. Then for every symmetric convex Borel set $C \subset B$ and $p \in B$:*

$$\mu(C) \geq \mu(C-p).$$

Now suppose (H,B,μ) is an abstract Wiener space and $L_0 \subset M_0$ are closed subspaces of H and L_0 has finite codimension in M_0. By taking $f = \mathbb{1}_E$ in (3.11) for some Borel subset E of B:

$$\mu_{M_0}(E) = \int_{K_0} \mu_{L_p}(E)\, d\gamma_{K_0}(p),$$

where γ_{K_0} is standard Gaussian measure on $K_0 := L_0^{\perp} \cap M_0$. If C is a symmetric convex Borel subset of B, Anderson's inequality gives us:

$$\mu_{L_0}(C) \geq \mu_{L_0}(C-p) = \mu_{L_p}(C),$$

for all $p \in L_0^{\perp}$. Then:

$$\mu_{M_0}(C) \leq \int_{K_0} \mu_{L_0}(C)\, d\gamma_{K_0}(p) = \mu_{L_0}(C),$$

for all symmetric convex Borel subsets C of B. For every real t, the set:

$$[x \in B : |x| \leq t]$$

is symmetric, convex and Borel in B, so:

$$\mu_{M_0}[x \in B : |x| \leq t] \leq \mu_{L_0}[x \in B : |x| \leq t].$$

Then for all $p > 0$:

$$p\int_0^\infty t^{p-1}\mu_{L_0}[x \in B : |x| > t]\, dt \leq p\int_0^\infty t^{p-1}\mu_{M_0}[x \in B : |x| > t]\, dt,$$

which is equivalent to:

$$\int_B |x|^p\, d\mu_{L_0}(x) \leq \int_B |x|^p\, d\mu_{M_0}(x).$$

We have just proved the following:

Theorem 3.12. *Let (H,B,μ) be an abstract Wiener space and $L_0 \subset M_0$ be closed subspaces of H such that $dim(L_0^{\perp} \cap M_0) < \infty$. Then:*

$$\int_B |x|^p\, d\mu_{L_0}(x) \leq \int_B |x|^p\, d\mu_{M_0}(x), \tag{3.22}$$

for all $p > 0$.

3.2 The Gaussian Radon Transform and Conditional Expectation

In this section we show how the Gaussian Radon transform may be interpreted as a conditional expectation. We begin with a short review of conditional expectations.

Definition 3.1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\mathcal{G} \subset \mathcal{F}$ be a subfield and X be an integrable random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. The *conditional expectation of X given $\mathcal{G}$*, denoted

$$\mathbb{E}[X|\mathcal{G}],$$

is any random variable Y that is $\mathcal{G}$-measurable and satisfies:

$$\int_A Y\, d\mathbb{P} = \int_A X\, d\mathbb{P}, \text{ for all } A \in \mathcal{G}. \tag{3.23}$$

It can be shown that the conditional expectation $\mathbb{E}[X|\mathcal{G}]$ exists and is unique, in the sense that if Y and Y' are $\mathcal{G}$-measurable random variables satisfying (3.23) then $Y = Y'$ almost surely. Any such random variable is called a *version* of $\mathbb{E}[X|\mathcal{G}]$. Note that whenever we write:

$$\text{“}\mathbb{E}[X|\mathcal{G}] = Y,\text{”}$$

this is to be understood in the sense of equality a.s.

The expectation $\mathbb{E}[X]$ of a random variable X is often used as the “best guess” of the value of X, given no other information. However, if we do have some other information, this guess can be replaced by a more useful one - the conditional expectation. Intuitively, the subfield $\mathcal{G}$ above represents the information that we have available (that is, for every event A in $\mathcal{G}$, we know whether or not A occurred) and then $\mathbb{E}[X|\mathcal{G}]$ is our “best guess” for the value of X given our knowledge of $\mathcal{G}$.

The most common occurrence of conditional expectations is conditioning on another random variable (or more). Specifically, if $\{Y_t\}_{t\in T}$ is a collection of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ and X is integrable, we define:

$$\mathbb{E}\left[X|Y_t\, : t \in T\right] := \mathbb{E}\left[X|\sigma(Y_t\, : t \in T)\right], \tag{3.24}$$

where $\sigma(Y_t\, : t \in T)$ is the σ-algebra generated by the collection $\{Y_t\}_{t\in T}$.

Now suppose $Y_1, Y_2, \ldots, Y_n$ are random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, where n is a positive integer. In this case, there is a Borel measurable function $g : \mathbb{R}^n \to \mathbb{R}$ such that:

$$\mathbb{E}\left[X|Y_1, \ldots, Y_n\right] = g(Y_1, \ldots, Y_n),$$

almost surely (see [6, Theorem 20.1]). We denote the conditional expectation:

$$\mathbb{E}\left[X|Y_1 = y_1, \ldots, Y_n = y_n\right] = g(y_1, \ldots, y_n),$$

to illustrate this point.

There are many interesting and useful properties of conditional expectations, such as "conditional" versions of the major convergence theorems and inequalities of measure theory (see [6, Section 34] or [11, Section 4.1]). One such property that will be of particular interest to us is the geometric interpretation of conditional expectations in the Hilbert space setting of L^2-spaces. Specifically, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{G} \subset \mathcal{F}$ be a subfield. Then $L^2(\Omega, \mathcal{G}, \mathbb{P})$ is a closed subspace of $L^2(\Omega, \mathcal{F}, \mathbb{P})$, so for every $f \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ we may consider its orthogonal projection onto $L^2(\Omega, \mathcal{G}, \mathbb{P})$. This is the point Y in $L^2(\Omega, \mathcal{G}, \mathbb{P})$ that is "closest" to f:

$$\|f - Y\|_{L^2(\Omega,\mathcal{F},\mathbb{P})} = \inf_{g \in L^2(\Omega,\mathcal{G},\mathbb{P})} \|g - f\|_{L^2(\Omega,\mathcal{F},\mathbb{P})}.$$

It turns out that this is exactly the conditional expectation:

$$\mathbb{E}[f|\mathcal{G}] = P_{L^2(\Omega,\mathcal{G},\mathbb{P})} f \text{ in } L^2(\Omega, \mathcal{F}, \mathbb{P}). \tag{3.25}$$

For a proof of this fact, see [11, Section 4.1, Theorem (1.4)].

Now suppose that $X, Y_1, \ldots, Y_n$ are random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ and X is integrable. Then if $X', Y_1', \ldots, Y_n'$ are any other random variables such that $X = X'$ a.s. and $Y_k = Y_k'$ a.s. for $1 \leq k \leq n$, any version of $\mathbb{E}[X|Y_1, \ldots, Y_n]$ is also a version of $\mathbb{E}[X'|Y_1 \ldots, Y_n']$. In light of this fact and (3.25), conditioning over elements of $L^2(\Omega)$ - which are equivalence classes of functions, not functions - is usually defined as follows.

Definition 3.2. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $f, f_1, \ldots, f_n \in L^2(\Omega)$. Let $M(f_1, \ldots, f_n)$ be the closed subspace of $L^2(\Omega)$ consisting of all $g(f_1, \ldots, f_n)$ where $g : \mathbb{R}^n \to \mathbb{R}$ is Borel and $\mathbb{E}[g^2(f_1, \ldots, f_n)] < \infty$. The conditional expectation of f given $f_1, \ldots, f_n$ is defined as the element of $M(f_1, \ldots, f_n)$ that is closest to f:

$$\mathbb{E}[f|f_1, \ldots, f_n] = P_{M(f_1,\ldots,f_n)} f \text{ in } L^2(\Omega). \tag{3.26}$$

In this case, for $y_1, \ldots, y_n \in \mathbb{R}$, we write as before:

$$\mathbb{E}[f|f_1 = y_1, \ldots, f_n = y_n] \tag{3.27}$$

as a (more intuitive) notation for a function $g(y_1, \ldots, y_n)$ such that $g \circ (f_1, \ldots, f_n)$ is a version of $\mathbb{E}[f|f_1, \ldots, f_n]$.

We now turn to the relationship between the Gaussian Radon transform and conditional expectation.

Lemma 3.13. *Let (H, B, μ) be an abstract Wiener space, Q_0 be a closed subspace of finite codimension in H and $f \in L^2(B, \mu)$. For every $y_1, \ldots, y_n \in \mathbb{R}$ let:*

$$F(y_1, \ldots, y_n) := Gf(y_1 u_1 + \ldots + y_n u_n + Q_0),$$

where $\{u_1, \ldots, u_n\}$ is an orthonormal basis for $Q_0^\perp$. Then $F(Iu_1, \ldots, Iu_n)$ is a version of $\mathbb{E}[f|Iu_1, \ldots, Iu_n]$:

$$\mathbb{E}[f|Iu_1 = y_1, \ldots, Iu_n = y_n] = Gf(y_1 u_1 + \ldots + y_n u_n + Q_0). \tag{3.28}$$

Recall from Corollary 3.9 that if $f \in L^2(B,\mu)$ then the map $p \mapsto Gf(Q_p)$ belongs to $L^2(Q_0^\perp, \gamma_{Q_0^\perp})$, so the function F above is defined almost everywhere on $\mathbb{R}^n$.

Proof. For simplicity of notation, let $\widetilde{u_k}$ be a representative of Iu_k in $L^2(B,\mu)$ for every $1 \le k \le n$. Since $u_k \in Q_0^\perp$, by (3.17):

$$\text{For } \gamma_{Q_0^\perp}\text{-a.a. } p \in Q_0^\perp\text{: } \widetilde{u_k}(x) = \langle p, u_k\rangle \text{ for } \mu_{Q_p}\text{-a.a. } x \in B. \tag{3.29}$$

Let $g(\widetilde{u_1},\dots,\widetilde{u_n})$, where $g : \mathbb{R}^n \to \mathbb{R}$, such that $\mathbb{E}[g^2(\widetilde{u_1},\dots,\widetilde{u_n})] < \infty$. By (3.11) and (3.29):

$$\begin{aligned}\int_B g(\widetilde{u_1},\dots,\widetilde{u_n}) f \, d\mu &= \int_{Q_0^\perp} \left(\int_B g(\widetilde{u_1},\dots,\widetilde{u_n}) f \, d\mu_{Q_p} \right) d\gamma_{Q_0^\perp}(p) \\ &= \int_{Q_0^\perp} g\left(\langle p, u_1\rangle,\dots,\langle p, u_n\rangle\right) Gf(Q_p)\, d\gamma_{Q_0^\perp}(p).\end{aligned} \tag{3.30}$$

Now consider the map:

$$\widetilde{P_{Q_0^\perp}} : B \to Q_0^\perp;\ \widetilde{P_{Q_0^\perp}} := \widetilde{u_1}u_1 + \ldots + \widetilde{u_n}u_n$$

(defined μ-a.e. on B). Since $\widetilde{u_1},\dots,\widetilde{u_n}$ are independent standard Gaussian random variables on B, he distribution measure of $\widetilde{P_{Q_0^\perp}}$ on $Q_0^\perp$ is exactly standard Gaussian measure $\gamma_{Q_0^\perp}$. Then (3.30) yields:

$$\begin{aligned}\int_B g(\widetilde{u_1},\dots,\widetilde{u_n}) f \, d\mu &= \int_B g\left(\langle \widetilde{P_{Q_0^\perp}}, u_1\rangle,\dots,\langle \widetilde{P_{Q_0^\perp}}, u_n\rangle\right) Gf\left(\widetilde{P_{Q_0^\perp}} + Q_0\right) d\mu \\ &= \int_B g(\widetilde{u_1},\dots,\widetilde{u_n}) F(\widetilde{u_1},\dots,\widetilde{u_n})\, d\mu.\end{aligned}$$

Therefore:

$$\langle f - F(\widetilde{u_1},\dots,\widetilde{u_n}), g'\rangle_{L^2(B,\mu)} = 0,$$

for all $g' \in L^2(B, \sigma(\widetilde{u_1},\dots,\widetilde{u_n}))$. Since this holds for any representatives $\widetilde{u_1},\dots,\widetilde{u_n}$, by Definition 3.2 we have that $F(Iu_1,\dots,Iu_n)$ is indeed a version of $\mathbb{E}[f|Iu_1,\dots,Iu_n]$. ☐

Theorem 3.14. *Let (H,B,μ) be an abstract Wiener space, $f \in L^2(B,\mu)$ and linearly independent elements $h_1,\dots,h_n$ of H. For every $y_1,\dots,y_n \in \mathbb{R}^n$ let:*

$$F(y_1,\dots,y_n) = Gf\left(\bigcap_{k=1}^n [\langle h_k, \cdot\rangle = y_k]\right).$$

Then $F(Ih_1,\dots,Ih_n)$ is a version of $\mathbb{E}[f|Ih_1,\dots,Ih_n]$:

$$\mathbb{E}[f|Ih_1 = y_1,\dots,Ih_n = y_n] = Gf\left(\bigcap_{k=1}^n [\langle h_k, \cdot\rangle = y_k]\right). \tag{3.31}$$

Proof. Let $\{u_1, \ldots, u_n\}$ be an orthonormal basis for $M = \text{span}\{h_1, \ldots, h_n\}$. Then:

$$h_k = \alpha_1^k u_1 + \alpha_2^k u_2 + \ldots + \alpha_n^k u_n, \text{ for all } 1 \leq k \leq n,$$

where $\alpha_j^k = \langle h_k, u_j \rangle$ for all $1 \leq j, k \leq n$. Let:

$$A = \begin{bmatrix} \alpha_1^1 & \alpha_2^1 & \ldots & \alpha_n^1 \\ \vdots & & & \vdots \\ \alpha_1^n & \alpha_2^n & \ldots & \alpha_n^n \end{bmatrix}.$$

Note that A is invertible, since $h_1, \ldots, h_n$ are linearly independent. Then:

$$Ih_k = \alpha_1^k (Iu_1) + \alpha_2^k (Iu_2) + \ldots + \alpha_n^k (Iu_n), \text{ for all } 1 \leq k \leq n,$$

so for $y_1, \ldots, y_n \in \mathbb{R}$:

$$\begin{aligned} \mathbb{E}[f|Ih_1 = y_1, \ldots, Ih_n = y_n] &= \mathbb{E}[f|Iu_1 = p_1, \ldots, Iu_n = p_n] \\ &= Gf(p_1 u_1 + \ldots + p_n u_n + M^\perp), \end{aligned} \tag{3.32}$$

where the last equality follows from Lemma 3.13 and $p = [p_1 \ldots p_n]^T \in \mathbb{R}^n$ is given by $p = A^{-1} y$, with $y = [y_1 \ldots y_n]^T$. Then:

$$p_1 u_1 + \ldots + p_n u_n + M^\perp = \bigcap_{k=1}^{n} \left[\langle h_k, \cdot \rangle = y_k \right],$$

and so (3.31) follows readily from (3.33). $\square$

3.3 The Gaussian Radon Transform and the Classical Wiener Space

In this section we compute some concrete examples of the Gaussian Radon transform on the classical Wiener space. Recall from Example 1.1 that this is the triple:

$$(\mathcal{H}, \mathcal{C}, \mu),$$

where $\mathcal{H}$ is the space of all absolutely continuous functions $h : [0,1] \to \mathbb{R}$ with $h(0) = 0$ and $h' \in L^2[0,1]$, $\mathcal{C}$ is the space of all continuous functions $f : [0,1] \to \mathbb{R}$ with $f(0) = 0$, and μ is classical Wiener measure. The space $\mathcal{H}$ is a real separable infinite-dimensional Hilbert space with norm $\|\cdot\|$ induced by the inner-product:

$$\langle h_1, h_2 \rangle := \int_0^1 h_1'(x) h_2'(x) dx, \text{ for all } h_1, h_2 \in \mathcal{H}. \tag{3.33}$$

The measure μ results by completing $\mathcal{H}$ with respect to the supremum norm $\|\cdot\|_\infty$, a measurable norm on $\mathcal{H}$, and obtaining $\mathcal{C}$.

Consider for a moment the space $\mathcal{C}[a,b]$ of continuous real-valued functions on a closed interval $[a,b]$, a Banach space with the supremum norm. The dual space $\mathcal{C}^*[a,b]$ is isomorphic to the space

$$NBV[a,b]$$

of *normalized* functions of bounded variation on $[a,b]$, that is the space of all bounded variation functions $g : [a,b] \to \mathbb{R}$ that are *right-continuous* and satisfy $g(a) = 0$. Specifically, for every $\Lambda \in \mathcal{C}^*[a,b]$ there is a unique $g \in NBV[a,b]$ such that Λf is given by the Lebesgue-Stieltjes integral:

$$\Lambda f = \int_a^b f\, dg,$$

for all $f \in \mathcal{C}[a,b]$. Moreover, $\|\Lambda\| = V_a^b(g)$, where $V_a^b(g)$ denotes the total variation of g. See Chapter 13 of [3] for details.

Since $\mathcal{C} = \delta_0^{-1}(0)$, where:

$$\delta_0 : \mathcal{C}[0,1] \to \mathbb{R};\ \delta_0 f := f(0)$$

is a continuous linear functional on $\mathcal{C}[0,1]$, $\mathcal{C}$ is a closed subspace of $\mathcal{C}[0,1]$. Therefore the dual space $\mathcal{C}^*$ consists exactly of the restrictions of elements of $\mathcal{C}^*[0,1]$ to $\mathcal{C}$:

$$\mathcal{C}^* = \{\Lambda_g : g \in NBV[0,1]\}, \text{ where } (f, \Lambda_g) = \int_0^1 f\, dg, \text{ for all } f \in \mathcal{C}. \tag{3.34}$$

We know that to every $\Lambda_g \in \mathcal{C}^*$ there corresponds a unique element $h_{\Lambda_g} \in \mathcal{H}$ such that:

$$(h, \Lambda_g) = \langle h, h_{\Lambda_g} \rangle \text{ for all } h \in \mathcal{H}.$$

Then:

$$\int_0^1 h dg = \int_0^1 h'(x) h'_{\Lambda_g}(x)\, dx, \text{ for all } h \in \mathcal{H}. \tag{3.35}$$

Recall that if $f_1, f_2 : [a,b] \to \mathbb{R}$ are right-continuous functions of bounded variation that have no common points of discontinuity, then:

$$\int_a^b f_1\, df_2 + \int_a^b f_2\, df_1 = f_1(b) f_2(b) - f_1(a) f_2(a).$$

Since every $h \in \mathcal{H}$ is continuous on $[0,1]$:

$$\int_0^1 h\, dg + \int_0^1 g\, dh = h(1)g(1) - h(0)g(0),$$

and since $h(0) = 0$:

$$\int_0^1 h\, dg = h(1)g(1) - \int_0^1 g\, dh. \tag{3.36}$$

Now recall that if $f : [a, b] \to \mathbb{R}$ is absolutely continuous, then:

$$\int_a^x f'(t)\, dt = f(x) - f(a), \text{ for all } x \in [a, b].$$

Moreover, the Lebesgue-Stieltjes measure μ_f on $[a, b]$ induced by an absolutely continuous function f on $[a, b]$ is absolutely continuous with respect to Lebesgue measure l, and the Radon-Nikodym derivative is given by the derivative of f:

$$\frac{d\mu_f}{dl} = f'.$$

In other words:

$$\int_a^b f_1\, df = \int_a^b f_1(x) f'(x) dx,$$

whenever the Lebesgue-Stieltjes integral on the left hand side exists.

Therefore (3.35) and (3.36) yield, for all $h \in \mathcal{H}$:

$$\begin{aligned} \int_0^1 h'(x) h'_{\Lambda_g}(x)\, dx &= h(1)g(1) - \int_0^1 g\, dh \\ &= g(1) \int_0^1 h'(x)\, dx - \int_0^1 g(x) h'(x)\, dx \\ &= \int_0^1 h'(x) \left(g(1) - g(x)\right)\, dx, \end{aligned}$$

so:

$$\int_0^1 h'(x) \left(h'_{\Lambda_g}(x) - g(1) + g(x)\right)\, dx = 0, \text{ for all } h \in \mathcal{H}. \tag{3.37}$$

If we let:

$$G(x) := \int_0^x g(t)\, dt,$$

for all $x \in [0, 1]$, then G is absolutely continuous on $[0, 1]$ and:

$$G'(x) = g(x) \text{ a.e.}$$

Then (3.37) becomes:

$$\begin{aligned} 0 &= \int_0^1 h'(x) \left(h_{\Lambda_g}(x) - g(1)x + G(x)\right)'\, dx \\ &= \left\langle h, h_{\Lambda_g} - g(1)id + G \right\rangle, \end{aligned}$$

for all $h \in \mathcal{H}$, where id denotes the identity function. Then necessarily:

$$h_{\Lambda_g} = g(1)id - G,$$

for every $\Lambda_g \in \mathcal{C}^*$. Note that:

$$h'_{\Lambda_g}(x) = g(1) - g(x) \text{ a.e.}$$

and since constant functions have zero variation:

$$\int_0^1 f\,dh'_{\Lambda_g} = -\int_0^1 f\,dg.$$

We summarize these conclusions in the following:

Theorem 3.15. *Let $(\mathcal{H}, \mathcal{C}, \mu)$ be the classical Wiener space. Then the dual space $\mathcal{C}^* = \{\Lambda_g : g \in NBV[0,1]\}$ is continuously embedded as the dense subspace $\mathcal{H}_{\mathcal{C}^*}$ of $\mathcal{H}$:*

$$\mathcal{H}_{\mathcal{C}^*} = \{h_{\Lambda_g} : \Lambda_g \in \mathcal{C}^*\},$$

where for every $\Lambda_g \in \mathcal{C}^*$:

$$h_{\Lambda_g}(x) = g(1)x - \int_0^x g(t)\,dt = \int_0^x (g(1) - g(t))\,dt, \tag{3.38}$$

for all $x \in [0,1]$, and:

$$(f, \Lambda_g) = \int_0^1 f\,dg = -\int_0^1 f\,dh'_{\Lambda_g}, \tag{3.39}$$

for all $f \in \mathcal{C}$.

Example 3.1. For every $t \in [0,1]$, we let δ_t denote the Dirac functional on $\mathcal{C}$:

$$\delta_t : \mathcal{C} \to \mathbb{R};\ \delta_t(f) = f(t), \text{ for all } f \in \mathcal{C}.$$

Clearly $\delta_t \in \mathcal{C}^*$ for every $t \in [0,1]$. For every $t \in (0,1]$, the function $g := 1\!\!1_{[t,1]}$ belongs to $NBV[0,1]$ and:

$$\delta_t(f) = f(t) = \int_0^1 f\,dg,$$

for all $f \in \mathcal{C}$. From (3.38), δ_t corresponds to the element $h_t := h_{\delta_t} \in \mathcal{H}_{\mathcal{C}^*}$, given by:

$$h_t(s) = g(1)s - \int_0^s g(x)\,dx = s - \int_0^s 1\!\!1_{[t,1]}\,dx,$$

for all $s \in [0,1]$. If $s \le t$, then $h_t(s) = s$, and if $s > t$ then $h_t(s) = s - (s-t) = t$. Therefore:

$$h_t(s) = s \wedge t, \text{ for all } s \in [0,1], \tag{3.40}$$

where $s \wedge t$ denotes $\min\{s,t\}$. If $t = 0$ then $\delta_0(f) = f(0) = 0$ for all $f \in \mathcal{C}$, so $h_0 = 0 = s \wedge 0$ for all $s \in [0,1]$, and (3.40) holds for all $t \in [0,1]$.

We make a few observations about these functionals. First note that for every $t \in [0,1]$:

$$h'_t = 1\!\!1_{(0,t)} \text{ a.e.}$$

so:

$$\|h_t\|^2 = \int_0^1 (h'_t(x))^2\,dx = \int_0^1 1\!\!1_{(0,t)}\,dx = t.$$

Then δ_t, as a random variable on $(\mathcal{C}, \mu)$, is centered Gaussian with variance $\|h_t\|^2 = t$:

$$\delta_t \sim \mathcal{N}(0, t) \text{ on } (\mathcal{C}, \mu).$$

Now suppose $s, t \in [0, 1]$. Then:

$$\langle h_s, h_t \rangle = \int_0^1 \mathbb{1}_{(0,s)} \mathbb{1}_{(0,t)} \, dx,$$

so:

$$\mathrm{Cov}(\delta_s, \delta_t) = \langle h_s, h_t \rangle = s \wedge t,$$

for all $s, t \in [0, 1]$.

In fact, as we show next, the functionals δ_t define a *Brownian motion* on $(\mathcal{C}, \mu)$. Recall that a stochastic process $\{B(t, \omega) : t \in [0, \infty), \omega \in \Omega\}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a Brownian motion provided that it satisfies the following conditions:

i. The process starts at 0, that is $B(0) = 0$ almost everywhere.

ii. For any $0 \leq s < t$, the random variable $B(t) - B(s)$ is centered Gaussian with variance $t - s$.

iii. The process has independent increments, that is for any $0 \leq t_1 < t_2 < \ldots < t_n$, the random variables:

$$B(t_1), B(t_2) - B(t_1), \ldots, B(t_n) - B(t_{n-1})$$

are independent.

iv. With probability 1, the function $t \mapsto B(t, \omega)$ is continuous in t.

Proposition 3.16. *Let $(\mathcal{H}, \mathcal{C}, \mu)$ be the classical Wiener space and $\delta_t \in \mathcal{C}^*$ be the Dirac functional for every $t \in [0, 1]$. Then:*

$$\{\delta_t : 0 \leq t \leq 1\}$$

is a Brownian motion on $(\mathcal{C}, \mu)$.

Proof. Since $\delta_0 = 0$, the process starts at 0. Let $0 < s < t \leq 1$. By our previous calculations:

$$\delta_s \sim \mathcal{N}(0, s), \; \delta_t \sim \mathcal{N}(0, t), \text{ and } \mathrm{Cov}(\delta_s, \delta_t) = s.$$

Then:

$$\delta_t - \delta_s \sim \mathcal{N}(0, t + s - 2\rho\sqrt{ts}),$$

where:

$$\rho = \mathrm{Corr}(\delta_s, \delta_t) = \frac{\mathrm{Cov}(\delta_s, \delta_t)}{\sqrt{st}} = \frac{s}{\sqrt{st}} = \sqrt{\frac{s}{t}}.$$

So indeed:

$$\delta_t - \delta_s \sim \mathcal{N}(0, t - s), \text{ for all } 0 \leq s < t \leq 1.$$

Now let $0 \le t_1 < t_2 < \ldots < t_n \le 1$. For any $j < k$:

$$\begin{aligned}
\mathrm{Cov}(\delta_{t_k} - \delta_{t_{k-1}}, \delta_{t_j} - \delta_{t_{j-1}}) &= \mathbb{E}[\delta_{t_k}\delta_{t_j} - \delta_{t_k}\delta_{t_{j-1}} - \delta_{t_{k-1}}\delta_{t_j} + \delta_{t_{k-1}}\delta_{t_{j-1}}] \\
&= t_j - t_{j-1} - t_j + t_{j-1} \\
&= 0.
\end{aligned}$$

So the increments $\delta_{t_1}, \delta_{t_2} - \delta_{t_1}, \ldots, \delta_{t_n} - \delta_{t_{n-1}}$ are pairwise independent, and since they are also jointly Gaussian, they are mutually independent. Finally, continuity of paths follows trivially since $t \mapsto f(t)$ is continuous in t for every $f \in \mathcal{C}$. □

Remark 3.2. Originally, the classical Wiener measure μ on $\mathcal{C}$ was defined as follows: first define μ on all cylinder subsets of $\mathcal{C}$, that is for every subset $A \subset \mathcal{C}$ of the form:

$$A = \{f \in \mathcal{C} : (f(t_1), f(t_2), \ldots, f(t_n)) \in U\}, \tag{3.41}$$

where $0 < t_1 < t_2 < \ldots < t_n \le 1$ and $U \in \mathcal{B}(\mathbb{R}^n)$, define:

$$\mu(A) := \int_U \left(\prod_{k=1}^{n} \frac{1}{\sqrt{2\pi(t_k - t_{k-1})}} e^{-\frac{(x_k - x_{k-1})^2}{2(t_k - t_{k-1})}} \right) dx_1\, dx_2 \ldots dx_n, \tag{3.42}$$

where $t_0 := 0$ and $x_0 := 0$. The collection $\mathcal{R}$ of all cylinder subsets of $\mathcal{C}$ forms an algebra, but not a σ-algebra. However, the σ-algebra $\sigma(\mathcal{R})$ generated by $\mathcal{R}$ is the Borel σ-algebra $\mathcal{B}(\mathcal{C})$ of $\mathcal{C}$. In 1923, Wiener proved that μ is a countably additive measure on $\mathcal{R}$ (see [30]), and therefore μ extends uniquely to a measure, also denoted μ, on $\mathcal{B}(\mathcal{C})$. Wiener called the space $(\mathcal{C}, \mu)$ the "differential space".

To see that the measure μ induced by $\mathcal{H}$ on $\mathcal{C}$ in the abstract Wiener space sense coincides with the original definition, note that any cylinder subset $A \subset \mathcal{C}$ of the form (3.41) can be expressed as:

$$A = [(\delta_{t_1}, \delta_{t_2}, \ldots, \delta_{t_n}) \in U].$$

Since the increments $\delta_{t_1}, \delta_{t_2} - \delta_{t_1}, \ldots, \delta_{t_n} - \delta_{t_{n-1}}$ are independent, their joint density function is the product of the individual density functions, so:

$$\begin{aligned}
\mu(A) &= \mu[(\delta_{t_1}, \delta_{t_2}, \ldots, \delta_{t_n}) \in U] \\
&= \mu[(\delta_{t_1}, \delta_{t_2} - \delta_{t_1}, \ldots, \delta_{t_n} - \delta_{t_{n-1}}) \in T(U)] \\
&= \int_{T(U)} \left(\prod_{k=1}^{n} \frac{1}{\sqrt{2(t_k - t_{k-1})}} e^{-\frac{y_k^2}{2(t_k - t_{k-1})}} \right) dy_1 \ldots dy_n,
\end{aligned} \tag{3.43}$$

where $T(x_1, \ldots, x_n) = (x_1, x_2 - x_1, \ldots, x_n - x_{n-1})$ and $t_0 = y_0 = 0$. Performing the change of variables:

$$y_1 = x_1, y_2 = x_2 - x_1, \ldots, y_n = x_n - x_{n-1},$$

we obtain a Jacobian matrix with determinant 1 and (3.43) becomes exactly the expression in (3.42), proving that μ coincides with the original definition of classical Wiener measure.

Let us now look at some examples of the Gaussian Radon transform on the classical Wiener space $(\mathcal{H}, \mathcal{C}, \mu)$.

Example 3.2. Consider $h_t \in \mathcal{H}_{\mathcal{C}^*}$, given by $h_t(s) = s \wedge t$, corresponding to the Dirac functional $\delta_t \in \mathcal{C}^*$ for a fixed $t \in (0, 1]$. Let $\alpha \in \mathbb{R}$ and:

$$M_0 := h_t^{\perp} \subset \mathcal{H}.$$

Then:

$$M_{\alpha,t} := \alpha h_t + M_0 = \alpha h_t + h_t^{\perp}$$

is a hyperplane in $\mathcal{H}$, and, since $h_t \in \mathcal{H}_{\mathcal{C}^*}$, its closure $\overline{M_{\alpha,t}}$ in $\mathcal{C}$ is a hyperplane in $\mathcal{C}$ (see the discussion following Theorem 2.6). In fact, since $h \perp h_t$ in $\mathcal{H}$ if and only if $0 = \langle h, h_t\rangle = (h, \delta_t) = h(t)$:

$$M_{\alpha,t} = \{h \in \mathcal{H} : h(t) = \alpha t\},$$

and by Lemma 2.5:

$$\overline{M_{\alpha,t}} = \{f \in \mathcal{C} : f(t) = \alpha t\}.$$

The measure $\mu_{M_{\alpha,t}}$ resulting from Theorem 2.4 is concentrated on the hyperplane $\overline{M_{\alpha,t}}$ in $\mathcal{C}$ and has characteristic function:

$$\int_{\mathcal{C}} e^{i\Lambda_g}\, d\mu_{M_{\alpha,t}} = e^{i\alpha\langle h_t, h_{\Lambda_g}\rangle - \frac{1}{2}\|P_{M_0} h_{\Lambda_g}\|^2},$$

for all $\Lambda_g \in \mathcal{C}^*$. Consider now $\delta_s \in \mathcal{C}^*$ for some $s \in [0, 1]$. Then, with respect to $\mu_{M_{\alpha,t}}$:

$$\delta_s \sim \mathcal{N}\left(\alpha\langle h_t, h_s\rangle, \|P_{M_0} h_s\|^2\right).$$

Since $\langle h_t, h_s\rangle = s \wedge t$ and $\|h_r\|^2 = r$ for all $r \in [0, 1]$:

$$\|P_{M_0} h_s\|^2 = \|h_s\|^2 - \|P_{M_0^{\perp}} h_s\|^2 = s - \frac{(s \wedge t)^2}{t},$$

so:

$$\delta_s \sim \mathcal{N}\left(\alpha(s \wedge t), s - \frac{(s \wedge t)^2}{t}\right), \tag{3.44}$$

with respect to $\mu_{M_{\alpha,t}}$. We then obtain:

$$G\delta_s\left(\alpha h_t + h_t^{\perp}\right) = \alpha(s \wedge t), \tag{3.45}$$

and:

$$G\delta_s^2\left(\alpha h_t + h_t^{\perp}\right) = s - \frac{(s \wedge t)^2}{t} + \alpha^2(s \wedge t)^2. \tag{3.46}$$

Note that if we consider $\mathbb{E}[\delta_s|\delta_t = y]$ for some $y \in \mathbb{R}$, where the conditional expectation is with respect to classical Wiener measure μ, by Theorem 3.14:

$$\begin{aligned}
\mathbb{E}[\delta_s|\delta_t = y] &= G\delta_s\left([\langle h_t, \cdot\rangle = y]\right) \\
&= G\delta_s\left(\frac{y}{t} h_t + h_t^{\perp}\right) \\
&= \frac{y}{t}(s \wedge t),
\end{aligned}$$

where the last equality follows from (3.45). We showed that:

$$\mathbb{E}[\delta_s|\delta_t = y] = y\frac{s \wedge t}{t}.$$

Now let $s_1, s_2 \in [0,1]$. By (3.2):

$$\mathrm{Cov}(\delta_{s_1}, \delta_{s_2}) = \langle P_{M_0} h_{s_1}, P_{M_0} h_{s_2} \rangle,$$

where we are once again working on $(\mathcal{C}, \mu_{M_{\alpha,t}})$. Now for every $s \in [0,1]$:

$$P_{M_0} h_s = h_s - \frac{\langle h_s, h_t \rangle}{\|h_t\|^2} h_t = h_s - \frac{s \wedge t}{t} h_t.$$

Then:

$$\mathrm{Cov}(\delta_{s_1}, \delta_{s_2}) = (s_1 \wedge s_2) - \frac{(s_1 \wedge t)(s_2 \wedge t)}{t}. \tag{3.47}$$

In turn, this gives us the Gaussian Radon transform of the product $\delta_{s_1}\delta_{s_2}$ on $M_{\alpha,t}$:

$$G(\delta_{s_1}\delta_{s_2})\left(\alpha h_t + h_t^{\perp}\right) = (s_1 \wedge s_2) + (s_1 \wedge t)(s_2 \wedge t)\left(\alpha^2 - \frac{1}{t}\right). \tag{3.48}$$

Next, we obtain a *Brownian bridge* on $(\mathcal{C}, \mu_{M_{\alpha,t}})$ as an almost immediate consequence of the above calculations.

Definition 3.3. Let $a, b \in \mathbb{R}$ and $L > 0$ be fixed. A *Brownian bridge of length L from a to b* is a Gaussian process $\{X(t) : 0 \leq t \leq L\}$ that has continuous paths, starts at a (that is, $X(0) = a$ a.s.) and satisfies:

$$\mathbb{E}[X(t)] = a + (b-a)\frac{t}{L} \text{ and } \mathrm{Cov}(X(s), X(t)) = (s \wedge t) - \frac{st}{L}$$

for all $s, t \in [0, L]$.

Proposition 3.17. *Let $(\mathcal{H}, \mathcal{C}, \mu)$ be the classical Wiener space and:*

$$M_{\alpha,t} = \alpha h_t + h_t^{\perp},$$

for some $\alpha \in \mathbb{R}$. Then the process:

$$\{\delta_s : 0 \leq s \leq t\},$$

considered on $(\mathcal{C}, \mu_{M_{\alpha,t}})$, is a Brownian bridge of length t from 0 to αt.

Proof. Clearly $\{\delta_s : 0 \leq s \leq t\}$ is a Gaussian process with continuous paths, and $\delta_0 = 0$. By (3.44):

$$\mathbb{E}[\delta_s] = \alpha s = (\alpha t)\frac{s}{t},$$

and by (3.47):

$$\mathrm{Cov}(\delta_{s_1}, \delta_{s_2}) = (s_1 \wedge s_2) - \frac{s_1 s_2}{t},$$

for all $s_1, s_2 \in [0, t]$. □

Example 3.3. To generalize the results in Example 3.2, let $0 < t_1 < t_2 < \ldots < t_n \leq 1$. Then the functions $h_{t_1}, h_{t_2}, \ldots, h_{t_n}$ are independent in $\mathcal{H}$, and if we let:

$$M_0 := [\mathrm{span}\{h_{t_1}, \ldots, h_{t_n}\}]^{\perp},$$

then:

$$\left\{\frac{1}{\sqrt{t_k - t_{k-1}}}(h_{t_k} - h_{t_{k-1}})\right\}_{1\leq k\leq n}$$

is an orthonormal basis for $M_0^{\perp}$, where $t_0 := 0$. To see this, recall from Proposition 3.16 that $\{\delta_{t_1}, \delta_{t_2} - \delta_{t_1}, \ldots, \delta_{t_n} - \delta_{t_{n-1}}\}$ are independent in $L^2(\mathcal{C}, \mu)$, so:

$$\langle h_{t_k} - h_{t_{k-1}}, h_{t_j} - h_{t_{j-1}}\rangle = \langle \delta_{t_k} - \delta_{t_{k-1}}, \delta_{t_j} - \delta_{t_{j-1}}\rangle_{L^2(\mathcal{C},\mu)} = \begin{cases} 0 & \text{, if } j \neq k \\ t_k - t_{k-1} & \text{, if } j = k \end{cases}$$

For any $s \in [0, 1]$:

$$\langle h_s, h_{t_k} - h_{t_{k-1}} = (s \wedge t_k) - (s \wedge t_{k-1}),$$

so:

$$P_{M_0^{\perp}} h_s = \sum_{k=1}^{n} \frac{(s \wedge t_k) - (s \wedge t_{k-1})}{t_k - t_{k-1}}(h_{t_k} - h_{t_{k-1}}).$$

If we let:

$$M_{\alpha,t} := \alpha_1 h_{t_1} + \ldots + \alpha_n h_{t_n} + M_0, \tag{3.49}$$

for some real $\alpha_1, \ldots, \alpha_n$, a closed subspace of codimension n in $\mathcal{C}$, then with respect to $\mu_{M_{\alpha,t}}$:

$$\delta_s \sim \mathcal{N}\left(\sum_{k=1}^{n}(s \wedge t_k), s - \sum_{k=1}^{n} \frac{((s \wedge t_k) - (s \wedge t_{k-1}))^2}{t_k - t_{k-1}}\right),$$

for all $s \in [0, 1]$. More precisely:

$$\delta_s \sim \begin{cases} \mathcal{N}\left(\sum_{k=1}^{j-1} \alpha_k t_k + \sum_{k=j}^{n} \alpha_k s,\ t_{j-1} + \frac{(s-t_{j-1})^2}{t_j - t_{j-1}}\right) & \text{, if } s \in [t_{j-1}, t_j),\ j = 1, \ldots, n \\ \mathcal{N}\left(\sum_{k=1}^{n} \alpha_k s,\ t_n\right) & \text{, if } s \in [t_n, 1]. \end{cases}$$

For any $s_1, s_2 \in [0, 1]$:

$$\begin{aligned} \mathrm{Cov}(\delta_{s_1}, \delta_{s_2}) &= \langle P_{M_0} h_{s_1}, P_{M_0} h_{s_2}\rangle \\ &= \langle h_{s_1} - P_{M_0^{\perp}} h_{s_1}, h_{s_2} - P_{M_0^{\perp}} h_{s_2}\rangle, \end{aligned}$$

so, with respect to $\mu_{M_{\alpha,t}}$:

$$\mathrm{Cov}(\delta_{s_1}, \delta_{s_2}) = (s_1 \wedge s_2) - \sum_{k=1}^{n} \frac{((s_1 \wedge t_k) - (s_1 \wedge t_{k-1}))\,((s_2 \wedge t_k) - (s_2 \wedge t_{k-1}))}{t_k - t_{k-1}}.$$

Example 3.4. Now consider $\Lambda_{\mathrm{id}} \in \mathcal{C}^*$ given by $\Lambda_{\mathrm{id}} f = \int_0^1 f(x)\,dx$ for all $f \in \mathcal{C}$. This corresponds to $h_{\mathrm{id}} \in \mathcal{H}_{\mathcal{C}^*}$ given by:

$$h_{\mathrm{id}}(x) = \int_0^x (1 - t)\,dt = x - \frac{x^2}{2},$$

for all $x \in [0,1]$. Then $h'_{\text{id}}(x) = 1 - x$ and $\|h_{\text{id}}\|^2 = \frac{1}{3}$. For every $s \in [0,1]$:

$$\langle h_{\text{id}}, h_s \rangle = (h_{\text{id}}, \delta_s) = s - \frac{s^2}{2},$$

so if $M_{\alpha,t}$ is as in (3.49), then, with respect to $\mu_{M_{\alpha,t}}$:

$$\Lambda_{\text{id}} \sim \mathcal{N}\left(\sum_{k=1}^{n} \alpha_k \left(t_k - \frac{t_k^2}{2}\right), \frac{1}{3} - t_n + t_n^2 - \frac{1}{4}t_n^3 - \frac{1}{4}\sum_{k=1}^{n} t_{k-1}t_k(t_k - t_{k-1})\right).$$

Example 3.5. Let $P_0 = h_{\text{id}}^{\perp}$ and consider the hyperplane in $\mathcal{H}$:

$$P_{\alpha,\text{id}} = \alpha h_{\text{id}} + h_{\text{id}}^{\perp},$$

for some $\alpha \in \mathbb{R}$. Remark that:

$$P_0 = \{h \in \mathcal{H} : \int_0^1 h(x)\, dx = 0\} \text{ and } P_{\alpha,\text{id}} = \{h \in \mathcal{H} : \int_0^1 h(x)\, dx = \frac{\alpha}{3}\}.$$

For any $s \in [0,1]$:

$$\|P_{P_0} h_s\|^2 = \|h_s\|^2 - \frac{|\langle h_s, h_{\text{id}} \rangle|^2}{\|h_{\text{id}}\|^2} = s - 3s^2 + 3s^3 - \frac{3s^4}{4}.$$

So, with respect to $\mu_{P_{\alpha,\text{id}}}$:

$$\delta_s \sim \mathcal{N}\left(\alpha\left(s - \frac{s^2}{2}\right), s - 3s^2 + 3s^3 - \frac{3s^4}{4}\right),$$

for all $s \in [0,1]$, and:

$$\text{Cov}(\delta_{s_1}, \delta_{s_2}) = (s_1 \wedge s_2) - 3\left(s_1 - \frac{s_1^2}{2}\right)\left(s_2 - \frac{s_2^2}{2}\right).$$

Example 3.6. The continuous linear functional $\Lambda_{\text{sqrt}} \in \mathcal{C}^*$ given by:

$$\Lambda_{\text{sqrt}} f = \int_0^1 f(x)\, d\sqrt{x},$$

for all $f \in \mathcal{C}$ corresponds to:

$$h_{sqrt}(x) = x - \frac{2}{3}x^{3/2}, \text{ for all } x \in [0,1]$$

in $\mathcal{H}_{\mathcal{C}^*}$. Then:

$$\langle h_{\text{sqrt}}, h_{\text{id}} \rangle = (h_{\text{sqrt}}, \Lambda_{\text{id}}) = \int_0^1 (x - \frac{2}{3}x^{3/2})\, dx = \frac{7}{30},$$

so:

$$\|P_{P_0} h_{\text{sqrt}}\|^2 = \|h_{\text{sqrt}}\|^2 - \frac{|\langle h_{\text{sqrt}}, h_{\text{id}} \rangle|^2}{\|h_{\text{id}}\|^2} = \frac{1}{300}.$$

Thus, with respect to $\mu_{P_{\alpha,\text{id}}}$:

$$\Lambda_{\text{sqrt}} \sim \mathcal{N}\left(\frac{7\alpha}{30}, \frac{1}{300}\right).$$

3.4 An Inversion Procedure for the Gaussian Radon Transform

The focus of this section is to develop a way to recover a function f from its Gaussian Radon transform. Our procedure will involve the Segal-Bargmann transform for abstract Wiener spaces, which we review next. We begin with the classical Segal-Bargmann transform for finite-dimensional spaces.

Let $f \in L^2(\mathbb{R}^n, \gamma_n)$, where γ_n is standard Gaussian measure on $\mathbb{R}^n$. The *Segal-Bargmann transform* of f is the function $Sf : \mathbb{C}^n \to \mathbb{C}$ given by:

$$(Sf)(z) := e^{-\frac{1}{2}(z,z)} \int_{\mathbb{R}^n} e^{(z,x)} f(x)\, d\gamma_n(x), \text{ for all } z \in \mathbb{C}^n, \tag{3.50}$$

where:

$$(z, w) := \sum_{k=1}^{n} z_k w_k,$$

for all $z, w \in \mathbb{C}^n$.

Next, we introduce the *Segal-Bargmann space* over $\mathbb{C}^n$, denoted by:

$$\mathcal{H}L^2(\mathbb{C}^n),$$

and defined as the space of all holomorphic functions on $\mathbb{C}^n$ that are square-integrable with respect to the measure λ_n on $\mathbb{C}^n$ given by:

$$d\lambda_n(z) := \frac{1}{\pi^n} e^{-|z|^2}\, dz,$$

where dz is $2n$-dimensional Lebesgue measure. Then $\mathcal{H}L^2(\mathbb{C}^n)$ is a closed subspace of $L^2(\mathbb{C}^n, \lambda_n)$, therefore a Hilbert space itself, and:

$$S : L^2(\mathbb{R}^n, \gamma_n) \to \mathcal{H}L^2(\mathbb{C}^n); f \mapsto Sf$$

is a unitary operator, that is $SS^* = S^*S = id$. For more details, see [26], [4], or [5].

Now let (H, B, μ) be an abstract Wiener space and $H_{\mathbb{C}} := H \oplus iH$ be the complexification of H. This is a complex Hilbert space with inner-product:

$$\langle h + ik, u + iv \rangle_{\mathbb{C}} = (\langle h, u\rangle + \langle k, v\rangle) + i\left(\langle k, u\rangle - \langle h, v\rangle\right), \text{ for all } h, k, u, v \in H,$$

and norm $\|h+ik\|^2_{\mathbb{C}} = \|h\|^2 + \|k\|^2$. For any $z_1, z_2 \in H_{\mathbb{C}}$ let (z_1, z_2) denote the complex bilinear extension of the H inner-product:

$$(h + ik, u + iv) = \langle h, u\rangle + i\langle h, v\rangle + i\langle k, u\rangle - \langle k, v\rangle, \text{ for all } h, k, u, v \in H.$$

For every $f \in L^2(B, \mu)$ the *Segal-Bargmann transform* of f is the function $S_B f : H_{\mathbb{C}} \to \mathbb{C}$ defined by:

$$(S_B f)(z) := e^{-\frac{1}{2}(z,z)} \int_B e^{(Iz)(x)} f(x)\, d\mu(x), \tag{3.51}$$

where $Iz := Ih + iIk$ for all $z = h + ik$ in $H_{\mathbb{C}}$ and $I : H \to L^2(B, \mu)$ is the map defined in (1.21).

Let $\mathcal{J}(H_{\mathbb{C}})$ be the collection of all finite-dimensional subspaces of $H_{\mathbb{C}}$. We say that a function $g : H_{\mathbb{C}} \to \mathbb{C}$ is *holomorphic* provided that g is locally bounded and the restriction $g|_F$ of g to any $F \in \mathcal{J}(H_{\mathbb{C}})$ is holomorphic. Consider for every $F \in \mathcal{J}(H_{\mathbb{C}})$ the Gaussian probability measure λ_F on F given by:

$$d\lambda_F(z) := \frac{1}{\pi^n} e^{-\|z\|_{\mathbb{C}}^2} \, dz,$$

where n =dim(F) and dz is Lebesgue measure on F. With these notations, we are ready to define the *Segal-Bargmann space* over $H_{\mathbb{C}}$, denoted by:

$$\mathcal{H}L^2(H_{\mathbb{C}}),$$

as the space of all holomorphic functions g on $H_{\mathbb{C}}$ that satisfy:

$$\|g\|_{SB}^2 := \sup_{F \in \mathcal{J}(H_{\mathbb{C}})} \int_F |g(z)|^2 \, d\lambda_F(z) < \infty.$$

Then $\|\cdot\|_{SB}$ as defined above is a complete inner-product norm on $\mathcal{H}L^2(H_{\mathbb{C}})$, which is thus a complex Hilbert space. As in the classical case, the map:

$$S_B : L^2(B, \mu) \to \mathcal{H}L^2(H_{\mathbb{C}}); f \mapsto S_B f$$

is unitary. For details and proofs of these facts, see [26], [9], [14], or [8].

The next theorem, our inversion procedure, shows that for a function $f \in L^2(B, \mu)$ the Segal-Bargmann transform $S_B f$ coincides with the finite-dimensional Segal - Bargmann transform of $Gf(p + Q_0)$ on the complexification of $Q_0^{\perp}$, for any closed subspace Q_0 of finite codimension in H.

Theorem 3.18. *Let (H, B, μ) be an abstract Wiener space, $f \in L^2(B, \mu)$ and Q_0 be a closed subspace of finite codimension in H. Consider the function $G_{Q_0} f$ defined on $Q_0^{\perp}$ by:*

$$G_{Q_0} f(p) := Gf(p + Q_0), \text{ for all } p \in Q_0^{\perp}.$$

Then:

$$\left(S_{Q_0^{\perp}}(G_{Q_0} f) \right)(z) = (S_B f)(z), \text{ for all } z \in (Q_0^{\perp})_{\mathbb{C}}, \tag{3.52}$$

where $S_{Q_0^{\perp}}$ and S_B are the Segal-Bargmann transforms on $L^2(Q_0^{\perp}, \gamma_{Q_0^{\perp}})$ and $L^2(B, \mu)$, respectively. In other words:

$$S_{Q_0^{\perp}}(G_{Q_0} f) = (S_B f)|_{(Q_0^{\perp})_{\mathbb{C}}}. \tag{3.53}$$

Proof. Recall that, from (3.14), $G_{Q_0} f \in L^2(B, \mu)$, so we may consider the Segal-Bargmann transform $S_{Q_0^{\perp}}$ of $G_{Q_0} f$. From (3.17), if $h \in Q_0^{\perp}$ then for $\gamma_{Q_0^{\perp}}$-almost all $p \in Q_0^{\perp}$:

$$Ih = \langle h, p \rangle, \ \mu_{Q_p}\text{-almost everywhere.}$$

Then if $z = h + ik \in (Q_0^\perp)_\mathbb{C}$ for some $h, k \in Q_0^\perp$, we have for $\gamma_{Q_0^\perp}$-almost all $p \in Q_0^\perp$:

$$Iz = Ih + iIk = \langle h, p\rangle + i\langle k, p\rangle = (z, p),\ \mu_{Q_p}\text{-almost everywhere.} \tag{3.54}$$

For any $z \in (Q_0^\perp)_\mathbb{C}$:

$$\begin{aligned}
\left(S_{Q_0^\perp}(G_{Q_0}f)\right)(z) &= e^{-\frac{1}{2}(z,z)} \int_{Q_0^\perp} e^{(z,p)} G_{Q_0}f(p)\, d\gamma_{Q_0^\perp}(p) \\
&= e^{-\frac{1}{2}(z,z)} \int_{Q_0^\perp} \int_B e^{(z,p)} f(x)\, d\mu_{Q_p}(x)\, d\gamma_{Q_0^\perp}(p) \\
&= e^{-\frac{1}{2}(z,z)} \int_{Q_0^\perp} \int_B e^{Iz(x)} f(x)\, d\mu_{Q_p}(x)\, d\gamma_{Q_0^\perp}(p) && \text{(by (3.54))} \\
&= e^{-\frac{1}{2}(z,z)} \int_B e^{Iz(x)} f(x)\, d\mu(x) && \text{(by Theorem 3.8)} \\
&= (S_B f)(z),
\end{aligned}$$

which proves our claim. □

Now suppose $u \in H$ is a unit vector. Taking $Q_0 = u^\perp$ in Theorem 3.18 we obtain:

$$S_{\mathbb{R}u}(G_{u^\perp} f) = (S_B f)_{\mathbb{C}u}.$$

In particular:

$$\left(S_{\mathbb{R}u}(G_{u^\perp} f)\right)(tu) = (S_B f)(tu), \text{ for all } t \in \mathbb{R},$$

which translates to:

$$\left(S_{\mathbb{R}}(G_{u^\perp} f)\right)(t) = (S_B f)(tu), \text{ for all } t \in \mathbb{R}, \tag{3.55}$$

where $S_\mathbb{R}$ is the Segal-Bargmann transform on $L^2(\mathbb{R}, \gamma_1)$. Recall that any hyperplane in H is of the form:

$$\gamma_{p,u} = pu + u^\perp,$$

for a unit vector u and $p \geq 0$. The equation (3.55) tells us that if we know the Gaussian Radon transform of f:

$$Gf(\gamma_{p,u}) = G_{u^\perp} f(p),$$

for all hyperplanes in H, then we know $(S_B f)(h)$ for all $h \in H$. Taking the holomorphic extension to $H_\mathbb{C}$, we know $S_B f$ and can then obtain f using the inverse Segal-Bargmann transform.

Chapter 4
The Gaussian Radon Transform and Machine Learning

Suppose we are observing an experiment, recording the outputs corresponding to certain inputs. The central task of machine learning is to predict the outputs corresponding to future, yet unobserved, inputs. More precisely, suppose that all input values are contained in a set $\mathcal{X}$, called the *input space*, or *sample space*. We will assume that all outputs are real numbers. If there is a finite (or countable) number of possible outcomes, this problem is known as *classification*; otherwise, it is known as *regression*. Moreover, suppose we have collected a set:

$$D = \{(t_1, y_1), \ldots, (t_n, y_n)\} \subset \mathcal{X} \times \mathbb{R}$$

of input values t_k together with their respective output values y_k, known as the *training data*. A classical example of a classification problem is handwriting recognition. For instance, say we feed a computer a large number of handwritten digits together with their corresponding labels "0", "1", "2" and so on; the goal of a learning algorithm would then be for the computer to correctly label new images of handwritten digits.

So the goal is to use the training data to find a "prediction" function $f : \mathcal{X} \to \mathbb{R}$ such that $f(t)$ is a close approximation of the output y resulting from a future input t. An important point to make is that we are not trying to find a function that matches the training data exactly, but one that yields good approximations of future outputs. In fact, modeling the training data too closely is known as *overfitting*; the "quality" of a prediction model is determined by its accuracy in predicting future outcomes, and not by its accuracy in fitting the training data.

Some of the most popular learning methods are *support vector machines* (SVM's), and a crucial assumption of these methods is that one searches for the prediction function within a special kind of Hilbert space of functions, known as a reproducing kernel Hilbert space (RKHS). We review this next. First, recall that a function $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is said to be *positive definite* provided that:

$$K(s,t) = K(t,s), \text{ for all } s, t \in \mathcal{X},$$

and:

$$\sum_{i,j=1}^{n} \alpha_i \alpha_j K(t_i, t_j) \geq 0,$$

for all $n \in \mathbb{N}$ and any choice of $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ and $t_1, \ldots, t_n \in \mathcal{X}$. The following important result, known as the Moore-Aronszajn Theorem, may be found in Chapter 4 of [28].

Theorem 4.1. *Let $\mathcal{X}$ be a non-empty set and $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a positive definite function. Then there is a unique Hilbert space H, whose elements are functions $f : \mathcal{X} \to \mathbb{R}$, such that:*

i. The function $K_t := K(t, \cdot)$ is contained in H for every $t \in \mathcal{X}$.

ii. For every function $f \in H$:

$$f(t) = \langle K_t, f\rangle, \text{ for all } t \in \mathcal{X}, \tag{4.1}$$

where $\langle \cdot, \cdot \rangle$ is the inner-product in H.

Moreover, the linear span of $\{K_t : t \in \mathcal{X}\}$ is dense in H.

The Hilbert space above is called the *reproducing kernel Hilbert space* over $\mathcal{X}$ with *reproducing kernel* K. Note that (4.1) yields:

$$K(s, t) = \langle K_s, K_t\rangle, \text{ for all } s, t \in \mathcal{X},$$

and also implies that the Dirac functional:

$$\delta_t : H \to \mathbb{R}; \delta_t(f) := f(t),$$

is continuous for every $t \in \mathcal{X}$:

$$|\delta_t(f)| = |f(t)| = |\langle K_t, f\rangle| \leq \|K_t\|\|f\|.$$

Another interesting, often useful, property of RKHS's is that norm convergence implies pointwise convergence. Specifically:

$$\text{If } f_n \to f \text{ in } H, \text{ then } \lim_{n\to\infty} f_n(t) = f(t), \text{ for all } t \in \mathcal{X}.$$

This follows readily from continuity of δ_t.

4.1 Ridge Regression

Suppose we have the training data $D = \{(t_1, y_1), \ldots, (t_n, y_n)\} \subset \mathcal{X} \times \mathbb{R}$ and a RKHS H with reproducing kernel $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. The method known as *ridge regression* seeks to find a prediction function in H by minimizing the quantity:

$$\mathcal{R}_{\lambda,D}(f) := \sum_{j=1}^{n} (y_j - f(t_j))^2 + \lambda\|f\|^2, \text{ for all } f \in H, \tag{4.2}$$

where $t_1, \ldots, t_n$ are the given input values in D, $y_1, \ldots, y_n$ are their corresponding collected outputs, and $\lambda > 0$ is a *regularization parameter.* The main role of this parameter is to avoid modeling the training data too closely, or overfitting.

The following result, showing that a unique solution to this problem exists, is a well-known result in machine learning theory. For completeness, we include here a geometrical proof of this result, with roots in the works [19], [20] of Kimeldorf and Wahba.

Theorem 4.2. *Let $\mathcal{X}$ be a non-empty set,*

$$D = \{(t_1, y_1), \ldots, (t_n, y_n)\} \subset \mathcal{X} \times \mathbb{R}$$

be a finite subset of $\mathcal{X} \times \mathbb{R}$, and H be a RKHS over $\mathcal{X}$ with reproducing kernel $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. Then for every $\lambda > 0$ there is a unique element $\hat{f}_{\lambda,D} \in H$ such that:

$$\mathcal{R}_{\lambda,D}(\hat{f}_{\lambda,D}) = \inf_{f \in H} \mathcal{R}_{\lambda,D}(f) = \inf_{f \in H} \left(\sum_{j=1}^{n} (y_j - f(t_j))^2 + \lambda \|f\|^2 \right).$$

Specifically, $\hat{f}_{\lambda,D}$ is given by:

$$\hat{f}_{\lambda,D} = \sum_{j=1}^{n} \hat{c}_j K_{t_j}, \tag{4.3}$$

where K_t denotes the function $K(t, \cdot)$ for all $t \in \mathcal{X}$, and the vector $\hat{c} \in \mathbb{R}^n$ is given by:

$$\hat{c} = (K_D + \lambda I_n)^{-1} y, \tag{4.4}$$

with K_D being the $n \times n$ matrix with entries $[K_D]_{i,j} = K(t_i, t_j)$, I_n the $n \times n$ identity matrix, and $y = [y_1 \ldots y_n] \in \mathbb{R}^n$.

Proof. Let H_λ denote the space H with the scaled inner-product:

$$\langle f, g \rangle_{H_\lambda} := \lambda \langle f, g, \rangle, \text{ for all } f, g \in H.$$

Consider the linear map:

$$T : \mathbb{R}^n \to H_\lambda;\ e_j \mapsto T(e_j) = \frac{1}{\lambda} K_{t_j}, \text{ for all } 1 \leq j \leq n,$$

where $\{e_1, \ldots, e_n\}$ is the standard orthonormal basis of $\mathbb{R}^n$. Then for all $f \in H_\lambda$:

$$\langle T^* f, e_j \rangle_{\mathbb{R}^n} = \langle f, T e_j \rangle_{H_\lambda} = \lambda \left\langle f, \frac{1}{\lambda} K_{t_j} \right\rangle = f(t_j)$$

for every $1 \leq j \leq n$, so:

$$T^* f = \sum_{j=1}^{n} f(t_j) e_j.$$

Then we may express $\mathcal{R}_{\lambda,D}$ as:

$$\begin{aligned} \mathcal{R}_{\lambda,D}(f) &= \sum_{j=1}^{n} (y_j - f(t_j))^2 + \lambda \|f\|^2 \\ &= \|y - T^* f\|_{\mathbb{R}^n}^2 + \|f\|_{H_\lambda}^2. \end{aligned}$$

If we consider the direct sum of Hilbert spaces $H_\lambda \oplus \mathbb{R}^n$, the norm in this space is:

$$\|(f, c)\|_{H_\lambda \oplus \mathbb{R}^n} = \|c\|_{\mathbb{R}^n}^2 + \|f\|_{H_\lambda}^2$$

for all $c \in \mathbb{R}^n$ and $f \in H_\lambda$, so $\mathcal{R}_{\lambda,D}(f)$ may be viewed geometrically as the distance between the points (f, T^*f) and $(0, y)$ in $H_\lambda \oplus \mathbb{R}^n$:

$$\mathcal{R}_{\lambda,D} = \operatorname{dist}\left((f, T^*f), (0, y)\right) \text{ in } H_\lambda \oplus \mathbb{R}^n.$$

In other words, minimizing $\mathcal{R}_{\lambda,D}(f)$ is the same as finding the point closest to $(0, y)$ on the subspace $\{(f, T^*f) : f \in H_\lambda\}$, which is simply the graph of T^*:

$$\mathrm{Gr}(T^*) = \{(f, T^*f) : f \in H_\lambda\} \subset H_\lambda \oplus \mathbb{R}^n.$$

Now note that for any $c \in \mathbb{R}^n$:

$$\begin{aligned} \langle (f, T^*f), (-Tc, c) \rangle_{H_\lambda \oplus \mathbb{R}^n} &= \langle f, -Tc \rangle_{H_\lambda} + \langle T^*f, c \rangle_{\mathbb{R}^n} \\ &= -\langle f, Tc \rangle_{H_\lambda} + \langle f, Tc \rangle_{H_\lambda} \\ &= 0. \end{aligned}$$

Conversely, if $f \in H_\lambda$ and $d \in \mathbb{R}^n$ are such that $\langle (f, d), (-Tc, c) \rangle_{H_\lambda \oplus \mathbb{R}^n} = 0$ for all $c \in \mathbb{R}^n$, then $\langle f, Tc \rangle_{H_\lambda} = \langle d, c \rangle_{\mathbb{R}^n}$, so:

$$\langle T^*f, c \rangle = \langle d, c \rangle, \text{ for all } c \in \mathbb{R}^n,$$

meaning that $d = T^*f$. Therefore:

$$\mathrm{Gr}(T^*) = \{(-Tc, c) : c \in \mathbb{R}^n\}^\perp. \tag{4.5}$$

So $\mathrm{Gr}(T^*)$ is a closed subspace of $H_\lambda \oplus \mathbb{R}^n$, and then there is unique point $(\hat{f}_{\lambda,D}, T^*\hat{f}_{\lambda,D})$ in $\mathrm{Gr}(T^*)$ that is closest to $(0, y)$. This point is of the form $(0, y) + (f, c)$, where $(f, c) \in H_\lambda \oplus \mathbb{R}^n$ is orthogonal to $\mathrm{Gr}(T^*)$. This is pictured below in Figure 4.1.

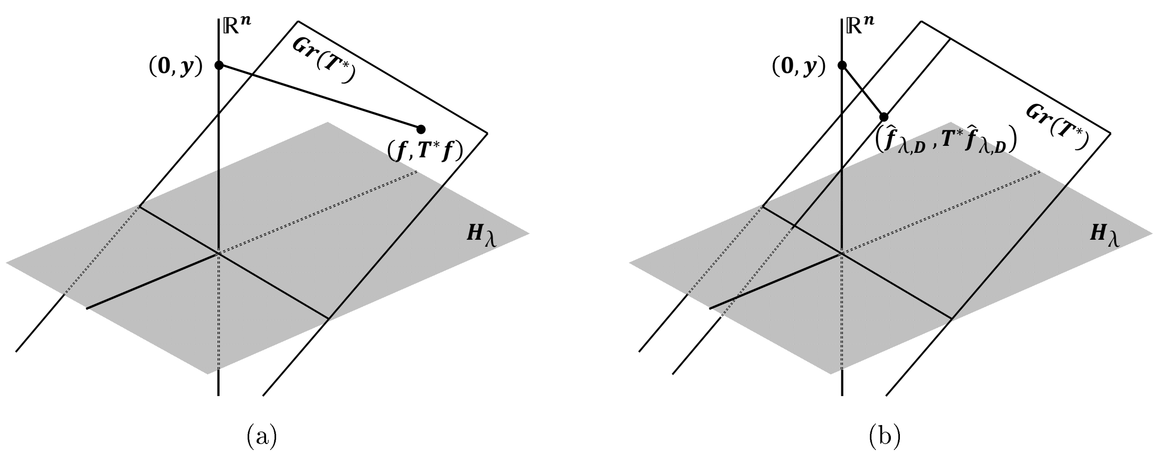


Figure 4.1: A geometric interpretation of Theorem 4.2.

From (4.5):

$$(\hat{f}_{\lambda,D}, T^*\hat{f}_{\lambda,D}) = (-Tc, y + c),$$

for some $c \in \mathbb{R}^n$. Then $\hat{f}_{\lambda,D} = -Tc$ and:

$$y + c = T^* \hat{f}_{\lambda,D} = -T^*Tc,$$

so:

$$y = -(T^*T + I_n)c. \tag{4.6}$$

Note that $(T^*T + I_n)d = 0$ if and only if $d = 0$, since $\langle (T^*T + I_n)d, d\rangle_{\mathbb{R}^n} \geq \|d\|^2_{\mathbb{R}^n}$ for all $d \in \mathbb{R}^n$, so the operator $T^*T + I_n$ is invertible. So we may solve (4.6) for c:

$$c = -(T^*T + I_n)^{-1}y.$$

Since $\hat{f}_{\lambda,D} = -Tc$:

$$\hat{f}_{\lambda,D} = T\left[(T^*T + I_n)^{-1}y\right].$$

For any $1 \leq i, j \leq n$:

$$\begin{aligned}
\langle (T^*T)e_i, e_j\rangle_{\mathbb{R}^n} &= \langle Te_i, Te_j\rangle_{H_\lambda} \\
&= \lambda \left\langle \frac{1}{\lambda}K_{t_i}, \frac{1}{\lambda}K_{t_j} \right\rangle \\
&= \frac{1}{\lambda}K(t_i, t_j) \\
&= \frac{1}{\lambda}[K_D]_{i,j},
\end{aligned}$$

so $T^*T = \frac{1}{\lambda}K_D$. Then:

$$\begin{aligned}
\hat{f}_{\lambda,D} &= T\left[\sum_{i,j=1}^{n}\left[\left(\frac{1}{\lambda}K_D + I_n\right)^{-1}\right]_{i,j} y_i e_j\right] \\
&= \sum_{j=1}^{n}\left(\sum_{i=1}^{n}\lambda\left[(K_D + \lambda I_n)^{-1}\right]_{i,j} y_i\right)\frac{1}{\lambda}K_{t_j} \\
&= \sum_{j=1}^{n}\left(\sum_{i=1}^{n}\left[(K_D + \lambda I_n)^{-1}\right]_{i,j} y_i\right)K_{t_j} \\
&= \sum_{j=1}^{n}\hat{c}_j K_{t_j},
\end{aligned}$$

where $\hat{c}$ is as in (4.4). □

4.2 Probabilistic Interpretations

Remark that the ridge regression problem described above contains no randomness at all. Our work in this area is motivated by a recent increase in the machine learning literature interest in probabilistic interpretations of support vector machines - see for instance [27], [24], [2], [23], [31], or [18].

Suppose we are performing an experiment and $\mathcal{S}$ is the set of all possible outcomes, a set known as the *sample space*. The outcome of the experiment follows an unknown probability distribution on $\mathcal{S}$. A *statistical model* is a collection of probability distributions on $\mathcal{S}$, where we suspect the "true" distribution lies. In particular, a *parametric statistical model* is a collection $\{\rho(y|\theta)\}_{\theta\in\Theta}$ of probability distributions on $\mathcal{S}$, indexed by a *parameter set* Θ. For example, a Gaussian model might take $\Theta = \{(m,\sigma) : m \in \mathbb{R}, \sigma > 0\}$ and for every $\theta = (m,\sigma) \in \Theta$ the Gaussian distribution on $\mathbb{R}$:

$$\rho(y|\theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(y-m)^2}.$$

The principal assumption of *Bayesian inference* is that the parameters themselves are also considered random, that is we assume that Θ is also equipped with a fixed probability distribution $p(\theta)$, called the *prior distribution*. The goal is then to use the collected data to "update" our knowledge of the best parameter, by finding the conditional distribution $p(\theta|y)$ of θ given the data y, a quantity known as the *posterior distribution*. These quantities are all connected by *Bayes' formula*:

$$p(\theta|y) = \frac{\rho(y|\theta)p(\theta)}{m(y)},$$

where $m(y) = \int_\Theta \rho(y|\theta)p(\theta)\,d\theta$ is the *marginal distribution*.

The *maximum a posteriori* (MAP) estimator seeks the value $\hat{\theta}_{MAP}$ of θ that maximizes the posterior:

$$\hat{\theta}_{MAP} = \arg\max_{\theta\in\Theta} p(\theta|y) = \arg\max_{\theta\in\Theta} \rho(y|\theta)p(\theta),$$

where the last equality follows because $m(y)$ is a positive quantity that does not depend on θ, so it has no influence on the maximum over Θ of $p(\theta|y)$.

Let us now give a Bayesian perspective on the ridge regression problem: suppose our sample space is $\mathbb{R}$ and our parameter space is a RKHS H over the input space $\mathcal{X}$, with reproducing kernel $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. If H is *finite-dimensional*, we may equip H with standard Gaussian measure γ_H as our *prior distribution*:

$$p(f) = \frac{1}{\sqrt{2\pi}^d} e^{-\frac{1}{2}\|f\|^2}, \text{ for all } f \in H,$$

where d is the dimension of H. Let $\tilde{f}$ denote the continuous linear functional:

$$\tilde{f} := \langle f, \cdot\rangle$$

on H for every $f \in H$. Then, with respect to γ_H, every $\tilde{f}$ is Gaussian with mean 0 and variance $\|f\|^2$, and $\mathrm{Cov}(\tilde{f}, \tilde{g}) = \langle f, g\rangle$ for all $f, g \in H$.

Recall that H contains the functions $K_t = K(t,\cdot)$ for all $t \in \mathcal{X}$ and also:

$$f(t) = \langle K_t, f\rangle = \tilde{K}_t(f), \text{ for all } f \in H, t \in \mathcal{X}.$$

Then $\{\tilde{K}_t\}_{t\in\mathcal{X}}$ is a centered Gaussian process on H with covariance function K:

$$\mathrm{Cov}(\tilde{K}_t, \tilde{K}_s) = \langle K_t, K_s\rangle = K(t,s), \text{ for all } t, s \in \mathcal{X}.$$

Now suppose $D = \{(t_1, y_1), \ldots, (t_n, y_n)\} \subset \mathcal{X} \times \mathbb{R}$ is the training data we collected. The relationship $f(t) = \tilde{K}_t(f)$ suggests that we could model our data y as arising from:

$$\tilde{y}_t = \tilde{K}_t(f), \text{ for every } t \in \mathcal{X}.$$

But suppose every measurement y_j contains some measurement error which we model as Gaussian noise. We would need a Gaussian process $\{\epsilon_1, \ldots, \epsilon_n\}$ on H, with covariance:

$$\mathrm{Cov}(\epsilon_i, \epsilon_j) = \lambda \delta_{i,j}$$

for some parameter $\lambda > 0$, which is also independent of $\{\tilde{K}_{t_1}, \ldots, \tilde{K}_{t_n}\}$. To achieve this, we consider an orthonormal set $\{e_1, \ldots, e_n\} \subset H$ such that:

$$\{e_1, \ldots, e_n\} \subset [\mathrm{span}\{K_{t_1}, \ldots, K_{t_n}\}]^{\perp}.$$

Then $\{\tilde{K}_{t_j}\}_{1 \leq j \leq n}$ and $\{\tilde{e}_j\}_{1 \leq j \leq n}$ are independent, and we model our data as:

$$\tilde{y}_j = \tilde{K}_{t_j}(f) + \sqrt{\lambda}\tilde{e}_j, \text{ for all } 1 \leq j \leq n,$$

for every $f \in H$, where $\lambda > 0$ is a fixed parameter. Then $\tilde{y}_j$ is Gaussian with mean $f(t_j)$ and variance λ, and $\{\tilde{y}_1, \ldots, \tilde{y}_n\}$ are independent. This gives rise, for every $f \in H$, to the statistical model of distributions on $\mathbb{R}^n$ given by:

$$\rho_{\lambda,D}(x|f) = \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi\lambda}} e^{-\frac{1}{2\lambda}(x_j - f(t_j))^2},$$

for every $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$.

Replacing x with the vector $y = [y_1 \ldots y_n]$ of observed values, the posterior distribution is then proportional to:

$$e^{-\frac{1}{2\lambda}\sum_{j=1}^{n}(y_j - f(t_j))^2} e^{-\frac{1}{2}\|f\|^2} = e^{-\frac{1}{2\lambda}\left[\sum_{j=1}^{n}(y_j - f(t_j))^2 + \lambda\|f\|^2\right]},$$

so finding the MAP estimator is equivalent to minimizing $\mathcal{R}_{\lambda,D}$ in (4.2). Therefore:

$$\hat{f}_{MAP} = \hat{f}_{\lambda,D},$$

where $\hat{f}_{\lambda,D}$ is the ridge regression solution in (4.3).

This Bayesian approach clearly depends on H being finite-dimensional; however, reproducing kernel Hilbert spaces used in practice are often infinite-dimensional (such as those arising from Gaussian RBF kernels). The ridge regression SVM previously discussed goes through regardless of the dimensionality of the RKHS, and there is still a need for a valid stochastic interpretation of the infinite-dimensional case.

We now explore another stochastic approach to ridge regression, which is equivalent to the Bayesian one, but which can be carried over in a sense to the infinite-dimensional case, as we shall see later. Suppose again that H is a finite-dimensional RKHS over $\mathcal{X}$, with reproducing kernel K, and equipped with standard Gaussian measure. Recall that $\{\tilde{K}_t\}_{t \in \mathcal{X}}$ is a centered Gaussian process on H with covariance function K. If we assume that the data arises from some unknown function in H, then the relationship $f(t) = \tilde{K}_t(f)$ again suggests that the random variable $\tilde{K}_t$ is a good model for the outputs. Moreover, the training data $D = \{(t_1, y_1), \ldots, (t_n, y_n)\}$ provides some previous knowledge of the random variables $\tilde{K}_{t_j}$, which we can use to refine our estimation of $\tilde{K}_t$ by taking conditional expectations. In other words, our first instinct would be to estimate the output of a future input $t \in \mathcal{X}$ by:

$$\mathbb{E}[\tilde{K}_t | \tilde{K}_{t_1} = y_1, \ldots, \tilde{K}_{t_n} = y_n]. \tag{4.7}$$

But if we want to include some possible noise in the measurements, we will again "attach" to $\{\tilde{K}_{t_1}, \ldots, \tilde{K}_{t_n}\}$ an independent centered Gaussian process.

So fix $t \in \mathcal{X}$, a future input whose output we'd like to predict. To take measurement error into account, we let again an orthonormal set $\{e_1, \ldots, e_n\} \subset H$ such that:

$$\{e_1, \ldots, e_n\} \subset [\operatorname{span}\{K_{t_1}, \ldots, K_{t_n}, K_t\}]^{\perp},$$

and $\lambda > 0$, and set:

$$y_j = \tilde{K}_{t_j} + \sqrt{\lambda}\tilde{e}_j, \text{ for all } 1 \leq j \leq n.$$

Then we estimate the output $\hat{y}(t)$ as the conditional expectation:

$$\hat{y}(t) = \mathbb{E}[\tilde{K}_t | \tilde{K}_{t_j} + \sqrt{\lambda}\tilde{e}_j = y_j, 1 \leq j \leq n].$$

As shown in Lemma 4.3 below:

$$\hat{y}(t) = a_1 y_1 + \ldots + a_n y_n,$$

where $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$ is:

$$a = A^{-1}\left[\operatorname{Cov}(\tilde{K}_t, \tilde{K}_{t_j} + \sqrt{\lambda}\tilde{e}_j)\right]_{1 \leq j \leq n},$$

with:

$$\begin{aligned}
[A]_{i,j} &= \operatorname{Cov}(\tilde{K}_{t_j} + \sqrt{\lambda}\tilde{e}_j, \tilde{K}_{t_i} + \sqrt{\lambda}\tilde{e}_i) \\
&= K(t_i, t_j) + \lambda\delta_{i,j} \\
&= [K_D + \lambda I_n]_{i,j}
\end{aligned}$$

for all $1 \leq i, j \leq n$. Moreover:

$$\operatorname{Cov}(\tilde{K}_t, \tilde{K}_{t_j} + \sqrt{\lambda}\tilde{e}_j) = \langle K_t, K_{t_j} \rangle = K_{t_j}(t).$$

Note this last relationship is why we required that $\{e_1, \ldots, e_n\}$ also be orthogonal to K_t. This yields:

$$\hat{y}(t) = \sum_{j=1}^{n} [(K_D + \lambda I_n)^{-1} y]_j K_{t_j}(t),$$

showing that:

$$\hat{y}(t) = \hat{f}_{\lambda,D}(t).$$

Lemma 4.3. *Let Z_0, Z_1, ..., Z_n be centered jointly Gaussian random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $A \in \mathbb{R}^{n\times n}$ be the matrix with entries:*

$$[A]_{i,j} = Cov(Z_i, Z_j), \text{ for all } 1 \leq i, j \leq n.$$

If A is invertible, then:

$$\mathbb{E}[Z_0|Z_1, \ldots, Z_n] = a_1 Z_1 + \ldots + a_n Z_n,$$

where $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$ is given by:

$$a = A^{-1} \left[Cov(Z_0, Z_1)\ Cov(Z_0, Z_2)\ \ldots\ Cov(Z_0, Z_n)\right]$$

Proof. Let $Y = Z_0 - Z_\perp$, where $Z_\perp$ is the orthogonal projection in $L^2(\Omega, \mathbb{P})$ of Z_0 onto the linear span of $Z_1, \ldots, Z_n$. Then $Y \perp Z_j$ for all $1 \leq j \leq n$, so $Y, Z_1, \ldots, Z_n$ are jointly Gaussian and Y is independent of the random vector $(Z_1, \ldots, Z_n)$. Then if S is any set in the σ-algebra $\sigma(Z_1, \ldots, Z_n)$ generated by $Z_1, \ldots, Z_n$:

$$\begin{aligned} \int_S Z_0 \, d\mathbb{P} &= \int_S (Z_\perp + Y) \, d\mathbb{P} \\ &= \int_S Z_\perp \, d\mathbb{P} + \int_S Y \, d\mathbb{P} \\ &= \int_S Z_\perp \, d\mathbb{P}. \end{aligned}$$

Since this holds for all $S \in \sigma(Z_1, \ldots, Z_n)$ and $Z_\perp$ is $\sigma(Z_1, \ldots, Z_n)$-measurable (being a linear combination of $Z_1, \ldots, Z_n$), the conditional expectation $\mathbb{E}[Z_0|Z_1, \ldots, Z_n]$ is exactly $Z_\perp$:

$$\mathbb{E}[Z_0|Z_1, \ldots, Z_n] = Z_\perp = a_1 Z_1 + \ldots + a_n Z_n,$$

for some $a_1, \ldots, a_n \in \mathbb{R}$. Note that, since all variables have mean 0:

$$\mathbb{E}[Z_0 Z_j] = \mathbb{E}[Z_\perp Z_j] = \sum_{i=1}^{n} \mathbb{E}[Z_i Z_j] a_i = [Aa]_j,$$

for all $1 \leq j \leq n$. So:

$$Aa = \left[\mathrm{Cov}(Z_0, Z_1)\ \mathrm{Cov}(Z_0, Z_2)\ \ldots\ \mathrm{Cov}(Z_0, Z_n)\right],$$

which proves the claim. □

If H is infinite-dimensional, the absence of standard Gaussian measure prevents us from having the Gaussian process $\{\tilde{K}_t\}_{t\in\mathcal{X}}$ directly on H. In what follows, we show that the Gaussian Radon transform offers a stochastic interpretation to the ridge regression problem when H is infinite-dimensional, by considering the Gaussian process and conditional expectation approach on an abstract Wiener space rather than on the Hilbert space itself.

Specifically, suppose $\mathcal{X}$ is a separable topological space and H is a real infinite-dimensional RKHS over $\mathcal{X}$ with reproducing kernel K. The assumption that $\mathcal{X}$ is a separable topological space ensures that H is also separable. Complete H with respect to a measurable norm and obtain an abstract Wiener space (H, B, μ). Then for every $t \in \mathcal{X}$, consider:

$$\tilde{K}_t := IK_t \in L^2(B, \mu),$$

where $I : H \to L^2(B, \mu)$ is the map described in (1.21). Since $\tilde{K}_t(f) = \langle K_t, f\rangle = f(t)$ for all $f \in H$, we choose $\tilde{K}_t$ as the model random variable for the outputs. Note that since I is an isometry, $\{\tilde{K}_t\}_{t\in\mathcal{X}}$ is a centered Gaussian process with covariance function K:

$$\mathrm{Cov}(\tilde{K}_t, \tilde{K}_s) = \langle K_t, K_s\rangle = K(t, s), \text{ for all } t, s \in \mathcal{X}.$$

As before, we account for noise in the training set $D = \{(t_1, y_1), \ldots, (t_n, y_n)\} \subset \mathcal{X} \times \mathbb{R}$ by taking an orthonormal set $\{e_1, \ldots, e_n\} \subset H$ such that:

$$\{e_1, \ldots, e_n\} \subset [\mathrm{span}\{K_{t_1}, \ldots, K_{t_n}, K_t\}]^{\perp}, \tag{4.8}$$

where $t \in \mathcal{X}$ is the future input whose output we want to predict. For a fixed parameter $\lambda > 0$, we model the data as:

$$y_j = \tilde{K}_{t_j} + \sqrt{\lambda}\tilde{e}_j, \text{ for all } 1 \le j \le n,$$

and then estimate the output $\hat{y}(t)$ corresponding to t by the conditional expectation:

$$\hat{y}(t) = \mathbb{E}[\tilde{K}_t | \tilde{K}_{t_j} + \sqrt{\lambda}\tilde{e}_j = y_j, 1 \le j \le n],$$

where $\tilde{e}_j = Ie_j$ for all $1 \le j \le n$.

By our assumption in (4.8) and the isometric property of I:

$$\begin{aligned}
\mathrm{Cov}(\tilde{K}_{t_j} + \sqrt{\lambda}\tilde{e}_j, \tilde{K}_{t_i} + \sqrt{\lambda}\tilde{e}_i) &= \langle K_{t_j} + \sqrt{\lambda}e_j, K_{t_i} + \sqrt{\lambda}e_i\rangle \\
&= K(t_i, t_j) + \lambda\delta_{i,j},
\end{aligned}$$

for every $1 \le i, j \le n$. Similarly:

$$\mathrm{Cov}(\tilde{K}_t, \tilde{K}_{t_j} + \sqrt{\lambda}\tilde{e}_j) = K(t_j, t) = K_{t_j}(t).$$

By Lemma 4.3:

$$\begin{aligned}
\mathbb{E}[\tilde{K}_t | \tilde{K}_{t_j} + \sqrt{\lambda}\tilde{e}_j = y_j, 1 \le j \le n] &= [K_{t_1}(t) \ldots K_{t_n}(t)](K_D + \lambda I_n)^{-1} y \\
&= \sum_{j=1}^{n} \hat{c}_j K_{t_j}(t), \text{ where } \hat{c} = (K_D + \lambda I_n)^{-1} y \\
&= \hat{f}_{\lambda,D}(t), , \qquad (4.9)
\end{aligned}$$

where $\hat{f}_{\lambda,D}$ is the ridge regression solution in (4.3).

Now note that $\{K_{t_1} + \sqrt{\lambda}e_1, \ldots, K_{t_n} + \sqrt{\lambda}e_n\}$ are linearly independent. For suppose $a_1, \ldots, a_n \in \mathbb{R}$ are such that:

$$a_1 K_{t_1} + \sqrt{\lambda}a_1 e_1 + \ldots + a_n K_{t_n} + \sqrt{\lambda}a_n e_n = 0.$$

Then:

$$a_1 K_{t_1} + \ldots + a_n K_{t_n} = -\sqrt{\lambda}(a_1 e_1 + \ldots + a_n e_n),$$

but from (4.8), both sides of the equality above must be 0. So $a_1 e_1 + \ldots + a_n e_n = 0$, and since $\{e_1, \ldots, e_n\}$ are orthonormal, $a_1 = \ldots = a_n = 0$. By Theorem 3.14, the conditional expectation above may be expressed as the Gaussian Radon transform of $\tilde{K}_t$ on the closed affine subspace determined by $\langle K_{t_j} + \sqrt{\lambda} e_j, \cdot \rangle = y_j$ for $1 \le j \le n$:

$$\mathbb{E}[\tilde{K}_t | \tilde{K}_{t_j} + \sqrt{\lambda}\tilde{e}_j = y_j, 1 \le j \le n] = G\tilde{K}_t \left(\bigcap_{j=1}^{n} \left[\langle K_{t_j} + \sqrt{\lambda} e_j, \cdot \rangle = y_j \right] \right).$$

Combined with (4.9), we see that the value $\hat{f}_{\lambda,D}(t)$ predicted by ridge regression can be expressed in terms of the Gaussian radon transform. We summarize these findings in the following theorem.

Theorem 4.4. *Let H be a RKHS over a separable topological space $\mathcal{X}$, with reproducing kernel $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, and B be the completion of H with respect to a measurable norm, with Wiener measure μ. Let $D = \{(t_1, y_1), \ldots, (t_n, y_n)\} \subset \mathcal{X} \times \mathbb{R}$ and $t \in \mathcal{X}$ be fixed. Let $\{e_1, \ldots, e_n\} \subset H$ be an orthonormal set such that:*

$$\{e_1, \ldots, e_n\} \subset [span\{K_{t_1}, \ldots, K_{t_n}, K_t\}]^{\perp}, \tag{4.10}$$

where $K_s = K(s, \cdot) \in H$ for all $s \in \mathcal{X}$. Then for every $\lambda > 0$:

$$\begin{aligned} \hat{f}_{\lambda,D} &= \mathbb{E}[\tilde{K}_t | \tilde{K}_{t_j} + \sqrt{\lambda}\tilde{e}_j = y_j, 1 \le j \le n] && (4.11) \\ &= G\tilde{K}_t \left(\bigcap_{j=1}^{n} \left[\langle K_{t_j} + \sqrt{\lambda} e_j, \cdot \rangle = y_j \right] \right), && (4.12) \end{aligned}$$

where $\tilde{K}_s = IK_s$ for all $s \in \mathcal{X}$ and $\hat{f}_{\lambda,D}$ is the ridge regression solution in (4.3).

Let us know briefly go back to our "first instinct" approximation in (4.7) and take simply:

$$\mathbb{E}[\tilde{K}_t | \tilde{K}_{t_1} = y_1, \ldots, \tilde{K}_{t_n} = y_n] \tag{4.13}$$

as our estimation of the output corresponding to $t \in \mathcal{X}$, basically ignoring noise and taking $\lambda = 0$ above. Note that we are assuming that we are still in the setting of an abstract Wiener space (H, B, μ), where H is an infinite-dimensional separable RKHS with reproducing kernel K, and $\tilde{K}_t = IK_t$ above. Then, again by Theorem 3.14, assuming that the functions $K_{t_1}, \ldots, K_{t_n}$ are linearly independent, the quantity in (4.13) may be expressed as:

$$\mathbb{E}[\tilde{K}_t | \tilde{K}_{t_1} = y_1, \ldots, \tilde{K}_{t_n} = y_n] = G\tilde{K}_t \left(\bigcap_{j=1}^{n} \left[\langle K_{t_j}, \cdot \rangle = y_j \right] \right). \tag{4.14}$$

Applying Lemma 4.3:

$$\mathbb{E}[\tilde{K}_t | \tilde{K}_{t_1} = y_1, \ldots, \tilde{K}_{t_n} = y_n] = \sum_{j=1}^{d} \hat{d}_j K_{t_j}(t), \tag{4.15}$$

where $\hat{d} = [\hat{d}_1 \ldots \hat{d}_n] = K_D^{-1}y$, where K_D is again the $n \times n$ matrix given by the covariances: $[K_D]_{i,j} = K(t_i, t_j)$, for all $1 \leq j \leq n$.

As shown below in Theorem 4.5, the quantity to the right of (4.15) is in fact the element $\hat{f}_s$ of H of minimal norm which satisfies the interpolation conditions:

$$f(t_j) = y_j, \text{ for all } 1 \leq j \leq n,$$

which is exactly the setup in the more traditional spline theory (see [19], [20]). In other words, the spline solution $\hat{f}_s$ may also be expressed as a Gaussian Radon transform:

$$\hat{f}_s(t) = G\tilde{K}_t\left(\bigcap_{j=1}^{n} [\langle K_{t_j}, \cdot\rangle = y_j]\right).$$

The fact that we may obtain the predicted value through a conditional expectation and the Gaussian Radon transform suggests that a broader class of prediction problems could be approached in this fashion. For instance, suppose one is interested not in predicting the value at a particular input t, but in predicting the maximum or minimum value attained on a set of future inputs. The predicted value would be:

$$GF(L), \tag{4.16}$$

where L is the closed affine subspace of the RKHS reflecting the training data, and F is, for instance, a function of the form:

$$F(x) = \sup_{t \in S} \tilde{K}_t(x),$$

for some given set $S \subset \mathcal{X}$ of future inputs one is interested in. Note that the prediction in (4.16) is generally not the same as taking the supremum over the individual predicted values, that is not the same as:

$$\sup_{t \in S} G\tilde{K}_t(L),$$

where $G\tilde{K}_t(L)$ is the SVM prediction as in Theorem 4.4.

We now return to the spline setting result discussed earlier; this is a known result in the literature, but we include a proof here for completeness.

Theorem 4.5. *Let H be a real RKHS over a non-empty set $\mathcal{X}$, with reproducing kernel K, and $D = \{(t_1, y_1), \ldots, (t_n, y_n)\} \subset \mathcal{X} \times \mathbb{R}$ be such that the functions $K_{t_1}, \ldots, K_{t_n}$ are linearly independent. Then the element of H of minimal norm which satisfies $f(t_j) = y_j$ for all $1 \leq j \leq n$ is given by:*

$$\hat{f}_s = \sum_{j=1}^{n} \hat{d}_j K_{t_j}, \tag{4.17}$$

where $\hat{d} = [\hat{d}_1 \ldots \hat{d}_n] = K_D^{-1}y$, with $y = [y_1 \ldots y_n]$ and K_D the $n \times n$ matrix with entries $[K_D]_{i,j} = K(t_i, t_j)$ for every $1 \leq i, j \leq n$.

Proof. The proof will be similar to the geometrical flavor of Theorem 4.2. Once again, we let $T : \mathbb{R}^n \to H$ be the linear operator that takes $e_j \mapsto T(e_j) = K_{t_j}$ for every $1 \leq j \leq n$, where $\{e_1, \ldots, e_n\}$ is the standard orthonormal basis of $\mathbb{R}^n$. The adjoint is then given by:

$$T^* f = \sum_{j=1}^{n} \langle f, K_{t_j} \rangle e_j = \sum_{j=1}^{n} f(t_j) e_j, \text{ for all } h \in H.$$

Then:

$$\{f \in H : f(t_1) = y_1, \ldots, f(t_n) = y_n\} = (T^*)^{-1}(y).$$

The assumption that the functions K_{t_j} are linearly independent is crucial because it guarantees that the affine subspace above is non-empty. For assume that $(a_1, \ldots, a_n) \in \mathbb{R}^n$ is orthogonal to the range of T^*. Then:

$$a_1 f(t_1) + \ldots + a_n f(t_n) = 0, \text{ for every } f \in H,$$

which is equivalent to $a_1 K_{t_1} + \ldots + a_n K_{t_n} = 0$. But since $\{K_{t_1}, \ldots, K_{t_n}\}$ are linearly independent, $a_j = 0$ for all $1 \leq i \leq n$, so $\mathrm{Im}(T^*) = \mathbb{R}^n$ and then $(T^*)^{-1}(y)$ is non-empty.

Since $(T^*)^{-1}(y)$ is a translate of the closed subspace $\mathrm{Ker}(T^*)$ of H, there is an element $\hat{f}_s$ in $(T^*)^{-1}(y)$ of minimal norm, specifically the point on $(T^*)^{-1}(y)$ that is orthogonal to $\mathrm{Ker}(T^*)$. This is pictured below in Figure 4.2.

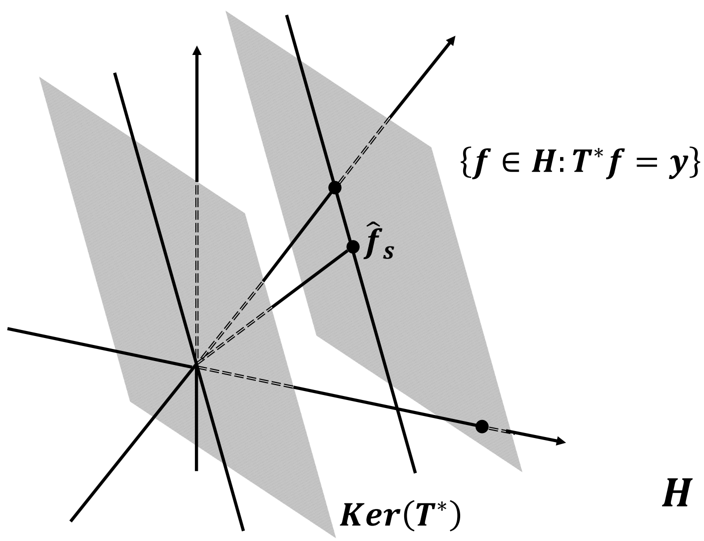


Figure 4.2: The point on $(T^*)^{-1}(y)$ closest to the origin.

Since:

$$\mathrm{Ker}(T^*) = [\mathrm{Im}(T)]^{\perp},$$

the orthogonal complement $[\mathrm{Ker}(T^*)]^{\perp}$ is the closure of the subspace $\mathrm{Im}(T)$. But $\mathrm{Im}(T)$ is a finite-dimensional subspace of H and is therefore closed, so:

$$[\mathrm{Ker}(T^*)]^{\perp} = \mathrm{Im}(T).$$

Therefore our point $\hat{f}_s \in \mathrm{Im}(T) \cap (T^*)^{-1}(y)$. Then $\hat{f}_s = Tc$ for some $c \in \mathbb{R}^n$ and $(T^*T)c = y$, so:

$$\hat{f}_s = Tc = T(T^*T)^{-1}y.$$

It follows readily that T^*T is given by the matrix K_D, so indeed:

$$\hat{f}_s = \sum_{i,j=1}^{n} [(K_D)^{-1}]_{i,j} y_i K_{t_j}.$$

□

4.3 Direct Sums of Abstract Wiener Spaces

One disadvantage of Theorem 4.4 is that, given the training data:

$$D = \{(t_1, y_1), \ldots, (t_n, y_n)\} \subset \mathcal{X} \times \mathbb{R},$$

for every input $t \in \mathcal{X}$ whose outcome we'd like to estimate, we must choose an orthonormal set $\{e_1, \ldots, e_n\} \in H$ such that every e_j is not only orthogonal to each K_{t_1}, …, K_{t_n}, but also to K_t. In other words, our choice of $\{e_1, \ldots, e_n\}$ could change with every training set and every future input $t \in \mathcal{X}$. Since $\mathrm{span}\{K_t : t \in \mathcal{X}\}$ is dense in H, we cannot find a set $\{e_1, \ldots, e_n\}$ that would "universally" work. This suggests that we would like to "attach" a Hilbert space to H, which could be our "repository" for errors, that is orthogonal to H. This is precisely the idea behind direct sums of Hilbert spaces.

Let H_1 and H_2 be Hilbert spaces with inner-products $\langle \cdot, \cdot \rangle_1$ and $\langle \cdot, \cdot \rangle_2$, respectively. The *orthogonal direct sum* of H_1 and H_2 is the space

$$H_1 \oplus H_2 := \{(h_1, h_2) : h_1 \in H_1, h_2 \in H_2\},$$

which is a Hilbert space with the inner-product:

$$\langle (h_1, h_2), (g_1, g_2) \rangle := \langle h_1, g_1 \rangle_1 + \langle h_2, g_2 \rangle_2,$$

for all $h_1, g_1 \in H_1$ and $h_2, g_2 \in H_2$. Note that H_1 and H_2 are continuously embedded as subspaces of $H_1 \oplus H_2$ through the maps:

$$H_1 \to H_1 \oplus H_2; h_1 \mapsto (h_1, 0) \text{ and } H_2 \to H_1 \oplus H_2; h_2 \mapsto (0, h_2),$$

and $H_1 \oplus H_2$ is the orthogonal direct sum of these subspaces.

Next, we investigate whether the abstract Wiener space construction over a direct sum of Hilbert spaces yields a direct sum of Banach spaces, where if B_1 and B_2 are Banach spaces with norms $|\cdot|_1$ and $|\cdot|_2$, respectively, their direct sum:

$$B_1 \oplus B_2 := \{(x_1, x_2) : x_1 \in B_1, x_2 \in B_2\}$$

is a Banach space with the norm:

$$|(x_1, x_2)| := |x_1|_1 + |x_2|_2,$$

for all $x_1 \in B_1$ and $x_2 \in B_2$.

Proposition 4.6. *Let (H_1, B_1, μ_1) and (H_2, B_2, μ_2) be abstract Wiener spaces, where $(H_1, \|\cdot\|_1)$ and $(H_2, \|\cdot\|_2)$ are real separable infinite-dimensional Hilbert spaces and B_1, B_2 are the Banach spaces obtained by completing H_1, H_2 with respect to measurable norms $|\cdot|_1$, $|\cdot|_2$, respectively. Then:*

$$(H_1 \oplus H_2, B_1 \oplus B_2, \mu_1 \times \mu_2)$$

is an abstract Wiener space.

Proof. Let $x^* \in (B_1 \oplus B_2)^*$. Then define for every $x_1 \in B_1$ and every $x_2 \in B_2$:

$$(x_1, x_1^*) := ((x_1, 0), x^*) \text{ and } (x_2, x_2^*) := ((0, x_2), x^*)\,.$$

Since x^* is continuous on $B_1 \oplus B_2$:

$$|(x_1, x_1^*)| = |((x_1, 0), x^*)| \leq c|(x_1, 0)| = c|x_1|_1,$$

for some $c > 0$, and x_1^* is clearly linear, so $x_1^* \in B_1^*$. Similarly, $x_2^* \in B_2^*$. Thus every continuous linear functional on $B_1 \oplus B_2$ is of the form $x^* = (x_1^*, x_2^*)$:

$$((x_1, x_2), (x_1^*, x_2^*)) = (x_1, x_1^*) + (x_2, x_2^*), \text{ for all } x_1 \in B_1, x_2 \in B_2,$$

for some $x_1^* \in B_1^*$ and $x_2^* \in B_2^*$.

Now recall that the measurable norms $|\cdot|_1$, $|\cdot|_2$ are weaker than the Hilbert norms $\|\cdot\|_1$, $\|\cdot\|_2$ on H_1, H_2, respectively, so there is $C > 0$ such that:

$$|h_1|_1 \leq C\|h_1\|_1 \text{ and } |h_2|_2 \leq C\|h_2\|_2,$$

for all $h_1 \in H_1$ and $h_2 \in H_2$. Then:

$$\begin{aligned} |(h_1, h_2)|^2 &= (|h_1|_1 + |h_2|_2)^2 \\ &\leq C^2(\|h_1\|_1 + \|h_2\|_2)^2 \\ &\leq 2C^2(\|h_1\|_1^2 + \|h_2\|_2^2), \end{aligned}$$

so:

$$|(h_1, h_2)| \leq \sqrt{2}C\|(h_1, h_2)\|, \text{ for all } h_1 \in H_1, h_2 \in H_2,$$

which shows that $|(\cdot, \cdot)|$ is a weaker norm than $\|(\cdot, \cdot)\|$ on $H_1 \oplus H_2$. Consequently, we may associate to every $x^* = (x_1^*, x_2^*)$ in $(B_1 \oplus B_2)^*$ a unique element $h_{x^*} \in H_1 \oplus H_2$ such that:

$$((h_1, h_2), x^*) = \langle (h_1, h_2), h_{x^*} \rangle\,, \text{ for all } (h_1, h_2) \in H_1 \oplus H_2.$$

This element is exactly $h_{x^*} = (h_{x_1^*}, h_{x_2^*})$, where $h_{x_1^*} \in (H_1)_{B_1^*}$ and $h_{x_2^*} \in (H_2)_{B_2^*}$:

$$\begin{aligned} ((h_1, h_2), (x_1^*, x_2^*)) &= (h_1, x_1^*) + (h_2, x_2^*) \\ &= \langle h_1, h_{x_1^*} \rangle_1 + \langle h_2, h_{x_2^*} \rangle_2 \\ &= \langle (h_1, h_2), (h_{x_1^*}, h_{x_2^*}) \rangle, \end{aligned}$$

for all $h_1 \in H_1$ and $h_2 \in H_2$.

Therefore the characteristic functional of the product measure $\mu_1 \times \mu_2$ on $B_1 \oplus B_2$ is:

$$\begin{aligned}\int_{B_1 \oplus B_2} e^{i(x_1^*, x_2^*)}\, d\mu_1 \times \mu_2 &= \int_{B_1 \times B_2} e^{i(x_1, x_1^*) + i(x_2, x_2^*)}\, d\mu_2(x_2)\, d\mu_1(x_1) \\ &= e^{-\frac{1}{2}(\|h_{x_1^*}\|_1^2 + \|h_{x_2^*}\|_2^2)} \\ &= e^{-\frac{1}{2}\|(h_{x_1^*}, h_{x_2^*})\|^2},\end{aligned}$$

for all $x_1^* \in B_1^*$, $x_2^* \in B_2^*$. Therefore $\mu_1 \times \mu_2$ is a centered non-degenerate Gaussian measure on $B_1 \oplus B_2$ with covariance operator:

$$R_{\mu_1 \times \mu_2}\left((x_1^*, x_2^*), (y_1^*, y_2^*)\right) = \left\langle (h_{x_1^*}, h_{x_2^*}), (h_{y_1^*}, h_{y_2^*}) \right\rangle,$$

for all $x_1^*, y_1^* \in B_1^*$ and $x_2^*, y_2^* \in B_2^*$. We then consider the Cameron-Martin space H of $(B_1 \oplus B_2, \mu_1 \times \mu_2)$:

$$H := \left\{ (x_1, x_2) \in B_1 \oplus B_2 : \|(x_1, x_2)\|' := \sup_{\substack{x_1^* \in B_1^*, x_2^* \in B_2^* \\ (x_1^*, x_2^*) \neq (0,0)}} \frac{|((x_1, x_2)(x_1^*, x_2^*))|}{\sqrt{\|h_{x_1^*}\|^2 + \|h_{x_2^*}\|^2}} \right\}.$$

Then note that for any $h_1 \in H_1$ and $h_2 \in H_2$:

$$\begin{aligned}|((h_1, h_2), (x_1^*, x_2^*))|^2 &= (|\langle h_1, h_{x_1^*}\rangle_1| + |\langle h_2, h_{x_2^*}\rangle_2|)^2 \\ &\leq (\|h_1\|_1 \|h_{x_1^*}\|_1 + \|h_2\|_2 \|h_{x_2^*}\|_2)^2 \\ &\leq (\|h_1\|_1^2 + \|h_2\|_2^2)(\|h_{x_1^*}\|_1^2 + \|h_{x_2^*}\|_2^2),\end{aligned}$$

for any x_1^*, x_2^*, so $H_1 \oplus H_2 \subset H$ and:

$$\|(h_1, h_2)\|' \leq \|(h_1, h_2)\|, \text{ for all } h_1 \in H_1, h_2 \in H_2.$$

Conversely, suppose $(x_1, x_2) \in H$. Then by letting $x_2^* = 0$ we have that:

$$\sup_{\substack{x_1^* \in B_1^* \\ x_1^* \neq 0}} \frac{|(x_1, x_1^*)|}{\|h_{x_1^*}\|_1} < \infty,$$

therefore x_1 belongs to the Cameron-Martin space of (B_1, μ_1) - which is H_1. Similarly, $x_2 \in H_2$, so $H = H_1 \oplus H_2$ as sets. Now to see that the norms are the same, note that for any $y_1^* \in B_1^*$ and $y_2^* \in B_2^*$, not both 0:

$$\begin{aligned}\|(h_{y_1^*}, h_{y_2^*})\|' &= \sup_{\substack{x_1^* \in B_1^*, x_2^* \in B_2^* \\ (x_1^*, x_2^*) \neq (0,0)}} \frac{|\langle h_{y_1^*}, h_{x_1^*}\rangle_1 + \langle h_{y_2^*}, h_{x_2^*}\rangle_2|}{\sqrt{\|h_{x_1^*}\|^2 + \|h_{x_2^*}\|^2}} \\ &\geq \frac{|\langle h_{y_1^*}, h_{y_1^*}\rangle_1 + \langle h_{y_2^*}, h_{y_2^*}\rangle_2|}{\sqrt{\|h_{y_1^*}\|^2 + \|h_{y_2^*}\|^2}} \\ &= \|(h_{y_1^*}, h_{y_2^*})\|,\end{aligned}$$

so:

$$\|(h_{x^*})\|' \geq \|h_{x^*}\|, \text{ for all } 0 \neq x^* \in (B_1 \oplus B_2)^*. \tag{4.18}$$

To every $x^* = (x_1^*, x_2^*) \in (B_1 \oplus B_2)^*$ we may associate a unique $h'_{x^*} = (h'_{x_1^*}, h'_{x_2^*}) \in H = H_1 \oplus H_2$ such that $(h_1, x_1^*) + (h_2, x_2^*) = \langle h_1, h'_{x_1^*}\rangle_1 + \langle h_2, h_{x_2^*}\rangle_2$ for all $h_1 \in H_1$, $h_2 \in H_2$, and $\{h'_{x^*} : x^* \in (B_1 \oplus B_2)^*\}$ is dense in H. But since $(h_1, x_1^*) + (h_2, x_2^*) = \langle h_1, h_{x_1^*}\rangle_1 + \langle h_2, h_{x_2^*}\rangle_2$ for all h_1, h_2, it follows that $h'_{x_1^*} = h_{x_1^*}$ and $h'_{x_2^*} = h_{x_2^*}$ for all x_1^*, x_2^*. Then, since $\{h_{x^*} : x^* \in (B_1 \oplus B_2)^*\}$ is dense in both H and $H_1 \oplus H_2$, (4.18) yields:

$$\|(h_1, h_2)\|' \geq \|(h_1, h_2)\|,$$

which proves that H and $H_1 \oplus H_2$ are the same as Hilbert spaces, and so $H_1 \oplus H_2$ is the Cameron-Martin space of $B_1 \oplus B_2$. □

Now consider the map $I : H_1 \oplus H_2 \to L^2(B_1 \oplus B_2, \mu_1 \times \mu_2)$ described in (1.21). Let $\{(x_n^1)^*\}_{n\in\mathbb{N}} \subset B_1^*$ and $\{(x_n^2)^*\}_{n\in\mathbb{N}} \subset B_2^*$ be such that:

$$h_{(x_n^1)^*} \to h_1 \text{ in } H_1 \text{ and } h_{(x_n^2)^*} \to h_2 \text{ in } H_2.$$

Then:

$$\left(h_{(x_n^1)^*}, h_{(x_n^2)^*}\right) \to (h_1, h_2) \text{ in } H_1 \oplus H_2,$$

so $I(h_1, h_2)$ is the $L^2(B_1 \oplus B_2, \mu_1 \times \mu_2)$-limit of $\{((x_n^1)^*, (x_n^2)^*)\}_{n\in\mathbb{N}}$. For every $k = 1, 2$, $I_k h_k$ is the $L^2(B_k, \mu_k)$-limit of $\{(x_n^k)^*\}_{n\in\mathbb{N}}$. Let (I_1h_1, I_2h_2) denote the map:

$$B_1 \oplus B_2 \ni (x_1, x_2) \mapsto (I_1h_1)(x_1) + (I_2h_2)(x_2).$$

Then:

$$\begin{aligned}
&\|(I_1h_1, I_2h_2) - ((x_n^1)^*, (x_n^2)^*)\|^2_{L^2(B_1\oplus B_2, \mu_1\times\mu_2)} \\
= \;& \int_{B_1\oplus B_2} \left[(I_1h_1 - (x_n^1)^*)(x_1) + (I_2h_2 - (x_n^2)^*)(x_2)\right] \, d\mu_1 \times \mu_2(x_1, x_2) \\
\leq \;& \|I_1h_1 - (x_n^1)^*\|^2_{L^2(B_1,\mu_1)} + \|I_2h_2 - (x_n^2)^*\|^2_{L^2(B_2,\mu_2)} \\
& +2\|I_1h_1 - (x_n^1)^*\|_{L^2(B_1,\mu_1)}\|I_2h_2 - (x_n^2)^*\|_{L^2(B_2,\mu_2)} \xrightarrow{n\to\infty} 0.
\end{aligned}$$

So:

$$I(h_1, h_2) = (I_1h_1, I_2h_2).$$

Let us re-analyze the ridge regression problem from this perspective. So let (H, B, μ) be an abstract Wiener space, where H is a real infinite-dimensional RKHS over a separable topological space $\mathcal{X}$, with reproducing kernel K. Let

$$D = \{(t_1, y_1), \ldots, (t_n, y_n)\} \subset \mathcal{X} \times \mathbb{R}$$

be our training data. Recall that we would like a "repository" for the measurment errors which is orthogonal to H, so let (H', B', μ') be another abstract Wiener space, where H' is a real separable infinite-dimensional Hilbert space.

For every $t \in \mathcal{X}$ let $K_t = K(t, \cdot) \in H$ and:

$$\tilde{K}_t := I_\oplus(K_t, 0) \in L^2(B \oplus B', \mu \times \mu'),$$

where $I_{\oplus} : H \oplus H' \to L^2(B \oplus B', \mu \times \mu')$ is the Paley-Wiener map in (1.21). As previously noted:

$$\tilde{K}_t(x, x') = (IK_t)(x),$$

where $I : H \to L^2(B, \mu)$ is the Paley-Wiener map for (H, B, μ). Let $\{e_j\}_{j\in\mathbb{N}}$ be an orthonormal basis for H' and $\lambda > 0$. For every $j \in \mathbb{N}$, let:

$$\tilde{e}_j := I_{\oplus}(0, e_j); \tilde{e}_j(x, x') = (I'e_j)(x'),$$

where $I' : H' \to L^2(B', \mu')$ is the Paley-Wiener map for (H', B', μ'). Then:

$$I_{\oplus}(K_t, \sqrt{\lambda}e_j) = \tilde{K}_t + \sqrt{\lambda}\tilde{e}_j,$$

for all $t \in \mathcal{X}$ and $j \in \mathbb{N}$. Then Lemma 4.3 and Theorem 3.14 yield:

$$\begin{aligned} \hat{f}_{\lambda,D}(t) &= \mathbb{E}[\tilde{K}_t | \tilde{K}_{t_j} + \sqrt{\lambda}\tilde{e}_j = y_j, 1 \leq j \leq n] && (4.19) \\ &= G\tilde{K}_t \left(\bigcap_{j=1}^{n} [\langle (K_{t_j}, \sqrt{\lambda}e_j), \cdot \rangle = y_j] \right), && (4.20) \end{aligned}$$

where $\hat{f}_{\lambda,D}$ is the ridge regression solution in (4.3) and both the conditional expectation and the Gaussian Radon transform above are on $B \oplus B'$. Remark that, in this approach, for any n training points and any future input $t \in \mathcal{X}$, we can just work with the same $\{\tilde{e}_1, \ldots, \tilde{e}_n\}$, and we no longer need to choose $\{e_1, \ldots, e_n\}$ based on the training set and the input t.

Bibliography

[1] Robert J. Adler, *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes*, Institute of Mathematical Statistics Lecture Notes âĂŤ Monograph Series, vol. 12, Institute of Mathematical Statistics, Hayward, CA, 1990.

[2] Aleksandr Y. Aravkin, Bradley M. Bell, James V. Burke, and Gianluigi Pillonetto, *The connection between Bayesian estimation of a Gaussian random field and RKHS*, Submitted to IEEE Transactions on Neural Networks and Learning Systems, available at `http://arxiv.org/pdf/1301.5288v3.pdf`.

[3] George Bachman and Lawrence Narici, *Functional Analysis*, Dover Publications, Inc., Mineola, NY, 2000.

[4] John C. Baez, Irving E. Segal, and Zhengfang Zhou, *Introduction to Algebraic and Constructive Quantum Field Theory*, Princeton University Press, Princeton, NJ, 1992.

[5] Valentine Bargmann, *On a Hilbert space of analytic functions and an associated integral transform, Part I*, Comm. Pure Appl. Math. **14** (1961), 187–214. MR0157250 (28 #486)

[6] Patrick Billingsley, *Probability and measure*, 3rd ed., Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons Inc., New York, 1995. A Wiley-Interscience Publication. MR1324786 (95k:60001)

[7] Vladimir I. Bogachev, *Gaussian measures*, Mathematical Surveys and Monographs, vol. 62, American Mathematical Society, Providence, RI, 1998.

[8] Bruce K. Driver and Maria Gordina, *Square integrable holomorphic functions on infinite-dimensional Heisenberg type groups*, Probab. Theory Related Fields **147** (2010), no. 3-4, 481–528, DOI 10.1007/s00440-009-0213-y. MR2639713 (2011k:58052)

[9] Bruce K. Driver and Brian C. Hall, *The energy representation has no non-zero fixed vectors*, (Leipzig, 1999), CMS Conf. Proc., vol. 29, Amer. Math. Soc., Providence, RI, 2000, pp. 143–155. MR1803410 (2002f:22034)

[10] Bruce K. Driver, *Probability Tools with Examples*, available at `http://www.math.ucsd.edu/~bdriver/Cornell%20Summer%20Notes%202010/Lecture_Notes/Probability%20Tools%20with%20Examples.pdf`.

[11] Rick Durrett, *Probability. Theory and examples*, 3rd ed., Duxbury Avanced Series, Thompson Brooks/Cole, Belmont, CA, 2005.

[12] Nate Eldredge, *Math 7770: Analysis and Probability on Infinite-Dimensional Spaces*, Online class notes (2012), available at `http://www.math.cornell.edu/~neldredge/7770/7770-lecture-notes.pdf`.

[13] Leonard Gross, *Abstract Wiener spaces*, Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66), Vol. II: Contributions to Probability Theory, Part 1, Univ. California Press, Berkeley, Calif., 1967, pp. 31–42. MR0212152 (35 #3027)

[14] Brian C. Hall, *The heat operator in infinite dimensions*, Infinite dimensional stochastic analysis, QP–PQ: Quantum Probab. White Noise Anal., vol. 22, World Sci. Publ., Hackensack, NJ, 2008, pp. 161–174. MR2412888 (2010a:58044)

[15] Sigurdur Helgason, *The Radon transform*, 2nd ed., Progress in Mathematics, vol. 5, Birkhäuser Boston Inc., Boston, MA, 1999. MR1723736 (2000m:44003)

[16] Irina Holmes and Ambar N. Sengupta, *A Gaussian Radon transform for Banach spaces*, J. Funct. Anal. **263** (2012), no. 11, 3689–3706. MR2984081

[17] Irina Holmes, *An inversion formula for the Gaussian Radon transform for Banach spaces*, Infin. Dimens. Anal. Quantum Probab. Relat. Top. **16** (2013), no. 4, available at `http://arxiv.org/pdf/1308.1392.pdf`.

[18] Ferenc Huszár and Simon Lacoste-Julien, *A Kernel Approach to Tractable Bayesian Nonparametrics*, arXiv preprint (2011), available at `http://arxiv.org/pdf/1103.1761v3.pdf`.

[19] George S. Kimeldorf and Grace Wahba, *A correspondence between Bayesian estimation on stochastic processes and smoothing by splines*, Ann. Math. Statist. **41** (1970), 495–502, available at `http://www.stat.wisc.edu/~wahba/ftp1/oldie/kw70bayes.pdf`. MR0254999 (40 #8206)

[20] ———, *Some Results on Tchebycheffian Spline Functions*, Journal of Mathematical Analysis and Applications **33** (1971), no. 1, 82–95, available at `http://www.stat.wisc.edu/~wahba/ftp1/oldie/kw71.pdf`.

[21] Hui Hsiung Kuo, *Gaussian measures in Banach spaces*, Lecture Notes in Mathematics, Vol. 463, Springer-Verlag, Berlin, 1975. MR0461643 (57 #1628)

[22] Vochita Mihai and Ambar N. Sengupta, *The Radon-Gauss transform*, Soochow J. Math. **33** (2007), no. 3, 415–433. MR2344371 (2009a:44006)

[23] Umut Ozertem and Deniz Erdogmus, *RKHS Bayes Discriminant: A Subspace Constrained Nonlinear Feature Projection for Signal Detection*, Neural Networks, IEEE Transactions on **20** (2009), no. 7, 1195-1203.

[24] Natesh S. Pillai, Qiang Wu, Feng Liang, Sayan Mukherjee, and Robert L. Wolpert, *Characterizing the function space for Bayesian kernel models*, J. Mach. Learn. Res. **8** (2007), 1769–1797, available at `http://jmlr.org/papers/volume8/pillai07a/pillai07a.pdf`. MR2332448

[25] Johann Radon, *Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten*, Computed tomography (Cincinnati, Ohio, 1982), Proc. Sympos. Appl. Math., vol. 27, Amer. Math. Soc., Providence, R.I., 1982, pp. 71–86 (German). MR692055 (84f:01040)

[26] Irving E. Segal, *The complex-wave representation of the free boson field*, Topics in functional analysis (essays dedicated to M. G. Kreĭn on the occasion of his 70th birthday), Adv. in Math. Suppl. Stud., vol. 3, Academic Press, New York, 1978, pp. 321–343. MR538026 (82d:81069)

[27] Peter Sollich, *Bayesian Methods for Support Vector Machines: Evidence and Predictive Class Probabilities*, Mach. Learn. **46** (2002), no. 1-3, 21–52, DOI 10.1023/A:1012489924661.

[28] Ingo Steinwart and Andreas Christmann, *Support vector machines*, Information Science and Statistics, Springer, New York, 2008. MR2450103 (2010f:62002)

[29] Daniel W. Stroock, *Abstract Wiener Spaces, Revisited*, Communications on Stochastic Analysis **2** (2008), no. 1, 145-151.

[30] Norbert Wiener, *Differential Space*, Journal of Mathematics and Physics **2** (1923), 131-174.

[31] Zhihua Zhang, Guang Dai, Donghui Wang, and Michael I. Jordan, *Bayesian Generalized Kernel Models*, Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (AISTATS-10), 2010, pp. 972-979.

Appendix A
Category Theory and the Kolmogorov Existence Theorem

A.1 Kolmogorov's Existence Theorem

Let T be an infinite index set and $\mathcal{J}$ be the collection of all finite non-empty subsets of T. Suppose that

$$\{(\Omega_t, \mathcal{F}_t)\}_{t \in T}$$

is an indexed family of measurable spaces. For every non-empty set $J \subset T$ let Ω_J denote the product:

$$\Omega_J := \prod_{t \in J} \Omega_t = \{x : J \to \cup_{t \in J} \Omega_t : x(t) = \omega_t \in \Omega_t, \forall t \in J\}.$$

We denote such elements x of Ω_J as $\omega = (\omega_t)_{t \in J}$. In particular, let Ω' denote Ω_T:

$$\Omega' := \prod_{t \in T} \Omega_t.$$

For any non-empty subsets $I \subset J \subset T$ let π_I^J denote the projection map from Ω_J onto Ω_I:

$$\pi_I^J : \Omega_J \to \Omega_I;\ \pi_I^J\left((\omega_t)_{t \in J}\right) = (\omega_t)_{t \in I}. \tag{A.1}$$

Note that if $\emptyset \neq H \subset I \subset J \subset T$:

$$\pi_H^I \circ \pi_I^J = \pi_H^J. \tag{A.2}$$

For every non-empty subset $J \subset T$ define the collection $\mathcal{E}_J$ as follows:

$$\mathcal{E}_J := \left\{ \prod_{t \in J} B_t : B_t \in \mathcal{F}_t, \forall t \in J, \text{ and } B_t = \Omega_t \text{ for all but finitely many } t \in J \right\},$$

and let $\mathcal{F}_J = \sigma(\mathcal{E}_J)$ be the σ-algebra generated by $\mathcal{E}_J$. In particular, let $\mathcal{E}'$ denote $\mathcal{E}_T$ and $\mathcal{F}' = \sigma(\mathcal{E}')$. Then $(\Omega_J, \mathcal{F}_J)$ is a measurable space for every $\emptyset \neq J \subset T$, called the product of the family $\{(\Omega_t, \mathcal{F}_t)\}_{t \in J}$ and denoted:

$$(\Omega_J, \mathcal{F}_J) = \otimes_{t \in J} (\Omega_t, \mathcal{F}_t).$$

Note that $\mathcal{F}_J$ is the smallest σ-algebra on Ω_J with respect to which $\pi_{\{t\}}^J$ is $\mathcal{F}_J$ - $\mathcal{F}_{\{t\}}$ measurable for every $t \in J$.

Now for every non-empty finite subset $F \in \mathcal{J}$ of T let the collection $\mathcal{C}_F$ be defined as:

$$\mathcal{C}_F := \{\pi_F^{-1}(B) : B \in \mathcal{F}_F\} \subset \Omega,$$

where π_F denotes the projection map π_F^T. Elements of $\mathcal{C}_F$ are called *F-cylinders.* Finally, define the collection $\mathcal{C}$ of *cylinders* by:

$$\mathcal{C} := \bigcup_{F \in \mathcal{J}} \mathcal{C}_F.$$

The collection $\mathcal{C}$ is an algebra of subsets of Ω', and in fact it generates $\mathcal{F}'$:

$$\mathcal{F}' = \sigma(\mathcal{C}).$$

With these notations, we may now give the following definition.

Definition A.1. Suppose that μ_F is a probability measure on $(\Omega_F, \mathcal{F}_F)$ for every $F \in \mathcal{J}$. We say that $(\mu_F)_{F \in \mathcal{J}}$ is a *projective family* of probability measures if:

$$\pi_F^G(\mu_G) = \mu_F, \text{ for all } F, G \in \mathcal{J} \text{ with } F \subset G, \tag{A.3}$$

where the left-hand side is the image measure of μ_G under π_F^G:

$$\pi_F^G(\mu_G)(B) = \mu_G\left[(\pi_F^G)^{-1}(B)\right], \text{ for all } B \in \mathcal{F}_F.$$

A probability measure μ' on $(\Omega', \mathcal{F}')$ is called a *projective limit* of the projective family $(\mu_F)_{F \in \mathcal{J}}$ if:

$$\pi_F(\mu') = \mu_F, \text{ for all } F \in \mathcal{J}, \tag{A.4}$$

where, once again, the left-hand side is the image measure of μ' under π_F. In this case, we write:

$$\mu' = \varprojlim_{F \in \mathcal{J}} \mu_F.$$

It is easily shown that any projective family of probability measures has at most one projective limit. Perhaps the most important example of a projective limit is the product measure: suppose $(\Omega_t, \mathcal{F}_t, \mu_t)$ is a probability space for every $t \in T$. We can quickly form a projective family by letting $\mu_F = \otimes_{t \in F} \mu_t$ be the finite product measure for every $F \in \mathcal{J}$. Then $(\mu_F)_{F \in \mathcal{J}}$ has a projective limit μ', which is the unique measure on Ω such that:

$$\mu'(B) = \prod_{t \in T} \mu_t(B_t), \text{ for all } B = \prod_{t \in T} B_t \in \mathcal{E}'.$$

In general, it may not be possible to determine whether or not an arbitrary projective family of probability measures has a projective limit. However, if each Ω_t is a *Polish space* - a separable topological space which is metrizable by means of a complete metric - and $\mathcal{F}_t$ is the Borel σ-algebra of Ω_t, the celebrated *Kolmogorov Extension Theorem* shows that this is indeed possible. We state this result next.

Theorem A.1. *Let T be an infinite index set, $\mathcal{J}$ denote the collection of all non-empty finite subsets of T, and $\mathcal{X}_t$ be a Polish space with its Borel σ-algebra $\mathcal{B}_t$ for every $t \in T$. Then any projective family $(\mu_F)_{F \in \mathcal{J}}$ of probability measures on (Ω_F, μ_F) for $F \in \mathcal{J}$ has a projective limit.*

Next, we will "translate" this fundamental result in the language of category theory. Before we proceed, we review some of the basic concepts of this theory.

A.2 Category Theory

Suppose we have a collection *Obj* of *objects* and a collection *Mor* of *morphisms*, or *arrows*, where to each morphism f there is associated an object $s(f)$ called its *domain*, or *source*, and an object $t(f)$ called its *codomain*, or *target*. We express this as:

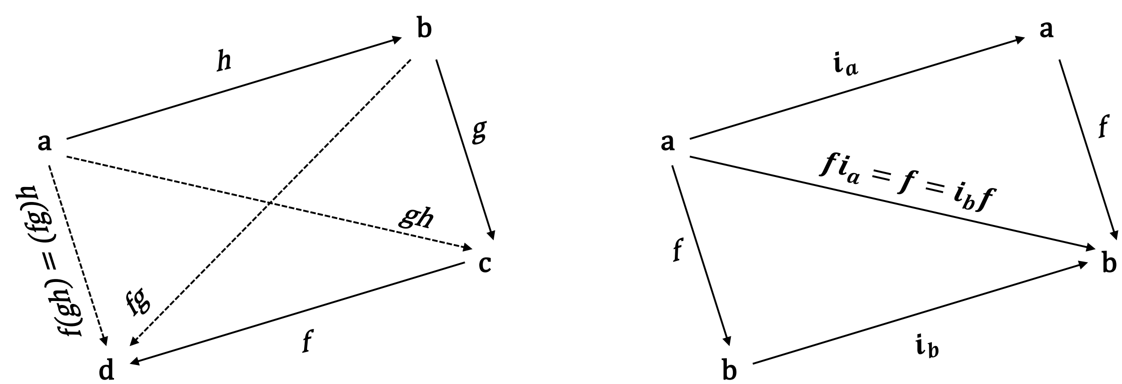
$$s(f) \xrightarrow{f} t(f) \text{ or } f : s(f) \to t(f).$$

We let $Mor(a, b)$ denote the set of all morphisms with source a and target b, for any objects a and b. A *composition rule* is defined by associating to each $f \in Mor(a, b)$ and $g \in Mor(b, c)$ a morphism $gf \in Mor(a, c)$, for all objects a, b, c.

We say that *Obj* and *Mor* specify a *category* provided that the composition rule satisfies the following axioms (pictured in Figure A.1):

i. *Associativity:* $f(gh) = (fg)h$ for all $h \in Mor(a, b)$, $g \in Mor(b, c)$ and $f \in Mor(c, d)$ and all objects a, b, c, d.

ii. *Identity Morphism:* For each object a there is an arrow $i_a \in Mor(a, a)$ such that $fi_a = i_b f = f$ for any $f \in Mor(a, b)$.



(a) Associativity. (b) Identity Morphism.

Figure A.1: Axioms of category theory.

An arrow $f \in Mor(a, b)$ is said to be an *isomorphism* if there is an arrow $g \in Mor(b, a)$ such that $gf = i_a$ and $fg = i_b$. An *initial object* is an object o with the property that for any object a there is a unique arrow $o \to a$. Similarly, a *final object* is an object $\hat{o}$ such that for any object a there is a unique arrow $a \to \hat{o}$. It is easily shown that any two initial (final) objects are isomorphic. Let us look at some basic examples of categories.

Example A.1. The category *Set*:

- <ins>Objects</ins>: Sets.

- Morphisms: Functions.

The initial object of this category is the empty set $\emptyset$ (with the empty functions $\emptyset \to A$ as morphisms) and every singleton set $\{a\}$ is a final object (with morphisms the functions that map all the elements of the source set to the single target element).

Example A.2. The category $Vect_{\mathbb{F}}$:

- Objects: All vector spaces over a fixed field $\mathbb{F}$.
- Morphisms: Linear transformations.

The zero vector space $\{0\}$ is both the initial object (with morphisms $0 \mapsto 0$) and the final object (with morphisms $v \mapsto 0$) in this category.

Example A.3. The category *Top*:

- Objects: Topological spaces.
- Morphisms: Continuous functions.

The initial object in this category is $\emptyset$ (as a topological space) and any singleton (as the topology $\{\emptyset, \{a\}\}$) is a final object.

Example A.4. *A category on a poset:* Let Γ be a non-empty set equipped with a partial order $\leq$, that is:

i. $\alpha \leq \alpha$ for all $\alpha \in \Gamma$.

ii. If $\alpha \leq \beta$ and $\beta \leq \alpha$ for some $\alpha, \beta \in \Gamma$, then $\alpha = \beta$.

iii. If $\alpha \leq \beta$ and $\beta \leq \kappa$ for some $\alpha, \beta, \kappa \in \Gamma$, then $\alpha \leq \kappa$.

We form a category as follows:

- Objects: The elements of Γ.
- Morphisms: We let $\alpha \to \beta$ denote $\alpha \leq \beta$.

Note that for any $\alpha, \beta \in \Gamma$ we either have a unique arrow $\alpha \to \beta$ or there is no arrow from α to β. A rule of composition is provided by transitivity of the partial order. Any minimal object is an initial object, and any maximal elements is a final object in this category.

Next, we define a special type of map between categories, called a functor.

Definition A.2. Suppose that C_1 and C_2 are categories. A *functor* $F : C_1 \to C_2$ associates to every object $a \in Obj(C_1)$ an object $F(a) \in Obj(C_2)$ and to each morphism $f : a \to b$ in C_1 a morphism $F(f) : F(a) \to F(b)$ in C_2 such that:

i. $F(fg) = F(f)F(g)$, whenever fg is defined.

ii. $F(i_a) = i_{F(a)}$, for all objects a in C_1.

Similarly, a *contravariant functor* $F : C_1 \to C_2$ associates to each object $a \in Obj(C_1)$ an object $F(a) \in Obj(C_2)$, but to each morphism $f : a \to b$ in C_1 a morphism $F(f) : F(b) \to F(a)$ in C_2 such that:

i. $F(fg) = F(g)F(f)$, whenever fg is defined.

ii. $F(i_a) = i_{F(a)}$, for all objects a in C_1.

Another important concept for our discussion is that of diagram.

Definition A.3. A *diagram* in a category C is a non-empty set D of objects and arrows between these objects in C such that D is itself a category, that is D contains composites whenever they are defined in C and D also contains i_a for all objects a in D. We say that a diagram is *commutative* if $f_1 f_2 \dots f_n = g_1 g_2 \dots g_m$ for all arrows f_i, g_j in D for which these composites are defined - in other words, all directed paths within the diagram D that have the same endpoints lead to the same result by composition.

An important type of diagram is an *indexed diagram.* Suppose Γ is a non-empty set equipped with a partial order $\leq$ and C is a category. Consider a contravariant functor $F : \Gamma \to C$, where Γ is considered a category as in Example A.4. Then F associates to every $\alpha \in \Gamma$ an object F_α in C, and to every $\alpha \leq \beta$ in Γ a morphism $f_{\alpha\beta} : F_\beta \to F_\alpha$ that satisfies:

$$f_{\alpha\beta} f_{\beta\kappa} = f_{\alpha\kappa}, \text{ for all } \alpha \leq \beta \leq \kappa \text{ in } \Gamma.$$

Note that F associates to $\alpha \leq \alpha$ in Γ the identity morphism $f_{\alpha\alpha} = i_{F_\alpha}$. Therefore the objects F_α along with the morphisms $f_{\alpha\beta}$ form a commutative diagram in C. We denote such an indexed diagram by:

$$D(F),$$

whenever $F : \Gamma \to C$ is a contravariant functor from a poset into the category C.

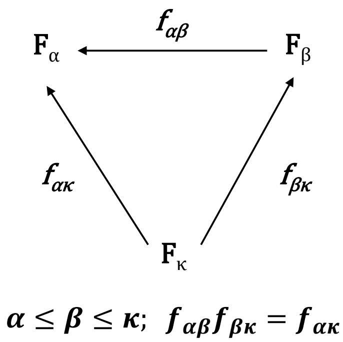
Finally, we discuss the concept of cone in a category.

Definition A.4. Let D be a commutative diagram in a category C and v be an object in C. A *cone* with vertex v and base D is specified by arrows $p_b : v \to b$ for every object b in D such that the combined diagram of D along with the object v and all the arrows p_b, is commutative. We denote such a cone by $v \star D$.

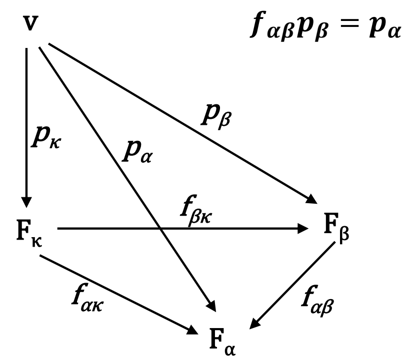
The collection of all cones with base D can be thought of as the objects of a category we denote $Cone_D$. The morphisms in this category are defined as follows: if $v \star D$ and $w \star D$ are cones with base D and vertices v and w, respectively, a morphism of cones $f : (v \star D) \to (w \star D)$ is an arrow $f : v \to w$ in the category C such that the combined diagram $[(v \star D), f, (w \star D)]$ is commutative. We are now ready to define the concept of projective limit in the sense of category theory.

Definition A.5. Let D be a commutative diagram in a category C. A *projective limit* of the diagram D is a final object in the category $Cone_D$, that is a cone $\hat{o} \star D$ such that for any cone $v \star D$ there is a unique morphism of cones $v \star D \to \hat{o} \star D$.

We will deal in particular with cones whose base is an indexed diagram $D(F)$, where $F : \Gamma \to C$ is a contravariant functor from a poset Γ to the category C. Then, as pictured in Figure A.2, the cone $v \star D(F)$ is specified by morphisms $p_\alpha : v \to F_\alpha$ for all $\alpha \in \Gamma$ that satisf $f_{\alpha\beta} p_\beta = p_\alpha$ for all $\alpha \leq \beta$ in Γ.



(a) An indexed diagram.



(b) An indexed cone.

Figure A.2: An indexed diagram and an indexed cone.

If $v \star D(F)$ and $w \star D(F)$ are indexed cones with base $D(F)$, specified by morphisms $p_\alpha : v \to F_\alpha$ and $q_\alpha : w \to F_\alpha$, respectively, a morphism $f : v \star D(F) \to w \star D(F)$ is an arrow $f : v \to w$ in C that satisfies $q_\alpha f = p_\alpha = f_{\alpha\beta} p_\beta$ for all $\alpha \leq \beta$ in Γ, as pictured in Figure A.3.

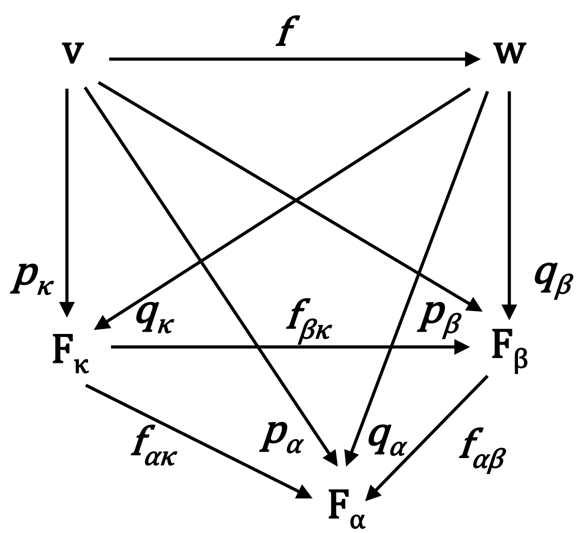


Figure A.3: A morphism of indexed cones.

A.3 A Categorical Interpretation of Kolmogorov's Existence Theorem

In an attempt to "translate" Kolmogorov's Existence Theorem into the language of category theory, we posed the question of forming a category of Polish probability spaces - that is, probability spaces $(\mathcal{F}, \mathcal{B}, \mu)$ where $\mathcal{X}$ is a Polish space and $\mathcal{B}$ is its Borel σ-algebra - and investigate whether or not every diagram has a projective limit. However, one issue quickly arises: while countable products of Polish spaces are also Polish spaces, arbitrary products of Polish spaces are not necessarily so. We therefore worked within the larger category $\mathbb{P}$ whose objects are probability spaces and whose morphisms are measurable functions, with regular composition of functions. We work with a special type of diagram in this category, which we call a Kolmogorov diagram and define next. Essentially, this is an indexed diagram whose objects are Polish probability spaces and some extra assumptions are made on its morphisms.

Definition A.6. Let Γ be an infinite index set equipped with a *directed* partial order $\leq$, that is for every $\alpha, \beta \in \Gamma$ there is $\kappa \in \Gamma$ such that $\alpha \leq \kappa$ and $\beta \leq \kappa$. Let $F : \Gamma \to \mathbb{P}$ be a contravariant functor, so for every $\alpha \in \Gamma$ we have a probability space $(\Omega_\alpha, \mathcal{B}_\alpha, \mu_\alpha)$ and for every $\alpha \leq \beta$ in Γ there are measurable functions $p_{\alpha\beta} : \Omega_\beta \to \Omega_\alpha$ such that:

$$p_{\alpha\beta} p_{\beta\kappa} = p_{\alpha\kappa}, \text{ for all } \alpha \leq \beta \leq \kappa \text{ in } \Gamma. \tag{A.5}$$

We say that the resulting indexed diagram $D(F)$ is a **Kolmogorov diagram** provided that the following conditions hold:

i. For every $\alpha \in \Gamma$, Ω_α is a Polish space and $\mathcal{B}_\alpha$ is its Borel σ-algebra.

ii. Each map $p_{\alpha\beta}$ is *continuous* and *surjective*.

iii. The maps $p_{\alpha\beta}$ satisfy the **Kolmogorov Consistency Condition**:

$$\mu_\beta \left(p_{\alpha\beta}^{-1}(B) \right) = \mu_\alpha(B), \text{ for all } \alpha \leq \beta \text{ in } \Gamma \text{ and } B \in \mathcal{B}_\alpha. \tag{A.6}$$

Suppose now that $D(F)$ is a Kolmogorov diagram as in Definition A.6 and let:

$$\Omega' := \prod_{\alpha \in \Gamma} \Omega_\alpha.$$

The set Ω' is unfortunately too large to be truly useful, so we concentrate instead on a special subset Ω of Ω', defined by:

$$\Omega := \left\{ (\omega_\alpha)_{\alpha \in \Gamma} \in \Omega' : p_{\alpha\beta}(\omega_\beta) = \omega_\alpha, \forall \alpha \leq \beta \text{ in } \Gamma \right\}. \tag{A.7}$$

If we let $\pi_\alpha : \Omega' \to \Omega_\alpha$; $\omega \mapsto \omega_\alpha$ denote the projection map for every $\alpha \in \Gamma$ and

$$p_\alpha = \pi_\alpha|_\Omega$$

denote the restriction of each projection map to Ω, the condition in the definition of Ω becomes:

$$p_{\alpha\beta}p_\beta = p_\alpha, \text{ for all } \alpha \leq \beta \text{ in } \Gamma. \tag{A.8}$$

As pictured in Figure A.4, this makes us think of a cone. However, we do not yet have a cone, because we need a probability space for the vertex. So our goal will be to construct a σ-algebra and a probability measure on Ω. Before we proceed though, one important issue remains: Ω could be empty! To avoid this possibility, we impose the Sequential Maximality Condition, first introduced by Bochner. As we shall see below, this condition ensures that Ω is non-empty.

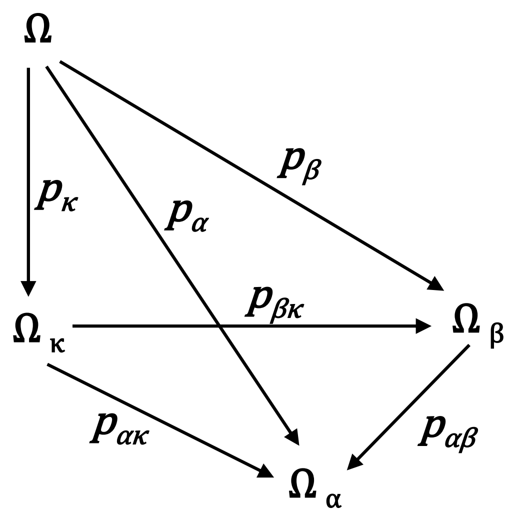


Figure A.4: A possible cone based on a Kolmogorov diagram.

Definition A.7. Let $D(F)$ be a Kolmogorov diagram as in Definition A.6 and Ω be the set defined in (A.7). We say that $D(F)$ satisfies the **Sequential Maximality Condition** provided that for every increasing sequence $\alpha_1 \leq \alpha_2 \leq \ldots$ in Γ and every sequence $\{\omega_n\}_{n\in\mathbb{N}}$ with $\omega_n \in \Omega_{\alpha_n}$ for every n such that:

$$\omega_n = p_{\alpha_n\alpha_{n+1}}(\omega_{n+1}), \text{ for all } n \in \mathbb{N},$$

there exists $\omega \in \Omega$ such that $p_{\alpha_n}(\omega) = \omega_n$ for all n.

The next result shows that imposing these conditions on a diagram in $\mathbb{P}$ leads to a projective limit in the sense of category theory. Moreover, we will see how this result implies Kolmogorov's Existence Theorem.

Theorem A.2. *Let $D(F)$ be a Kolmogorov diagram, as in Definition A.6, which satisfies the Sequential Maximality Condition. Then:*

i. The set Ω defined in (A.7) *is non-empty. Moreover, the projection maps $p_\alpha : \Omega \to \Omega_\alpha$ are surjective for all $\alpha \in \Gamma$.*

ii. For every $\alpha \in \Gamma$ consider the σ-algebra of subsets of Ω generated by p_α:

$$\mathcal{B}'_\alpha := p_\alpha^{-1}(\mathcal{B}_\alpha) = \{p_\alpha^{-1}(B) : B \in \mathcal{B}_\alpha\},$$

and let:

$$\mathcal{A} := \bigcup_{\alpha \in \Gamma} \mathcal{B}'_\alpha.$$

Then $\mathcal{A}$ is an algebra of subsets of Ω and:

$$\mu'\left(p_\alpha^{-1}(B)\right) := \mu_\alpha(B), \text{ for all } \alpha \in \Gamma, B \in \mathcal{B}_\alpha, \tag{A.9}$$

defines a countably additive probability measure on $\mathcal{A}$.

iii. Let μ denote the extension of μ' to $\mathcal{F} := \sigma(\mathcal{A})$. Then the cone $\Omega \star D(F)$, with vertex $(\Omega, \mathcal{F}, \mu)$ and base $D(F)$, is the projective limit of the diagram $D(F)$.

Proof. i. Let $\alpha_1 \leq \alpha_2 \leq \ldots$ be an increasing sequence in Γ and pick for every $n \in \mathbb{N}$ a non-empty Borel set $B_n \in \mathcal{B}_{\alpha_n}$ such that:

$$\inf_{n \in \mathbb{N}} \mu_{\alpha_n}(B_n) = \epsilon,$$

for some $\epsilon > 0$. For instance, we can choose $B_n = \Omega_{\alpha_n}$ for all n and $\epsilon = 1$. Now recall that any finite Borel measure on a Polish space is *regular*, so for every n we may choose a non-empty compact set $D_n \subset B_n$ such that:

$$\mu_{\alpha_n}(B_n \setminus D_n) < \frac{\epsilon}{2^{n+1}}. \tag{A.10}$$

Let:

$$C_n := p_{\alpha_1\alpha_n}^{-1}(D_1) \cap p_{\alpha_2\alpha_n}^{-1}(D_2) \cap \ldots \cap p_{\alpha_n\alpha_n}^{-1}(D_n), \text{ for all } n \in \mathbb{N}. \tag{A.11}$$

Recall that $p_{\alpha_n\alpha_n} = 1\!\!1_{\Omega_{\alpha_n}}$, so $C_n \subset D_n$. Since every map $p_{\alpha_j\alpha_n}$ is continuous, C_n is a closed subset of the compact space D_n, therefore C_n is compact.

To see that each C_n is non-empty, note that for all n:

$$\begin{aligned} B_n \setminus C_n &= \bigcup_{j=1}^{n} \left(B_n \setminus p_{\alpha_j\alpha_n}^{-1}(D_j)\right) \\ &= \bigcup_{j=1}^{n} p_{\alpha_j\alpha_n}^{-1}(B_j \setminus D_j), \end{aligned}$$

so:

$$\begin{aligned} \mu_{\alpha_n}(B_n \setminus C_n) &\leq \sum_{j=1}^{n} \mu_{\alpha_n}\left(p_{\alpha_j\alpha_n}^{-1}(B_j \setminus D_j)\right) \\ &= \sum_{j=1}^{n} \mu_{\alpha_j}(B_j \setminus D_j) \qquad \text{(by (A.6))} \\ &< \sum_{j=1}^{\infty} \frac{\epsilon}{2^{n+1}} \\ &= \frac{\epsilon}{2}. \end{aligned}$$

Then:

$$\mu_{\alpha_n}(C_n) > \mu_{\alpha_n}(B_n) - \frac{\epsilon}{2} \geq \frac{\epsilon}{2} > 0,$$

which proves that C_n cannot be empty.

Now let $n \leq m$ be positive integers. Then:

$$p_{\alpha_n\alpha_m}(C_m) = p_{\alpha_n\alpha_m}\left(\bigcap_{j=1}^{m} p^{-1}_{\alpha_j\alpha_n}(D_j)\right) \subset \bigcap_{j=1}^{n} p_{\alpha_n\alpha_m}p^{-1}_{\alpha_j\alpha_m}(D_j).$$

If $1 \leq j \leq n$, then $\alpha_j \leq \alpha_n \leq \alpha_m$, so $p_{\alpha_n\alpha_n}p_{\alpha_n\alpha_m} = p_{\alpha_j\alpha_m}$. Then:

$$p_{\alpha_n\alpha_m}p^{-1}_{\alpha_j\alpha_m}(D_j) = p^{-1}_{\alpha_j\alpha_n}(D_j),$$

by surjectivity of $p_{\alpha_n\alpha_m}$. So:

$$p_{\alpha_n\alpha_m}(C_m) \subset \bigcap_{j=1}^{n} p^{-1}_{\alpha_j\alpha_n}(D_j) = C_n.$$

We have shown that:

$$p_{\alpha_n\alpha_m}(C_m) \subset C_n, \text{ for all } n \leq m. \tag{A.12}$$

Consider now:

$$C := \prod_{n\in\mathbb{N}} C_n \subset \Omega',$$

a compact set by Tychonoff's theorem. For every n set:

$$K_n := \left\{\omega \in C : p_{\alpha_n\alpha_{n+1}}(\omega_{n+1}) = \omega_n\right\}.$$

Remark that each K_n is non-empty: by (A.12), we may simply pick $\omega_{n+1} \in C_{n+1}$ and $\omega_n = p_{\alpha_n\alpha_{n+1}}(\omega_{n+1}) \in C_n$. For every n consider the map:

$$f_n : C \to C_n \times C_n;\ f_n(\omega) = \left(\omega_n, p_{\alpha_n\alpha_{n+1}}(\omega_{n+1})\right).$$

Then f_n is continuous, so the diagonal Δ_n of f_n is closed. Therefore $K_n = f_n^{-1}(\Delta_n)$ is closed in C, so K_n is compact. If we can show that the collection $\{K_n\}_{n\in\mathbb{N}}$ has the Finite Intersection Property, we will have that $\bigcap_{n\in\mathbb{N}} K_n \neq \emptyset$. In turn, this gives us a sequence $\{\omega_n\}$ with $\omega_n \in C_n \subset \Omega_{\alpha_n}$ and:

$$p_{\alpha_n\alpha_{n+1}}(\omega_{n+1}) = \omega_n$$

for all n. Since our diagram satisfies the Sequential Maximality Condition, there is $\omega \in \Omega$ such that $p_{\alpha_n}(\omega) = \omega_n$ for all n, proving that Ω is non-empty.

So consider positive integers $n_1 < n_2 < \ldots < n_m$, pick any

$$u_{n_m+1} \in C_{n_m+1},$$

and let:

$$u_i = p_{\alpha_i\alpha_{n_j+1}}(u_{n_j+1}) \in C_i, \text{ for all } n_{j-1}+1 \leq i \leq n_j, 1 \leq j \leq m,$$

where $n_0 = 0$. Let $\omega \in C$ be given by:

$$\omega_i = \begin{cases} u_i, \text{ if } i \leq n_m + 1 \\ \text{arbitrary in } C_i, \text{ if } i > n_m + 1. \end{cases}$$

Then for any $1 \leq j \leq m$:

$$p_{\alpha_j \alpha_{n_j+1}}(\omega_{n_j+1}) = u_{n_j} = \omega_{n_j},$$

which proves that $\omega \in K_j$ for all $1 \leq j \leq m$, so $\{K_n\}_n$ has the Finite Intersection Property, as desired.

To prove that the projection maps are surjective, let $\alpha \in \Gamma$ and $\omega'_\alpha \in \Omega_\alpha$. Choose $\alpha_2 \geq \alpha_1 := \alpha$ in Γ. Since $p_{\alpha_1\alpha_2}$ is surjective, choose $\omega_2 \in \Omega_{\alpha_2}$ such that $\omega_1 := \omega'_\alpha = p_{\alpha_1\alpha_2}(\omega_2)$. Continuing this process inductively, we obtain a sequence $\alpha_1 \leq \alpha_2 \leq \ldots$ in Γ and $\omega_n \in \Omega_{\alpha_n}$ for all n such that $\omega_n = p_{\alpha_n\alpha_{n+1}}(\omega_{n+1})$. Then from the Sequential Maximality Condition there is $\omega \in \Omega$ such that $p_\alpha(\omega) = \omega'_\alpha$, or $p_\alpha(\omega) = \omega'_\alpha$.

ii. We begin by noting that if $\alpha \leq \beta$ in Γ and $E = p_\alpha^{-1}(B)$ for some $B \in \mathcal{B}_\alpha$, we may express E as:

$$E = (p_{\alpha\beta} p_\beta)^{-1}(B) = p_\beta^{-1}\left(p_{\alpha\beta}^{-1}(B)\right),$$

where we used (A.8). Since $p_{\alpha\beta}$ is measurable, this means that $E \in \mathcal{B}'_\beta$:

$$\mathcal{B}'_\alpha \subset \mathcal{B}'_\beta, \text{ for all } \alpha \leq \beta \text{ in } \Gamma. \tag{A.13}$$

Now $\mathcal{A}$ is clearly closed under complementation. If $A, B \in \mathcal{A}$ then there are $\alpha, \beta \in \Gamma$ such that $A \in \mathcal{B}'_\alpha$ and $B \in \mathcal{B}'_\beta$. But since Γ is directed, there is $\kappa \in \Gamma$ such that $\alpha \leq \kappa$ and $\beta \leq \kappa$. From (A.13), both A and B are in $\mathcal{B}'_\kappa$, so $A \cup B \in \mathcal{B}'_\kappa \subset \mathcal{A}$. So indeed $\mathcal{A}$ is an algebra.

Let $E \in \mathcal{A}$ and $\alpha \in \Gamma$ such that $E \in \mathcal{B}'_\alpha$. Note that since p_α is surjective, $B = p_\alpha(E)$ is the unique element of $\mathcal{B}_\alpha$ such that $E = p_\alpha^{-1}(B)$. Now suppose $\beta \in \Gamma$ is such that $E \in \mathcal{B}'_\beta$ as well. Pick $\kappa \in \Gamma$ such that $\alpha \leq \kappa$ and $\beta \leq \kappa$. As above, we may express:

$$\begin{aligned} \mu'(E) &= \mu'\left(p_\kappa^{-1}(p_{\alpha\kappa}^{-1}(B))\right) \\ &= \mu_\kappa(p_{\alpha\beta}^{-1}(B)) \\ &= \mu_\alpha(B), \end{aligned}$$

where the last equality follows from the Kolmogorov Consistency Condition. Similarly, we see that $\mu'(E) = \mu_\beta(C)$, where $C \in \mathcal{B}'_\beta$ is such that $E = p_\beta^{-1}(C)$. So μ' is well-defined. To see that μ' is a finitely additive measure, note that:

$$\mu'(\emptyset) = \mu_\alpha(\emptyset) = 0 \text{ and } \mu'(\Omega) = \mu_\alpha(\Omega_\alpha) = 1, \text{ for any } \alpha \in \Gamma,$$

and if $A, B \in \mathcal{A}$ with $A \cap B = \emptyset$ then there is $\alpha \in \Gamma$ such that $A = p_\alpha^{-1}(C)$ and $B = p_\alpha^{-1}(D)$, for disjoint elements C, D in $\mathcal{B}_\alpha$, so:

$$\mu'(A \cup B) = \mu_\alpha(C \cup D) = \mu_\alpha(C) + \mu_\alpha(D) = \mu'(A) + \mu'(B).$$

Finally, to see that μ' is countably additive on $\mathcal{A}$, it suffices to prove that if $E_1 \supset E_2 \supset \ldots$ is a decreasing sequence of non-empty sets in $\mathcal{A}$ such that:

$$\inf_{n\in\mathbb{N}} \mu'(E_n) = \epsilon > 0,$$

then $\cap_n E_n$ is non-empty. Because Γ is directed, we may produce a sequence $\alpha_1 \leq \alpha_2 \leq \ldots$ in Γ such that $E_n \in \mathcal{B}'_{\alpha_n}$ for al n. So let $B_n \in \mathcal{B}_{\alpha_n}$ be such that $E_n = p_{\alpha_n}^{-1}(B_n)$ for all n. We use regularity of μ_{α_n} to choose for every n a non-empty compact subset $D_n \subset B_n$ such that $\mu_{\alpha_n}(B_n \setminus D_n) < {}^{\epsilon}/_{2^{n+1}}$ and let C_n be defined as in (A.11). Then we proceed exactly as in the proof of part *i.* of this theorem, and produce an element $\omega \in \Omega$ with $p_{\alpha_n}(\omega) \in C_n \subset B_n$. Then $\omega \in \cap_n E_n$, which proves the claim.

iii. Now that we proved μ' is countably additive on $\mathcal{A}$, the Hopf Extension Theorem shows that there is a probability measure μ on $\mathcal{F} = \sigma(\mathcal{A})$ such that:

$$\mu\left(p_\alpha^{-1}(B)\right) = \mu_\alpha(B), \text{ for all } \alpha \in \Gamma, B \in \mathcal{B}_\alpha. \tag{A.14}$$

We now have a legitimate cone in the category $\mathbb{P}$, as in A.4.

Suppose that $(\mathcal{X}, \mathcal{B}, \lambda)$ is the vertex of another cone in $\mathbb{P}$ with base $D(F)$, with maps $q_\alpha : \mathcal{X} \to \Omega_\alpha$ for every $\alpha \in \Gamma$. For every $x \in \mathcal{X}$, consider:

$$f(x) = y := (q_\alpha(x))_{\alpha\in\Gamma} \in \Omega'.$$

Then $p_{\alpha\beta}(y_\beta) = p_{\alpha\beta}(q_\beta(x)) = q_\alpha(x) = y_\alpha$ for any $\alpha \leq \beta$ in Γ, so $f(x) \in \Omega$. So:

$$f : (\mathcal{X}, \mathcal{B}, \lambda) \to (\Omega, \mathcal{F}, \mu); x \mapsto (q_\alpha(x))_{\alpha\in\Gamma}$$

is well-defined. Moreover:

$$p_\alpha f(x) = q_\alpha(x), \text{ for all } x \in \mathcal{X}, \alpha \in \Gamma,$$

so the combined diagram $(\Omega \star D(F), f, \mathcal{X} \star D(F))$ is commutative. Therefore f is a morphism of cones with base $D(F)$ in $\mathbb{P}$.

If $g : \mathcal{X} \to \Omega$ is another such morphism of cones, then for all $x \in \mathcal{X}$ and $\alpha \in \Gamma$: $p_\alpha g(x) = q_\alpha(x)$, so $g(x) = (q_\alpha(x))_{\alpha\in\Gamma} = f(x)$. Thus for every cone $\mathcal{X} \star D(F)$ with base $D(F)$ there exists a unique morphism $(\mathcal{X}, \mathcal{B}, \lambda) \to (\Omega, \mathcal{F}, \mu)$, proving that $\Omega \star D(F)$ is indeed the projective limit of $D(F)$.

$\square$

Let us see how this result implies Kolmogorov's Existence Theorem. So let T be an infinite index set, $\mathcal{J}$ be the collection of all finite subsets of T, and suppose $(\mathcal{X}_t, \mathcal{B}_t)$ is a Polish space with its Borel σ-algebra for every $t \in T$. Now let μ_F be a probability measure on every finite product $(\mathcal{X}_F, \mathcal{B}_F)$, where $F \in \mathcal{J}$, such that $(\mu_F)_{F\in\mathcal{J}}$ is a projective family, that is:

$$\pi_F^G(\mu_G) = \mu_F, \tag{A.15}$$

where $\pi_F^G : \mathcal{X}_G \to \mathcal{X}_F$ is the projection map for every $F \subset G$ in $\mathcal{J}$. Note that inclusion is a directed partial order on $\mathcal{J}$ and for every $F \in \mathcal{J}$ we have a Polish probability space $(\mathcal{X}_F, \mathcal{B}_F, \mu_F)$ and surjective maps $p_{FG} = \pi_F^G$ given by projection for all $F \subset G$ in $\mathcal{J}$, that satisfy:

$$\pi_F^G \pi_G^H = \pi_F^H, \text{ for all } F \subset G \subset H \text{ in } \mathcal{J}.$$

Finally, (A.15) translates to the Kolmogorov Consistency Condition in (A.6), so the collection $D_{\mathcal{J}}$ of objects $(\mathcal{X}_F, \mathcal{B}_F, \mu_F)$ for $F \in \mathcal{J}$, together with the maps π_F^G, forms a Kolmogorov diagram in the category $\mathbb{P}$.

If $\omega = (\omega_t)_{t \in T}$ is an element of the product space Ω':

$$\pi_F^G(\omega_G) = \pi_F^G((\omega_t)_{t \in G}) = (\omega_t)_{t \in F} = \omega_F,$$

so in this case the space Ω in Theorem A.2 is the whole product space Ω', and the Sequential Maximality Condition is trivially satisfied. Moreover, the algebra $\mathcal{A}$ in Theorem A.2 is the algebra of cylinder subsets of Ω', so $\sigma(\mathcal{A}) = \mathcal{F}'$ is the classical σ-algebra of subsets of Ω' mentioned in Section A.1. The measure μ obtained in Theorem A.2 satisfies:

$$\mu\left(\pi_F^{-1}(B)\right) = \mu_F(B), \text{ for all } F \in \mathcal{J}, B \in \mathcal{B}_F,$$

which translates exactly to (A.4). So μ is the projective limit of the family $(\mu_F)_{F \in \mathcal{J}}$ in the sense of Definition A.1, and the cone $\Omega' \star D_{\mathcal{J}}$ with vertex $(\Omega', \mathcal{F}', \mu)$ and base $D_{\mathcal{J}}$ is the projective limit of the diagram $D_{\mathcal{J}}$ in the sense of category theory.

Appendix B
Permission for Use

The main results presented in this thesis have previously appeared in the journal articles [16] and [17].

The first article, [16], appeared in the Elsevier Inc. *Journal of Functional Analysis*. According to the "Author Rights" webpage of Elsevier:

http://www.elsevier.com/journal-authors/author-rights-and-responsibilities,

authors may use "either their accepted author manuscript or final published article" for a range of purposes, which include "inclusion in a thesis or dissertation."

The second article, [17], appeared in the World Scientific journal *Infinite Dimensional Analysis, Quantum Probability and Related Topics*. According to the World Scientific "Author Rights" webpage:

http://www.worldscientific.com/page/authors/author-rights,

authors "may post the preprint anywhere at any time," provided that it is accompanied by the acknowledgement that the printed version appears in the particular World Scientific journal.

Vita

Irina Holmes was born in 1986 in Bucharest, Romania. After she earned her high school diploma from the Mihai Viteazul National College (Bucharest) in June 2004, she came to Louisiana State University in Baton Rouge, Louisiana, where she earned her Bachelor of Science degree in Mathematics, Summa cum Laude, in May 2008. Holmes continued her graduate studies at Louisiana State University and earned the Master of Science in Mathematics in May 2010. Her adviser is Professor Ambar Sengupta. Currently she is a candidate for the degree of Doctor of Philosophy in Mathematics, which will be awarded in August 2014. Later in August 2014, Holmes will begin a three year Jack Hale Postdoctoral Fellowship at Georgia Tech in Atlanta, Georgia.