Daniel W. Stroock

Some Properties of Gaussian Measures

May 15, 2022

Springer

Notation

General

α^+ & α^-	The positive and negative parts of $\alpha \in \mathbb{R}$ (§1.2).
$\lfloor t \rfloor$	The integer part of $t \in \mathbb{R}$.
$f \restriction S$	The restriction of the function f to the set S.
$\\| \cdot \\|_{\mathrm{u}}$	The uniform (supremum) norm.
i	The imaginary number $\sqrt{-1}$.
$o\big(g(t)\big)$	A function f for which $\frac{f(t)}{g(t)}$ tends to 0 as t tends to a limit.
$B_E(a, r)$	The ball of radius r around a in the metric space E.
$\mathring{\Gamma}$ & $\bar{\Gamma}$,	The interior, and closure of a subset in a topological space.
$\langle x, \xi \rangle$	The action of $\xi \in B^*$ on $x \in B$.

Sets, Functions, and Spaces

$\mathbb{C}$	The complex numbers.
$\mathbb{Z}$ & $\mathbb{Z}^+$	The integers and the positive integers.
$\mathbb{N}$	The non-negative integers: $\mathbb{N} = \{0\} \cup \mathbb{Z}^+$.
$\mathbb{Q}$	The set of rational numbers.
$A^{\complement}$	The complement of the set A.
$\mathbb{S}^{N-1}$	The unit sphere in $\mathbb{R}^N$.
$\mathbf{1}_A$	The indicator function of the set A.
$\mathrm{sgn}(x)$	The signum function: equal to 1 if $x \geq 0$ and -1 if $x < 0$.
$C_{\mathrm{b}}\big(E;\mathbb{R}\big)$ or $C_{\mathrm{b}}\big(E;\mathbb{C}\big)$	The space of bounded continuous functions from a topological space E into $\mathbb{R}$ or $\mathbb{C}$.
$\mathscr{S}(\mathbb{R}^N;\mathbb{R})$ or $\mathscr{S}(\mathbb{R}^N;\mathbb{C})$	The Schwartz test function space of smooth $\mathbb{R}$ or $\mathbb{C}$-valued functions with rapidly decreasing derivatives.
$C_{\mathrm{b}}(E;\mathbb{R})$ or $C_{\mathrm{b}}(E;\mathbb{C})$	The space of bounded, continuous, $\mathbb{R}$ or $\mathbb{C}$-valued functions on E.
$C^n(E;\mathbb{R}^N)$	The space of $f : G \longrightarrow \mathbb{R}^N$ with $n \in \mathbb{N} \cup \{\infty\}$ continuous derivatives.
$C^n_{\mathrm{c}}(E;\mathbb{R}^N)$	The space of $f \in C^n(G;\mathbb{R}^N)$ that vanish off of a compact set.
$\mathscr{S}(\mathbb{R}^N;\mathbb{R})$ or $\mathscr{S}(\mathbb{R}^N;\mathbb{C})$	The Schwartz space of smooth $\mathbb{R}$ or$\mathbb{C}$-valued functions with rapidly decreasing derivatives.

<u>*Measure Theoretic*</u>

$(\text{a.e.}, \mu)$	To be read *almost everywhere with respect to* μ.
$\sigma(\mathcal{C})$	The σ-algebra generated by $\mathcal{C}$ (§ 2.1.2).
$\mathcal{B}_E$	The Borel σ-algebra $\sigma\big(\mathfrak{G}(E)\big)$ over E.
$\overline{\mathcal{B}}^{\mu}$	The completion of the σ-algebra $\mathcal{B}$ with respect to the measure μ.
δ_a	The unit point mass at a.
λ_S	Lebesgue measure on the set $S \in \mathcal{B}_{\mathbb{R}^N}$.
$M_1(E)$	Space of probability Borel measures on E.
$\mu \ll \nu$	μ is absolutely continuous with respect to ν.
$\mu \perp \nu$	μ is singular to ν.
$\Phi_*\mu$	The pushforward (image) of μ under Φ.
$\int_\Gamma f(x)\,dx$	Equivalent to the Lebesgue integral $\int_\Gamma f\,d\lambda_{\mathbb{R}^N}$ of f on Γ.
$L^p(\mu; E)$ or $L^p(\mu; E)$	The Lebesgue space of E-valued or $\mathbb{C}$-valued functions f for which $\|f\|_E^p$ is μ-integrable.
$\langle \varphi, \mu \rangle$	The integral of φ with respect to μ.
$(\text{R})\int_{[a,b]} \varphi(t)\,d\psi(t)$	The Riemann–Stieltjes integral of φ on $[a,b]$ with respect to ψ.
$\mathbb{E}$ & $\mathbb{E}^\mu$	The expected value and expected value with respect to μ.

Contents

Chapter 1
Characteristic Functions

This chapter is devoted to a few facts, several of which I learned from an interesting set of notes by W. Bryc [1], about and related to what probabilists call *characteristic* functions.

1.1 Some Basic Facts

Given a Borel probability measure μ on $\mathbb{R}^N$, its characteristic function is its Fourier transform $\hat{\mu}$ defined by

$$\hat{\mu}(\xi) = \int e^{i(\xi,x)_{\mathbb{R}^N}}\,\mu(dx) \quad \text{for } \xi \in \mathbb{R}^N.$$

Lemma 1.1.1 *For every $\mu \in M_1(\mathbb{R}^N)$, $\hat{\mu}$ is a a uniformly continuous $\mathbb{C}$-valued function such that $\|\mu\|_{\mathrm{u}} \le 1 = \hat{\mu}(0)$ and $\hat{\mu}(-\xi) = \overline{\hat{\mu}(\xi)}$. Moreover, if $\varphi \in C_{\mathrm{b}}(\mathbb{R}^N;\mathbb{C}) \cap L^1(\lambda_{\mathbb{R}^N};\mathbb{C})$ for which $\hat{\varphi} \in L^1(\lambda_{\mathbb{R}^N};\mathbb{C})$, then*

$$\int \varphi\,d\mu = (2\pi)^{-N} \int \hat{\varphi}(\xi)\hat{\mu}(-\xi)\,d\xi. \tag{1.1.1}$$

In particular, $\mu = \nu$ if and only if $\hat{\mu} = \hat{\nu}$.

Proof. That $\|\hat{\mu}\|_{\mathrm{u}} \le 1 = \hat{\mu}(0)$ is obvious, as is the continuity of $\hat{\mu}$. To see that $\hat{\mu}$ is uniformly continuous, note that

$$|\hat{\mu}(\eta) - \hat{\nu}(\xi)|^2 \le \int \left|1 - e^{i(\eta-\xi,x)_{\mathbb{R}^N}}\right|^2 \mu(dx) = 2\int \big(1 - \cos(\eta-\xi,x)_{\mathbb{R}^N}\big)\,\mu(dx),$$

and therefore that

$$|\hat{\mu}(\eta) - \hat{\mu}(\xi)|^2 \le 2\big(1 - \Re\hat{\mu}(\eta-\xi)\big). \tag{1.1.2}$$

Turning to (1.1.1), note that, because φ and $\hat{\varphi}$ are $\lambda_{\mathbb{R}^N}$-integrable, the L^1-Fourier inversion formula says that

$$\varphi(x) = (2\pi)^{-N} \int e^{-i(\xi,x)_{\mathbb{R}^N}} \hat{\varphi}(\xi)\, d\xi.$$

Hence, by Fubini's theorem

$$\begin{aligned}(2\pi)^N \int \varphi\, d\mu &= \int \left(\int e^{-i(\xi,x)_{\mathbb{R}^N}} \hat{\varphi}(\xi)\, d\xi \right) \mu(dx) \\ &= \int \hat{\varphi}(\xi) \left(\int e^{-i(\xi,x)_{\mathbb{R}^N}}\, \mu(dx) \right) d\xi = \int \hat{\varphi}(\xi) \hat{\mu}(-\xi)\, d\xi.\end{aligned}$$

Finally, to see that $\mu = \nu$ if $\hat{\mu} = \hat{\nu}$, remember that the Fourier transform takes the Schwartz space $\mathscr{S}(\mathbb{R}^N;\mathbb{C})$ (cf. § 3.6.1) into itself. Thus, if $\varphi \in \mathscr{S}(\mathbb{R}^N;\mathbb{C})$, then both φ and $\hat{\varphi}$ are in $L^1(\lambda_{\mathbb{R}^N};\mathbb{C})$, and so, by (1.1.1),

$$\hat{\mu} = \hat{\nu} \implies \int \varphi\, d\mu = \int \varphi\, d\nu \text{ for all } \varphi \in \mathscr{S}(\mathbb{R}^N;\mathbb{C}),$$

which means that $\mu = \nu$. □

Lemma 1.1.2 *If μ is the distribution of a random variable X, and, for some $n \ge 1$, $\mathbb{E}[|X|^n] < \infty$, then $\hat{\mu} \in C^n(\mathbb{R};\mathbb{C})$ and*

$$\mathbb{E}\big[X^m e^{i\xi X}\big] = (-i)^m \partial^m \hat{\mu}(\xi) \text{ for } 0 \le m \le n.$$

Proof. This is trivial when $m = 0$. If it holds for some $0 \le m < n$, then

$$i^m \frac{\partial^m \hat{\mu}(\xi+\eta) - \partial^m \hat{\mu}(\xi)}{\eta} = \mathbb{E}\left[X^m e^{i\xi X} \frac{e^{i\eta X} - 1}{\eta} \right].$$

Since $\left| X^m e^{i\xi X} \frac{e^{i\eta X}-1}{\eta} \right| \le |X|^{m+1}$, the result follows from Lebesgue's Dominated Convergence theorem. □

Define $\Delta_t \varphi(\xi) = \frac{\varphi(\xi+\frac{t}{2}) - \varphi(\xi-\frac{t}{2})}{t}$ for functions $\varphi : \mathbb{R} \longrightarrow \mathbb{C}$. If φ vanishes at 0 faster than $|\xi|^n$ for some $n \ge 1$, then $\lim_{t\searrow 0} \Delta_t^n \varphi(0) = 0$. Thus, for f's which are n times differentiable in a neighborhood of 0, $\lim_{t\searrow 0} \Delta_t^n f(0) = \lim_{t\searrow 0} \Delta_t^n P_n(0)$, where P_n is the Taylor polynomial $\sum_{m=0}^n \frac{f^{(m)}(0)}{m!} \xi^m$. Therefore we will know that $\lim_{t\searrow 0} \Delta_t^n f(0) = f^{(n)}(0)$ once we show that $\Delta_t^n \xi^m = n!\delta_{m,n}$ for $0 \le m \le n$, which comes down to checking that $\Delta_t^m \xi^m = m!$ for all $m \ge 1$. That $\Delta_t \xi = 1$ is obvious. Now assume the result for $1 \le m < M$. Then

$$
\begin{aligned}
t\Delta_t^M \xi^M &= \Delta_t^{M-1}\Big(\big(\xi+\tfrac{t}{2}\big)^M - \big(\xi-\tfrac{t}{2}\big)^M\Big) \\
&= \sum_{m=0}^{M}\binom{M}{m}\big(\Delta_t^{M-1}\xi^{M-m}\big)\big(\tfrac{t}{2}\big)^m\big(1+(-1)^{m+1}\big) = Mt\Delta_t^{M-1}\xi^{M-1} = tM!.
\end{aligned}
$$

Lemma 1.1.3 *If $n \geq 1$ and there exists a sequence $t_k \searrow 0$ such that $\sup_{k\geq 1}|\Delta_{t_k}^{2n}\hat{\mu}(0)| < \infty$, then $\mathbb{E}[X^{2n}] < \infty$. In particular, this will be the case if $\hat{\mu}$ is $2n$ times continuously differentialble in a neighborhood of 0.*

Proof. Set $\mathfrak{e}_x(\xi) = e^{ix\xi}$, and observe that

$$\Delta_t^2\mathfrak{e}_x(\xi) = 2\mathfrak{e}_x(x)\frac{\cos xt - 1}{t^2} = -4t^{-2}\mathfrak{e}_x(\xi)\sin^2\frac{xt}{2},$$

and therefore

$$\Delta_t^{2n}\mathfrak{e}_x(\xi) = (-4)^n t^{-2n}\mathfrak{e}_x(\xi)\sin^{2n}\frac{xt}{2}.$$

Thus

$$
\begin{aligned}
\big|\Delta_{t_k}^{2n}\hat{\mu}(0)\big| &= 4^n t_k^{-2n}\mathbb{E}\left[\sin^{2n}\frac{Xt_k}{2}\right] \\
&\geq 4^n t_k^{-2n}\mathbb{E}\left[\sin^{2n}\frac{Xt_k}{2},\, |X| \leq \frac{\pi}{2t_k}\right] \geq 2^{-n}\mathbb{E}\left[X^{2n},\, |X| \leq \frac{\pi}{2t_k}\right].
\end{aligned}
$$

□

Theorem 1.1.4 *If $\hat{\mu}$ admits an extension as an analytic function on $B_{\mathbb{C}}(0,r)$, then $\mathbb{E}\big[e^{\alpha|X|}\big] < \infty$ for $\alpha \in [0,r)$, and so*

$$\zeta \in \{z \in \mathbb{C} :\, \Re\zeta \in (-r,r)\} \longmapsto M(\zeta) = \mathbb{E}\big[e^{\zeta X}\big] \in \mathbb{C}$$

is analytic.

Proof. Since $\hat{\mu} \in C^\infty(\mathbb{R};\mathbb{C})$, Lemma 1.1.3 implies that $\mathbb{E}[|X|^n] < \infty$ and, by Lemma 1.1.2, $\partial^n\hat{\mu}(0) = i^n\mathbb{E}[X^n]$ for all $n \geq 1$. Hence,

$$\hat{\mu}(\xi) = \sum_{n=0}^{\infty}\frac{\mathbb{E}[X^n]}{n!}(i\xi)^n$$

where the radius of convergence of the series is at least r. In particular, for $\alpha \in [0,r)$,

$$\mathbb{E}\big[e^{\alpha X} + e^{-\alpha X}\big] = 2\sum_{n=0}^{\infty}\frac{\mathbb{E}[X^{2n}]}{(2n)!}\alpha^{2n} < \infty,$$

and so $\mathbb{E}\big[e^{\alpha|X|}\big] < \infty$. □

Knowing that the map $\mu \in M_1(\mathbb{R}^N) \longmapsto \hat{\mu} \in C_{\mathrm{b}}(\mathbb{R}^N;\mathbb{C})$ is one-to-one, one should ask about its continuity properties. For that purpose, introduce the notation $\langle \varphi, \mu \rangle$ to denote the integral of a function φ with respect to a (not necessarily finite) measure μ, and say that a sequence $\{\mu_n : n \geq 1\} \subseteq M_1(\mathbb{R}^N)$ *converges weakly* to $\mu \in M_1(\mathbb{R}^N)$ if, for every $\varphi \in C_{\mathrm{b}}(\mathbb{R}^N;\mathbb{C})$, $\langle \varphi, \mu_n \rangle \longrightarrow \langle \varphi, \mu \rangle$. I will write $\mu_n \xrightarrow{\mathrm{w}} \mu$ to mean that $\{\mu_n : n \geq 1\}$ converges weakly to μ.

It is clear that $\hat{\mu}_n \longrightarrow \hat{\mu}$ pointwise if $\mu_n \xrightarrow{\mathrm{w}} \mu$. In fact, as we will see, the convergence of the $\hat{\mu}_n$'s is uniform on compact. More important, we will show that uniform convergence on compacts of the $\hat{\mu}_n$'s implies weak convergence of the μ_n's.

Lemma 1.1.5 *Let $\{\mu_n : n \geq 1\} \cup \{\mu\} \subseteq M_1(\mathbb{R}^N)$ be given, and assume that $\langle \varphi, \mu_n \rangle \longrightarrow \langle \varphi, \mu \rangle$ for all $\varphi \in C_{\mathrm{c}}^\infty(\mathbb{R}^N;\mathbb{R})$. Then, for any $\psi \in C\big(\mathbb{R}^N;[0,\infty)\big)$, $\langle \psi, \mu \rangle \leq \underline{\lim}_{n\to\infty} \langle \psi, \mu_n \rangle$. Moreover, if $\psi \in C\big(\mathbb{R}^N;[0,\infty)\big)$ is μ_n-integrable for each $n \in \mathbb{Z}$ and if $\langle \psi, \mu_n \rangle \longrightarrow \langle \psi, \mu \rangle \in [0,\infty)$, then for any sequence $\{\varphi_n : n \geq 1\} \subseteq C(\mathbb{R}^N;\mathbb{C})$ that converges uniformly on compacts to a $\varphi \in C(\mathbb{R}^N;\mathbb{C})$ and satisfies $|\varphi_n| \leq C\psi$ for some $C < \infty$ and all $n \geq 1$, $\langle \varphi_n, \mu_n \rangle \longrightarrow \langle \varphi, \mu \rangle$.*

Proof. Choose $\rho \in C_{\mathrm{c}}^\infty\big(\mathbb{R}^N;[0,\infty)\big)$ with total $\lambda_{\mathbb{R}}$-integral 1, and set $\rho_\epsilon(x) = \epsilon^{-N}\rho(\epsilon^{-1}x)$ for $\epsilon > 0$. Also, choose $\eta \in C_{\mathrm{c}}^\infty\big(\mathbb{R}^N;[0,1]\big)$ so that $\eta = 1$ on $\overline{B_{\mathbb{R}^N}(0,1)}$ and 0 off of $B_{\mathbb{R}^N}(0,2)$, and set $\eta_R(x) = \eta(R^{-1}x)$ for $R > 0$.

Begin by noting that $\langle \varphi, \mu_n \rangle \longrightarrow \langle \varphi, \mu \rangle$ for all $\varphi \in C_{\mathrm{c}}^\infty(\mathbb{R}^N;\mathbb{C})$. Next, suppose that $\varphi \in C_{\mathrm{c}}(\mathbb{R}^N;\mathbb{C})$, and, for $\epsilon > 0$, set $\varphi_\epsilon = \rho_\epsilon \star \varphi$, the convolution

$$\int_{\mathbb{R}^N} \rho_\epsilon(x-y)\varphi(y)\,dy$$

of ρ_ϵ with φ. Then, for each $\epsilon > 0$, $\varphi_\epsilon \in C_{\mathrm{c}}^\infty(\mathbb{R}^N;\mathbb{C})$ and therefore $\langle \varphi_\epsilon, \mu_n \rangle \longrightarrow \langle \varphi_\epsilon, \mu \rangle$. In addition, there is an $R > 0$ such that $\operatorname{supp}(\varphi_\epsilon) \subseteq B_{\mathbb{R}^N}(0,R)$ for all $\epsilon \in (0,1]$. Hence,

$$\overline{\lim_{n\to\infty}} \big|\langle \varphi, \mu_n \rangle - \langle \varphi, \mu \rangle\big| \leq 2\langle \eta_R, \mu \rangle \|\varphi_\epsilon - \varphi\|_{\mathrm{u}}.$$

Since $\lim_{\epsilon \searrow 0} \|\varphi_\epsilon - \varphi\|_{\mathrm{u}} = 0$, we have now shown that $\langle \varphi, \mu_n \rangle \longrightarrow \langle \varphi, \mu \rangle$ for all $\varphi \in C_{\mathrm{c}}(\mathbb{R}^N;\mathbb{C})$.

Now suppose that $\psi \in C\big(\mathbb{R}^N;[0,\infty)\big)$, and set $\psi_R = \eta_R \psi$, where η_R is as above. Then, for each $R > 0$, $\langle \psi_R, \mu \rangle = \lim_{n\to\infty} \langle \psi_R, \mu_n \rangle \leq \underline{\lim}_{n\to\infty} \langle \psi, \mu_n \rangle$. Hence, by Fatou's Lemma, $\langle \psi, \mu \rangle \leq \underline{\lim}_{R\to\infty} \langle \psi_R, \mu \rangle \leq \underline{\lim}_{n\to\infty} \langle \psi, \mu_n \rangle$.

Finally, suppose that $\psi \in C\big(\mathbb{R}^N;[0,\infty)\big)$ is μ_n-integrable for each $n \in \mathbb{Z}$ and that $\langle \psi, \mu_n \rangle \longrightarrow \langle \psi, \mu \rangle \in [0,\infty)$. Given $\{\varphi_n : n \geq 1\} \subseteq C(\mathbb{R}^N;\mathbb{C})$ satisfying $|\varphi_n| \leq C\psi$ and converging uniformly on compacts to φ, one has

$$\big|\langle \varphi_n, \mu_n \rangle - \langle \varphi, \mu \rangle\big| \leq \big|\langle \varphi_n - \varphi, \mu_n \rangle\big| + \big|\langle \varphi, \mu_n \rangle - \langle \varphi, \mu \rangle\big|.$$

Moreover, for each $R > 0$,

$$
\begin{aligned}
&\varlimsup_{n\to\infty}\big|\langle\varphi_n-\varphi,\mu_n\rangle\big|\\
&\le \varlimsup_{n\to\infty}\sup_{x\in B_{\mathbb{R}^N}(0,2R)}|\varphi_n(x)-\varphi(x)|\langle\eta_R,\mu_n\rangle+\varlimsup_{n\to\infty}\big|\langle(1-\eta_R)(\varphi_n-\varphi),\mu_n\rangle\big|\\
&\le 2C\varlimsup_{n\to\infty}\langle(1-\eta_R)\psi,\mu_n\rangle=\varlimsup_{n\to\infty}2C\big(\langle\psi,\mu_n\rangle-\langle\eta_R\psi,\mu_n\rangle\big)\\
&=2C\langle(1-\eta_R)\psi,\mu\rangle,
\end{aligned}
$$

and similarly

$$
\begin{aligned}
&\varlimsup_{n\to\infty}\big|\langle\varphi,\mu_n\rangle-\langle\varphi,\mu\rangle\big|\\
&\le \varlimsup_{n\to\infty}\big|\langle\eta_R\varphi,\mu_n\rangle-\langle\eta_R\varphi,\mu\rangle\big|+C\varlimsup_{n\to\infty}\langle(1-\eta_R)\psi,\mu_n\rangle+C\langle(1-\eta_R)\psi,\mu\rangle\\
&=2C\langle(1-\eta_R)\psi,\mu\rangle.
\end{aligned}
$$

Finally, because ψ is μ-integrable, $\langle(1-\eta_R)\psi,\mu\rangle\longrightarrow 0$ as $R\to\infty$ by Lebesgue's Dominated Convergence Theorem, and so we are done. □

In the proof of the following theorem and elsewhere, a function $\varphi:\mathbb{R}^N\longrightarrow\mathbb{C}$ is said to be *rapidly decreasing* if $\sup_{x\in\mathbb{R}^N}(1+x^{2n})|\varphi(x)|<\infty$ for all $n\ge 0$.

Theorem 1.1.6 *Given a sequence $\{\mu_n : n\ge 1\}\subseteq M_1(\mathbb{R}^N)$ and a $\mu\in M_1(\mathbb{R}^N)$, $\widehat{\mu_n}\longrightarrow\hat\mu$ uniformly on compacts if $\mu_n\xrightarrow{\mathrm{w}}\mu$. Conversely, if $\widehat{\mu_n}(\xi)\longrightarrow\hat\mu(\xi)$ pointwise, then $\langle\varphi_n,\mu_n\rangle\longrightarrow\langle\varphi,\mu\rangle$ whenever $\{\varphi_n : n\ge 1\}$ is a uniformly bounded sequence in $C_{\mathrm{b}}(\mathbb{R}^N;\mathbb{C})$ that tends to φ uniformly on compacts. In particular, $\hat\mu_n\longrightarrow\hat\mu$ uniformly on compacts.*

Proof. Since $e^{i(\xi_n,x)_{\mathbb{R}^N}}\longrightarrow e^{i(\xi,x)_{\mathbb{R}^N}}$ uniformly for x in compacts if $\xi_n\to\xi$, Lemma 1.1.5 says that, if $\mu_n\xrightarrow{\mathrm{w}}\mu$, then $\hat\mu_n(\xi_n)\longrightarrow\hat\mu(\xi)$ if $\xi_n\to\xi$, and therefore that $\mu_n\xrightarrow{\mathrm{w}}\mu$ implies that $\hat\mu_n\to\mu$ uniformly on compacts.

Turning to the second part of the theorem, suppose that $\widehat{\mu_n}\longrightarrow\hat\mu$ pointwise. By Lemma 1.1.5, we need only check that $\langle\varphi,\mu_n\rangle\longrightarrow\langle\varphi,\mu\rangle$ when $\varphi\in C_{\mathrm{c}}^\infty\big(\mathbb{R}^N;\mathbb{R}\big)$. But, for such a φ, $\hat\varphi$ is smooth and rapidly decreasing, and therefore the result follows immediately from (1.1.1) together with Lebesgue's Dominated Convergence Theorem. □

The following is a generalization to $\mathbb{R}^N$ of the classical Helly-Bray Theorem. It was further generalized to complete separable metric spaces by Y. Prohorov, and, with only minor changes (cf. Theorem 9.1.9 in [10]), the treatment here can be used to prove Prohorov's result.

Lemma 1.1.7 *A subset S of $M_1(\mathbb{R}^N)$ is sequentially relatively compact in the weak topology if and only if*

$$\lim_{R\to\infty}\sup_{\mu\in S}\mu\big(B_{\mathbb{R}^N}(0,R)\complement\big)=0. \tag{1.1.3}$$

Proof. The first step is to recognize that there is a countable set $\{\varphi_k : k\in\mathbb{Z}^+\}\subseteq C_{\mathrm{c}}(\mathbb{R}^N;\mathbb{R})$ of linear independent functions whose span is dense, with respect to uniform convergence, in $C_c(\mathbb{R}^N;\mathbb{R})$. To see this, choose $\eta \in C_c\big(\mathbb{R}^N;[0,1]\big)$ so that $\eta = 1$ on $\overline{B_{\mathbb{R}^N}(0,1)}$ and 0 off $B_{\mathbb{R}^N}(0,2)$, and set $\eta_R(y) = \eta(R^{-1}y)$ for $R > 0$. Next, for each $\ell \ge 1$, apply the Stone–Weierstrass Theorem to choose a countable dense subset $\{\psi_{j,\ell} : j \in \mathbb{Z}\}$ of $C\big(\overline{B_{\mathbb{R}^N}(0,2\ell)};\mathbb{R}\big)$, and set $\varphi_{j,\ell} = \eta_\ell \psi_{j,\ell}$. Clearly $\{\varphi_{j,\ell} : (j,\ell) \in (\mathbb{Z}^+)^2\}$ is dense in $C_c(\mathbb{R}^N;\mathbb{R})$. Finally, using lexicographic ordering of $(\mathbb{Z}^+)^2$, extract a linearly independent subset $\{\varphi_k : k \ge 1\}$ by taking $\varphi_k = \varphi_{j_k,\ell_k}$, where $(j_1,\ell_1) = (1,1)$ and (j_{k+1},ℓ_{k+1}) is the first (j,ℓ) such that $\varphi_{j,\ell}$ is linearly independent of $\{\varphi_1,\dots,\varphi_k\}$.

Given a sequence $\{\mu_n : n \ge 1\} \subseteq S$, we can use a diagonalization procedure to find a subsequence $\{\mu_{n_m} : m \ge 1\}$ such that $a_k = \lim_{m\to\infty}\langle \varphi_k, \mu_{n_m}\rangle$ exists for every $k \in \mathbb{Z}^+$. Next, define the linear functional Λ on the span of $\{\varphi_k : k \ge 1\}$ so that $\Lambda(\varphi_k) = a_k$. Notice that if $\varphi = \sum_{k=1}^K \alpha_k \varphi_k$, then

$$\big|\Lambda(\varphi)\big| = \lim_{m\to\infty}\left|\sum_{k=1}^K \alpha_k \langle \varphi_k, \mu_{n_m}\rangle\right| = \lim_{m\to\infty}\big|\langle \varphi, \mu_{n_m}\rangle\big| \le \|\varphi\|_{\mathrm{u}},$$

and similarly that $\Lambda(\varphi) = \lim_{m\to\infty}\langle\varphi,\mu_{n_m}\rangle \ge 0$ if $\varphi \ge 0$. Therefore Λ admits a unique extension as a non-negativity preserving linear functional on $C_c(\mathbb{R}^N;\mathbb{R})$ that satisfies $|\Lambda(\varphi)| \le \|\varphi\|_{\mathrm{u}}$ for all $\varphi \in C_c(\mathbb{R}^N;\mathbb{R})$.

Now assume that (1.1.3) holds. For each $\ell \in \mathbb{Z}$, apply the Riesz Representation Theorem to produce a non-negative Borel measure ν_ℓ supported on $\overline{B_{\mathbb{R}^N}(0,2\ell)}$ so that $\langle \varphi, \nu_\ell\rangle = \Lambda(\eta_\ell \varphi)$ for $\varphi \in C_c(\mathbb{R}^N;\mathbb{R})$. Since $\langle \varphi, \nu_{\ell+1}\rangle = \Lambda(\varphi) = \langle \varphi, \nu_\ell\rangle$ whenever φ vanishes off of $B_{\mathbb{R}^N}(0,\ell)$, it is clear that

$$\nu_{\ell+1}\big(\Gamma \cap B_{\mathbb{R}^N}(0,\ell+1)\big) \ge \nu_{\ell+1}\big(\Gamma \cap B_{\mathbb{R}^N}(0,\ell)\big) = \nu_\ell\big(\Gamma \cap B_{\mathbb{R}^N}(0,\ell)\big)$$

for all $\Gamma \in \mathcal{B}_{\mathbb{R}^N}$. Hence the limit

$$\mu(\Gamma) \equiv \lim_{\ell\to\infty} \nu_\ell\big(\Gamma \cap B_{\mathbb{R}^N}(0,\ell)\big) = \sum_{\ell=1}^\infty \nu_\ell\Big(\Gamma \cap \big(B_{\mathbb{R}^N}(0,\ell) \setminus B_{\mathbb{R}^N}(0,\ell-1)\big)\Big),$$

exists and determines a non-negative Borel measure μ on $\mathbb{R}^N$ whose restriction to $B_{\mathbb{R}^N}(0,\ell)$ is ν_ℓ for each $\ell \in \mathbb{Z}^+$. In particular, $\mu(\mathbb{R}^N) \le 1$ and $\langle \varphi, \mu\rangle = \lim_{m\to\infty}\langle \varphi, \mu_{n_m}\rangle$ for every $\varphi \in C_c(\mathbb{R}^N;\mathbb{R})$. Thus, by Lemma 1.1.5, all that remains is to check that $\mu(\mathbb{R}^N) = 1$. But

$$\begin{aligned}\mu(\mathbb{R}^N) \ge \langle \eta_\ell, \mu\rangle = \lim_{m\to\infty}\langle \eta_\ell, \mu_{n_m}\rangle &\ge \varlimsup_{m\to\infty} \mu_{n_m}\big(\overline{B_{\mathbb{R}^N}(0,\ell)}\big)\\ &= 1 - \varliminf_{m\to\infty} \mu_{n_m}\big(B_{\mathbb{R}^N}(0,\ell)\complement\big),\end{aligned}$$

and, by (1.1.3), the final term tends to 0 as $\ell \to \infty$.

To prove the converse assertion, suppose that S is sequentially relatively compact. If (1.1.3) failed, then we could find an $\theta \in (0,1)$ and, for each $n \in \mathbb{Z}$, a $\mu_n \in S$ such that $\mu_n\big(B_{\mathbb{R}^N}(0,n)\big) \le \theta$. By sequential relative compactness, this would mean that there is a subsequence $\{\mu_{n_m} : m \geq 1\} \subseteq S$ and a $\mu \in M_1(\mathbb{R}^N)$ such that $\mu_{n_m} \xrightarrow{\mathrm{w}} \mu$ and $\mu_{n_m}\big(B_{\mathbb{R}^N}(0, n_m)\big) \leq \theta$. On the other hand, for any $R > 0$,

$$\mu\big(B_{\mathbb{R}^N}(0,R)\big) = \lim_{m\to\infty} \langle \eta_R, \mu_{n_m}\rangle \leq \varlimsup_{m\to\infty} \mu_{n_m}\big(B_{\mathbb{R}^N}(0,n_m)\big) \leq \theta,$$

and so we would arrive at the contradiction $1 = \lim_{R\to\infty} \mu\big(B_{\mathbb{R}^N}(0,R)\big) \leq \theta$. □

Lemma 1.1.8 *Let $\mu \in M_1(\mathbb{R}^N)$. Then, for all $(r,R) \in [0,\infty)^2$ and $e \in \mathbb{S}^{N-1}$,*

$$\big|1 - \hat{\mu}(re)\big| \leq rR + \mu\big(\{x \in \mathbb{R}^N : |(x,e)_{\mathbb{R}^N}| \geq R\}\big). \tag{1.1.4}$$

Next, define

$$s(r) = \sup_{\theta \geq r}\left(1 - \frac{\sin\theta}{\theta}\right) \quad \textit{for } r > 0.$$

Then $0 < s(r) \leq Cr$ for some $C < \infty$, and

$$\mu\big(\{x \in \mathbb{R}^N : |x| \geq R\}\big) \leq \max_{|\xi|\leq r} \frac{N|1-\hat{\mu}(\xi)|}{s(N^{-\frac{1}{2}}rR)} \quad \textit{for all } (r,R) \in (0,\infty)^2. \tag{1.1.5}$$

Proof. The facts about $s(r)$ are easily checked using elementary calculus.

To prove (1.1.4), simply observe that, since $|1 - e^{i\xi}| \leq |\xi| \wedge 1$,

$$\big|1 - \hat{\mu}(re)\big| \leq \int \big|1 - e^{ir(x,e)_{\mathbb{R}^N}}\big|\,\mu(dx) \leq rR + \mu\big(\{x : |(x,e)| \geq R\}\big).$$

Turning to (1.1.5), note that, for $t \geq 0$,

$$\big|1 - \hat{\mu}(te)\big| \geq \int \big(1 - \cos(tx,e)_{\mathbb{R}^N}\big)\,d\mu.$$

Now integrate this inequality with respect to $t \in [0,r]$, and divide by r to see that

$$\begin{aligned}\max_{|\xi|\leq r}|1-\hat{\mu}(\xi)| &\geq \frac{1}{r}\int_0^r \big|1-\hat{\mu}(te)\big|\,dt \geq \int \left(1 - \frac{\sin(re,x)_{\mathbb{R}^N}}{(re,x)_{\mathbb{R}^N}}\right)\mu(dx) \\ &\geq \int\limits_{|(e,x)_{\mathbb{R}^N}|\geq R} \left(1 - \frac{\sin(re,x)_{\mathbb{R}^N}}{(re,x)_{\mathbb{R}^N}}\right)\mu(dx) \geq s(rR)\mu\big(\{x : |(e,x)_{\mathbb{R}^N}| \geq R\}\big).\end{aligned}$$

Finally, use the fact that

$$\mu\big(\{x : |x| \geq R\}\big) \leq N \sup_{e\in\mathbb{S}^{N-1}} \mu\big(\{x : |(e,x)_{\mathbb{R}^N}| \geq N^{-\frac{1}{2}}R\}\big)$$

to arrive at (1.1.5) □

The following theorem answers the question posed above about the relationship between characteristic functions and weak convergence. It was proved by P. Lévy and is called *Lévy's Continuity Theorem.*

Theorem 1.1.9 *Let* $\{\mu_n : n \geq 1\} \subseteq M_1(\mathbb{R}^N)$, *and assume that* $f(\xi) = \lim_{n\to\infty} \hat{\mu}_n(\xi)$ *exists for each* $\xi \in \mathbb{R}^N$. *Then* f *is the characteristic function of a* $\mu \in M_1(\mathbb{R}^N)$ *if and only if there is a* $\delta > 0$ *for which*

$$\lim_{n\to\infty} \sup_{|\xi|\leq\delta} \left|\hat{\mu}_n(\xi) - f(\xi)\right| = 0,$$

in which case $\mu_n \xrightarrow{\mathrm{w}} \mu$.

Proof. The only assertion not already covered by Lemmas 1.1.5 and 1.1.1 is the "if" part of the equivalence. But, if $\hat{\mu}_n \longrightarrow f$ uniformly in a neighborhood of 0, then it is easy to check that $\sup_{n\geq 1} |1 - \hat{\mu}_n(\xi)|$ must tend to zero as $|\xi| \to 0$. Hence, by (1.1.5) and Lemma 1.1.7, we know that there exists a μ and a subsequence $\{\mu_{n_m} : m \geq 1\}$ such that $\mu_{n_m} \xrightarrow{\mathrm{w}} \mu$. Since $\hat{\mu}$ must equal f, Lemma 1.1.1 says that $\mu_n \xrightarrow{\mathrm{w}} \mu$. □

Bochner found an interesting characterization of characteristic functions. To describe his result, say that a function $f : \mathbb{R}^N \longrightarrow \mathbb{C}$ is *non-negative definite* if the matrix

$$\big(\big(f(\xi_j - \xi_k)\big)\big)_{1\leq j,k\leq n}$$

is non-negative definite for all $n \geq 2$ and $\xi_1, \dots, \xi_n \in \mathbb{R}^N$, which is equivalent to saying

$$\sum_{j,k=1}^{n} f(\xi_j - \xi_k)\alpha_j\overline{\alpha_k} \geq 0$$

for all $\alpha_1, \dots, \alpha_n \in \mathbb{C}$.

Theorem 1.1.10 *A function* $f : \mathbb{R}^N \longrightarrow \mathbb{C}$ *is a characteristic function if and only if* f *is continuous,* $f(0) = 1$, *and* f *is non-negative definite.*

Proof. Assume that $f = \hat{\mu}$ for some $\mu \in M_1(\mathbb{R}^N)$. Then it is obvious that f is continuous and that $f(0) = 1$. To see that it is non-negative definite, observe that

$$\begin{aligned}\sum_{j,k=1}^{n} f(\xi_j - \xi_k)\alpha_j\overline{\alpha_k} &= \int \left(\sum_{j,k=1}^{n} e^{i(\xi_j-\xi_k)x}\alpha_j\overline{\alpha_k}\right) \mu(dx) \\ &= \int \left|\sum_{j,k=1}^{n} e^{i\xi_j x}\alpha_j\right|^2 \mu(dx) \geq 0.\end{aligned}$$

Now assume that f is a continuous, non-negative definite function with $f(0) = 1$. Because

$$A \equiv \begin{pmatrix} 1 & f(\xi) \\ f(-\xi) & 1 \end{pmatrix}$$

is non-negative definite, $\Im\big(f(\xi) + f(-\xi)\big)$ and $\Im\big(if(\xi) - if(-\xi)\big)$ are both 0, and therefore $f(\xi) = \overline{f(-\xi)}$. Thus A is Hermitian, and because it is non-negative definite, $1 - |f(\xi)|^2 \geq 0$. Therefore $|f(\xi)| \leq 1$. Next, let $\psi \in \mathscr{S}(\mathbb{R}^N;\mathbb{R})$, and use Riemann approximations to see that

$$\iint f(\xi-\eta)\hat{\psi}(\xi)\overline{\hat{\psi}(\eta)}\,d\xi d\eta \geq 0.$$

Assume for the moment that $f \in L^1(\lambda_{\mathbb{R}^N};\mathbb{C})$, and set

$$g(x) = (2\pi)^{-N}\int e^{-i(\xi,x)}f(\xi)\,d\xi.$$

By Fubini's Theorem and the fact that $\overline{\hat{\psi}(\xi)} = \hat{\psi}(-\xi)$,

$$\begin{aligned}(2\pi)^N \int g(x)\psi(x)^2\,dx &= \int f(\xi)\widehat{\psi^2}(-\xi)\,d\xi = \int f(\xi)\big(\hat{\psi} * \hat{\psi}\big)(-\xi)\,d\xi \\ &= \iint f(\xi)\overline{\hat{\psi}(\xi+\eta)}\hat{\psi}(\eta)\,d\xi d\eta = \iint f(\xi-\eta)\overline{\hat{\psi}(\xi)}\hat{\psi}(\eta)\,d\xi d\eta \geq 0.\end{aligned}$$

Hence, since g is continuous, it follows that $g \geq 0$. In addition, $f = \hat{g}$ and so $\int g(x)\,dx = f(0) = 1$ and f is the Fourier transform of the probability measure $d\mu = g\,d\lambda_{\mathbb{R}^N}$.

To remove the assumption that f is integrable, choose a non-negative, even $\rho \in \mathscr{S}(\mathbb{R}^N;\mathbb{R})$ for which $\langle \rho, \lambda_{\mathbb{R}^N}\rangle = 1$, and set $\rho_t(x) = t^{-\frac{N}{2}}\rho(t^{-1}x)$ for $t > 0$. Then $\hat{\rho} \in \mathscr{S}(\mathbb{R}^N;\mathbb{R})$ and $\widehat{\rho_t}(\xi) = \hat{\rho}(t\xi)$. Therefore $f_t \equiv \widehat{\rho_t}f$ is a continuous, $\lambda_{\mathbb{R}^N}$-integrable function that is 1 at 0. To see that f_t is non-negative definite, note that

$$\begin{aligned}\sum_{j,k=1}^n f_t(\xi_j - \xi_k)\alpha_j\overline{\alpha_k} &= \sum_{j,k=1}^n f(\xi_j-\xi_k)\alpha_j\overline{\alpha_k}\int e^{i(\xi_j-\xi_k)x}\,\rho_t(x)\,dx \\ &= \int\left(\sum_{j,k=1}^n f(\xi_j-\xi_k)\big(\alpha_j e^{i\xi_j x}\big)\big(\overline{\alpha_k e^{i\xi_k x}}\big)\right)\rho_t(x)\,dx \geq 0.\end{aligned}$$

Thus $f_t = \widehat{\mu_t}$ for some $\mu_t \in M_1(\mathbb{R}^N)$, and so, since $f_t \longrightarrow f$ uniformly on compacts, Lévy's Continuity Theorem implies that μ_t tends weakly to a $\mu \in M_1(\mathbb{R}^N)$ for which $f = \hat{\mu}$. □

Because it is difficult to check whether a function is non-negative definite, it is the more or less trivial necessity part of Bochner's Theorem that turns out in practice to be more useful than the sufficiency conditions.

Exercise 1.1.1

(**i**) By combining Theorem 1.1.10 with (1.1.2), one sees that if f is a continuous, non-negative function for which $f(0) = 1$, then $|f(\xi)| \le 1$ and $|f(\eta) - f(\xi)|^2 \le 2\big(1 - \Re f(\eta - \xi)\big)$. Show that these inequalities hold even if one drops the continuity assumption.

Hint: Use the non-negative definiteness of the matrices

$$\begin{pmatrix} 1 & f(-\xi) \\ f(\xi) & 1 \end{pmatrix} \text{ and } \begin{pmatrix} 1 & f(-\xi) & f(-\eta) \\ f(\xi) & 1 & f(\xi - \eta) \\ f(\eta) & f(\eta - \xi) & 1 \end{pmatrix}$$

to see that $f(-\xi) = \overline{f(\xi)}$ and that

$$1 + 2\alpha\big(1 - \Re f(\eta - \xi)\big) + 2\alpha^2\big|f(\eta) - f(\xi)\big| \ge 0 \text{ for all } \alpha \in \mathbb{R}.$$

(**ii**) Show that if f_1 and f_2 are continuous, non-negative definite functions, then so are $f_1 f_2$ and, for any $a, b \ge 0$, $af_1 + bf_2$. See if you can drop the continuity assumption.

(**iii**) Suppose that $f : \mathbb{R}^N \longrightarrow \mathbb{C}$ is a function for which $f(0) = 1$. Show that if $\lim_{|x|\searrow 0} \frac{1-f(x)}{|x|^2} = 0$, then f cannot be a characteristic function. In particular, if $\alpha > 2$, then $e^{-|\xi|^\alpha}$ is not a characteristic function.

(**iv**) Given a finite signed Borel measure μ on $\mathbb{R}^N$, define $\hat{\mu}(\xi) = \int e^{i(\xi,x)_{\mathbb{R}^N}}\, \mu(dx)$, and show that $\hat{\mu} = 0$ if and only if $\mu = 0$.

Hint: Use the Hahn Decomposition Theorem to write μ as the difference of two, mutually singular, non-negative Borel measures on $\mathbb{R}^N$.

(**v**) Suppose that $f : \mathbb{R} \longrightarrow \mathbb{C}$ is a non-constant, twice continuously differentiable characteristic function. Show that $f''(0) < 0$ and that $\frac{f''}{f''(0)}$ is again a characteristic function. In addition, show that $\|f'\|_{\mathrm{u}} \vee \|f''\|_{\mathrm{u}} \le |f''(0)|^{\frac{1}{2}}$ and that $|f(\eta) - f(\xi)| \le \big(|f'(0)| + |f''(0)|^{\frac{1}{2}}\big)|\eta - \xi|$.

(**vi**) Let $\mu_n \in M_1(\mathbb{R})$ be the measure for which $\frac{d\mu_n}{d\lambda_{\mathbb{R}}} = (2n)^{-1}\mathbf{1}_{[-n,n]}$. Show that $\widehat{\mu_n} \longrightarrow \mathbf{1}_{\{0\}}$ pointwise, and conclude that $\{\mu_n : n \ge 1\}$ has no weak limits. This example demonstrates the important role that continuity plays in Bochner's and Lévy's theorems.

1.2 Infinitely Divisible Laws

Except for Lemma 1.2.1, the contents of this subsection will not be used below.

Because sums of independent, identically distributed random variables play a prominent role in probability theory, Lévy and Khinchine asked what are the distributions of random variables that, for every $n \geq 1$, can be written as the sum of n independent, identically distributed random variables. To express this in terms of measures, remember that the convolution of $\mu, \nu \in M_1(\mathbb{R}^N)$ is the measure $\mu * \nu \in M_1(\mathbb{R}^N)$ given by

$$\mu * \nu(\Gamma) = \iint \mathbf{1}_\Gamma(x+y)\,\mu(dx)\nu(dy)$$

and that, if μ and ν are the distributions of independent random variables X and Y, then $\mu * \nu$ is the distribution of $X+Y$. Thus, what they were asking is which $\mu \in M_1(\mathbb{R}^N)$ have the property that, for every $n \geq 2$, μ admits nth root $\mu_{\frac{1}{n}} \in M_1(\mathbb{R}^N)$ with respect to convolution. That is,

$$\mu = (\mu_{\frac{1}{n}})^{*n} = \underbrace{\mu_{\frac{1}{n}} * \cdots * \mu_{\frac{1}{n}}}_{n},$$

which, since $\widehat{\mu * \nu} = \hat{\mu}\hat{\nu}$, is equivalent to $\hat{\mu} = (\widehat{\mu_{\frac{1}{n}}})^n$.

Denote by $\mathcal{I}(\mathbb{R}^N)$ the set of infinitely divisible $\mu \in M_1(\mathbb{R}^N)$. What Lévy and Khinchine proved is that $\mu \in \mathcal{I}(\mathbb{R}^N)$ if and only if

$$\begin{aligned}\hat{\mu}(\xi) = \exp\Bigg(& i(b,\xi)_{\mathbb{R}^N} - \frac{1}{2}\big(\xi, A\xi\big)_{\mathbb{R}^N} \\ &+ \int \Big(e^{i(\xi,y)_{\mathbb{R}^N}} - 1 - i\mathbf{1}_{B_{\mathbb{R}^N}(0,1)}(y)(\xi,y)_{\mathbb{R}^N}\Big)\, M(dy)\Bigg),\end{aligned} \tag{1.2.1}$$

for some $b \in \mathbb{R}^N$, non-negative definite, symmetric $A \in \mathrm{Hom}(\mathbb{R}^N;\mathbb{R}^N)$, and Borel measure M on $\mathbb{R}^N$ such that $M(\{0\}) = 0$ and $\int \frac{|y|^2}{1+|y|^2}\, M(dy) < \infty$. The expression in (1.2.1) is called the *Lévy–Khinchine formula*, a measure M satisfying the stated conditions is called a *Lévy measure*, and the triple (b, A, M) is called a *Lévy system*. It is clear that if the right hand side of (1.2.1) is a characteristic function for every Lévy system, then these are characteristic functions of infinitely divisible laws. Namely, if μ corresponds to (b, A, M) and $\mu_{\frac{1}{n}}$ corresponds to $\big(\frac{b}{n}, \frac{A}{n}, \frac{M}{n}\big)$, then $\hat{\mu} = (\widehat{\mu_{\frac{1}{n}}})^n$.

Proving that the function $f_{(b,A,M)}$ on the right hand side of (1.2.1) is a characteristic function is a relatively easy. Indeed, $f_{(0,I,0)} = \hat{\gamma}$ (cf. (2.1.1)), where $\gamma(dx) = (2\pi)^{-\frac{N}{2}} e^{-\frac{|x|^2}{2}}\, \lambda_{\mathbb{R}^N}(dx)$, and so it is easy to check that $f_{b,A,0}$ is the characteristic function of the distribution of $b + A^{\frac{1}{2}}x$ under γ. Also, if the Lévy measure M is finite and π_M is the Poisson measure given by

$$\pi_M = e^{-M(\mathbb{R}^N)} \sum_{n=0}^{\infty} \frac{M^{*n}}{n!},$$

then $f_{b_M,0,M} = \widehat{\pi_M}$, where

$$b_M = \int_{B_{\mathbb{R}^N}(0,1)} y\, M(dy).$$

Hence, when M is finite, $f_{(b,A,M)}$ is the characteristic function of $\gamma_{b-b_M,A} * \pi_M$. Finally, for general Lévy measures M, set $M_k(dy) = \mathbf{1}_{[\frac{1}{k},\infty)}(|y|)M(dy)$. Then M_k is finite, and so $f_{(b,A,M_k)}$ is a characteristic function. Therefore, since $f_{(b,A,M_k)} \longrightarrow f_{(b,A,M)}$ uniformly on compacts, Theorem 1.1.9 says that $f_{(b,A,M)}$ is a characteristic function.

There are several proofs that (1.2.1) describes the characteristic function of every $\mu \in \mathcal{I}(\mathbb{R}^N)$, but none of them is simple. Nonetheless, it is possible to explain the idea on which all approaches are based. Namely, elements of $\mathcal{I}(\mathbb{R}^N)$ have nth roots for all $n \geq 1$ in the Abelian group $M_1(\mathbb{R}^N)$ with convolution. Thus, for any $m,\, n \geq 1$, a $\mu \in \mathcal{I}(\mathbb{R}^N)$ has an $\frac{m}{n}$th root $\mu_{\frac{m}{n}} = (\mu_{\frac{1}{n}})^{*m}$. Hence one should expect that one can take the logarithm of μ. Equivalently, one should expect that the limit

$$\ell_\mu(\xi) = \lim_{n\to\infty} n\big(\widehat{\mu_{\frac{1}{n}}}(\xi) - 1\big)$$

exists and that $\hat{\mu} = e^{\ell_u}$. As we will show, the existence of this limit is quite easy, but the proof that

$$\ell_\mu(\xi) = i(b,\xi)_{\mathbb{R}^N} - \frac{1}{2}\big(\xi, A\xi\big)_{\mathbb{R}^N} + \int \Big(e^{i(\xi,y)_{\mathbb{R}^N}} - 1 - i\mathbf{1}_{B_{\mathbb{R}^N}(0,1)}(y)(\xi,y)_{\mathbb{R}^N}\Big)\, M(dy)$$

for some Lévy system (b, A, M) is harder.

To prove that ℓ_μ exists, we will need the following elementary fact about $\mathbb{C}$-valued functions. In its statement, log is the principle branch of the logarithm function on $\mathbb{C} \setminus (-\infty, 0]$. That is, $\log(re^{i\theta}) = \log r + \theta$ for $r > 0$ and $\theta \in (-\pi, \pi)$. In particular,

$$\log(\zeta) = -\sum_{n=0}^{\infty} \frac{(1-\zeta)^n}{n} \quad \text{when } |1-\zeta| < 1.$$

Lemma 1.2.1 *Given $R > 0$, suppose that $f : \overline{B_{\mathbb{R}^N}(0,R)} \longrightarrow \mathbb{C}$ is a continuous function that equals 1 at 0 and never vanishes. Then there exists a unique continuous function $\ell_f : \overline{B_{\mathbb{R}^N}(0,R)} \longrightarrow \mathbb{C}$ such that $\ell_f(0) = 0$ and $f = e^{\ell_f}$ on $\overline{B_{\mathbb{R}^N}(0,R)}$. Moreover, if $\xi \in \overline{B_{\mathbb{R}^N}(0,R)}$ and $r > 0$, then*

$$\sup\left\{\left|1 - \frac{f(\eta)}{f(\xi)}\right| : \eta \in \overline{B_{\mathbb{R}^N}(\xi,r) \cap B_{\mathbb{R}^N}(0,R)}\right\} < 1$$
$$\implies \ell_f(\eta) = \ell_f(\xi) + \log\frac{f(\eta)}{f(\xi)} \text{ for } \eta \in \overline{B_{\mathbb{R}^N}(\xi,r) \cap B_{\mathbb{R}^N}(0,R)},$$

and so

$$\left|1 - \frac{f(\eta)}{f(\xi)}\right| \leq \frac{1}{2} \implies \big|\ell_f(\eta) - \ell_f(\xi)\big| \leq 2\left|1 - \frac{f(\eta)}{f(\xi)}\right|.$$

Finally, if $g : \overline{B_{\mathbb{R}^N}(0,R)} \longrightarrow \mathbb{C}$ *is a second function with the same properties as* f*, then*

$$\sup_{\xi \in \overline{B_{\mathbb{R}^N}(0,R)}} \left|1 - \frac{g(\xi)}{f(\xi)}\right| \le \frac{1}{2} \implies |\ell_g(\xi) - \ell_f(\xi)| \le 2\left|1 - \frac{g(\xi)}{f(\xi)}\right|.$$

Proof. To prove the existence and uniqueness of ℓ_f, begin by observing that there exists an $M \in \mathbb{Z}$ and $0 = r_0 < r_1 < \cdots < r_M = R$ such that

$$\left|1 - \frac{f(\xi)}{f\left(\frac{r_{m-1}\xi}{|\xi|}\right)}\right| \le \frac{1}{2} \quad \text{for } 1 \le m \le M \text{ and } \xi \in \overline{B_{\mathbb{R}^N}(0,r_m)} \setminus \overline{B_{\mathbb{R}^N}(0,r_{m-1})}.$$

Thus we can define a function ℓ_f on $\overline{B_{\mathbb{R}^N}(0,R)}$ so that $\ell_f(0) = 0$ and

$$\ell_f(\xi) = \ell_f\left(\frac{r_{m-1}\xi}{|\xi|}\right) + \log \frac{f(\xi)}{f\left(\frac{r_{m-1}\xi}{|\xi|}\right)}$$
$$\text{if } 1 \le m \le M \text{ and } \xi \in \overline{B_{\mathbb{R}^N}(0,r_m)} \setminus \overline{B_{\mathbb{R}^N}(0,r_{m-1})}.$$

Furthermore, working by induction on $1 \le m \le M$, one sees that this ℓ_f is continuous and satisfies $f = e^{\ell_f}$. Finally, for any $\ell \in C\big(\overline{B_{\mathbb{R}^N}(0,R)};\mathbb{C}\big)$ satisfying $\ell(0) = 0$ and $f = e^{\ell}$, $(i2\pi)^{-1}(\ell - \ell_f)$ is a continuous, $\mathbb{Z}$-valued function that vanishes at 0, and therefore $\ell = \ell_f$.

Next suppose that $\xi \in B_{\mathbb{R}^N}(0,R)$ and that

$$\left|1 - \frac{f(\eta)}{f(\xi)}\right| < 1 \quad \text{for all } \eta \in \overline{B_{\mathbb{R}^N}(\xi,r) \cap B_{\mathbb{R}^N}(0,R)}.$$

Set

$$\ell(\eta) = \ell_f(\xi) + \log \frac{f(\eta)}{f(\xi)} \quad \text{for } \eta \in B_{\mathbb{R}^N}(\xi,r) \cap B_{\mathbb{R}^N}(0,R),$$

and check that $\eta \rightsquigarrow (i2\pi)^{-1}\big(\ell(\eta) - \ell_f(\eta)\big)$ is a continuous, $\mathbb{Z}$-valued function that vanishes at ξ. Hence, $\ell = \ell_f$ on $B_{\mathbb{R}^N}(0,R) \cap B_{\mathbb{R}^N}(\xi,r)$, and therefore on $\overline{B_{\mathbb{R}^N}(0,R) \cap B(\xi,r)}$. Since $|\log(1-\zeta)| \le 2|\zeta|$ if $|\zeta| \le \frac{1}{2}$, this completes the proof of the asserted properties of ℓ_f.

Turning to the comparison between ℓ_g and ℓ_f when $\left|1 - \frac{g(\xi)}{f(\xi)}\right| \le \frac{1}{2}$ for all $\xi \in \overline{B_{\mathbb{R}^N}(0,R)}$, set $\ell(\xi) = \ell_f(\xi) + \log \frac{g(\xi)}{f(\xi)}$, check that $\ell(0) = 0$ and $g = e^{\ell}$, and conclude that $\ell_g - \ell_f = \log \frac{g}{f}$. From this, the asserted estimate for $|\ell_g - \ell_f|$ is immediate. □

Lemma 1.2.2 *If* $\mu \in \mathcal{I}(\mathbb{R}^N)$*, then* $\hat{\mu}$ *never vanishes.*

Proof. Choose $r > 0$ so that $|1 - \hat{\mu}(\xi)| \le \frac{1}{2}$ for $|\xi| \le r$. By Lemma 1.2.1, there is a unique continuous $\ell : \overline{B_{\mathbb{R}^N}(0,r)} \longrightarrow \mathbb{C}$ such that $\ell(0) = 0$ and $\hat{\mu} = e^{\ell}$

on $\overline{B_{\mathbb{R}^N}(0,r)}$. Now choose $\mu_{\frac{1}{n}}$ so that $\hat{\mu} = (\widehat{\mu}_{\frac{1}{n}})^n$. Then $\widehat{\mu}_{\frac{1}{n}}$ doesn't vanish on $\overline{B_{\mathbb{R}^N}(0,r)}$, and there is an $\ell_n : \overline{B_{\mathbb{R}^N}(0,r)} \longrightarrow \mathbb{C}$ such that $\ell_n(0) = 0$ and $\widehat{\mu_{\frac{1}{n}}} = e^{\ell_n}$ on $\overline{B_{\mathbb{R}^N}(0,r)}$. Hence, $\hat{\mu} = e^{n\ell_n}$, and so, by uniqueness, $\ell = n\ell_n$ and therefore $\widehat{\mu_{\frac{1}{n}}} = e^{\frac{\ell}{n}}$ on $\overline{B_{\mathbb{R}^N}(0,r)}$.

Next, because $\ell = \log \hat{\mu}$ and $|1-\hat{\mu}| \le \frac{1}{2}$ on $\overline{B_{\mathbb{R}^N}(0,r)}$, $|\ell| \le 2$ there. Therefore, since $\Re \ell \le 0$, $|1-\widehat{\mu_{\frac{1}{n}}}| = |1-e^{\frac{\ell}{n}}| \le \frac{2}{n}$ on $\overline{B_{\mathbb{R}^N}(0,r)}$. Now apply (1.1.5) to see that, for any $\rho > 0$,

$$\mu_n\big(\{x : |x| \ge \rho\}\big) \le \frac{2N}{ns(N^{-\frac{1}{2}}\rho r)}$$

and therefore, by (1.1.4), that

$$|1-\widehat{\mu_{\frac{1}{n}}}(\xi)| \le \rho R + \frac{2N}{ns(N^{-\frac{1}{2}}r\rho)} \quad \text{for } \rho,\ R > 0 \text{ and } |\xi| \le R.$$

Finally, take $\rho = \frac{1}{4R}$ and $n \ge \frac{4N}{s(N^{-\frac{1}{2}}r\rho)}$, and conclude that $\widehat{\mu_{\frac{1}{n}}}$, and therefore $\hat{\mu}$, doesn't vanish on $\overline{B_{\mathbb{R}^N}(0,R)}$ for any $R > 0$. □

As a consequence of Lemmas 1.2.2 and 1.2.1, we know that if $\mu \in \mathcal{I}(\mathbb{R}^N)$, then $\hat{\mu} = e^{\ell_\mu}$ for a unique continuous $\ell_\mu : \mathbb{R}^N \longrightarrow \mathbb{C}$ which vanishes at 0. Further, by the argument at the beginning of the preceding proof, $\mu = (\mu_{\frac{1}{n}})^{*n} \implies \widehat{\mu_{\frac{1}{n}}} = e^{\frac{\ell_\mu}{n}}$. Thus ℓ_μ is the limit which was predicted above.

The challenge now is to show that

$$\ell_\mu(\xi) = \ell_{(b,A,M)}(\xi) \equiv i(b,\xi)_{\mathbb{R}^N} - \frac{1}{2}\big(\xi, A\xi\big)_{\mathbb{R}^N} + \int \Big(e^{i(\xi,y)_{\mathbb{R}^N}} - 1 - i\mathbf{1}_{B_{\mathbb{R}^N}(0,1)}(y)(\xi,y)_{\mathbb{R}^N}\Big)\, M(dy)$$

for some Lévy system (b, A, M). Here is an outline of one approach. Using the estimates in Lemma 1.1.8, one can show that $\ell_\mu(\xi)$ has at most quadratic growth, thereby justifying

$$\int \hat{\varphi}(\xi)\ell_\mu(-\xi)\,d\xi = \lim_{t\searrow 0} t^{-1}\int \hat{\varphi}(\xi)\big(e^{t\ell_\mu(-\xi)}-1\big)\,d\xi = (2\pi)^N \lim_{t\searrow 0}\frac{\langle \varphi, \mu_t\rangle - \varphi(0)}{t}$$

for $\varphi \in \mathscr{S}(\mathbb{R}^N;\mathbb{C})$. Next, define the linear functional $\mathcal{A}_\mu$ on $\mathscr{S}(\mathbb{R};\mathbb{R})$ by

$$\mathcal{A}_\mu \varphi = \lim_{t\searrow 0}\frac{\langle \varphi, \mu_t\rangle - \varphi(0)}{t}.$$

At this point the problem becomes that of showing that

$$\mathcal{A}_\mu\varphi = \big(b, \nabla\varphi(0)\big)_{\mathbb{R}^N} + \frac{1}{2}\sum_{j,k=1}^N A_{j,k}\partial_j\partial_k\varphi(0) \\ + \int \Big(\varphi(x) - \varphi(0) - \mathbf{1}_{B_{\mathbb{R}^N}(0,1)}(x)\big(x, \nabla\varphi(0)\big)_{\mathbb{R}^N}\Big)\, M(dx)$$

for some Lévy system (b, A, M). Indeed, define the *Lévy operator* $\mathcal{L}_{(b,A,M)}$ on $C^2_{\mathrm{b}}(\mathbb{R}^N;\mathbb{C})$ by

$$\mathcal{L}_{(b,A,M)}\varphi(x) = \big(b_j, \nabla\varphi(x)\big)_{\mathbb{R}^N} + \frac{1}{2}\sum_{j,k=1}^N A_{j,k}\partial_j\partial_k\varphi(x) \\ + \int \Big(\varphi(x+y) - \varphi(x) - \mathbf{1}_{B_{\mathbb{R}^N}(0,1)}(x)\big(y, \nabla\varphi(x)\big)_{\mathbb{R}^N}\Big)\, M(dy).$$

Using Fubini's theorem and elementary Fourier theory, one can check that

$$(2\pi)^{-N}\int \hat{\varphi}(\xi)\ell_{(b,A,M)}(-\xi)\,d\xi = \mathcal{L}_{(b,A,M)}\varphi(0),$$

and therefore one would know that

$$\int \hat{\varphi}(\xi)\ell_\mu(-\xi)\,d\xi = \int \hat{\varphi}(\xi)\ell_{(b,A,M)}(-\xi)\,d\xi \quad \text{for } \varphi \in \mathscr{S}(\mathbb{R};\mathbb{C}),$$

which is possible only if $\ell_\mu = \ell_{(b,A,M)}$.

The proof that $\mathcal{A}_\mu\varphi = \mathcal{L}_{(b,A,M)}\varphi(0)$ for some Lévy system (b, A, M) relies on two facts, the most crucial of which is the simple observation that $\mathcal{A}_\mu\varphi \ge 0$ if $\varphi \in \mathscr{S}(\mathbb{R}^N;\mathbb{R})$ satisfies $\varphi \ge \varphi(0)$, a property that is reminiscent of the minimum principle for second order elliptic operators and is obvious from original expression for $\mathcal{A}_\mu\varphi$. The second required fact is that $\mathcal{A}_\mu$ has a quasi-local property. Namely, if $\varphi \in \mathscr{S}(\mathbb{R};\mathbb{C})$ and $\varphi_R(x) = \varphi\big(\frac{x}{R}\big)$, then $\mathcal{A}_\mu\varphi_R \longrightarrow 0$ as $R \to \infty$. Verifying this property is most easily done by using the expression for the action of $\mathcal{A}_\mu$ in terms of ℓ_μ. Finally, based on these two properties alone, one can show that $\mathcal{A}_\mu\varphi = \mathcal{L}_{(b,A,M)}\varphi(0)$ for some (b, A, M). See § 3.1 in [10] for more details.

Exercise 1.2.1

(**i**) Show that if $\mu \in \mathcal{I}(\mathbb{R}^N)$, then $e^{t\ell_\mu}$ is a characteristic function for each $t > 0$.

(**ii**) Let $\mu \in M_1(\mathbb{R}^N)$ and $\{n_k : k \ge 1\} \subseteq \mathbb{Z}^+$ is a sequence that increases to ∞. Show that if, for each $k \ge 1$, there is an $\mu_{\frac{1}{n_k}} \in M_1(\mathbb{R}^N)$ such that $\mu = (\mu_{\frac{1}{n_k}})^{*n_k}$, then $\mu \in \mathcal{I}(\mathbb{R}^N)$.

(**iii**) Show that for each $\alpha \in (0, 2]$ there is a $\mu \in \mathcal{I}(\mathbb{R}^N)$ for which $e^{-|\xi|^\alpha}$ is the characteristic function. Thus, when combined with (**iii**) in Exercise 1.1.1, this proves that, for any $t > 0$, $e^{-t|\xi|^\alpha}$ is a non-negative definite function if and only $\alpha \in (0, 2]$.

(**iv**) The Poisson kernel for the upper halfspace $\mathbb{R} \times (0, \infty)$ is the function $p_y(x) = \frac{y}{\pi(x^2+y^2)}$. After checking that $\int p_y(x)\, dx = 1$, take $d\mu_y = p_y\, d\lambda_{\mathbb{R}}$, and show that $\widehat{\mu_y}(\xi) = e^{-y|\xi|}$.

Hint: Show that $\partial_y^2 p_y(x) = -\partial_x^2 p_y(x)$ and therefore that $\partial_y^2 \widehat{p_y}(\xi) = \xi^2 \widehat{p_y}(\xi)$. Next, show that $\lim_{y \searrow 0} \widehat{p_y}(\xi) = 1$, and use the fact that $|\widehat{p_y}(\xi)| \leq 1$ to conclude that $\widehat{p_y}(\xi) = e^{-y|\xi|}$.

Chapter 2
Gaussian Measures and Families

This chapter deals with some of the properties of Gaussian measures and the construction of families of Gaussian random variables.

2.1 Gaussian Measures on $\mathbb{R}$

The *standard Gaussian measure* on $\mathbb{R}$ is the Borel probability measure $\gamma_{0,1} \ll \lambda_{\mathbb{R}}$ with Randon–Nikodym derivative

$$\frac{d\gamma_{0,1}}{d\lambda_{\mathbb{R}}}(x) = (2\pi)^{-\frac{1}{2}} e^{-\frac{x^2}{2}} \lambda_{\mathbb{R}}(dx).$$

If $b \in \mathbb{R}$ and $a \geq 0$, then the $\gamma_{b,a} \in M_1(\mathbb{R}^N)$ is the distribution of $b + a^{\frac{1}{2}} x$ under $\gamma_{0,1}$. Thus, $\gamma_{b,0} = \delta_b$, the unit point mass at b, and if $a > 0$, then $\gamma_{b,a}(dx) = (2\pi a)^{-\frac{1}{2}} e^{-\frac{(x-b)^2}{2a}} \lambda_{\mathbb{R}}(dx)$. It is easy to check that a real valued random variable with distribution $\gamma_{b,a}$ has mean value b and variance a. Some people say that such a random variable is Gaussian and others say it is normal, and we will sometimes use one term and sometimes the other term. Finally, the set $N(b,a)$ will denote the set of normal random variables with mean b and variance a.

The following lemma contains a few useful facts about Gaussian random variables.

Lemma 2.1.1 *If $a > 0$ and $X \in N(0,a)$, then*

$$\mathbb{E}\big[e^{\zeta X}\big] = e^{\frac{a\zeta^2}{2}} \textit{ for all } \zeta \in \mathbb{C}, \tag{2.1.1}$$

$$\mathbb{E}\big[e^{\frac{\alpha X^2}{2}}\big] = (1-\alpha a)^{-\frac{1}{2}} \quad \textit{for } \alpha \in \big[0, a^{-1}\big), \tag{2.1.2}$$

$$\mathbb{E}\big[|X|^\alpha\big] = C_\alpha a^{\frac{\alpha}{2}} \text{ where } C_\alpha = \int |x|^\alpha \, \gamma_{0,1}(dx) \text{ for } \alpha \in [0,\infty), \tag{2.1.3}$$

and, for $n \geq 1$,

$$\mathbb{E}\big[X^{2n-1}\big] = 0 \text{ and } \mathbb{E}\big[X^{2n}\big] = \frac{(2n)!}{2^n n!} a^n = a^n \prod_{m=1}^{n} (2m-1). \tag{2.1.4}$$

Moreover, if $\{X_n : n \geq 1\}$ is a sequence of Gaussian random variables and X_n converges in probability to X, then X is Gaussian and $\mathbb{E}\big[|X_n - X|^p\big] \longrightarrow 0$ for all $p \in [1,\infty)$.

Proof. When $\zeta \in \mathbb{R}$, (2.1.1) follows by writing $\zeta x - \frac{x^2}{2a}$ as

$$-\frac{(x - a\zeta)^2}{2a} + \frac{a\zeta^2}{2}$$

and using the translation invariance of $\lambda_{\mathbb{R}}$. To handle general $\zeta \in \mathbb{C}$, note that both sides of the equation are entire functions that agree on $\mathbb{R}$ and therefore on $\mathbb{C}$.

To prove (2.1.2), set $c = \frac{a}{1-\alpha a}$, and observe that

$$\mathbb{E}\Big[e^{\frac{\alpha X^2}{2}}\Big] = (2\pi a)^{-\frac{1}{2}} \int \exp\Big(-\frac{x^2}{2c}\Big)\, dx = \Big(\frac{c}{a}\Big)^{\frac{1}{2}}.$$

Since X has the same distribution as $a^{\frac{1}{2}}$ times a standard normal random variable, (2.1.3) is trivial and (2.1.4) reduces to the case when $a = 1$. But, when $a = 1$, Theorem 1.1.4 justifies

$$\sum_{m=0}^{\infty} \frac{\alpha^m}{m!} \mathbb{E}\big[X^m\big] = \mathbb{E}\big[e^{\alpha X}\big] = e^{\frac{\alpha^2}{2}} = \sum_{m=0}^{\infty} \frac{\alpha^{2m}}{2^m m!},$$

from which (2.1.4) with $a = 1$ follows immediately.

Now suppose that $X_n \in N(b_n, a_n)$ and that $X_n \longrightarrow X$ in probability. Then, for all $\xi \in \mathbb{R}$,

$$e^{ib_n\xi - \frac{a_n\xi^2}{2}} = \mathbb{E}\big[e^{i\xi X_n}\big] \longrightarrow \mathbb{E}\big[e^{i\xi X}\big].$$

Choose $r > 0$ so that $\mathbb{E}\big[e^{i\xi X}\big] \neq 0$ for $|\xi| \leq r$. Then, by Lemma 1.2.1, there is a unique continuous $\ell : [-r, r] \longrightarrow \mathbb{C}$ such that $\ell(0) = 0$ and $\mathbb{E}\big[e^{i\xi X}\big] = e^{\ell(\xi)}$, and so, by that lemma, $ib_n - \frac{a\xi^2}{2} \longrightarrow \ell(\xi)$ for $|\xi| \leq r$. Hence, there exist $b \in \mathbb{R}$ and $a \geq 0$ such that $b_n \longrightarrow b$ and $a_n \longrightarrow a$, and so $X \in N(b,a)$. Furthermore, since $b_n \longrightarrow b$, to see that $\mathbb{E}\big[|X_n - X|^p\big] \longrightarrow 0$, it suffices to do so when $b_n = 0$ for all n. But, since $\{a_n : n \geq 1\}$ is bounded, in that case, (2.1.2) says that there exists and $\alpha > 0$ such that

$$\mathbb{E}\big[e^{\frac{\alpha X^2}{2}}\big] \le \sup_{n\ge 1} \mathbb{E}\big[e^{\frac{\alpha X_n^2}{2}}\big] < \infty,$$

and so $\big\{|X_n - X|^p : n \ge 1\big\}$ is a sequence of uniformly integrable functions that tend to 0 in probability. □

2.2 Cramér–Lévy Theorem

The following remarkable theorem was discovered by Cramér and Lévy. So far as I know, there is no truly probabilistic or real analytic proof of it.

Theorem 2.2.1 *If X and Y are independent random variables whose sum is Gaussian, then each of them is Gaussian.*

Proof. Without loss in generality, assume that $X + Y \in N(0,1)$.

Choose $r > 0$ so that $\mathbb{P}(|X| \ge r) \vee \mathbb{P}(|Y| \ge r) \le \frac{1}{2}$. Then

$$\mathbb{P}(|X| \ge r+R) \le 2\mathbb{P}(|X| \ge r+R \ \&\ |Y| \le r) \le 2\mathbb{P}(|X+Y| \ge R) \le 4e^{-\frac{R^2}{2}},$$

and similarly $\mathbb{P}(|Y| \ge r+R) \le 4e^{-\frac{R^2}{2}}$. Therefore $\mathbb{E}\big[e^{\alpha X^2}\big] \vee \mathbb{E}\big[e^{\alpha Y^2}\big] < \infty$ for some $\alpha > 0$. Knowing that X and Y are integrable, one can reduce to the case when they have mean 0, and so we will proceed under that assumption.

Set $f(\zeta) = \mathbb{E}\big[e^{\zeta X}\big]$ and $g(\zeta) = \mathbb{E}\big[e^{\zeta Y}\big]$ for $\zeta \in \mathbb{C}$. Both f and g are entire functions whose product equals $e^{\frac{\zeta^2}{2}}$. In particular, neither of them vanishes anywhere, and so, by Lemma 1.2.1, there is an entire function θ such that $\theta(0) = 0$, $f(\zeta) = e^{\theta(\zeta)}$, and $g(\zeta) = \exp\big(\frac{\zeta^2}{2} - \theta(\zeta)\big)$. Furthermore, since $\mathbb{E}[X] = 0$ and therefore $\theta'(0) = 0$,

$$\theta(\zeta) = \sum_{n=2}^{\infty} c_n \zeta^n \quad \text{where } n! c_n = \partial_\xi^n \log \mathbb{E}\big[e^{\xi X}\big]\,\big|_{\xi=0} \in \mathbb{R}.$$

Observe that, by Hölder's inequality, $\log \circ f$ and $\log \circ g$ are convex functions of $\xi \in \mathbb{R}$. Thus, since they and their first derivatives vanish at $\xi = 0$, they must be non-negative on $\mathbb{R}$. That is,

$$\theta(\xi) \ge 0 \le \frac{\xi^2}{2} - \theta(\xi).$$

Therefore, for $\zeta = \xi + i\eta$,

$$e^{\Re\theta(\zeta)} = \big|e^{\theta(\zeta)}\big| \le \mathbb{E}\big[\big|e^{\zeta} X\big|\big]\big| = e^{\theta(\xi)} \le e^{\frac{\xi^2}{2}}$$

and

$$e^{\frac{\xi^2-\eta^2}{2} - \Re\theta(\zeta)} \le e^{\frac{\xi^2}{2} - \theta(\xi)},$$

which means that $-\eta^2 \le 2\Re\theta(\zeta) \le \xi^2$ and therefore that $\big|\Re\theta(\zeta)\big| \le \frac{|\zeta|^2}{2}$.

Finally, by Cauchy's Theorem, for $n > 2$ and any $r > 0$,

$$2\pi c_n = r^{-n}\int_0^{2\pi}\theta(re^{it})e^{-int}\,dt = 2r^{-n}\int_0^{2\pi}\Re\theta(re^{it})e^{-int}\,dt$$

since

$$\int_0^{2\pi}\overline{\theta(re^{it})}e^{-int}\,dt = \overline{\int_0^{2\pi}\theta(re^{it})e^{int}\,dt} = 0.$$

Thus $2\pi|c_n| \le 2\pi r^{2-n}$ for all $r > 0$, which means that $c_n = 0$ for $n > 2$. Therefore $f(\zeta) = e^{c_n\zeta^2}$, which is possible only if $c_n \le 0$. □

In the following and elsewhere, I will employ the technique of *symmetrization* of a $\mu \in M_1(\mathbb{R})$. To describe this technique, define $\tilde{\mu}$ by $\tilde{\mu}(\Gamma) = \mu(-\Gamma)$. Then the symmetrizaton of μ is the measure $\mu * \tilde{\mu}$. Equivalently, $\widehat{\mu * \tilde{\mu}} = |\hat{\mu}|^2$, and if X and Y are independent random variables with distribution μ, then $X-Y$ has distribution $\mu*\tilde{\mu}$. Obviously, $\mu*\tilde{\mu}$ is even in the sense that it assigns $-\Gamma$ the same measure as it does Γ. Less obvious is the fact that integrability properties of X are intimately related to those of $X - Y$. Specifically, let α be a median of X (i.e., $\mathbb{P}(X \ge \alpha) \ge \frac{1}{2} \le \mathbb{P}(X \le \alpha)$). For any $R > 0$,

$$\mathbb{P}(X \ge R + \alpha) \le 2\mathbb{P}(X \ge R + \alpha \;\&\; Y \le \alpha) \le 2\mathbb{P}(X - Y \ge R),$$

and similarly $\mathbb{P}(X \le -R + \alpha) \le 2\mathbb{P}(X - Y \le -R)$. Therefore

$$\mathbb{P}(|X - \alpha| \ge R) \le 2\mathbb{P}(|X - Y| \ge R),$$

and so

$$\mathbb{E}\big[|X|^p\big]^{\frac{1}{p}} \le 2^{\frac{1}{p}}\mathbb{E}\big[|X - Y|^p\big]^{\frac{1}{p}} + \alpha$$

for all $p \in [1,\infty)$.

The following corollary, which is closely related to (**iii**) in Exercise 1.1.1 and 1.2.1, was proved originally by J. Marcinkiewicz.

Corollary 2.2.2 *If $\hat{\mu}$ is the exponential of a polynomial, then μ is Gaussian.*

Proof. By symmetrization and the Lévy–Cramér Theorem, we may assume that μ is even, and therefore that $\hat{\mu}(\xi) = e^{P(\xi)}$, where $P(\xi) = \sum_{m=0}^{n} c_m\xi^{2m}$ with $c_m \in \mathbb{R}$ for all $0 \le m \le n$. If $n = 0$, there is noting to do, and so we will assume that $n \ge 1$ and $c_n \ne 0$, in which case, since $|\hat{\mu}(\xi)| \le 1$, it is clear that $c_n < 0$, and, by Theorem 1.1.4, we know that

$$\mathbb{E}\big[e^{\zeta X}\big] = e^{P(-i\zeta)} \text{ for } \zeta \in \mathbb{C}.$$

Now let X be a random variable with distribution μ, and take $\zeta_K = iKe^{\frac{\pi}{2n}}$, where $K > 0$. Then, on the one hand,

$$\big|\mathbb{E}\big[e^{\zeta_K X}\big]\big| = \exp\big(\Re P\big(Ke^{i\frac{\pi}{2n}}\big)\big) = \exp\big(-c_nK^{2n} + o(K^{2n})\big),$$

and, on the other hand,

$$\begin{aligned}\left|\mathbb{E}\left[e^{\zeta_K X}\right]\right| &\le \mathbb{E}\left[e^{\Re\zeta_K X}\right] = \mathbb{E}\left[e^{-K\sin\frac{\pi}{2n}X}\right]\\ &= e^{P(iK\sin\frac{\pi}{2n})} = \exp\Big((-1)^n c_n K^{2n}\sin^{2n}\frac{\pi}{2n} + o(K^{2n})\Big).\end{aligned}$$

Hence $|c_n| \le (-1)^n c_n \sin^{2n}\frac{\pi}{2n}$, which is possible only if $n = 1$. □

By combinining Corollary 2.2.2 with Theorem 1.1.10, one sees that the only non-negative definite functions which are the exponential of a polynomial are, apart from a multiplicative constant, characteristic functions of Gaussian random variables.

2.2.1 Cauchy's Equation

A function $f : \mathbb{R} \longrightarrow \mathbb{R}$ is *additive* if $f(x+y) = f(x) + f(y)$ for all $x, y \in \mathbb{R}$. Cauchy asked which additive functions are linear. The following lemma provides a way to answer Cauchy's question.

Lemma 2.2.3 *Let $\mu \in M_1(\mathbb{R})$ and α, $\beta \in (0,1)$ with $\alpha^2 + \beta^2 = 1$ be given. Then μ is a centered Gaussian if and only if $\hat\mu(\xi) = \hat\mu(\alpha\xi)\hat\mu(\beta\xi)$.*

Proof. The necessity part is trivial.

Now assume the sufficiency condition. Using induction, one sees that, for any $n \ge 0$,

$$\hat\mu(\xi) = \prod_{m=0}^{n} \hat\mu(\alpha^m\beta^{n-m}\xi)^{\binom{n}{m}}.$$

In particular, $\hat\mu$ never vanishes and so there is a unique continuous choice of $\log\hat\mu(\xi)$ which vanishes at $\xi = 0$.

First, suppose the $t \equiv \int x^2\,\mu(dx) < \infty$. Then

$$\int x\,\mu(dx) = \iint \big(\alpha x + \beta y\big)\,\mu(dx)\mu(dy) = (\alpha+\beta)\int x\,\mu(dx),$$

which, since $\alpha+\beta > 1$, means that $\int x\,\mu(dx) = 0$. In addition, for each $\xi \in \mathbb{R}$,

$$\begin{aligned}\log\hat\mu(\xi) &= \sum_{m=0}^{n}\binom{n}{m}\log\big(\hat\mu(\alpha^m\beta^{n-m}\xi)\big)\\ &= -\sum_{m=0}^{n}\binom{n}{m}\left(\frac{t\alpha^{2m}\beta^{2(n-m)}\xi^2}{2} + o\big(\alpha^{2m}\beta^{2(n-m)}\big)\right) \longrightarrow -\frac{t\xi^2}{2}.\end{aligned}$$

as $n \to \infty$. Thus, it suffices to show that $\int x^2\,\mu(dx) < \infty$.

Next suppose that μ is symmetric. Then $\hat\mu > 0$ everywhere and

$$\log \hat{\mu}(1) = \sum_{m=0}^{n} \binom{n}{m} \log\left(\int \cos(\alpha^m \beta^{n-m} x)\, \mu(dx)\right).$$

Hence, since $1 - t \le -\log t$ for $t \in (0,1]$,

$$-\log \hat{\mu}(1) \ge \sum_{m=0}^{n} \binom{n}{m} \int \bigl(1 - \cos(\alpha^m \beta^{n-m} x)\bigr)\, \mu(dx).$$

Because

$$0 \le \sum_{m=0}^{n} \binom{n}{m} \bigl(1 - \cos(\alpha^m \beta^{n-m} x)\bigr) \longrightarrow \frac{x^2}{2},$$

Fatou's lemma implies that $\int x^2\, \mu(dx) \le -2 \log\bigl(\hat{\mu}(1)\bigr) < \infty$.

Finally, for general μ's, one can use either the Theorem 2.2.1 or observe that the symmetrization of μ will again satisfy the hypothesis and therefore have a second moment. □

Say that a function $f : \mathbb{R} \longrightarrow \mathbb{R}$ is *a.e. additive* if $f(x+y) = f(x) + f(y)$ for $\lambda_{\mathbb{R}^2}$-almost every $(x,y) \in \mathbb{R}^2$.

Theorem 2.2.4 *Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be a Borel measurable function. If f is additive, then $f(x) = f(1)x$ for all $x \in \mathbb{R}$. If f is a.e. additive, then there is an $a \in \mathbb{R}$ such that $f(x) = ax$ for $\lambda_{\mathbb{R}}$-almost every $x \in \mathbb{R}$.*

Proof. Assume that f is additive. It is easy to check that $f(qx) = qf(x)$ for all $q \in \mathbb{Q}$ and $x \in \mathbb{R}$. Thus, if f is continuous, then $f(x) = f(1)x$. Now assume that f is locally $\lambda_{\mathbb{R}}$-integrable, and choose $\rho \in C_c^\infty(\mathbb{R};\mathbb{R})$ with total integral 1. Then $f * \rho$ is smooth and $f * \rho(x) = f(x) + \int f(-y)\rho(y)\, dy$. Thus f is smooth and therefore it is linear. In general, let μ be the distribution of f under the standard Gaussian measure $\gamma_{0,1}$, and set $q_1 = \frac{3}{5}$ and $q_2 = \frac{4}{5}$. Then, because $q_1^2 + q_2^2 = 1$,

$$\begin{aligned}\hat{\mu}(q_1\xi)\hat{\mu}(q_2\xi) &= \iint e^{\imath\xi(q_1 f(x) + q_2 f(y))}\, \gamma_{0,1}(dx)\gamma_{0,1}(dy) \\ &= \iint e^{\imath\xi f(q_1 x + q_2 y)}\, \gamma_{0,1}(dx)\gamma_{0,1}(dy) = \hat{\mu}(\xi),\end{aligned}$$

and therefore, by Lemma 2.2.3, μ is a centered Gaussian measure. In particular,

$$\int f(x)^2\, \gamma_{0,1}(dx) = \int x^2\, \mu(dx) < \infty,$$

and so f is locally integrable and therefore linear.

Now assume that f is a.e. additive. The first step is to show that, for all $q_1, q_2 \in \mathbb{Q}^+$, $f(q_1 x + q_2 y) = q_1 f(x) + q_2 f(y)$ (a.e.,$\lambda_{\mathbb{R}^2}$). To this end, let $n \ge 1$, and assume that $f(nx + y) = nf(x) + f(y)$ (a.e.,$\lambda_{\mathbb{R}^2}$). Then, because the distribution of $(x, x+y)$ under $\lambda_{\mathbb{R}^2}$ is equivalent to that of (x,y),

$$\begin{aligned} f\big((n+1)x+y\big) &= f\big(nx+(x+y)\big) = nf(x)+f(x+y) \\ &= nf(x)+f(x)+f(y) = (n+1)f(x)+f(y) \end{aligned}$$

for $\lambda_{\mathbb{R}^2}$-a.e. $(x,y) \in \mathbb{R}^2$. Hence, by induction, for all $n \geq 1$, $f(nx+y) = nf(x)+f(y)$ for $\lambda_{\mathbb{R}^2}$-a.e. $(x,y) \in \mathbb{R}^2$. At the same time, because the distribution of (nx,y) under $\lambda_{\mathbb{R}^2}$ is equivalent to that of (x,y),

$$f(nx)+f(y) = f(nx+y) = nf(x)+f(y) \quad \text{for } \lambda_{\mathbb{R}^2}\text{-a.e. } (x,y) \in \mathbb{R}^2,$$

and so, by Fubini's Theorem, $f(nx) = nf(x)$ (a.e.,$\lambda_{\mathbb{R}}$). Similarly, because the $\lambda_{\mathbb{R}^2}$-distribution of $\big(\frac{x}{n},y\big)$ is equivalent of that of (x,y), for all m, $n \in \mathbb{Z}^+$, $f\big(\frac{m}{n}x\big) = \frac{m}{n}f(x)$ (a.e.,$\lambda_{\mathbb{R}}$). Finally, given q_1, $q_2 \in \mathbb{Q}^+$, the $\lambda_{\mathbb{R}^2}$-distribution of (q_1x, q_2y) is equivalent to that of (x,y), and therefore

$$f(q_1x+q_2y) = f(q_1x)+f(q_2y) = q_1f(x)+q_2f(y) \quad \text{for } \lambda_{\mathbb{R}^2}\text{-a.e. } (x,y) \in \mathbb{R}^2.$$

To complete the proof, again take $q_1 = \frac{3}{5}$ and $q_2 = \frac{4}{5}$, and again consider the distribution μ of f under $\gamma_{0,1}$. Then, just as before, $\hat{\mu}(\xi) = \hat{\mu}(q_1\xi)\hat{\mu}(q_2\xi)$, and therefore f is locally $\lambda_{\mathbb{R}}$-integrable. In addition, if $\rho \in C_{\mathrm{c}}^{\infty}(\mathbb{R};\mathbb{R})$ has total integral 1, then

$$\tilde{f}(x) \equiv f*\rho(x) - \int f(-y)\rho(y)\,dy = f(x) \quad \text{for } \lambda_{\mathbb{R}}\text{-a.e. } x \in \mathbb{R}.$$

In particular, for each y, $\tilde{f}(x+y) = f(x+y)$ for $\lambda_{\mathbb{R}}$-a.e. $x \in \mathbb{R}$, and from this and Fubini's Theorem, it follows that $\tilde{f}$ is a smooth additive function. Hence, $\tilde{f}$ is linear, and so $f(x) = \tilde{f}(1)x$ for $\lambda_{\mathbb{R}}$-a.e. $x \in \mathbb{R}$. □

Extentions of these results and ideas can be found in [2].

2.3 Gaussian Spectral Properties

The first goal here is to prove

$$\int \big(\varphi - \langle \varphi, \gamma_{0,1}\rangle\big)^2\,d\gamma_{0,1} \leq \|\varphi'\|^2_{L^2(\gamma_{0,1};\mathbb{R})} \tag{2.3.1}$$

for $\varphi \in C^1(\mathbb{R};\mathbb{R})$. A closely related inequality was used to great effect by H. Poincaré, and such inequalities have ever since been called a *Poincaré inequality*.

In the proof of (2.3.1), we will make use of the function

$$p(t,x,y) = \big(2\pi(1-e^{-t})\big)^{-\frac{1}{2}} \exp\left(-\frac{(y-e^{-\frac{t}{2}}x)^2}{2(1-e^{-t})}\right) \text{ for } (t,x,y) \in (0,\infty)\times\mathbb{R}\times\mathbb{R}.$$

Observe that

$$\int p(t,x,y)\,dy = 1 \text{ and } p(s+t,x,y) = \int p(s,x,\xi)p(t,\xi,y)\,d\xi$$

and

$$\partial_t p(t,x,y) = \tfrac{1}{2}\big(\partial_s^2 - x\partial_x\big)p(t,x,y).$$

Hence, if P_t is the operator on $C_{\mathrm{b}}(\mathbb{R};\mathbb{C})$ given by

$$P_t\varphi(x) = \int \varphi(y)p(t,x,y)\,dy, \tag{2.3.2}$$

then it is a contraction on $C_{\mathrm{b}}(\mathbb{R};\mathbb{C})$ into itself, and $P_{s+t} = P_t \circ P_s$. In addition,

$$\partial_t P_t\varphi = LP_t\varphi \text{ where } \mathcal{L} = \tfrac{1}{2}\big(\partial_x^2 - x\partial_x\big), \tag{2.3.3}$$

and so, if $\varphi \in C^2_{\mathrm{b}}(\mathbb{R};\mathbb{C})$,

$$\lim_{t\searrow 0} \frac{P_t\varphi - \varphi}{t} = \mathcal{L}\varphi, \tag{2.3.4}$$

where the convergence is uniform on compacts. Probabilists call $\mathcal{L}$ the *Ornstein–Uhlenbeck operator*.

Note that

$$p(t,x,y)\gamma_{0,1}(dx) = \big(2\pi(1-e^{-t})^{\frac{1}{2}}\big)^{-1} \exp\left(-\frac{x^2 - 2e^{-\frac{t}{2}}xy + y^2}{2(1-e^{-t})}\right)\lambda_{\mathbb{R}}(dx),$$

and so

$$\big(\varphi, P_t\psi\big)_{L^2(\gamma_{0,1};\mathbb{R})} = \big(P_t\psi,\varphi\big)_{L^2(\gamma_{0,1};\mathbb{R})}.$$

In particular,

$$\int P_t\varphi\,d\gamma_{0,1} = \big(\mathbf{1}, P_t\varphi\big)_{L^2(\gamma_{0,1};\mathbb{R})} = \big(P_t\mathbf{1},\varphi\big)_{L^2(\gamma_{0,1};\mathbb{R})} = \int \varphi\,d\gamma_{0,1}.$$

Next, by Jensen's inequality, $|P_t\varphi|^p \le P_t|\varphi|^p$, and therefore

$$\|P_t\varphi\|_{L^p(\gamma_{0,1};\mathbb{R})} \le \|\varphi\|_{L^p(\gamma_{0,1};\mathbb{R})}$$

for $p \in [1,\infty]$. Hence, P_t extends to $L^2(\gamma_{0,1};\mathbb{R})$ as a self-adjoint contraction, and so $\{P_t : t > 0\}$ can be viewed as a strongly continuous semigroup of self-adjoint contractions on $L^2(\gamma_{0,1};\mathbb{R})$. Finally, note that

$$\lim_{t\to\infty} P_t\varphi(x) = \int \varphi\,d\gamma_{0,1} \text{ and } \lim_{t\searrow 0} P_t\varphi = \varphi,$$

first uniformly on compacts for $\varphi \in C_{\mathrm{b}}(\mathbb{R};\mathbb{C})$ and then in $L^p(\gamma_{0,1};\mathbb{R})$ for $p \in [1,\infty)$ and $\varphi \in L^p(\gamma_{0,1};\mathbb{R})$.

The estimate (2.3.1) is equivalent to

$$\|\varphi\|^2_{L^2(\gamma_{0,1};\mathbb{R})} \le \|\varphi'\|^2_{L^2(\gamma_{0,1};\mathbb{R})} + \langle \varphi, \gamma_{0,1}\rangle^2,$$

and it suffices to check it when $\varphi \in \mathscr{S}(\mathbb{R};\mathbb{R})$. Thus let $\varphi \in \mathscr{S}(\mathbb{R};\mathbb{R})$ be given, and observe that

$$(P_t\varphi)'(x) = \partial_x \int \varphi(y + e^{-\frac{t}{2}}x)p(t,0,y)\,dy = e^{-\frac{t}{2}} P_t\varphi'(x).$$

Using (2.3.3), integration by parts, and the preceding, conclude that

$$\begin{aligned}\frac{d}{dt}\big(\varphi, P_t\varphi\big)_{L^2(\gamma_{0,1};\mathbb{R})} &= \big(\varphi, \mathcal{L}P_t\varphi\big)_{L^2(\gamma_{0,1};\mathbb{R})} \\ &= -\tfrac{1}{2}\big(\varphi', (P_t\varphi)'\big)_{L^2(\gamma_{0,1};\mathbb{R})} = -\frac{e^{-\frac{t}{2}}}{2}\big(\varphi', P_t\varphi'\big)_{L^2(\gamma_{0,1};\mathbb{R})}.\end{aligned}$$

Since

$$\big|\big(\varphi', P_t\varphi'\big)_{L^2(\gamma_{0,1};\mathbb{R})}\big| \le \|\varphi'\|_{L^2(\gamma_{0,1};\mathbb{R})}\|P_t\varphi'\|_{L^2(\gamma_{0,1};\mathbb{R})} \le \|\varphi'\|^2_{L^2(\gamma_{0,1};\mathbb{R})},$$

$$-\frac{d}{dt}\big(\varphi, P_t\varphi\big)_{L^2(\gamma_{0,1};\mathbb{R})} \le \frac{e^{-\frac{t}{2}}}{2}\|\varphi'\|^2_{L^2(\gamma_{0,1};\mathbb{R})}.$$

Integrating the preceding over $t \in (0,\infty)$, one arrives at (2.3.1).

An important application of (2.3.1) is the following ergodic property of $\{P_t : t > 0\}$. Namely, suppose that $\langle \varphi, \gamma_{0,1}\rangle = 0$. Then

$$\frac{d}{dt}\|P_t\varphi\|^2_{L^2(\gamma_{0,1};\mathbb{R})} = 2\big(P_t\varphi, \mathcal{L}P_t\varphi\big)_{\gamma_{0,1}} = -\|(P_t\varphi)'\|^2_{L^2(\gamma_{0,1};\mathbb{R})} \le -\|P_t\varphi\|^2_{L^2(\gamma_{0,1};\mathbb{R})},$$

where the final inequality comes from (2.3.1) and the fact that $\langle P_t\varphi, \gamma_{0,1}\rangle = 0$. Hence, $\|P_t\varphi\|^2_{L^2(\gamma_{0,1};\mathbb{R})} \le e^{-t}\|\varphi\|^2_{L^2(\gamma_{0,1};\mathbb{R})}$, and so, even if $\langle \varphi, \gamma_{0,1}\rangle \ne 0$,

$$\begin{aligned}\|P_t\varphi - \langle \varphi, \gamma_{0,1}\rangle\|^2_{L^2(\gamma_{0,1};\mathbb{R})} &\le e^{-t}\|\varphi - \langle \varphi, \gamma_{0,1}\rangle\|^2_{L^2(\gamma_{0,1};\mathbb{R})} \\ &\le e^{-t}\|\varphi\|^2_{L^2(\gamma_{0,1};\mathbb{R})}.\end{aligned} \tag{2.3.5}$$

A feature of Poincaré inequalities like (2.3.1) is their behavior when one takes products. Namely, suppose that $\mu_j \in M_1(\mathbb{R}^{N_j})$, $j \in \{1,2\}$, and that

$$\|\varphi\|^2_{L^2(\mu_j;\mathbb{R})} \le c\||\nabla_j\varphi|\|^2_{L^2(\mu_j;\mathbb{R})} + \langle \varphi, \mu_j\rangle^2,$$

where ∇_j is the gradient for functions on $\mathbb{R}^{N_j}$. Then, if $N = N_1 + N_2$ and $\mu = \mu_1 \times \mu_2$ on $\mathbb{R}^N$, an application of Fubini's Theorem shows that

$$\|\varphi\|_{L^2(\mu)}^2 \le c\||\nabla\varphi|\|_{L^2(\mu)}^2 + \langle\varphi,\mu\rangle^2,$$

where ∇ is the gradient for functions on $\mathbb{R}^N$. As a consequence, we now see that (2.3.1) implies that, for any $N \in \mathbb{Z}^+$,

$$\|\varphi - \langle\varphi,\gamma_{0,1}^N\rangle\|_{L^2(\gamma_{0,1}^N;\mathbb{R})}^2 \le \||\nabla\varphi|\|_{L^2(\gamma_{0,1}^N;\mathbb{R})}^2. \tag{2.3.6}$$

Define

$$P_t^{(N)}\varphi(x) = \int_{\mathbb{R}^N} \varphi(y)\prod_{j=1}^N p(t,x_j,y_j)\lambda_{\mathbb{R}^N}(dy)$$

for $\varphi \in C_{\mathrm{b}}(\mathbb{R}^N;\mathbb{C})$. Then it is easy to check from (2.3.3) that $\partial_t P_t^{(N)}\varphi = \mathcal{L}^{(N)}P_t^{(N)}\varphi$, where $\mathcal{L}^{(N)}\psi(x) = \frac{1}{2}\big(\Delta\psi(x) - \big(x,\nabla\psi(x)\big)_{\mathbb{R}^N}\big)$. Thus, proceeding as we did in the derivation of (2.3.5), one finds that

$$\begin{aligned}\|P_t^{(N)}\varphi - \langle\varphi,\gamma_{0,1}^N\rangle\|_{L^2(\gamma_{0,1}^N)}^2 \\ \le e^{-t}\|\varphi - \langle\varphi,\gamma_{0,1}^N\rangle\|_{L^2(\gamma_{0,1}^N)}^2 \le e^{-t}\|\varphi\|_{L^2(\gamma_{0,1}^N)}^2.\end{aligned} \tag{2.3.7}$$

2.3.1 A Logarithmic Sobolev Inequality

It turns out that one can prove a slightly stronger inequality than (2.3.1). Namely,

$$\int \varphi^2 \log\frac{\varphi^2}{\|\varphi\|_{L^2(\gamma_{0,1};\mathbb{R})}^2}\,d\gamma_{0,1} \le 2\|\varphi'\|_{L^2(\gamma_{0,1};\mathbb{R})}^2. \tag{2.3.8}$$

This inequality was proved first by L. Gross [5] who called it a *logarithmic Sobolev inequality*.

Obviously, (2.3.8) looks a lot like (2.3.1), especially when one rewrites it as

$$\int \varphi^2\log\varphi^2\,d\gamma_{0,1} \le 2\|\varphi'\|_{L^2(\gamma_{0,1};\mathbb{R})}^2 + \|\varphi\|_{L^2(\gamma_{0,1};\mathbb{R})}^2\log\|\varphi\|_{L^2(\gamma_{0,1};\mathbb{R})}^2.$$

However, as Gross showed, it has regularity consequences similar to, albeit weaker than, a classical Sobolev inequality. To prove (2.3.8), let φ be a strictly positive element of $\mathscr{S}(\mathbb{R};\mathbb{R})$, and set (cf. (2.3.2)) $\varphi_t = P_t\varphi$. Then, since $\langle\partial_t\varphi_t,\gamma_{0,1}\rangle = \partial_t\langle\varphi_t,\gamma_{0,1}\rangle = 0$,

$$\begin{aligned}\frac{d}{dt}\big\langle\varphi_t\log\varphi_t,\gamma_{0,1}\big\rangle &= \big\langle\mathcal{L}\varphi_t\log\varphi_t,\gamma_{0,1}\big\rangle \\ &= -\frac{1}{2}\left\langle\frac{(\varphi_t')^2}{\varphi_t},\gamma_{0,1}\right\rangle = -\frac{e^{-t}}{2}\left\langle\frac{(P_t\varphi')^2}{P_t\varphi},\gamma_{0,1}\right\rangle.\end{aligned}$$

Observe that, by Schwarz's inequality,

$$(P_t\varphi')^2 = \left(P_t\left(\frac{\varphi'}{\varphi^{\frac12}}\varphi^{\frac12}\right)\right)^2 \le P_t\left(\frac{(\varphi')^2}{\varphi}\right)P_t\varphi,$$

and therefore

$$\frac{d}{dt}\big\langle\varphi_t\log\varphi_t,\gamma_{0,1}\big\rangle \ge -\frac{e^{-t}}{2}\left\langle P_t\left(\frac{(\varphi')^2}{\varphi}\right),\gamma_{0,1}\right\rangle = -\frac{e^{-t}}{2}\left\langle\frac{(\varphi')^2}{\varphi},\gamma_{0,1}\right\rangle.$$

After integrating over $t\in(0,\infty)$, one has that

$$\big\langle\varphi\log\varphi,\gamma_{0,1}\big\rangle \le \frac12\left\langle\frac{(\varphi')^2}{\varphi},\gamma_{0,1}\right\rangle + \langle\varphi,\gamma_{0,1}\rangle\log\langle\varphi,\gamma_{0,1}\rangle.$$

Knowing this inequality for strictly positive φ's in $\mathscr{S}(\mathbb{R};\mathbb{R})$, it is obvious that it extends to non-negative φ's in $C^1(\mathbb{R};\mathbb{R})$. Finally, given a $\varphi\in C^1(\mathbb{R};\mathbb{R})$, apply the inequality to φ^2 and thereby arrive at (2.3.8).

To see that (2.3.8) is optimal, suppose that it holds with $2\|\varphi'\|^2_{L^2(\gamma_{0,1};\mathbb{R})}$ replaced by $c\|\varphi'\|^2_{L^2(\gamma_{0,1};\mathbb{R})}$, and take $\varphi = e^{\alpha\frac{x^2}{4}}$ for $\alpha\in(0,1)$. Show that

$$\big\langle\varphi^2\log\varphi^2\gamma_{0,1}\big\rangle = \frac{\alpha}{2(1-\alpha)^{\frac32}},\ \langle\varphi^2,\gamma_{0,1}\rangle = (1-\alpha)^{-\frac12},\ \|\varphi'\|^2_{L^2(\gamma_{0,1};\mathbb{R})} = \frac{\alpha^2}{4(1-\alpha)^{\frac32}},$$

and conclude that $\alpha \le \frac{c}{2}\alpha^2 - \frac{(1-\alpha)}{2}\log(1-\alpha)$. Now let $\alpha\nearrow 1$ to see that $c\ge 2$.

The regularity result alluded to above was discovered by E. Nelson [8] and is called *hypercontraction*. It states that the semigroup $\{P_t : t>0\}$ is smoothing in the sense that

$$\begin{gathered}\|P_t\varphi\|_{L^{q(t)}(\gamma_{0,1};\mathbb{R})} \le \|\varphi\|_{L^p(\gamma_{0,1};\mathbb{R})}\\ \text{when } p\in(1,\infty) \text{ and } q(t) = 1+(p-1)e^t\end{gathered} \tag{2.3.9}$$

Gross's proof of this from (2.3.8) is the following. Let a uniformly positive $\varphi\in C^1_{\mathrm{b}}(\mathbb{R};\mathbb{R})$ be given, and set $\varphi_t = P_t\varphi$ and $F(t) = \|\varphi_t\|_{L^{q(t)}(\gamma_{0,1})}$. Then, with $\dot q = \frac{dq}{dt}$, one sees that

$$\begin{aligned}\frac{dF}{dt} &= -\frac{\dot q}{q^2}F\log F^q + \frac{\dot q F^{1-q}}{q}\langle\varphi_t^q\log\varphi_t,\gamma_{0,1}\rangle + F^{1-q}\langle\varphi_t^{q-1}L\varphi_t,\gamma_{0,1}\rangle\\ &= \frac{F^{1-q}}{q^2}\left[\dot q\left\langle\varphi_t^q\log\frac{\varphi_t^q}{\langle\varphi_t^q,\gamma_{0,1}\rangle},\gamma_{0,1}\right\rangle - \frac{q^2(q-1)}{2}\langle\varphi_t^{q-2}(\varphi_t')^2,\gamma_{0,1}\rangle\right].\end{aligned}$$

Since $\dot q = q-2$ and $\varphi_t^{q-1}(\varphi_t')^2 = \frac{4}{q^2}\big(\psi_t'\big)^2$ where $\psi_t = \varphi_t^{\frac q2}$, the bracketed expression equals

$$(q-1)\left[\left\langle \psi_t^2 \log \frac{\psi_t^2}{\|\psi_t\|^2_{L^2(\gamma_{0,1};\mathbb{R})}}, \gamma_{0,1}\right\rangle - 2\|\psi_t\|^2_{L^2(\gamma_{0,1};\mathbb{R})}\right],$$

which, by (2.3.8) is less than or equal to 0. Hence F is non-increasing, and so (2.3.9) holds.

Notice that, because $P_{s+t} = P_t \circ P_s$ and $q(s+t) = 1 + \big(q(s)-1\big)e^t$, (2.3.9) implies that $\|P_t\varphi\|_{L^{q(t)}(\gamma_{0,1})}$ is a non-decreasing function of $t \geq 0$, and therefore Gross's computation shows that (2.3.9) is equivalent to (2.3.8). In particular, $q(t)$ cannot be replaced by $1+(p-1)e^{ct}$ for any $c > 1$. In fact, by direct computation of $\|\varphi\|_{L^p(\gamma_{0,1};\mathbb{R})}$ and $\|\varphi\|_{L^q(\gamma_{0,1})}$ when $\varphi = e^{\alpha x}$, one can show that, for any $q > q(t)$ there is a $\varphi \in L^p(\gamma_{0,1};\mathbb{R})$ for which $\|P_t\varphi\|_{L^q(\gamma_{0,1};\mathbb{R})} = \infty$.

Just as was the case for the Poincaré inequality, one can use Fubini's theorem to check that (2.3.8) implies

$$\int_{\mathbb{R}^N} \varphi^2 \log \frac{\varphi^2}{\|\varphi\|^2_{L^2(\gamma^N_{0,1};\mathbb{R})}} d\gamma^N_{0,1} \leq 2\||\nabla\varphi|\|^2_{L^2(\gamma^N_{0,1};\mathbb{R})} \tag{2.3.10}$$

for $\varphi \in C^2(\mathbb{R}^N;\mathbb{R})$. Thus, by Gross's argument, one knows that

$$\begin{gathered}\|P^{(N)}_t\varphi\|_{L^{q(t)}(\gamma^N_{0,1};\mathbb{R})} \leq \|\varphi\|_{L^p(\gamma^N_{0,1};\mathbb{R})} \\ \text{when } p \in (1,\infty) \text{ and } q(t) = 1+(p-1)e^t.\end{gathered} \tag{2.3.11}$$

Because his goal was to construct quantum fields on infinite dimensional spaces, Nelson's interest in these matters was their dimension independence. Given a Borel measurable function $V : \mathbb{R}^N \longrightarrow \mathbb{R}$ which is bounded either below or above, a key step in his program was to prove a dimension independent estimate for

$$\begin{aligned}\Lambda(V) = \sup\big\{&\langle V\varphi^2, \gamma^N_{0,1}\rangle^2 - \tfrac{1}{2}\|\nabla\varphi\|^2_{L^2(\gamma^N_{0,1};\mathbb{R}^N)} \\ &: \varphi \in C^1(\mathbb{R}^N;\mathbb{R}) \text{ with } \|\varphi\|_{L^2(\gamma_{0,1}*N;\mathbb{R})} = 1\big\},\end{aligned} \tag{2.3.12}$$

which is the upper bound for the spectrum of the Schrödinger operator

$$L^{(N)} + V = \frac{1}{2}\big(\Delta - (x,\nabla)_{\mathbb{R}^N}\big) + V$$

on $L^2(\gamma^N_{0,1};\mathbb{R})$, and what he showed is that

$$\Lambda(V) \leq \frac{1}{4}\log\big\langle e^{4V}, \gamma^N_{0,1}\big\rangle. \tag{2.3.13}$$

Although Nelson based his proof of (2.3.13) on (2.3.11), it can be seen as a direct consequence of (2.3.10) combined with the following interesting variational formula. Given a pair of probability measures μ and ν on a measurable space $(E, \mathcal{F})$, define the *relative entropy* $H(\nu|\mu)$ of ν with respect to μ by

$$H(\nu|\mu) = \begin{cases} \int f \log f \, d\mu & \text{if } \nu \ll \mu \text{ and } f = \frac{d\nu}{d\mu} \\ \infty & \text{if } \nu \not\ll \mu. \end{cases}$$

Because $t \log t$ is a convex function of $t \in [0, \infty)$, Jensen's inequality implies that, for each μ, $H(\nu|\mu)$ is a non-negative, convex function of ν and that it is 0 when $\mu = \mu$. Less obvious is the fact that

$$\|\nu - \mu\|_{\text{var}}^2 \leq 2H(\nu|\mu), \tag{2.3.14}$$

where $\|\cdot\|_{\text{var}}$ is the variation norm. One need only check this when $H(\nu|\mu) < \infty$, in which case $\|\nu - \mu\|_{\text{var}} = \|f - 1\|_{L^1(\mu;\mathbb{R})}$ where $f = \frac{d\nu}{d\mu}$. At this point one needs the inequality

$$3(t-1)^2 \leq (4 + 2t)\bigl(t \log t - t + 1\bigr) \text{ for } t \geq 0.$$

To prove this inequality, let $g(t)$ be the right hand side minus the left hand side, and check that g is a convex function for which $g(1) = g'(1) = 0$. Now apply this inequality to see that

$$\begin{aligned} 3\|f-1\|_{L^1(\mu)}^2 &\leq \|(4+2f)^{\frac{1}{2}}(f \log f - f + 1)^{\frac{1}{2}}\|_{L^1(\mu;\mathbb{R})}^2 \\ &\leq \langle 4 + 2f, \mu\rangle \langle f \log f - f + 1, \mu\rangle = 6H(\nu|\mu). \end{aligned}$$

Lemma 2.3.1 *Let $B(E;\mathbb{R})$ denote the space of bounded, $\mathcal{F}$-measurable $\mathbb{R}$-valued functions on E, and, for any probability measures μ and ν on $(E, \mathcal{F})$, set*

$$I_\mu(\nu) = \sup\bigl\{\langle \varphi, \nu\rangle - \log\langle e^\varphi, \mu\rangle : \varphi \in B(E;\mathbb{R})\bigr\}.$$

Then $H(\nu|\mu) = I_\mu(\nu)$.

Before proving Lemma 2.3.1, notice that (2.3.13) is a relatively easy consequence of (2.3.10) combined with Lemma 2.3.1 applied to $\mu = \gamma_{0,1}^N$. Namely, given $\varphi \in C^1(\mathbb{R};\mathbb{R})$ with $\|\varphi\|_{L^2(\gamma_{0,1}^N;\mathbb{R})} = 1$, set $f = \varphi^2$ and take $d\nu = f d\mu$. Then $H(\nu|\gamma_{0,1}^N) = \bigl\langle \varphi^2 \log \varphi^2, \gamma_{0,1}^N \bigr\rangle \leq 2\||\nabla\varphi|\|_{L^2(\gamma_{0,1}^N;\mathbb{R})}^2$, and so, if V is bounded, then

$$2\||\nabla\varphi|\|_{L^2(\gamma_{0,1}^N;\mathbb{R})}^2 \geq 4\bigl\langle V\varphi^2, \gamma_{0,1}^N\bigr\rangle - \log\bigl\langle e^{4V}, \gamma_{0,1}^N\bigr\rangle,$$

which means that

$$\tfrac{1}{4}\log\bigl\langle e^{4V}, \gamma_{0,1}^N\bigr\rangle \geq \bigl\langle V\varphi^2, \gamma_{0,1}^N\bigr\rangle - \tfrac{1}{2}\||\nabla\varphi|\|_{L^2(\gamma_{0,1}^N;\mathbb{R})}^2,$$

which proves (2.3.13) for bounded V's. Finally, knowing it for bounded V's, it is easy to extend it to V's which are bounded either above or below.

Proof of Lemma 2.3.1. First note that $\nu \rightsquigarrow I_\mu(\nu)$ is non-negative and convex. Further, because, by Jensen's inequality, $\langle\varphi,\mu\rangle \le \log\langle e^\varphi,\mu\rangle$, $I_\mu(\mu)=0$.

Next, given $\theta\in(0,1)$, set $\nu_\theta=(1-\theta)\nu+\theta\mu$. Then, because $H(\mu|\mu)=0$, convexity implies $H(\nu_\theta|\mu)\le(1-\theta)H(\nu|\mu)$. On the other hand,

$$\nu \not\ll \mu \implies \nu_\theta \not\ll \mu \implies H(\nu_\theta|\mu)=\infty,$$

and if $d\nu=f\,d\mu$, then $d\nu_\theta=f_\theta\,d\mu$, where $f_\theta=(1-\theta)f+\theta$, and therefore, since log is non-decreasing and concave,

$$\begin{aligned}H(\nu_\theta|\mu)&=\langle f_\theta\log f_\theta,\mu\rangle\\&=\theta\langle\log f_\theta,\mu\rangle+(1-\theta)\langle f\log f_\theta,\mu\rangle\ge\theta\log\theta+(1-\theta)^2H(\nu|\mu).\end{aligned}$$

Hence $H(\nu|\mu)=\lim_{\theta\searrow0}H(\nu_\theta|\mu)$.

We now show that $I_\nu(\mu)\le H(\nu|\mu)$, and clearly we need do so only when $H(\nu|\mu)<\infty$ and therefore $d\nu=f\,d\mu$. If f is everywhere positive, then, by Jensen's inequality,

$$\exp\bigl(\langle\varphi,\nu\rangle-H(\nu|\mu)\bigr)=\exp\bigl(\langle\varphi-\log f,\nu\rangle\bigr)\le\left\langle\frac{e^\varphi}{f},\nu\right\rangle=\langle e^\varphi,\mu\rangle,$$

and so $\langle\varphi,\nu\rangle-\log\bigl\langle e^\varphi,\mu\bigr\rangle\le H(\nu|\mu)$. If f can vanish, consider ν_θ, and, after letting $\theta\searrow0$, conclude that this inequality continues to hold, and therefore that $I_\mu(\nu)\le H(\nu|\mu)$.

In proving that $H(\nu|\mu)\le I_\mu(\nu)$, we will assume $I_\mu(\nu)<\infty$. The first step is to show that $\nu\ll\mu$. Thus, suppose $\Gamma\in\mathcal{F}$ with $\mu(\Gamma)=0$, and, given $r>0$, set $\varphi_r=r\mathbf{1}_\Gamma$. Then

$$I_\mu(\nu)\ge\langle\varphi_r,\nu\rangle-\log\bigl\langle e^{\varphi_r},\mu\bigr\rangle=r\nu(\Gamma),$$

for all $r>0$, and so $\nu(\Gamma)=0$. Now assume that $d\nu=f\,d\mu$. If f is bounded and uniformly positive, then $\log f\in B(E;\mathbb{R})$, and so

$$I_\mu(\nu)\ge\langle\log f,\nu\rangle-\log\bigl\langle f,\mu\bigr\rangle=H(\nu|\mu).$$

If f is uniformly positive and not necessarily bounded, set $f_n=f\wedge n$. Then, since $\lim_{n\to\infty}\langle f_n,\mu\rangle=1$,

$$H(\nu|\mu)=\lim_{n\to\infty}\langle\log f_n,\nu\rangle=\lim_{n\to\infty}\bigl(\langle\log f_n,\nu\rangle-\log\langle f_n,\mu\rangle\bigr)\le I_\mu(\nu)$$

Finally if f can vanish,

$$H(\nu_\theta|\mu)\le I_\mu(\nu_\theta)\le(1-\theta)I_\mu(\nu),$$

and so $H(\nu|\mu)\le I_\mu(\nu)$. $\square$

2.3.2 Hermite Polynomials

For $n \geq 0$, set $H_n(x) = (-1)^n e^{\frac{x^2}{2}} \partial_x^n e^{-\frac{x^2}{2}}$. Then $\partial H_n(x) = xH_n(x) - H_{n+1}(x)$, and so $H_{n+1} = a_+ H_n$, where $a_+ = x - \partial_x$ is the *raising operator*. Proceeding by induction, one sees that H_n is an nth order polynomial, known as the nth unnormalized *Hermite polynomial*, for which 1 is the coefficient of x^n and $H_n(-x) = (-1)^n H_n(x)$. Next, observe

$$\big(a_+\varphi, \psi\big)_{L^2(\gamma_{0,1};\mathbb{R})} = \big(\varphi, \partial\psi\big)_{L^2(\gamma_{0,1};\mathbb{R})} \tag{2.3.15}$$

for $\varphi, \psi \in C^1(\mathbb{R};\mathbb{C})$ whose derivatives have at most exponential growth. Hence, if $m \leq n$, then

$$\big(H_n, H_m\big)_{L^2(\gamma_{0,1};\mathbb{R})} = \big(a_+^n H_0, H_m\big)_{L^2(\gamma_{0,1};\mathbb{R})} = \big(H_0, \partial^n H_m\big)_{L^2(\gamma_{0,1};\mathbb{R})} = n!\delta_{m,n},$$

and therefore $\|H_n\|_{L^2(\gamma_{0,1};\mathbb{R})} = (n!)^{\frac{1}{2}}$ and $\{H_n : n \geq 0\}$ is an orthogonal sequence in $L^2(\gamma_{0,1};\mathbb{R})$. In addition, for $n \geq 1$,

$$\big(\partial H_n, H_m\big)_{L^2(\gamma_{0,1};\mathbb{R})} = \big(H_n, a_+ H_m\big)_{L^2(\gamma_{0,1};\mathbb{R})} = \begin{cases} n! & \text{if } m = n-1 \\ 0 & \text{if } m \neq n-1. \end{cases}$$

Since ∂H_n is in the span of $\{H_m : 0 \leq m < n\}$, $\partial H_n = \sum_{m=0}^{n-1} \alpha_m H_m$ for some $\{\alpha_m : 0 \leq m < n\} \subseteq \mathbb{R}$. By taking inner products in $L^2(\gamma_{0,1};\mathbb{R})$, one finds that $\alpha_m = 0$ for $m < n-1$ and $(n-1)!\alpha_{n-1} = n!$ and thereby concludes that $\partial H_n = nH_{n-1}$.

Summarizing, our results thus far,

$$(x - \partial)H_n = H_{n+1},\ \partial H_n = nH_{n-1},\ \big(H_n, H_m\big)_{L^2(\gamma_{0,1};\mathbb{R})} = n!\delta_{m,n}. \tag{2.3.16}$$

In particular,

$$LH_n = -\frac{n}{2}H_n, \tag{2.3.17}$$

where $L = \frac{1}{2}(\partial^2 - x\partial)$ is the Ornstein–Uhlenbeck operator.

To show that $\{H_n : n \geq 0\}$ is a basis in $L^2(\gamma_{0,1};\mathbb{R})$, use the exponential Taylor's series to see that

$$e^{\zeta x - \frac{\zeta^2}{2}} = \sum_{n=0}^{\infty} \frac{\zeta^n}{n!} H_n(x) \text{ for } (\zeta, x) \in \mathbb{C} \times \mathbb{R}, \tag{2.3.18}$$

where the convergence is uniform on compact subsets. In addition, using the preceding computations, one sees that, as a function of x, the convergence is in $L^2(\gamma_{0,1};\mathbb{R})$ uniformly for ζ is compacts. Now suppose that $\varphi \in L^2(\gamma_{0,1};\mathbb{R})$ is orthogonal to $\{H_n : n \geq 0\}$, and set $\psi(x) = (2\pi)^{-\frac{1}{2}} e^{-\frac{x^2}{2}} \varphi(x)$. Then, $\psi \in L^1(\lambda_{\mathbb{R}};\mathbb{R}) \cap L^2(\lambda_{\mathbb{R}};\mathbb{R})$, and, by (2.3.18) with $\zeta = i\xi$,

$$\hat{\psi}(\xi) = e^{-\frac{\xi^2}{2}} \sum_{n=0} \frac{(i\xi)^n}{n!} \big(\varphi, H_n\big)_{L^2(\gamma_{0,1};\mathbb{R})} = 0,$$

and so ψ, and therefore φ, vanish (a.e.,$\lambda_{\mathbb{R}}$). Hence, we now know that if $\tilde{H}_n = \frac{H_n}{\sqrt{n!}}$ is the normalized nth Hermite polynomial, then $\{\tilde{H}_n : n \geq 0\}$ is an orthonormal basis in $L^2(\gamma_{0,1};\mathbb{R})$ consisting of eigenfunctions for the Ornstein–Uhlenbeck operator.

Define the operators $\{P_t : t > 0\}$ as in (2.3.2), and recall that $\partial_t P_\varphi = LP_t\varphi$. In addition, using the semigroup property, check that $LP_t\varphi = P_t L\varphi$ for $\varphi \in C^2(\mathbb{R};\mathbb{C})$ whose second derivative has at most polynomial growth. Thus, by (2.3.17), $\partial_t P_t H_n = -\frac{n}{2} P_t H_n$, and so $P_t H_n = e^{-\frac{nt}{2}} H_n$. Now let $\varphi \in L^2(\gamma_{0,1};\mathbb{R})$ be given. Then

$$\varphi = \sum_{n=0}^{\infty} \frac{(\varphi, H_n)_{L^2(\gamma_{0,1};\mathbb{R})}}{n!} H_n,$$

where the convergence is in $L^2(\gamma_{0,1};\mathbb{R})$, and so, since P_t is self-adjoint on $L^2(\gamma_{0,1};\mathbb{R})$,

$$P_t\varphi = \sum_{n=0}^{\infty} \frac{e^{-\frac{nt}{2}} (\varphi, H_n)_{L^2(\gamma_{0,1};\mathbb{R})}}{n!} H_n \text{ for } \varphi \in L^2(\gamma_{0,1};\mathbb{R}), \tag{2.3.19}$$

where the convergence is in $L^2(\gamma_{0,1};\mathbb{R})$.

2.3.3 Hermite Functions

For $n \geq 0$, the nth unnormalized *Hermite function* h_n is given by $h_n(x) = e^{-\frac{x^2}{2}} H_n(2^{\frac{1}{2}}x)$. Using (2.3.16), one can check that

$$\begin{gathered}(x - \partial)h_n = 2^{\frac{1}{2}} h_{n+1},\ (x + \partial)h_n = 2^{\frac{1}{2}} n h_{n-1};\\ \text{and } \big(h_n, h_m\big)_{L^2(\lambda_{\mathbb{R}};\mathbb{R})} = \pi^{\frac{1}{2}} n! \delta_{m,n},\end{gathered} \tag{2.3.20}$$

and therefore that

$$\mathcal{H}h_n = -\big(n + \tfrac{1}{2}\big)h_n \quad \text{where } \mathcal{H} = \tfrac{1}{2}\big(\partial^2 - x^2\big). \tag{2.3.21}$$

Mathematicians call H the *Hermite operator*, and phycists call it the one dimensional harmonic oscillator Hamiltonian. Clearly, the h_n's are mutually orthogonal elements of $\mathscr{S}(\mathbb{R};\mathbb{R})$. To check that they form a basis in $L^2(\lambda_{\mathbb{R}};\mathbb{C})$, let $\varphi \in L^2(\lambda_{\mathbb{R}};\mathbb{C})$, set $\psi(x) = \pi^{-\frac{1}{2}} e^{-\frac{x^2}{4}} \varphi(2^{-\frac{1}{2}}x)$, and check that

$$(\varphi, h_n)_{L^2(\lambda_{\mathbb{R}};\mathbb{C})} = (\psi, H_n)_{L^2(\gamma_{0,1};\mathbb{C})}.$$

Hence $\varphi = 0$ if φ is orthogonal to $\{h_n : n \geq 0\}$, and so $\{h_n : n \geq 0\}$ is an orthogonal basis in $L^2(\lambda_{\mathbb{R}};\mathbb{R})$ consisting of eigenfunctions for H.

Lemma 2.3.2 *For each $n \geq 0$, $\widehat{h_n} = i^n (2\pi)^{\frac{1}{2}} h_n$.*

Proof. First note that $\widehat{h_0} = (2\pi)^{\frac{1}{2}} h_0$. Next, assume that $\widehat{h_n} = i^n (2\pi)^{\frac{1}{2}} h_n$, and conclude that

$$\begin{aligned} 2^{\frac{1}{2}} \widehat{h_{n+1}}(\xi) &= \int e^{i\xi x} x h_n(x)\,dx - \int e^{i\xi x} h_n'(x)\,dx = -i\partial_\xi \widehat{h_n}(\xi) + i\xi \widehat{h_n}(\xi) \\ &= i^{n+1}(2\pi)^{\frac{1}{2}}\big(\xi - \partial_\xi\big) h_n(\xi) = i^{n+1} 2^{\frac{1}{2}} (2\pi)^{\frac{1}{2}} h_{n+1}(\xi). \end{aligned}$$

□

After multiplying both sides of (2.3.21) by $2h_n$ and integrating, we know that

$$\|x h_n\|^2_{L^2(\lambda_{\mathbb{R}};\mathbb{R})} + \|h_n'\|^2_{L^2(\lambda_{\mathbb{R}};\mathbb{R})} = (2n+1)\|h_n\|^2_{L^2(\lambda_{\mathbb{R}};\mathbb{R})} = (2n+1)\pi^{\frac{1}{2}} n!,$$

and thedrefore

$$\int (1+x^2) h_n(x)^2\,dx \leq 2(n+1)\pi^{\frac{1}{2}} n!.$$

As a consequence of this and Schwarz's inequality,

$$\|h_n\|_{L^1(\lambda_{\mathbb{R}};\mathbb{R})} = \int (1+x^2)^{-\frac{1}{2}} (1+x^2)^{\frac{1}{2}} |h_n(x)|\,dx \leq \pi^{\frac{1}{2}} \big(2(n+1)\pi^{\frac{1}{2}} n!\big)^{\frac{1}{2}},$$

and so

$$\|h_n\|_{L^1(\lambda_{\mathbb{R}};\mathbb{R})} \leq 2^{\frac{1}{2}} \pi^{\frac{3}{4}} (n+1)^{\frac{1}{2}} (n!)^{\frac{1}{2}}.$$

At the same time, by Lemma 2.3.2, we have that

$$\|h_n\|_{\mathrm{u}} = (2\pi)^{-\frac{1}{2}} \|\widehat{h_n}\|_{\mathrm{u}} \leq (2\pi)^{-\frac{1}{2}} \|h_n\|_{L^1(\lambda_{\mathbb{R}};\mathbb{R})}.$$

Starting from this and using the fact coming from (2.3.20) that $h_n' = 2^{\frac{1}{2}}\big(n h_{n-1} - h_{n+1}\big)$, one sees that

$$\|h_n\|_{\mathrm{u}} \leq \pi^{\frac{1}{4}} \big((n+1)!\big)^{\frac{1}{2}} \text{ and } \|h_n'\|_{\mathrm{u}} \leq 2^{-\frac{1}{2}} \pi^{\frac{1}{4}} (2n+1)\big(n!\big)^{\frac{1}{2}}, \tag{2.3.22}$$

from which it follows that

$$|H_n(x)| \leq \pi^{\frac{1}{4}} \big((n+1)!\big)^{\frac{1}{2}} e^{\frac{x^2}{4}} \text{ and } |H_n'(x)| \leq \pi^{\frac{1}{4}} n \big(n!\big)^{\frac{1}{2}} e^{\frac{x^2}{2}}. \tag{2.3.23}$$

Define the operator Q_t for $t > 0$ and $\varphi \in L^2(\lambda_{\mathbb{R}};\mathbb{R})$ by

$$Q_t \varphi = \frac{e^{-\frac{t}{2}}}{\pi^{\frac{1}{2}}} \sum_{n=0}^{\infty} e^{-nt} \frac{(\varphi, h_n)_{L^2(\lambda_{\mathbb{R}};\mathbb{R})}}{n!} h_n, \tag{2.3.24}$$

and so

$$Q_t h_n = e^{-(n+\frac{1}{2})t} h_n. \tag{2.3.25}$$

Using the properties of $\{h_n : n \geq 0\}$ in (2.3.20), one can check that $\{Q_t : t > 0\}$ is a semigroup of self-adjoint operators and that $\|Q_t\varphi\|_{L^2(\lambda_{\mathbb{R}};\mathbb{R})} \leq e^{-\frac{t}{2}}\|\varphi\|_{L^2(\lambda_{\mathbb{R}};\mathbb{R})}$. In addition, if $\varphi \in \mathscr{S}(\mathbb{R};\mathbb{C})$, then (2.3.21) says that

$$\partial_t Q_t \varphi = \frac{e^{-\frac{t}{2}}}{\pi^{\frac{1}{2}}} \sum_{n=0}^{\infty} e^{-nt} \frac{(\varphi, \mathcal{H}h_n)_{L^2(\lambda_{\mathbb{R}};\mathbb{R})}}{n!} h_n.$$

and so

$$\partial_t Q_t \varphi = Q_t \mathcal{H}\varphi \text{ for } \varphi \in \mathscr{S}(\mathbb{R};\mathbb{C}). \tag{2.3.26}$$

Recall that for $\varphi \in C_{\mathrm{b}}(\mathbb{R};\mathbb{R})$,

$$\int p(t,x,y)\varphi(y)\,dy = \sum_{n=0}^{\infty} e^{-\frac{nt}{2}} \frac{(\varphi, H_n)_{L^2(\gamma_{0,1};\mathbb{R})}}{n!} H_n(x)$$

$$\text{where } p(t,x,y) = \big(2\pi(1-e^{-t})\big)^{-\frac{1}{2}} \exp\left(-\frac{(y-e^{-\frac{t}{2}}x)^2}{2(1-e^{-t})}\right).$$

Using the estimate (2.3.23), observe that, for each $\epsilon > 0$, the series

$$\sum_{n=0}^{\infty} e^{-\frac{nt}{2}} \frac{H_n(x)H_n(y)}{n!}$$

is absolutely convergent uniformly for $t \geq \epsilon$ and (x,y) in compacts. Thus

$$(1-e^{-t})^{-\frac{1}{2}} \exp\left(-\frac{(y-e^{-\frac{t}{2}}x)^2}{2(1-e^{-t})}\right) = \sum_{n=0}^{\infty} e^{-\frac{nt}{2}} \frac{H_n(x)H_n(y)}{n!} e^{-\frac{y^2}{2}},$$

and so,

$$(1-e^{-t})^{-\frac{1}{2}} \exp\left(-\frac{e^{-t}y^2 - 2e^{-\frac{t}{2}}xy + e^{-t}y^2}{2(1-e^{-t})}\right) = \sum_{n=0}^{\infty} e^{-\frac{nt}{2}} \frac{H_n(x)H_n(y)}{n!}.$$

Equivalently, if, for $\theta \in \{z : \Re z \in (-1,1)\}$,

$$M(\theta,x,y) = \big(2\pi(1-\theta^2)\big)^{-\frac{1}{2}} \exp\left(-\frac{(\theta x)^2 - 2\theta xy + (\theta y)^2}{2(1-\theta^2)}\right),$$

then

$$M(\theta,x,y) = \sum_{n=0}^{\infty} \theta^n \frac{H_n(x)H_n(y)}{n!}, \tag{2.3.27}$$

first for $\theta \in (0,1)$ and then, by analytic continuation, for $\theta \in \mathbb{C}$ with $\Re\theta \in (-1,1)$.

The function $M(\theta, x, y)$ is called the *Mehler kernel*, and (2.3.27) is one of the many formulas in which it appears. Another formula in which it plays a role is in connection with the semigroup $\{Q_t : t > 0\}$. Namely, notice that the estimate (2.3.22) guarantees that, each $\epsilon > 0$, the series

$$\sum_{n=0}^{\infty} e^{-nt} \frac{h_n(x)h_n(y)}{n!}$$

absolutely uniformly for $t \geq \epsilon$ and $(x,y) \in \mathbb{R}^2$. Moreover,

$$\begin{aligned}\sum_{n=0}^{\infty} e^{-nt} \frac{h_n(x)h_n(y)}{n!} &= M\big(e^{-t}, 2^{\frac{1}{2}}x, 2^{\frac{1}{2}}y\big)e^{-\frac{x^2+y^2}{2}} \\ &= (1-e^{-2t})^{-\frac{1}{2}} \exp\left(-\frac{(1+e^t)x^2 - 2e^{-t}xy + (1+e^t)y^2}{1-e^{-2t}}\right).\end{aligned}$$

Hence, for $\varphi \in L^1(\lambda_{\mathbb{R}}; \mathbb{C})$,

$$\begin{aligned}&Q_t\varphi(x) = \int q(t,x,y)\varphi(y)\,dy \quad \text{where } q(t,x,y) \text{ equals} \\ &\left(\frac{e^{-t}}{\pi(1-e^{-2t})}\right)^{\frac{1}{2}} \exp\left(-\frac{(1+e^t)x^2 - 2e^{-t}xy + (1+e^t)y^2}{2(1-e^{-2t})}\right) \\ &= (2\pi\sinh t)^{-\frac{1}{2}} \exp\left(-\frac{(\cosh t)x^2 - 2xy + (\cosh t)y^2}{2\sinh t}\right).\end{aligned} \tag{2.3.28}$$

Finally, notice that this expression for Q_t shows that Q_t maps $L^1(\lambda_{\mathbb{R}}; \mathbb{R})$ into $\mathscr{S}(\mathbb{R}; \mathbb{C})$ and therefore that, by (2.3.26), $\partial_s Q_s Q_t \varphi = Q_s H Q_t \varphi$. After $s \searrow 0$, this means that

$$\partial_t Q_t \varphi = H Q_t \varphi \quad \text{for } \varphi \in L^1(\lambda_{\mathbb{R}}; \mathbb{R}). \tag{2.3.29}$$

2.4 Gaussian Families

If X and Y are square integrable random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, then the *covariance* $\operatorname{cov}(X,Y)$ of X and Y is the number

$$\mathbb{E}\big[\big(X - \mathbb{E}[X]\big)\big(Y - \mathbb{E}[Y]\big)\big] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

If $X_1, \dots, X_n \in L^2(\mathbb{P}, \mathbb{R})$, then the covariance $\operatorname{cov}(X_1, \dots, X_n)$ of $\{X_1, \dots, X_n\}$ is the $n \times n$-matrix whose (k,ℓ)th entry is $\operatorname{cov}(X_k, X_\ell)$.

Lemma 2.4.1 *Given random variable $X_1, \dots, X_n \in L^2(\mathbb{P}; \mathbb{R})$, set*

$$X = (X_1, \ldots, X_n)^\top \text{ and } A = \operatorname{cov}(X_1, \ldots, X_n).$$

If $\mathbb{E}[X_k] = 0$ *for all* $1 \le k \le n$*, then the span* $\operatorname{span}\big(\{X_1, \ldots, X_n\}\big)$ *of* $\{X_1, \ldots, X_n\}$ *in* $L^2(\mathbb{P};\mathbb{R})$ *equals* $\operatorname{span}\big(\{(\xi, X)_{\mathbb{R}^n} : \xi \perp \operatorname{Null}(A)\}\big)$.

Proof. Let Π denote orthogonal projection onto $\operatorname{Null}(A)$. Then

$$X = (I - \Pi)X + \Pi X \text{ and } \mathbb{E}\big[|\Pi X|^2\big] = \operatorname{Trace}(\Pi A \Pi) = 0.$$

Hence, $X = (I - \Pi)X$ (a.s, $\mathbb{P}$). □

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a *Gaussian family* is a subspace $\mathcal{G}$ of $L^2(\mathbb{P};\mathbb{R})$ each of whose elements is Gaussian. A Gaussian family is said to be *centered* if all its elements have mean value 0. By Lemma 2.1.1, the L^2-closure of a Gaussian family is again a Gaussian family.

If $\mathcal{G}$ is a Gaussian family, define the function $m_{\mathcal{G}} : \mathcal{G} \longrightarrow \mathbb{R}$ by $m_{\mathcal{G}}(X) = \mathbb{E}[X]$ and the function $C_{\mathcal{G}} : \mathcal{G}^2 \longrightarrow \mathbb{R}$ by $C_{\mathcal{G}}(X, Y) = \operatorname{cov}(X, Y)$. The functions $m_{\mathcal{G}}$ and $C_{\mathcal{G}}$ are known as the mean and covariance of $\mathcal{G}$.

2.4.1 A Few Basic Facts

Set $\gamma_{0,I} = \gamma_{0,1}^N$. Given $b \in \mathbb{R}^N$ and a non-negative, symmetric linear operator A on $\mathbb{R}^N$, take $\gamma_{b,A}$ to be the distribution of $b + A^{\frac{1}{2}}x$ under $\gamma_{0,I}$. When A is non-degenerate, it is easy to check that $\gamma_{b,A} \ll \lambda_{\mathbb{R}^N}$ and

$$\begin{aligned} \frac{d\gamma_{b,A}}{d\lambda_{\mathbb{R}^N}}(x) &= g_{b,A}(x) \\ &\equiv \big((2\pi)^N \det(A)\big)^{-\frac{1}{2}} \exp\left(-\frac{\big(x - b, A^{-1}(x - b)\big)_{\mathbb{R}^N}}{2}\right). \end{aligned} \tag{2.4.1}$$

In keeping with the notation used in the real valued case, I will write $X \in N(b, A)$ to mean that X is an $\mathbb{R}^N$-valued random variable whose distrubution is $\gamma_{b,A}$.

Remark: When $N = 2$ and $A = \begin{pmatrix} 1 & \theta \\ \theta & 1 \end{pmatrix}$ for some $\theta \in (-1, 1)$, a remark that statisticians have found useful is the relation between $g_{0,A}$ and the Mehler kernel. Namely,

$$g_{0,A}(x_1, x_2) = (2\pi)^{-1} M(\theta, x_1, x_2) e^{-\frac{x_1^2 + x_2^2}{2}},$$

and so

$$\frac{d\gamma_{0,A}}{d\gamma_{0,I}}(x) = \sum_{n=0}^{\infty} \theta^n \frac{H_n(x_1) H_n(x_2)}{n!}.$$

Lemma 2.4.2 *$\gamma_{b,A}$ is the one and only $\mu \in M_1(\mathbb{R}^N)$ with the property that $\{(x,\xi)_{\mathbb{R}^N} : \xi \in \mathbb{R}^N\}$ is a Gaussian family $\mathcal{G}$ under μ such that*

$$m_{\mathcal{G}}\big((x,\xi)_{\mathbb{R}^N}\big) = (b,\xi)_{\mathbb{R}^N} \text{ and } C_{\mathcal{G}}\big((x,\xi)_{\mathbb{R}^N},(x,\xi)_{\mathbb{R}^N}\big) = (\xi, A\xi)_{\mathbb{R}^N} \text{ for all } \xi \in \mathbb{R}^N.$$

In addition,[1]

$$\int e^{(\zeta,y)_{\mathbb{R}^N}}, \gamma_{b,A}(dy) = \exp\left((b,\zeta)_{\mathbb{R}^N} + \frac{(\zeta, A\zeta)_{\mathbb{R}^N}}{2}\right) \text{ for } \zeta \in \mathbb{C}^N, \tag{2.4.2}$$

and, if $\alpha < \|A\|_{\mathrm{op}}^{-1}$, then

$$\int e^{\alpha|x|^2}\,\gamma_{0,A}(dx) = \big(\det(I-\alpha A)\big)^{-\frac{1}{2}}. \tag{2.4.3}$$

Proof. To prove (2.4.2), begin with the case when $b = 0$ and $A = I$, and therefore, since $\gamma_{0,I} = \gamma_{0,1}^N$, that the desired result follows from (2.1.1). To handle the general case, write X as $b + A^{\frac{1}{2}}Y$, where $Y \in N(0,I)$ and check that

$$\mathbb{E}\big[e^{(\zeta, b+A^{\frac{1}{2}}Y)_{\mathbb{R}^N}}\big] = e^{(\zeta,b)_{\mathbb{R}^N}}\,\mathbb{E}\big[e^{(A^{\frac{1}{2}}\zeta, Y)_{\mathbb{R}^N}}\big] = \exp\left((\zeta,b)_{\mathbb{R}^N} + \frac{(A^{\frac{1}{2}}\zeta, A^{\frac{1}{2}}\zeta)_{\mathbb{R}^N}}{2}\right).$$

From (2.4.2) we know that

$$\widehat{\gamma_{b,A}}(\xi) = \exp\left(i(b,\xi)_{\mathbb{R}^N} - \frac{(\xi, A\xi)_{\mathbb{R}^N}}{2}\right) \tag{2.4.4}$$

and that any $\mu \in M_1(\mathbb{R}^N)$ with the stated properties will have the same characteristic function as $\gamma_{b,A}$.

To prove (2.4.3), let $a_1, \dots, a_n$ be the eigenvalues of A, and observe that, by (2.1.2)

$$\int e^{\alpha|x|^2}\,\gamma_{0,A}(dx) = \prod_{k=1}^N \int_{\mathbb{R}} e^{\alpha a_k x_k^2}\,\gamma_{0,1}(dx) = \prod_{k=1}^N (1-\alpha a_k)^{-\frac{1}{2}} = \big(\det(I-\alpha A)\big)^{-\frac{1}{2}}$$

□

An alternative formulation of the first statement in Lemma 2.4.2 is that the joint distribution of $\{X_1, \dots, X_n\}$ is $\gamma_{b,A}$ with $b = m_{\mathcal{G}}(X_1, \dots, X_b)$ and $A = C_{\mathcal{G}}(X_1, \dots, X_n)$.

The following result is arguably the most important property of Gaussian families.

[1] Even though some of the the vectors involved are complex, the inner product here is the Euclidean one, not the Hermitian one.

Theorem 2.4.3 *Let $\mathcal{G}$ is a centered Gaussian family on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For any subset $\emptyset \neq S \subsetneq \mathcal{G}$, set $\mathcal{F}_S = \sigma(S)$, the smallest σ-algebra with respect to which all the elements of S are measurable, and take $S^\perp$ to be the perpendicular complement in $L^2(\mathbb{P}; \mathbb{R})$ of S Then $\mathcal{F}_S$ is independent of $\mathcal{F}_{S^\perp \cap \mathcal{G}}$.*

Proof. What we must show is that if $\{X_1, \dots, X_m\} \subseteq S$ and $\{Y_1, \dots, Y_n\} \subseteq S^\perp \cap \mathcal{G}$, then the distribution of $(X_1, \dots, X_m, Y_1, \dots, Y_n)^\top$ is the product of the distribution of $(X_1, \dots, X_m)^\top$ with the distribution of $(Y_1, \dots, Y_n)^\top$. To this end, set $A = \operatorname{cov}(X_1, \dots, X_m)$, $B = \operatorname{cov}(Y_1, \dots, Y_n)$, and observe that

$$C \equiv \operatorname{cov}(X_1, \dots, X_m, Y_1, \dots, Y_n) = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}.$$

Hence, since $\gamma_{0,A}$, $\gamma_{0,B}$, and $\gamma_{0,C}$ are, respectively, the distributions of $(X_1, \dots, X_m)^\top$, $(Y_1, \dots, Y_n)^\top$, and $(X_1, \dots, X_m, Y_1, \dots, Y_n)^\top$,

$$\begin{aligned} \widehat{\gamma_{0,C}} \begin{pmatrix} \xi \\ \eta \end{pmatrix} &= \exp\left(-\frac{(\xi, A\xi)_{\mathbb{R}^m} + (\eta, B\eta)_{\mathbb{R}^n}}{2} \right) \\ &= \widehat{\gamma_{0,A}}(\xi) \widehat{\gamma_{0,B}}(\eta) = \widehat{\gamma_{0,A} \times \gamma_{0,B}} \begin{pmatrix} \xi \\ \eta \end{pmatrix} \end{aligned}$$

for all $\xi \in \mathbb{R}^m$ and $\eta \in \mathbb{R}^n$, from which it follows that $\gamma_{0,C} = \gamma_{0,A} \times \gamma_{0,B}$. □

2.4.2 A Concentration Property of Gaussian Measures

In this subsection I will show that if a Gaussian measure gives positive measure to a set then it is nearly concentrated on a neighborhood of that set. The driving force behind the analysis here is the following beautiful result of B. Maurey and G. Pisier.

Theorem 2.4.4 *Let A be a strictly positive definite, symmetric transformation on $\mathbb{R}^N$, and let X be a $N(0, A)$-random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. If $f : \mathbb{R}^N \longrightarrow \mathbb{R}$ is a continuous function satisfying*

$$|f(y) - f(x)| \leq \lambda \left| A^{-\frac{1}{2}}(y - x) \right| \text{ for } x, y \in \mathbb{R}^N,$$

then

$$\mathbb{E}\left[e^{t(f(X) - \mathbb{E}[f(X)])} \right] \leq e^{\frac{\lambda^2 \pi^2 t^2}{8}} \quad \text{for } t \in \mathbb{R}.$$

Proof. First observe that, by replacing f by $\lambda^{-1}\big(f - \mathbb{E}[f(X)]\big)$, we can reduce to the case when $\lambda = 1$ and $\mathbb{E}\big[f(X)\big] = 0$ Thus we will proceed under these assumptions. In addition, without loss in generality, we will assume that there is a second $N(0,A)$-random variable Y which is independent of X.

Next note that, after applying a standard mollification procedure, we may assume that f is smooth and $|A^{\frac{1}{2}}\nabla f|\le 1$ everywhere.

Now let f be a smooth function satisfying $\mathbb{E}\big[f(X)\big]=0$ and $|A^{\frac{1}{2}}\nabla f|\le 1$ everywhere, Then, by Jensen's inequality applied to the convex function $x \rightsquigarrow e^{-x}$,

$$\mathbb{E}\big[e^{-tf(Y)}\big]\ge e^{-t\mathbb{E}^{\mathbb{P}}[f(Y)]}=1,$$

and so

$$\mathbb{E}\big[e^{t(f(X)-f(Y))}\big]=\mathbb{E}\big[e^{tf(X)}\big]\mathbb{E}\big[e^{-tf(Y)}\big]\ge\mathbb{E}\big[e^{tf(X)}\big]. \tag{$*$}$$

Next, for $\theta\in\mathbb{R}$, set $X(\theta)=X\cos\theta+Y\sin\theta$ and $Y(\theta)=-X\sin\theta+Y\cos\theta$. Using characteristic functions, it is easy to check that, for each θ, $X(\theta)$ and $Y(\theta)$ are again mutually independent, $N(0,A)$-random variables. Furthermore, by the Fundamental Theorem of Calculus,

$$f(X)-f(Y)=\int_0^{\frac{\pi}{2}}\big(\nabla f\big(X(\theta)\big),Y(\theta)\big)_{\mathbb{R}^N}\,d\theta,$$

and so, by Jensen's inequality,

$$e^{t(f(X)-f(Y))}=e^{t\int_0^{\frac{\pi}{2}}\left(\nabla f(X(\theta)),Y(\theta)\right)_{\mathbb{R}^N}\,d\theta}\le\frac{2}{\pi}\int_0^{\frac{\pi}{2}}e^{\frac{\pi t}{2}\left(\nabla f(X(\theta)),Y(\theta)\right)_{\mathbb{R}^N}}\,d\theta.$$

Hence, by Fubini's Theorem and $(*)$,

$$\mathbb{E}\big[e^{tf(X)}\big]\le\frac{2}{\pi}\int_0^{\frac{\pi}{2}}\mathbb{E}\Big[e^{\frac{\pi t}{2}\left(\nabla f(X(\theta)),Y(\theta)\right)_{\mathbb{R}^N}}\Big]\,d\theta.$$

Finally, because $Y(\theta)$ is independent of $X(\theta)$, (2.4.2) implies that

$$\mathbb{E}\Big[e^{\frac{\pi t}{2}\left(\nabla f(X(\theta)),Y(\theta)\right)_{\mathbb{R}^N}}\Big]=\mathbb{E}\left[\exp\left(\frac{\pi^2t^2\big|A^{\frac{1}{2}}\nabla f\big(X(\theta)\big)\big|^2}{8}\right)\right]\le e^{\frac{\pi^2t^2}{8}}.$$

□

As a more or less immediate consequence of Theorem 2.4.4, we have that

$$\begin{aligned}\mathbb{P}\big(f(X)-\mathbb{E}^{\mathbb{P}}[f(X)]\ge R\big)&\le e^{-\frac{2R^2}{\pi^2\lambda^2}}\\ \mathbb{P}\big(|f(X)-\mathbb{E}^{\mathbb{P}}[f(X)]|\ge R\big)&\le 2e^{-\frac{2R^2}{\pi^2\lambda^2}}\end{aligned}\quad\text{for } R>0. \tag{2.4.5}$$

Indeed, by Markov's inequality,

$$\mathbb{P}\big(f(X)-\mathbb{E}^{\mathbb{P}}[f(X)]\ge R\big)\le e^{-tR+\frac{\lambda^2\pi^2t^2}{8}}\quad\text{for all } t\ge 0,$$

and so the first of these follows when one takes $t = \frac{4R}{\lambda^2\pi^2}$. Further, given the first estimate, the second follows when the first one is applied to both f and $-f$ and the two are added.

Perhaps the most interesting aspect of these results is their dimension independence. That is, in any dimension, the distribution of a uniformly Lipschitz continuous function of a Gaussian random variable satisfies tail estimates that are remarkably like those of an $\mathbb{R}$-valued Gaussian random variable. To appreciate the significance of this dimensional independence, consider $f(x) = |x|$. Then

$$\lambda^2 = \|A\|_{\text{op}} \equiv \sup\bigl\{ \bigl(e, Ae\bigr)_{\mathbb{R}^N} : e \in \mathbb{S}^{N-1} \bigr\},$$

and so the second estimate in (2.4.5) says that

$$\mathbb{P}\bigl(\bigl||X| - \mathbb{E}[|X|]\bigr| \geq R\bigr) \leq 2e^{-\frac{2R^2}{\pi^2\|A\|_{\text{op}}}}$$

Hence, since

$$\operatorname{Var}(|X|) = \mathbb{E}\bigl[\bigl(|X| - \mathbb{E}^{[}|X|]\bigr)^2\bigr] = 2\int_0^\infty R\mathbb{P}\bigl(\bigl||X| - \mathbb{E}^{\mathbb{P}}[|X|]\bigr| \geq R\bigr)\, dR$$

we see that $\operatorname{Var}(|X|) \leq \frac{\pi^2\|A\|_{\text{op}}}{2}$. That is, independent of dimension, the variance of the length of an $\mathbb{R}^N$-valued, $N(0, A)$-random variable is dominated by a universal constant times the operator norm of A. What makes this conclusion somewhat surprising is that $\mathbb{E}^{\mathbb{P}}[|X|^2] = \operatorname{Trace}(A)$. Thus, for instance, if $A = I$, then $\mathbb{E}^{\mathbb{P}}\bigl[|X|^2\bigr] = N$ is tending to infinity as $N \to \infty$ and yet, because $\|I\|_{\text{op}} = 1$, $\operatorname{Var}(|X|)$ is bounded by $\frac{\pi^2}{2}$, independent of N. A closely related application shows that

$$\mathbb{P}\bigl(|X| \geq R\bigr) \leq e^{-\frac{R^2}{2\pi^2\|A\|_{\text{op}}}} \quad \text{if } R \geq 2\sqrt{\operatorname{Trace}(A)}.$$

To check this, note that $R \geq 2\sqrt{\operatorname{Trace}(A)} \implies \mathbb{E}^{\mathbb{P}}[|X|] \leq \frac{R}{2}$, and therefore, by the first part of (2.4.5),

$$\mathbb{P}\bigl(|X| \geq R\bigr) \leq \mathbb{P}\bigl(|X| - \mathbb{E}^{\mathbb{P}}[|X|] \geq \tfrac{R}{2}\bigr) \leq e^{-\frac{R^2}{2\pi^2\|A\|_{\text{op}}}}.$$

The preceding estimates say that the rate at which $\mathbb{P}\bigl(|X| \geq R\bigr)$ decays for $R \geq 2\sqrt{\text{Trace(A)}}$ is very fast and depends on $\|A\|_{\text{op}}$ but not on N. Equivalently, the distribution of X is very nearly concentrated on a ball of radius $R > 2\sqrt{\operatorname{Trace}(A)}$. The following theorem gives a more general statement of this *Gaussian concentration phenomenon.*

Theorem 2.4.5 *If X is an $\mathbb{R}^N$-valued, $N(0, A)$-random variable and $\Gamma \in \mathcal{B}_{\mathbb{R}^N}$, then*

$$\mathbb{P}\big(X \in \Gamma\big) \wedge \mathbb{P}\big(X \notin \Gamma^{(R)}\big) \le e^{-\frac{R^2}{2\pi^2\|A\|_{\mathrm{op}}}} \quad \textit{for } R \ge 0,$$

where $\Gamma^{(R)} = \{x \in \mathbb{R}^N : |x - \Gamma| \le R\}$. *Hence, if* $\epsilon \in (0,1)$ *and* $\mathbb{P}(X \in \Gamma) \ge \epsilon$, *then*

$$\mathbb{P}\big(X \notin \Gamma^{(R)}\big) \le e^{-\frac{R^2}{2\pi^2\|A\|_{\mathrm{op}}}} \quad \textit{for } R > \pi\sqrt{2\|A\|_{\mathrm{op}} \log \tfrac{1}{\epsilon}}.$$

Proof. Set $f(x) = |x - \Gamma|$. If $\mathbb{E}[f(X)] \le \frac{R}{2}$, then, by (2.4.5),

$$\mathbb{P}\big(X \notin \Gamma^{(R)}\big) \le \mathbb{P}\big(f(X) - \mathbb{E}^{\mathbb{P}}[f(X)] \ge \tfrac{R}{2}\big) \le e^{-\frac{R^2}{2\pi^2\|A\|_{\mathrm{op}}}}.$$

If $\mathbb{E}^{\mathbb{P}}[f(X)] \ge \frac{R}{2}$, then, by (2.4.5) applied to $-f$,

$$\mathbb{P}(X \in \Gamma) \le \mathbb{P}\big(\mathbb{E}[f(X)] - f(X) \ge \tfrac{R}{2}\big) \le e^{-\frac{R^2}{2\pi^2\|A\|_{\mathrm{op}}}}.$$

Hence, the first assertion is proved.

To prove the second assertion, let $R > \pi\sqrt{2\|A\|_{\mathrm{op}} \log \frac{1}{\epsilon}}$ be given. Then, because $\mathbb{P}(X \in \Gamma) \ge \epsilon > e^{-\frac{R^2}{2\pi^2\|A\|_{\mathrm{op}}}}$, $\mathbb{P}\big(X \notin \Gamma^{(R)}\big) \le e^{-\frac{R^2}{2\pi^2\|A\|_{\mathrm{op}}}}$. □

As a consequence of Theorem 2.4.5, one sees that if $\mathbb{P}(X \in \Gamma) \ge \epsilon$, then, with large probability, X lies within a distance on the order of $\sqrt{\|A\|_{\mathrm{op}} \log \frac{1}{\epsilon}}$ from Γ. In other words, once one knows that $\gamma_{0,A}(\Gamma) \ge \epsilon$, one knows that most of the mass of $\gamma_{0,A}$ is concentrated relatively nearby Γ, and the extent of this concentration depends only on $\|A\|_{\mathrm{op}}$ and not on dimension.

2.4.3 The Gaussian Isoperimetric Inequality

The goal in this subsection is to prove a result that can be thought of as a isoperimetric inequality for Gaussian measures and can be used to derive concentration results closely related to those in the proceding subsection. To describe this result, let $p(\tau) = (2\pi)^{-\frac{1}{2}} e^{-\frac{\tau^2}{2}}$ be the standard Gauss kernel on $\mathbb{R}$, and take $\Phi(x) = \int_{-\infty}^{x} p(\tau)\, d\tau$ to be the error function. Then the result states that, for $\Gamma \in \mathcal{B}_{\mathbb{R}^N}$ and $t \ge 0$,

$$\begin{gathered}\gamma_{0,I}\big(\Gamma^{(t)}\big) \ge \Phi\Big(\Phi^{-1}\big(\gamma_{0,I}(\Gamma)\big) + t\Big), \\ \text{where } \Gamma^{(t)} = \{x \in \mathbb{R}^N : |x - \Gamma| \le t\}.\end{gathered} \tag{2.4.6}$$

The sense in which (2.4.6) is an isoperimetric inequality is that for a half-space $H = \{x : (x,e)_{\mathbb{R}^N} \le a\}$ with $e \in \mathbb{S}^{N-1}$ and $a \in \mathbb{R}$, $\gamma_{0,I}\big(H^{(t)}\big) = \Phi(a+t)$. Thus, if $a = \Phi^{-1}\big(\gamma_{0,I}(\Gamma)\big)$, then $\gamma_{0,I}(H) = \gamma_{0,I}(\Gamma)$ and $\gamma_{0,I}\big(\Gamma^{(t)}\big) \ge$

$\gamma_{0,I}\big(H^{(t)}\big)$. In other words, among the sets $B \in \mathcal{B}_{\mathbb{R}^N}$ with $\gamma_{0,I}(B) = \gamma_{0,I}(\Gamma)$, all the ones for which the growth of $t \rightsquigarrow \gamma_{0,I}\big(B^{(t)}\big)$ is slowest are half-spaces.

The first derivations, given independently by C. Borell and by B. Tsirelson & V. Sudakov, of (2.4.6) were based on Lévy's isoperimetric inequality for spheres combined with the observation (originally made by F. Mehler and rediscovered by E. Borel) that $\gamma_{0,1}^N$ is the weak limit as $n \to \infty$ of the marginal distribution the first N-coordinates under the normalized surface measure on the n-sphere of radius $n^{\frac{1}{2}}$. The derivation that follows was given by S. Bobkov. It too requires realizing $\gamma_{0,I}$ as a weak limit, but this time of measures based on sums of Bernoulli random variables rather than the surface measure on spheres.

The first step is to rewrite (2.4.6) as

$$\Phi^{-1}\big(\gamma_{0,I}(\Gamma^{(t)})\big) - \Phi^{-1}\big(\gamma_{0,I}(\Gamma)\big) \geq t. \tag{2.4.7}$$

It is clear that (2.4.7) is equivalent to (2.4.6) and that it will hold for all $\Gamma \in \mathcal{B}_{\mathbb{R}^N}$ if it holds for closed ones. In addition, since $\Phi^{-1}(1) = \infty$ and $\Phi^{-1}(0) = -\infty$, we may and will assume that $0 < \gamma_{0,I}(\Gamma) \leq \gamma_{0,I}(\Gamma^{(t)}) < 1$. Now consider the right-continous, non-decreasing function $F(t) = \Phi^{-1}\big(\gamma_{0,I}(\Gamma^{(t)})\big)$ for $t \in [0, T)$, where $T = \sup\{t \geq 0 : \gamma_{0,I}(\Gamma^{(t)}) < 1\}$. Lebesgue's Differentiation Theorem says that

$$f(t) = \lim_{\tau \searrow 0} \frac{F(t+\tau) - F(t)}{\tau} \text{ exists for } \lambda_{\mathbb{R}}\text{-a.e. } t \in [0, T)$$

and that $F(t) - F(0) \geq \int_0^t f(\tau)\, d\tau$. Hence, (2.4.7) will follow once we show that $f \geq 1$ (a.e.,$\lambda_{\mathbb{R}}$) on $[0, T)$. Next observe that $(\Phi^{-1})' = \frac{1}{\Psi}$, where $\Psi = p \circ \Phi^{-1}$, and therefore

$$\varliminf_{\tau \searrow 0} \frac{F(t+\tau) - F(t)}{\tau} = \frac{1}{\Psi\big(\gamma_{0,I}(\Gamma^{(t)})\big)} \varliminf_{\tau \searrow 0} \frac{\gamma_{0,I}\big(\Gamma^{(t+\tau)}\big) - \gamma_{0,I}\big(\Gamma^{(t)}\big)}{\tau},$$

which means that it suffices to show that

$$\varliminf_{\tau \searrow 0} \frac{\gamma_{0,I}\big(\Gamma^{(t+\tau)}\big) - \gamma_{0,I}\big(\Gamma^{(t)}\big)}{\tau} \geq \Psi\big(\gamma_{0,I}(\Gamma^{(t)})\big).$$

In fact, because $\Gamma^{(t+\tau)} = (\Gamma^{(t)})^{(\tau)}$, we need only prove that

$$\varliminf_{\tau \searrow 0} \frac{\gamma_{0,I}\big(\Gamma^{(\tau)}\big) - \gamma_{0,I}\big(\Gamma\big)}{\tau} \geq \Psi\big(\gamma_{0,I}(\Gamma)\big). \tag{2.4.8}$$

The second step is to show that (2.4.8) will follow once we know that for all $\varphi \in C_{\mathrm{b}}^2\big(\mathbb{R}^N; [0,1]\big)$,

$$\Psi\big(\langle \varphi, \gamma_{0,I} \rangle\big) \leq \langle \Psi \circ \varphi, \gamma_{0,I} \rangle + \langle |\nabla \varphi|, \gamma_{0,I} \rangle. \tag{2.4.9}$$

To see this, let $\epsilon \in \left(0, \frac{1}{2}\right)$ be given, and define

$$\eta_\tau(x) = 1 - \frac{|x - \Gamma^{(\epsilon)}|}{(1-2\epsilon)\tau} \wedge 1.$$

Then η_τ is a $[0,1]$-valued, Lipschitz continuous function with Lipschitz constant equal to $\big((1-2\epsilon)\tau\big)^{-1}$ and satisfying

$$\eta_\tau(x) = \begin{cases} 1 & \text{if } |x - \Gamma| \le \epsilon\tau \\ 0 & \text{if } |x - \Gamma| \ge (1-\epsilon)\tau. \end{cases}.$$

Now choose $\rho \in C_{\mathrm{b}}^{\infty}\big(\mathbb{R}^N; [0,\infty)\big)$ so that $\rho = 0$ off $B_{\mathbb{R}^N}(0, \epsilon\tau)$ and $\int \rho \, d\lambda_{\mathbb{R}^N} = 1$, and define $\varphi_\tau = \rho * \eta_\tau$. Then $\varphi_\tau \in C^\infty\big(\mathbb{R}^N; [0,1]\big)$, $|\nabla\varphi_\tau| \le \big((1-2\epsilon)\tau\big)^{-1}$, and

$$\varphi_\tau(x) = \begin{cases} 1 & \text{for } x \in \Gamma \\ 0 & \text{for } x \notin \Gamma^{(\tau)}. \end{cases}$$

Hence, $\lim_{\tau\searrow 0} \varphi_\tau(x) = \mathbf{1}_\Gamma(x)$, and therefore $\lim_{\tau\searrow 0} \Psi \circ \varphi_\tau(x) = 0$ for each $x \in \mathbb{R}^N$. Further, since φ_τ achieves its minimum value off of $\Gamma^{(\tau)}$ and its maximum on Γ, $|\nabla\varphi_\tau| \le \frac{\mathbf{1}_{\Gamma^{(\tau)}} - \mathbf{1}_\Gamma}{(1-2\epsilon)\tau}$, and so, by (2.4.9),

$$\begin{aligned} \Psi\big(\gamma_{0,1}(\Gamma)\big) &= \lim_{\tau\searrow 0} \Psi\big(\langle \varphi_\tau, \gamma_{0,I}\rangle\big) \\ &\le \varliminf_{\tau\searrow 0} \langle |\nabla\varphi_\tau|, \gamma_{0,I}\rangle \le (1-2\epsilon)^{-1} \varliminf_{\tau\searrow 0} \frac{\gamma_{0,I}(\Gamma^{(\tau)}) - \gamma_{0,I}(\Gamma)}{\tau}. \end{aligned}$$

Thus (2.4.8) follows after one lets $\epsilon \searrow 0$.

For reasons that will become clear shortly, we will prove (2.4.9) by proving the slightly stronger inequality

$$\Psi\big(\langle \varphi, \gamma_{0,I}\rangle\big) \le \int_{\mathbb{R}^N} \big[(\Psi\circ\varphi)^2 + |\nabla\varphi|^2\big]^{\frac{1}{2}} \, d\gamma_{0,I}. \tag{2.4.10}$$

This inequality looks somewhat like a Poincaré inequality, and, like a Poincaré inequality, it is preserved under products. To see this, assume that it holds for $1 \le N \le M$, and let $\varphi \in C_{\mathrm{b}}^2\big(\mathbb{R}^{M+1}; [0,1]\big)$ be given. Writing $\gamma_{0,I}$ for $\mathbb{R}^{M+1}$ as the product of $\gamma_{0,I}$ for $\mathbb{R}^M$ and $\gamma_{0,1}$,

$$\begin{aligned} &\int_{\mathbb{R}^{M+1}} \big[(\Psi\circ\varphi)^2 + |\nabla\varphi|^2\big]^{\frac{1}{2}} \, d\gamma_{0,I} \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}^M} \Big[\big(\Psi\circ\varphi(x,y)\big)^2 + |\nabla_x\varphi(x,y)|^2 + |\partial_y\varphi(x,y)|^2\Big]^{\frac{1}{2}} \gamma_{0,I}(dx) \right) \gamma_{0,1}(dy). \end{aligned}$$

Because the triangle inequality implies that $(a,b) \in \mathbb{R}^2 \longmapsto (a^2+b^2)^{\frac{1}{2}} \in [0,\infty)$ is convex, Jensen's inequality implies that

$$\int_{\mathbb{R}^M} \left[\bigl(\Psi\circ\varphi(x,y)\bigr)^2 + |\nabla_x\varphi(x,y)|^2 + |\partial_y\varphi(x,y)|^2\right]^{\frac12}\gamma_{0,I}(dx)$$
$$\ge \left[\left(\int_{\mathbb{R}^M} \left[\Psi\circ\varphi(x,y)^2 + |\nabla_x\varphi(x,y)|^2\right]^{\frac12}\gamma_{0,I}(dx)\right)^2\right.$$
$$\left. + \left(\int_{\mathbb{R}^M} |\partial_y\varphi(x,y)|\,\gamma_{0,I}(dx)\right)^2\right]^{\frac12} \ge \left[\Psi\circ\psi(y)^2 + |\psi'(y)|^2\right]^{\frac12},$$

where $\psi(y) = \int_{\mathbb{R}^M}\varphi(x,y)\,\gamma_{0,I}(dx)$. Since

$$\int_{\mathbb{R}} \left[\bigl(\Psi\circ\psi(y)\bigr)^2 + |\psi'(y)|^2\right]^{\frac12}\gamma_{0,1}(dy) \ge \Psi\bigl(\langle\psi,\gamma_{0,1}\rangle\bigr) = \Psi\bigl(\langle\varphi,\gamma_{0,I}\rangle\bigr),$$

we now know that (2.4.10) holds for all $N\ge 1$ if it does when $N=1$. Thus what we need to show is that

$$\Psi\bigl(\langle\varphi,\gamma_{0,1}\rangle\bigr) \le \int_{\mathbb{R}} \left[\bigl(\Psi\circ\varphi\bigr)^2 + |\varphi'|^2\right]^{\frac12} d\gamma_{0,1} \tag{2.4.11}$$

for $\varphi\in C^2_{\mathrm{b}}\bigl(\mathbb{R};[0,1]\bigr)$.

In some ways, the next step is the most interesting. What we are going to show is that (2.4.11) follows from a discrete analog of itself. Namely, let β be the symmetric Bernoulli measure on $\{-1,1\}$ (i.e., $\beta(\{\pm1\}) = \frac12$), and, given a function $f:\{-1,1\}\longrightarrow[0,1]$, define $Df(\pm1) = \pm\frac{f(1)-f(-1)}{2}$. Then the analog of (2.4.11) in this setting is

$$\Psi\bigl(\langle f,\beta\rangle\bigr) \le \int_{\{-1,1\}} \left[\bigl(\Psi\circ f\bigr)^2 + |Df|^2\right]^{\frac12} d\beta. \tag{2.4.12}$$

To understand why (2.4.12) implies (2.4.11), observe that, by exactly to same argument as we used above, (2.4.12) is self-replicating under products. Thus, if $\mathbb{P} = \beta^{\mathbb{Z}^+}$ on $\Omega = \{-1,1\}^{\mathbb{Z}^+}$ and $\tilde S_n(\omega) = n^{-\frac12}\sum_{m=1}^n \omega(m)$, then, for any $n\ge 1$,

$$\Psi\bigl(\mathbb{E}[\varphi\circ\tilde S_n]\bigr) \le \mathbb{E}\left[\left(\bigl(\Psi\circ\varphi\circ\tilde S_n\bigr)^2 + \sum_{m=1}^n |D_m\varphi\circ\tilde S_n|^2\right)^{\frac12}\right],$$

where

$$D_m\varphi\circ\tilde S_n(\omega) = \frac{\varphi\bigl(\tilde S_n(\omega)\bigr) - \varphi\bigl(\tilde S_n(\omega) - 2n^{-\frac12}\omega(m)\bigr)}{2}.$$

Since

$$\left|D_m\varphi\bigl(\tilde S_n(\omega)\bigr) + n^{-\frac12}\omega(m)\varphi'\bigl(\tilde S_n(\omega)\bigr)\right| \le \frac{\|\varphi''\|_{\mathrm{u}}}{n}$$

and therefore

$$\left|\sum_{m=1}^{n}\big(D_m\varphi\circ\tilde{S}_n(\omega)\big)^2-\big(\varphi'\circ\tilde{S}_n(\omega)\big)^2\right|\le\frac{C}{n^{\frac{1}{2}}}$$

for some $C<\infty$, an application of Central Limit Theorem completes the proof that (2.4.11) follows from (2.4.12).

What remains to be done is give a proof of (2.4.12), an intricate but relatively elementary exercise in calculous. Given $f:\{-1,1\}\longrightarrow[0,1]$, set $c=\langle f,\beta\rangle=\frac{f(1)+f(-1)}{2}$ and $\xi=\frac{f(1)-f(-1)}{2}$. Then (2.4.12) becomes

$$\Psi(c)\le\frac{\big(\Psi(c+\xi)^2+\xi^2\big)^{\frac{1}{2}}+\big(\Psi(c-\xi)^2+\xi^2\big)^{\frac{1}{2}}}{2}. \qquad (*)$$

To verify $(*)$, we will make frequent use of the calculations in the following lemma.

Lemma 2.4.6 *For any $\theta\in[0,1]$, $\Phi^{-1}(1-\theta)=-\Phi^{-1}(\theta)$ and $\Psi(1-\theta)=\Psi(\theta)$. Moreover, for $c\in\big[0,\frac{1}{2}\big]$ and $\xi\in[0,c]$, $\Psi(c+\xi)\ge\Psi(c-\xi)$ and $|\Psi'(c+\xi)|\le\Psi'(c-\xi)$. Finally,*

$$(\Psi^2)''=2\big((\Psi')^2-1\big)\ \textit{and}\ \big((\Psi')^2\big)''=2\frac{(\Psi')^2+1}{\Psi^2}\ \textit{on}\ (0,1).$$

Proof. Once one checks that $\Psi'=-\Phi^{-1}$, the calculations of derivatives are simple applications of the product and chain rules.

Because

$$1-\Phi(t)=1-\gamma_{0,1}\big((-\infty,t]\big)=\gamma_{0,1}\big((t,\infty)\big)=\gamma_{0,1}\big((-\infty,-t)\big)=\Phi(-t),$$

$\Phi^{-1}(1-\theta)=-\Phi^{-1}(\theta)$, from which it follows that $\Psi(1-\theta)=\Psi(\theta)$.

Next observe that Ψ is non-decreasing on $\big[0,\frac{1}{2}\big]$. Now let $c\in\big[0,\frac{1}{2}\big]$ and $\xi\in[0,c]$ be given. Then $0\le c-\xi\le1-c-\xi\le\frac{1}{2}$, and so $\Psi(c+\xi)=\Psi(1-c-\xi)\ge\Psi(c-\xi)$. Similarly, because Ψ' is non-negative and non-increasing on $\big[0,\frac{1}{2}\big]$, $|\Psi'(c+\xi)|\le\Psi'(c-\xi)$. □

Returning to $(*)$, first note that there is nothing to do if $c\in\{0,1\}$ and second that it suffices to handle $\xi>0$. Further, since $\Psi(1-\theta)=\Psi(\theta)$ and therefore $(*)$ holds for f if it does for $1-f$, we may and will assume that $c\in\big(0,\frac{1}{2}\big]$ and $\xi\in(0,c]$. Now set $u(\xi)=\Psi(c+\xi)^2+\xi^2$, and, after rewriting $(*)$ in terms of u and squaring both sides, check that it is equivalent first to

$$4u(0)-\big(u(\xi)+u(-\xi)\big)\le2\big(u(\xi)u(-\xi)\big)^{\frac{1}{2}},$$

and then to

$$16u(0)^2+\big(u(\xi)-u(-\xi)\big)^2\le8u(0)\big(u(\xi)+u(-\xi)\big).$$

Thus, if $v(\xi)=u(\xi)-u(0)$, then $(*)$ is equivalent to

$$\big(\Psi(c+\xi)^2 - \Psi(c-\xi)^2\big)^2 \le 8\Psi(c)^2\big(v(\xi)+v(-\xi)\big). \tag{$**$}$$

Using the calculations in Lemma 2.4.6, one sees that $v'' = 2(\Psi')^2$ and therefore that

$$v''(\xi) + v''(-\xi) = 2\big(\Psi'(c+\xi)^2 + \Psi'(c-\xi)^2\big).$$

Since, by Lemma 2.4.6, $(\Psi')^2$ is convex, the right hand side of the preceding dominates $4\Psi'(c)^2$, and so, because $v(\xi)+v(-\xi)$ vanishes to first order at $\xi = 0$, $v(\xi)+v(-\xi) \ge 2\Psi'(c)^2\xi^2$. Therefore $(**)$ will hold if

$$\big(\Psi(c+\xi)^2 - \Psi(c-\xi)^2\big)^2 \le 16\Psi(c)^2\Psi'(c)^2\xi^2.$$

Because, again by Lemma 2.4.6, $\Psi(c+\xi) \ge \Psi(c-\xi)$ and $\Psi'(c) \ge 0$, what we need to show is that

$$\frac{\Psi(c+\xi)^2 - \Psi(c-\xi)^2}{\xi} \le 4\Psi(c)\Psi'(c) = 2(\Psi^2)'(c).$$

But, from Lemma 2.4.6, we know that

$$(\Psi^2)''(c+\xi) - (\Psi^2)''(c-\xi) = 2\big(\Psi'(c+\xi)^2 - \Psi'(c-\xi)^2\big) \le 0,$$

which means that $\xi \rightsquigarrow \Psi(c+\xi)^2 - \Psi(c-\xi)^2$ is concave and therefore that

$$\frac{\Psi(c+\xi)^2 - \Psi(c-\xi)^2}{2\xi} \le (\Psi^2)'(c).$$

As with the Maurey–Pisier estimate, the most important feature of (2.4.6) is its dimension independence. For instance it says that, independent of dimension,

$$\gamma_{0,I}(\Gamma) \ge \frac{1}{2} \implies \gamma_{0,I}\big(\mathbb{R}^N \setminus \Gamma^{(t)}\big) \le 1 - \Phi(t) \le e^{-\frac{t^2}{2}}.$$

To see how it can be used to prove estimates like the one in (2.4.5), consider a Lipschitz continuous function $f : \mathbb{R}^N \longrightarrow \mathbb{R}$ with Lipschitz constant λ, let m be a median of f under $\gamma_{0,I}$, and set $\Gamma_\pm(t) = \{\pm(f-m) \le \lambda t\}$. Then $\gamma_{0,I}\big(\Gamma_\pm(0)\big) \ge \frac{1}{2}$, $\big(\Gamma_\pm(0)\big)^{(t)} \subseteq \Gamma_\pm(t)$, and so

$$\gamma_{0,I}\big(\Gamma_\pm(t)\big) \ge \Phi(t).$$

Therefore

$$\gamma_{0,I}\big(\{|f-m| > \lambda t\}\big) \le 2\big(1-\Phi(t)\big) \le 2e^{-\frac{t^2}{2}}. \tag{2.4.13}$$

Finally, suppose that $g : \mathbb{R} \longrightarrow [0,\infty)$ is a continuous function with the property that $g(s) \le g(t)$ if $0 \le s \le t$ or $0 \ge s \ge t$. Then, starting from (2.4.13), one can show that

$$\int_{\mathbb{R}^N} g\big(f(x)-m\big)\,\gamma_{0,I}(dx) \le \int_{\mathbb{R}} g(\lambda t)\,\gamma_{0,1}(dt). \tag{2.4.14}$$

Indeed, it suffices to check this when g is continuously differentiable and $g(0)=0$, in which case

$$\begin{aligned}\int_{\mathbb{R}^N} g\big(f(x)-m\big)\,\gamma_{0,I}(dx) &= \int_0^\infty g'(t)\gamma_{0,I}\big(\{x:\, f(x)-m\ge t\}\big)\,dt\\ &\qquad + \int_0^\infty g'(-t)\gamma_{0,I}\big(\{x:\, f(x)-m\le -t\}\big)\,dt\\ &\le \int_0^\infty g'(t)\big(1-\Phi(\lambda^{-1}t)\big)\,dt + \int_0^\infty g'(-t)\big(1-\Phi(\lambda^{-1}t)\big)\,dt\\ &= \lambda\int_0^\infty g'(\lambda t)\big(1-\Phi(t)\big)\,dt + \lambda\int_0^\infty g'(-\lambda t)\big(1-\Phi(t)\big)\,dt\\ &= \int_{\mathbb{R}} g(\lambda t)\,\gamma_{0,1}(dt).\end{aligned}$$

In particular

$$\int \exp\left(\frac{\alpha|f(x)-m|^2}{2\lambda^2}\right)\gamma_{0,I}(dx) \le \frac{1}{(1-\alpha)^{\frac12}} \text{ for } \alpha\in[0,1).$$

2.5 Constructing Gaussian Families

Let $\mathcal{I}$ be a non-empty index set, and let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space. Given a subset $\{X(\xi):\,\xi\in\mathcal{I}\}$ of $L^2(\mathbb{P};\mathbb{R})$, set

$$\begin{aligned}c(\xi,\eta) = \operatorname{cov}\big(X(\xi),X(\eta)\big) &\equiv \mathbb{E}\big[\big(X(\xi)-\mathbb{E}[X(\xi)]\big)\big(X(\eta)-\mathbb{E}[X(\eta)]\big)\big]\\ &= \mathbb{E}\big[X(\xi)X(\eta)\big] - \mathbb{E}\big[X(\xi)\big]\mathbb{E}\big[X(\eta)\big].\end{aligned}$$

The function c is called the *covariance function* for $\{X(\xi):\,\xi\in\mathcal{I}\}$.

There are three obvious properties that such a covariance function possesses. Namely, it is $\mathbb{R}$-valued, symmetric (i.e., $c(\xi,\eta)=c(\eta,\xi)$), and non-negative definite in the sense that, for all $n\ge 1$, $\{\xi_1,\dots,\xi_n\}\subseteq\mathcal{I}$, and $\{s_1,\dots,s_n\}\subseteq\mathbb{R}$,

$$\sum_{1\le k,\ell\le n} c(\xi_k,\xi_\ell)s_k s_\ell \ge 0. \tag{2.5.1}$$

To check the last property, simply observe that

$$\sum_{1\le k,\ell\le n} c(\xi_k,\xi_\ell)s_k s_\ell = \operatorname{var}\left(\sum_{k=1}^n s_k X(t_k)\right) \ge 0.$$

Given a symmetric function $c : \mathcal{I}^2 \longrightarrow \mathbb{R}$ satisfying (2.5.1), the goal in this subsection is to show it is the covariance function for a family of centered Gaussian random variables. That is, we will show that there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which there is a collection $\{X(\xi) : \xi \in \mathcal{I}\}$ of centered random variables with the property that the span of $\{X(\xi) : \xi \in \mathcal{I}\}$ is a Gaussian family such that, for all $n \geq 1$ and $\{s_1, \dots, s_n\} \subseteq \mathbb{R}$, $\sum_{j=1}^n s_j X(\xi_j)$ is a centered Gaussian random variable with variance

$$\sum_{1 \leq j,k \leq n} c(\xi_j, \xi_k) s_j s_k.$$

Such a family is called a *Gaussian process* with covariance c.

In order to show this, we will use a famous theorem of A. Kolmogorov known as *Kolmogorov's Consistency Theorem.* To state his result, for each $\xi \in \mathcal{I}$, let (E_ξ, ρ_ξ) be a complete, separable metric space. Given $\emptyset \neq S \subseteq \mathcal{I}$, set $\Omega_S = \prod_{\xi \in S} E_\xi$, and take $\Omega = \Omega_{\mathcal{I}}$. Thinking of Ω_S as the set of all functions $\omega_S : S \longrightarrow \bigcup_{\xi \in S} E_\xi$ such that $\omega_S(\xi) \in E_\xi$ for each $\xi \in S$, define $\pi_S : \Omega \longrightarrow \Omega_S$ so that $\pi_S \omega = \omega \restriction S$. If $\emptyset \neq F \subset\subset \mathcal{I}$ (i.e., F is a non-empty, finite subset of $\mathcal{I}$), give Ω_F the product topology, and set $\mathcal{A}_F = \{\pi_F^{-1} \Gamma : \Gamma \in \mathcal{B}_{\Omega_F}\}$. Finally, take $\mathcal{F} = \sigma(\mathcal{A})$, where

$$\mathcal{A} = \bigcup \{\mathcal{A}_F : \emptyset \neq F \subset\subset \mathcal{I}\},$$

and note that $\mathcal{A}$ is an algebra of subsets of Ω.

If, for each $\emptyset \neq F \subset\subset \mathcal{I}$, $\mu_F \in M_1(\Omega_F)$, the family $\{\mu_F : \emptyset \neq F \subset\subset \mathcal{I}\}$ is said to be *consistent*, if

$$\mu_{F_1}(\Gamma) = \mu_{F_2}\big(\{\omega_{F_2} \in \Omega_{F_2} : \omega_{F_2} \restriction F_1 \in \Gamma\}\big)$$

for all $\emptyset \neq F_1 \subset F_2 \subset\subset \mathcal{I}$ and $\Gamma \in \mathcal{B}_{\Omega_{F_1}}$.

Theorem 2.5.1 *Referring to the preceding, if $\{\mu_F : \emptyset \neq F \subset\subset \mathcal{I}\}$ is a consistent family, then there is a unique probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ such that, for all $\emptyset \neq F \subset\subset \mathcal{I}$ and $\Gamma \in \mathcal{B}_{\Omega_F}$,*

$$\mathbb{P}\big(\{\omega : \omega \restriction F \in \Gamma\}\big) = \mu_F(\Gamma).$$

Proof. Uniqueness is trivial in any case, and, when $\mathcal{I}$ is finite, there is nothing to do. Next, suppose that $\mathcal{I}$ is countable, in which case we may assume that $\mathcal{I} = \mathbb{Z}^+$ and define

$$\rho(\omega, \omega') = \sum_{m=1}^{\infty} 2^{-j} \frac{\rho_j\big(\omega(m), \omega'(m)\big)}{1 + \rho_j\big(\omega(m), \omega'(m)\big)}$$

for $\omega, \omega' \in \Omega$. One can easily check that ρ is a complete, separable metric for Ω and that ρ-convergence of $\{\omega_k : k \geq 1\}$ in Ω is equivalent to ρ_m-convergence of $\{\omega_k(m) : k \geq 1\}$ in E_m for each $m \geq 1$. In particular, the

σ-algebra $\mathcal{F}$ described above coincides with the Borel field $\mathcal{B}_\Omega$ determined by ρ.

Set $F_n = \{1, \dots, n\}$, $\Omega_n = \Omega_{F_n}$, $\pi_n = \pi_{F_n}$, $\mathcal{A}_n = \mathcal{A}_{F_n}$, and $\mu_n = \mu_{F_n}$ for $n \geq 1$. From the consistency hypothesis, we know that for $1 \leq m < n$ and $\Gamma \in \mathcal{B}_{\Omega_m}$,

$$\mu_m(\Gamma) = \mu_n\big(\{\omega_{F_n} : \omega_{F_n} \restriction F_m \in \Gamma\}\big).$$

Next, for each $m \geq 1$, choose an element $e_m \in E_m$, and define $\Phi_n : \Omega_n \longrightarrow \Omega$ so that

$$\Phi_n(\omega_{F_n})(m) = \begin{cases} \omega_{F_n}(m) & \text{if } 1 \leq m \leq n \\ e_m & \text{if } m > n. \end{cases}$$

Clearly, Φ_n is continuous, and $\pi_n \circ \Phi_n$ is the identity map on Ω_n. Now define $\mathbb{P}_n \in M_1(\Omega)$ to be $(\Phi_n)_*\mu_n$. Then

$$\mathbb{P}_n\big(\{\omega : \omega \restriction F_n \in \Gamma\}\big) = \mu_n(\Gamma)$$

for $\Gamma \in \mathcal{B}_{\Omega_n}$. What we need show is that there exists a $\mathbb{P} \in M_1(\Omega)$ such that $\mathbb{P} \restriction \mathcal{A}_n = \mathbb{P}_n \restriction \mathcal{A}_n$ for all $n \geq 1$, and to do so it suffices that show that there is a $\mathbb{P} \in M_1(\Omega)$ to which $\{\mathbb{P}_n : n \geq 1\}$ converges in the sense that $\langle \varphi, \mathbb{P}_n \rangle \longrightarrow \langle \varphi, \mathbb{P} \rangle$ for all $\varphi \in C_{\mathrm{b}}(\Omega; \mathbb{R})$. That there can be at most one $\mathbb{P}$ is obvious, and so, by Prohorov's Theorem (cf. Lemma 1.1.7), it suffices to prove that, for each $\epsilon > 0$, there is a compact set $K \subseteq \Omega$ such that $\inf_{n \geq 1} \mathbb{P}_n(K) \geq 1 - \epsilon$. To this end, let $\epsilon > 0$ be given, and, using Ulam's Lemma, choose a compact $K_1 \ni e_1$ in E_1 such that $\mu_1\big(K_1\big) \geq 1 - \frac{\epsilon}{2}$, and, for $n \geq 2$, choose a compact $K_n \ni e_n$ in E_n so that $\mu_n\big(\Omega_{n-1} \times K_n\big) \geq 1 - \frac{\epsilon}{2^n}$. A standard diagonalization argument shows that $K = \{\omega : \omega(n) \in K_n \text{ for } n \geq 1\}$ is a compact subset of Ω. In addition, if $A_n = \{\omega : \omega(m) \notin K_m \text{ for some } 1 \leq m \leq n\}$, then $A_n \nearrow \Omega \setminus K$. Finally, $\mathbb{P}_1(A_1) = \mu_1(E_1 \setminus K_1) \leq \frac{\epsilon}{2}$, and, for each $n \geq 2$,

$$\begin{aligned}
\mathbb{P}_n(A_n) &\leq \sum_{m=1}^{n} \mathbb{P}_n\big(\{\omega : \omega(m) \notin K_m\}\big) \\
&= \mu_1(E_1 \setminus K_1) + \sum_{m=2}^{n} \mathbb{P}_m\big(\{\omega : \omega(m) \notin K_m\}\big) \\
&= \mu_1(E_1 \setminus K_1) + \sum_{m=1}^{n} \mu_m\big(\Omega_{m-1} \times (E_m \setminus K_m)\big) \leq \epsilon \sum_{m=1}^{n} 2^{-m} \leq \epsilon.
\end{aligned}$$

Therefore, for any $1 \leq m \leq n$, $\mathbb{P}_n(A_m) = \mathbb{P}_m(A_m) \leq \epsilon$. At the same time, if $1 \leq n < m$, then, because $e_j \in E_j$ for all $j \geq 1$, $\Phi_n^{-1}(A_m) = \Phi_n^{-1}(A_n)$ and therefore that $\mathbb{P}_n(A_m) = \mathbb{P}_n(A_n) \leq \epsilon$. It follows that $\mathbb{P}_n(\Omega \setminus K) = \lim_{m \nearrow \infty} \mathbb{P}_n(A_m) \leq \epsilon$ for all $n \geq 1$, which means that $\mathbb{P}$ exists.

It remains to treat the case when $\mathcal{I}$ is uncountable. For each countable subset $S \subset \mathcal{I}$, let $\mathbb{P}_S$ be the measure just constructed on Ω_S. Then, by uniqueness, if $S_1 \subseteq S_2$,

$$\mathbb{P}_{S_2}\big(\{\omega_{S_2} : \omega_{S_2} \restriction S_1 \in A\}\big) = \mathbb{P}_{S_1}(A)$$

for $A \in \mathcal{B}_{\Omega_S}$. Hence we can define a finitely additive function $\mathbb{P}$ on the algebra $\mathcal{A}$ by setting $\mathbb{P}(\pi_S^{-1}A) = \mathbb{P}_S(A)$ for $A \in \mathcal{B}_{\Omega_S}$. Furthermore, if $\{A_k : k \ge 1\} \subseteq \mathcal{A}$ and $A_k \searrow \emptyset$, then we can choose a countable $S \subseteq \mathcal{I}$ such that $\{\pi_S A_k : k \ge 1\} \subseteq \mathcal{B}_{\Omega_S}$, and clearly $\pi_S A_k \searrow \emptyset$. Hence, $\mathbb{P}(A_k) = \mathbb{P}_S\big(\pi_S A_k\big) \searrow 0$, and therefore, by the Daniell Extention Theorem, $\mathbb{P}$ admits an extention to $\mathcal{F} = \sigma(\mathcal{A})$ as a probability measure. □

Corollary 2.5.2 *Let $\mathcal{I}$ be a non-empty set and $c : \mathcal{I}^2 \longrightarrow \mathbb{R}$ a symmetric function satisfying* (2.5.1). *Then there is a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which there is a collection of random variables $\{X(\xi) : \xi \in \mathcal{I}\}$ whose span is a centered Gaussian family for which c is the covariance function.*

Proof. Take $\Omega = \mathbb{R}^{\mathcal{I}}$, the space of all maps $\omega : \mathcal{I} \longrightarrow \mathbb{R}$, and define the σ-algebra $\mathcal{F}$ accordingly. Given a $\emptyset \ne F \subset\subset \mathcal{I}$, take μ_F to be the centered Gaussian measure on $\mathbb{R}^F$ with covariance matrix $A_F = \big(\big(c(\xi_j, \xi_k)\big)\big)_{\xi_j,\xi_k \in F}$. Then it is easy to check that $\big\{\mu_F : \emptyset \ne F \subset\subset \mathcal{I}\big\}$ is a consistent family. Now apply Theorem 2.5.1 to produce a probability measure on $(\Omega, \mathcal{F})$ with the property that, for all $n \ge 1$ and $\xi_1, \dots, \xi_n \in \mathcal{I}$, the distribution of $\big(\omega(\xi_1), \dots, \omega(\xi_n)\big)^\top$ is $\mu_{\{\xi_1,\dots,\xi_n\}}$, and conclude that $\{\omega(\xi) : \xi \in \mathcal{I}\}$ are random variables with the required property. □

Exercise 2.5.1 Show that if c_1 and c_2 are covariance functions on $\mathcal{I}^2$, then $c_1 c_2$ is also a covariance function there.

Hint: Construct a probability space on which there are mutually independent families $\{X_1(\xi) : \xi \in \mathcal{I}\}$ and $\{X_2(\xi) : \xi \in \mathcal{I}\}$ of centered random variables, the first with covariance function c_1 and the second with covariance function c_2. Alternatively, see if you can find a more direct proof.

2.5.1 Continuity Considerations

Unless $\mathcal{I}$ is countable, there is an inherent weakness in the conclusion of Theorem 2.5.1. To understand this weakness, consider the case in which $\mathcal{I} = \mathbb{R}$ and, for all $\xi \in \mathbb{R}$, E_ξ is the same separable Banach space E. In this case, Ω is the set of all maps $\omega : \mathbb{R} \longrightarrow E$ and $\mathcal{F}$ is the sigma algebra generated for the maps $\omega \rightsquigarrow \omega(\xi)$. Thus, the only events to which the measure $\mathbb{P}$ can assign a probability are those which depend on the values of ω at a countable number of times. In particular, because, for any sequence $\{\xi_m : m \ge 1\} \subseteq \mathbb{R}$ and every $\omega \in C(\mathbb{R};\mathbb{R})$ there is a discontinuous $\omega' \in \Omega$ that equals ω at all the ξ_n's, the only $\Gamma \in \mathcal{F}$ contained in $C(\mathbb{R};E)$ is empty. Hence, $C(\mathbb{R};E)$ has inner $\mathbb{P}$-measure 0, and therefore, unless $C(\mathbb{R};E)$ has $\mathbb{P}$-outer measure 0, it cannot be $\mathbb{P}$-measurable.

When presented with an uncountable collection $\{X(\xi) : \xi \in \mathcal{I}\}$ of random variables on a probability space $(\Omega, \mathcal{F}.\mathbb{P})$, one way to overcome the kind of problem raised above is to ask whether there is another family $\{\tilde{X}(\xi) : \xi \in \mathcal{I}\}$ which has the same distribution as $\{X(\xi) : \xi \in \mathcal{I}\}$ and has the desired property. With this in mind, one says that $\{\tilde{X}(\xi) : \xi \in \mathcal{I}\}$ is a *version* of $\{X(\xi) : \xi \in \mathcal{I}\}$ if $\tilde{X}(\xi) = X(\xi)$ (a.s., $\mathbb{P}$) for each $\xi \in \mathcal{I}$. Clearly, any version of $\{X(\xi) : \xi \in \mathcal{I}\}$ will have the same distribution as $\{X(\xi) : \xi \in \mathcal{I}\}$. To see how this idea applies to questions like continuity, again consider the setting described at the end of the preceding paragraph, and suppose that there exists a version $\{\tilde{\omega}(\xi) : \xi \in \mathbb{R}\}$ of $\{\omega(\xi) : \xi\mathbb{R}\}$ with the property that $\xi \rightsquigarrow \tilde{\omega}(\xi)$ is always continuous. Then even though the inner $\mathbb{P}$-measure of $\{X(\,\cdot\,)$ is continuous$\}$ is 0, its outer $\mathbb{P}$-measure is 1. To see this, suppose $\Gamma \in \sigma\big(\{\omega(\xi_m) : m \geq 1\}\big)$ contains $C(\mathbb{R}; E)$, and set $A = \{\omega : \omega(\xi_m) = \tilde{\omega}_m(\xi_m)$ for all $m \geq 1\}$. Then, $\mathbb{P}(A) = 1$ and, because $\{\tilde{\omega} : \omega \in \Omega\} \subseteq A$ and therefore $A \subseteq \Gamma$, $\mathbb{P}(\Gamma) = 1$.

We will now apply the preceding considerations to the Gaussian processes constructed in Corollary 2.5.2. Suppose that $\mathcal{I}$ is a metric space and that c is a covariance function on E^2. If there is a choice of random variables $X(\xi)$, $\xi \in E$, that are continuous with respect to ξ and form a centered Gaussian process for which c is the covariance function, then c must be a continuous function on E^2. Indeed, suppose $(\xi_n, \eta_n) \longrightarrow (\xi, \eta)$ in E^2. Then, by Lemma 2.1.1, $X(\xi_n) \longrightarrow X(\xi)$ and $X(\eta_n) \longrightarrow X(\eta)$ in $L^2(\mathbb{P}; \mathbb{R})$, and so

$$c(\xi_n, \eta_n) - c(\xi, \eta) = \mathbb{E}\big[\big(X(\xi_n) - X(\xi)\big)X(\eta_n)\big] + \mathbb{E}\big[X(\xi)\big(X(\eta_n) - X(\eta)\big)\big] \longrightarrow 0.$$

However, the converse statement is false. That is, just because c is continuous on E^2, does not mean that there is a Gaussian process $\{X(\xi) : \xi \in E\}$ on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with c as its covariance function with the property that $\xi \in E \rightsquigarrow X(\xi, \omega)$ is a continuous function for every $\omega \in \Omega$. In fact, even when $E = \mathbb{R}$, the condition on c that guarantees the existence of a continuous process is very technical.[2] Nonetheless, the following theorem of Kolmogorov enables us to prove that if $E = \mathbb{R}^N$ and, for some $\beta \in (0, 1]$,

$$\left\{ \frac{c(\xi, \xi) + c(\eta, \eta) - 2c(\xi, \eta)}{|\eta - \xi|^\beta} : |\eta| \vee |\xi| \leq R \right\} < \infty \text{ for all } R > 0, \qquad (2.5.2)$$

then a continuous choice exists.

In the proof of the following theorem, we use the fact that if Q is a closed cube in $\mathbb{R}^N$ and, for each vertex $\mathbf{v}$ of Q, $a_{\mathbf{v}}$ is an element of a vector space E, then there is a unique function $f : Q \longrightarrow E$, known as the multilinear extention of $\mathbf{v} \rightsquigarrow a_{\mathbf{v}}$, such that $f(\mathbf{v}) = a_{\mathbf{v}}$ for each vertex $\mathbf{v}$ and f is an affine function of each coordinate. For example, if $Q = [0, 1]^2$, then

[2] It's sufficiency was proved by Richard Dudley, and its necessity was proved later by Michel Talagrand.

$$f(\xi_1,\xi_2) = (1-\xi_1)(1-\xi_2)a_{(0,0)} + (1-\xi_1)\xi_2 a_{(0,1)} + \xi_1(1-\xi_2)a_{(1,0)} + \xi_1\xi_2 a_{(1,1)}.$$

The general case can be proved by translation, scaling, and induction on N.

Theorem 2.5.3 *Suppose that, for some cube $Q = [a,b]^N \subseteq \mathbb{R}^N$, $\{X(\xi) : \xi \in Q\}$ is a family of random variables taking values in a Banach space E, and assume that, for some $p \in [1,\infty)$, $C < \infty$, and $r \in (0,1]$,*

$$\mathbb{E}\big[\|X(\eta) - X(\xi)\|_E^p\big]^{\frac{1}{p}} \le C|\eta - \xi|^{\frac{N}{p}+r} \quad \textit{for all } \xi,\, \eta \in Q.$$

Then there exists a version $\{\tilde{X}(\xi) : \xi \in Q\}$ of $\{X(\xi) : \xi \in Q\}$ such that $\xi \in Q \longmapsto \tilde{X}(\xi)(\omega) \in E$ is continuous for all $\omega \in \Omega$. In fact, for each $\alpha \in [0,r)$, there is a $K < \infty$, depending only on N, p, r, and α, such that

$$\mathbb{E}\left[\sup_{\substack{\xi,\eta\in[0,R]^N\\ \xi\ne\eta}} \left(\frac{\|\tilde{X}(\eta) - \tilde{X}(\xi)\|_E}{|\eta-\xi|^\alpha}\right)^p\right]^{\frac{1}{p}} \le KC(b-a)^{\frac{N}{p}+r-\alpha}.$$

Proof. Given $\xi \in \mathbb{R}^N$, define $\|\xi\|_\infty = \max_{1\le j\le N} |\xi_j|$.

First note that, by an elementary translation and rescaling argument, it suffices to treat the case when $Q = [0,1]^N$.

Given $n \ge 0$, set

$$\begin{aligned} M_n &= \max_{\substack{\mathbf{k},\mathbf{m}\in\mathbb{N}^N\cap[0,2^n]^N\\ \|\mathbf{m}-\mathbf{k}\|_\infty=1}} \big\|X(\mathbf{m}2^{-n}) - X\big(\mathbf{k}2^{-n}\big)\big\|_E \\ &\le \left(\sum_{\substack{\mathbf{k},\mathbf{m}\in\mathbb{N}^N\cap[0,2^n]^N\\ \|\mathbf{m}-\mathbf{k}\|_\infty=1}} \big\|X(\mathbf{m}2^{-n}) - X\big(\mathbf{k}2^{-n}\big)\big\|_E^p\right)^{\frac{1}{p}}, \end{aligned}$$

and observe that

$$\mathbb{E}\big[M_n^p\big]^{\frac{1}{p}} \le \left(\sum_{\substack{\mathbf{k},\mathbf{m}\in\mathbb{N}^N\cap[0,2^n]^N\\ \|\mathbf{m}-\mathbf{k}\|_\infty=1}} \mathbb{E}\big[\big\|X(\mathbf{m}2^{-n}) - X\big(\mathbf{k}2^{-n}\big)\big\|_E^p\big]\right)^{\frac{1}{p}} \le C2^{-nr+\frac{N}{p}}.$$

Let $n \ge 0$ be given, and take $X_n(\,\cdot\,)$ to be the function that equals $X(\,\cdot\,)$ at the vertices of and is multilinear on each cube $\mathbf{m}2^{-n} + [0,2^{-n}]^N$. Because $X_{n+1}(\xi) - X_n(\xi)$ is a multilinear function on $\mathbf{m}2^{-n-1} + [0,2^{-n-1}]^N$,

$$\sup_{\xi\in[0,1]^N}\|X_{n+1}(\xi)-X_n(\xi)\|_E$$
$$=\max_{\mathbf{m}\in\mathbb{N}^N\cap[0,2^{n+1}]^N}\|X_{n+1}(\mathbf{m}2^{-n-1})-X_n(\mathbf{m}2^{-n-1})\|_E.$$

Since $X_{n+1}(\mathbf{m}2^{-n-1})=X(\mathbf{m}2^{-n-1})$ and either $X_n(\mathbf{m}2^{-n-1})=X(\mathbf{m}2^{-n-1})$ or

$$X_n(\mathbf{m}2^{-n-1})=\sum_{\substack{\mathbf{k}\in\mathbb{N}^N\cap[0,2^{n+1}]\\ \|\mathbf{k}-\mathbf{m}\|_\infty=1}}\theta_{\mathbf{m},\mathbf{k}}X(\mathbf{k}2^{-n-1}),$$

where the $\theta_{\mathbf{m},\mathbf{k}}$'s are non-negative and sum to 1, it follows that

$$\sup_{\xi\in[0,1]^N}\|X_{n+1}(\xi)-X_n(\xi)\|_E\le M_{n+1}$$

and therefore that

$$\mathbb{E}\left[\sup_{\xi\in[0,1]^N}\|X_{n+1}(\xi)-X_n(\xi)\|_E^p\right]^{\frac{1}{p}}\le C2^{-nr+\frac{N}{p}}.$$

Hence, for $0\le n<n'$,

$$\mathbb{E}\left[\sup_{n'>n}\sup_{\xi\in[0,1]^N}\|X_{n'}(\xi)-X_n(\xi)\|_E^p\right]^{\frac{1}{p}}\le\frac{C2^{-nr+\frac{N}{p}}}{1-2^{-r}},$$

and so $\{X_n: n\ge 0\}$ converges in $C\big([0,1]^N;E\big)$ both $\mathbb{P}$-almost surely and in $L^p\big(\mathbb{P};C([0,1]^N;E)\big)$. Therefore there exists a measurable map $\tilde{X}:[0,1]^N\times\Omega\longrightarrow E$ such that $\xi\rightsquigarrow\tilde{X}(\xi,\omega)$ is continuous for each $\omega\in\Omega$ and

$$\mathbb{E}\left[\sup_{\xi\in[0,1]^N}\|\tilde{X}(\xi)-X_n(\xi)\|_E^p\right]^{\frac{1}{p}}\le\frac{C2^{-nr+\frac{N}{p}}}{1-2^{-r}}.$$

Furthermore, $\tilde{X}(\xi)=X(\xi)$ (a.s., $\mathbb{P}$) if $\xi=\mathbf{m}2^{-n}$ for some $n\ge 0$ and $\mathbf{m}\in\mathbb{N}^N\cap[0,2^n]^N$, and therefore, since $\xi\rightsquigarrow\tilde{X}(\xi)$ is continuous and

$$\mathbb{E}\big[\|X(\mathbf{m}2^{-n})-X(\xi)\|_E^p\big]^{\frac{1}{p}}\le C2^{-n(\frac{N}{p}+r)}$$
$$\text{if } m_j2^{-n}\le\xi_j<(m_j+1)2^{-n}\text{ for }1\le j\le N,$$

it follows that $X(\xi)=\tilde{X}(\xi)$ (a.s.,$\mathbb{P}$) for each $\xi\in[0,1]^N$.

To prove the final estimate, suppose that $2^{-n-1}<|\eta-\xi|\le 2^{-n}$. Then

$$\|X_n(\eta)-X_n(\xi)\|_E\le N^{\frac{1}{2}}2^n|\xi-\eta|M_n,$$

and so, $\mathbb{P}$-almost surely,

$$\|\tilde{X}(\eta) - \tilde{X}(\xi)\|_E \le 2 \sup_{\beta \in [0,1]^N} \|\tilde{X}(\beta) - X_n(\beta)\|_E + N^{\frac{1}{2}} 2^n |\xi - \eta| M_n.$$

Hence, by the preceding,

$$\mathbb{E}\left[\sup_{\substack{\xi,\eta \in [0,1]^N \\ 2^{-n-1} < |\eta-\xi| \le 2^{-n}}} \left(\frac{\|\tilde{X}(\eta) - \tilde{X}(\xi)\|_E}{|\eta - \xi|^\alpha}\right)^p\right]^{\frac{1}{p}} \le C\left(\frac{2^{\frac{N}{p}+1}}{1-2^{-r}} + 2^{\frac{N}{p}} N^{\frac{1}{2}}\right) 2^{-n(r-\alpha)},$$

and therefore

$$\mathbb{E}\left[\left(\sup_{\substack{\xi,\eta \in [0,1]^N \\ \eta \ne \xi}} \frac{\|\tilde{X}(\eta) - \tilde{X}(\xi)\|_E}{|\eta - \xi|^\alpha}\right)^p\right]^{\frac{1}{p}} \le KC,$$

where $K = 2^{\frac{N}{p}} \left(\frac{2^r}{2^r - 1} + 4N^{\frac{1}{2}+\frac{1}{p}}\right)(1 - 2^{-(r-\alpha)})^{-1}$. □

Corollary 2.5.4 *Assume that there is a $p \in [1, \infty)$, $\beta > \frac{N}{p}$, and $C < \infty$ such that*

$$\mathbb{E}\big[\|\tilde{X}(\eta) - \tilde{X}(\xi)\|_E^p\big]^{\frac{1}{p}} \le C|\eta - \xi|^\beta \quad \textit{for all } \xi, \eta \in \mathbb{R}^N.$$

Then, for each $\gamma > \beta$,

$$\lim_{|\xi| \to \infty} \frac{\|\tilde{X}(\xi) - \tilde{X}(0)\|_E}{|\xi|^\gamma} = 0 \quad (\text{a.}s., \mathbb{P}) \textit{ and in } L^p(\mathbb{P}; E).$$

Proof. Take $\alpha = 0$ in Theorem 2.5.3. Then, because

$$\sup_{2^{n-1} \le \|\xi\|_\infty \le 2^n} \frac{\|\tilde{X}(\xi) - \tilde{X}(0)\|_E}{|\xi|^\gamma} \le 2^{-(n-1)\gamma} \sup_{2^{n-1} \|\xi\|_\infty \le 2^n} \|\tilde{X}(\xi) - \tilde{X}(0)\|_E,$$

$$\mathbb{E}\left[\left(\sup_{2^{n-1} \le \|\xi\|_\infty \le 2^n} \frac{\|\tilde{X}(\xi) - \tilde{X}(0)\|_E}{|\xi|^\gamma}\right)^p\right]^{\frac{1}{p}} \le 2^{-\gamma(n-1)} \mathbb{E}\left[\sup_{2^{n-1} \le \|\xi\|_\infty \le 2^n} \|\tilde{X}(\xi) - \tilde{X}(0)\|_E^p\right]^{\frac{1}{p}} \le 2^{\beta+\gamma} KC 2^{(\beta-\gamma)n},$$

and so

$$\mathbb{E}\left[\left(\sup_{\|\xi\|_\infty \geq 2^{m-1}} \frac{\|\tilde{X}(\xi) - \tilde{X}(0)\|_E}{|\xi|^\gamma}\right)^p\right]^{\frac{1}{p}} \leq \frac{2^{\beta+\gamma} KC}{1 - 2^{\beta-\gamma}} 2^{(\beta-\gamma)m}.$$

□

Corollary 2.5.5 *Suppose that c is a covariance function on $\mathbb{R}^N \times \mathbb{R}^N$ that satisfies* (2.5.2)*. Then, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which there is a Gaussian process $\{X(\xi) : \xi \in \mathbb{R}^N\}$ with covariance function c such that $X(\,\cdot\,, \omega) \in C(\mathbb{R}^N; \mathbb{R})$ for each $\omega \in \Omega$. Moreover, for each $\alpha < \frac{\beta}{2}$, $X(\,\cdot\,, \omega)$ is Hölder continuous of order α on compact sets. Finally, if there is a $C < \infty$ such that*

$$c(\eta, \eta) + c(\xi, \xi) - 2c(\xi, \eta) \leq C|\eta - \xi|^\beta \text{ for all } \xi,\, \eta \in \mathbb{R}^N,$$

then, for each $\alpha > \frac{\beta}{2}$, $\frac{|X(\xi,\omega)|}{|\xi|^\alpha} \longrightarrow 0$ as $|\xi| \to \infty$.

Proof. Let $\{X(\xi) : \xi \in \mathbb{R}^N\}$ be a Gaussian process with covariance function c. Then, for each $R > 0$,

$$\mathbb{E}\big[|X(\eta) - X(\xi)|^2\big] = c(\eta, \eta) + c(\xi, \xi) - 2c(\xi, \eta) \leq C_R|\eta - \xi|^\beta$$
$$\text{for some } C_R < \infty \text{ and all } \xi,\, \eta \in [-R, R]^N,$$

and so, for any $p \in [1, \infty)$,

$$\mathbb{E}\big[|X(\eta) - X(\xi)|^p\big]^{\frac{1}{p}} \leq K_p C_R |\eta - \xi|^{\frac{\beta}{2}},$$

where $K_p^p = \int |x|^p \,\gamma_{0,1}(dx)$. Now let $0 \leq \alpha < \frac{\beta}{2}$ be given, choose $0 < r < \frac{\beta}{2} - \alpha$, and determine $p \in [1, \infty)$ by $\frac{N}{p} = \frac{\beta}{2} - r$. Then

$$\mathbb{E}\big[|X(\eta) - X(\xi)|^p\big]^{\frac{1}{p}} \leq K_p C_R |\eta - \xi|^{\frac{N}{p}+r},$$

and so the continuity assertion follows from Theorem 2.5.3. Similarly, when C_R can be chosen independent of R, the concluding growth estimate follows from Corollary 2.5.4. □

2.5.2 Some Examples

By far the most renowned Gaussian process parameterized by the real numbers is the one constructed originally by N. Wiener and usually called Brownian motion. The covariance function for this process is

$$w(s,t) = \begin{cases} |s| \wedge |t| & \text{if } st > 0 \\ 0 & \text{if } st \leq 0. \end{cases} \tag{2.5.3}$$

To check that w is a covariance function, it suffices to check that its restriction to $[0,\infty)^2$ is and to observe that

$$s \wedge t = \int_0^\infty \mathbf{1}_{[0,\infty)}(s-u)\mathbf{1}_{[0,\infty)}(t-u)\,du \text{ for } s,t \geq 0,$$

and therefore that

$$\sum_{j,k=1}^n s_j \wedge s_k \alpha_j \alpha_k = \int_0^\infty \left(\sum_{j=1}^n \mathbf{1}_{[0,\infty)}(s_j - u) \right)^2 du \geq 0$$

for all choices of $s_1, \ldots, s_n \in [0,\infty)$ and $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$.

Theorem 2.5.6 *There exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which there is a centered Gaussian process $\{B(t) : t \in \mathbb{R}\}$ with covariance function w and having the property that, for each $\omega \in \Omega$, $B(\,\cdot\,,\omega)$ a Hölder continuous function of every order $0 \leq \alpha < \frac{1}{2}$, and*

$$\lim_{|t|\to\infty} \frac{|B(t,\omega)|}{|t|^\beta} = 0 \quad \textit{for every } \beta > \frac{1}{2}.$$

Proof. Simply observe that $w(t,t) + w(s,s) - 2w(s,t) = |t-s|$, and apply Corollary 2.5.5. □

From now on, I will say that a collection $\{X(t) : t \in \mathbb{R}\}$ of random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a *Brownian motion* if it is a centered Gaussian process with covariance w and $X(\,\cdot\,,\omega)$ is continuous for $\mathbb{P}$-almost every $\omega \in \Omega$. Clearly, $\mathbb{P}$-almost all of the paths $X(\,\cdot\,,\omega)$ will possesses all the properties described in Theorem 2.5.6. Observe that if $\{X(t) : t \in \mathbb{R}\}$ is a Brownian motion, then $\sigma\big(\{X(t) : t \geq 0\}\big)$ is independent of $\sigma\big(\{X(-t) : t \geq 0\}\big)$.

Wiener constructed this process to provide a mathematically rigorous foundation for A. Einstein's model of a physical phenomenon, first reported by a botanist named Brown, on which he was basing his kinetic theory of gases. However, at least for mathematicians, its renown does not derive from its connection to Einstein or the frequency of its appearance in models of physical, engineering, and even financial phenomena, but to the sometimes startling properties it possesses. The following provides a few elementary examples of these exotic properties.

Theorem 2.5.7 *Let $\{B(t) : t \in \mathbb{R}\}$ be a Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P})$.*

(**i**) *$\sigma\big(\{B(-t) : t \geq 0\}\big)$ is independent of $\sigma\big(\{B(t) : t \geq 0\}\big)$, for each $s \in [0,\infty)$, $B(s)$ is independent of $\{B(t_2) - B(t_1) : s \leq t_1 \leq t_2\}$, and, for any $n \geq 1$ and $t_0 < \cdots < t_n$, $\{B(t_m) - B(t_{m-1}) : 1 \leq m \leq n\}$ are mutually independent, $N(0, t_m - t_{m-1})$-random variables.*

(**ii**) *Both* $\{-B(t) : t \in \mathbb{R}\}$ *and* $\{B(-t) : t \in \mathbb{R}\}$ *are Brownian motions, and, for any* $\alpha > 0$, $\{\alpha^{-\frac{1}{2}} B(\alpha t) : t \in \mathbb{R}\}$ *is again a Brownian motion. In addition, for any* $T \in \mathbb{R}$, $\{B(t+T) - B(T) : t \in \mathbb{R}\}$ *is a Brownian motion.*

(**iii**) *Set* $\tilde{B}(0) = 0$ *and* $\tilde{B}(t) = |t| B\left(\frac{1}{t}\right)$ *for* $t \neq 0$. *Then* $\{\tilde{B}(t) : t \in \mathbb{R}\}$ *is a Brownian motion.*

(**vi**) *Define*

$$V_n(t) = \sum_{m=0}^{\lfloor 2^n t \rfloor - 1} \Big(B\big((m+1)2^{-n}\big) - B\big(m2^{-n}\big)\Big)^2 \text{ for } t > 0.$$

Then,

$$\lim_{n\to\infty} \sup_{s\in[0,t]} |V_n(s) - s| = 0 \quad (\text{a.s.}, \mathbb{P}) \text{ for all } t > 0.$$

In particular, $\mathbb{P}$*-almost no path* $B(\,\cdot\,)$ *has locally bounded variation or is locally Hölder continuous of any order larger than* $\frac{1}{2}$.

Proof. By Theorem 2.4.3, to verify (**i**) it suffices to observe that $\mathbb{E}\big[B(t)B(-t)\big] = 0$ for all $t \in \mathbb{R}$, $\mathbb{E}\big[\big(B(t_2) - B(t_1)\big)B(s)\big] = 0$ if $0 \le s \le t_1 \le t_2$, and $\mathbb{E}\big[\big(B(t_2) - B(t_1)\big)\big(B(t_4) - B(t_3)\big)\big] = 0$ for $t_1 < t_2 < t_3 < t_4$.

Since all the processes described in (**ii**) are centered Gaussian processes with continuous paths, all that one needs to do is check that w is their covariance function. Similarly, the collection in (**iii**) is a centered Gaussian process with covariance function w, and so it suffices to check that its paths are $\mathbb{P}$-almost surely continuous, which comes down to showing that they are continuous at 0. But we know that $\lim_{|t|\to\infty} \frac{B(t)}{t} = 0$ (a.s.,$\mathbb{P}$), and so $\lim_{t\to 0} |t| B\left(\frac{1}{t}\right) = 0$ (a.s.,$\mathbb{P}$).

Turning to (**iv**), define $\Delta_{m,n} = B\big((m+1)2^{-n}\big) - B\big(m2^{-n}\big)$ for $m \ge 0$. Clearly, $\Delta_{m,n} \in N(0, 2^{-n})$, and, by (**i**), for each $n \ge 0$, the $\Delta_{m,n}$'s are mutually independent. Now set $Y_{m,n} = \Delta_{m,n}^2 - 2^{-n}$ and $S_n(m) = \sum_{k=0}^{m} Y_{k,n}$. Then, for each $n \ge 0$, the $Y_{m,n}$'s are mutually independent random variable with mean 0 and variance $2 \cdot 4^{-n}$, and

$$|V_n(t) - t| \le |V_n(t) - 2^{-n}\lfloor 2^n t \rfloor| + 2^{-n} = |S_n(\lfloor 2^n t \rfloor - 1)| + 2^{-n}.$$

Thus it suffices to show that $\sup\big\{|S_n(m)| : 0 \le m \le 2^n t\big\} \longrightarrow 0$. But, by Kolmogorov's inequality,

$$\mathbb{P}\left(\sup_{0\le m\le 2^n t} |S_n(m)| \ge \epsilon\right) \le \epsilon^{-2}\mathbb{E}\big[S_n\big(\lfloor 2^n t\rfloor\big)^2\big] \le \epsilon^{-2} 2^{1-n} t,$$

and so the proof of the first part of (**iv**) is complete. To prove the rest, use (**ii**) to see that it suffices to consider the paths restricted to $[0,1]$. Next observe that if $\varphi : [0,1] \longrightarrow \mathbb{R}$ is Hölder continuous of order $\alpha > \frac{1}{2}$, then

$$\sum_{m=0}^{2^n-1}\Big(\varphi\big((m+1)2^{-n}\big)-\varphi\big(m2^{-n}\big)\Big)^2 \le \|\varphi\|_{C^\alpha([0,1];\mathbb{R})}2^{n(1-2\alpha)} \longrightarrow 0.$$

Also, if φ is continuous and of bounded variation and ρ is the modulus of continuity for φ, then

$$\sum_{m=0}^{2^n-1}\Big(\varphi\big((m+1)2^{-n}\big)-\varphi\big(m2^{-n}\big)\Big)^2 \le \|\varphi\|_{\mathrm{var}}\rho(2^{-n}) \longrightarrow 0.$$

Hence, $\mathbb{P}$-almost no Brownian path can have either of these properties. □

Given a collection $\{X(t) : t \in \mathbb{R}\}$ of random variables, set $\mathcal{F}_s = \sigma\big(\{X(\tau) : \tau \in (-\infty, s]\}\big)$ for $s \in \mathbb{R}$. Then $\{X(t) : t \in \mathbb{R}\}$ is a *Markov process* if, for all $s < t$, the conditional distribution of $X(t)$ given $\sigma\big(\{X(s)\}\big)$ is the same as that of $X(t)$ given $\mathcal{F}_s$. If $\{X(t) : t \in \mathbb{R}\}$ is a Markov process and, for $s < t$, $x \in \mathbb{R} \rightsquigarrow P(s, x; t, \cdot\,) \in M_1(\mathbb{R})$ is a measurable map (i.e., $x \rightsquigarrow P(s, x; t, \Gamma)$ is measurable for all $\Gamma \in \mathcal{B}_{\mathbb{R}}$) such that

$$\mathbb{E}\big[\{X(t) \in \Gamma\}\,\big|\,\mathcal{F}_s\big] = P\big(s, X(s); t, \Gamma\big) \text{ for } \Gamma \in \mathcal{B}_{\mathbb{R}},$$

then $P(s, x; t, \cdot\,)$ is called the *transition probability function* for $\{X(t) : t \in \mathbb{R}\}$. Finally, $\{X(t) : t \in \mathbb{R}\}$ is said to be a *homogeneous Markov process* if it has a transition probability function that, as a function of $s < t$, depends only on $t - s$. That is, $P(s, x; t, \cdot\,) = P(t - s, x, \cdot\,) \equiv P(0, x; t - s, \cdot\,)$.

An important fact about a Brownian motion $\{B(t) : t \in \mathbb{R}\}$ is that $\{B(t^+) : t \in \mathbb{R}\}$ is a Markov process. Indeed, if $\mathcal{F}_s = \sigma\big(\{B(\tau^+) : \tau \in (-\infty, s]\}\big)$, then, for all $s < t$ and bounded measurable $\varphi : \mathbb{R} \longrightarrow \mathbb{R}$,

$$\mathbb{E}\big[\varphi\big(B(t^+)\big)\,\big|\,\mathcal{F}_s\big] = \int \varphi\big(B(s^+) + y\big)\,\gamma_{0,t^+-s^+}(dy)$$

since, by (**i**) in Theorem 2.5.7, $B(t^+) - B(s^+)$ is independent of $\mathcal{F}_s$ and is a centered Gaussian with variance $t^+ - s^+$. Therefore $\{B(t^+) : t \in \mathbb{R}\}$ is a homogeneous Markov process with transition probability γ_{x,t^+-s^+}.

Lemma 2.5.8 *Let $\{X(t) : t \in \mathbb{R}\}$ be a Gaussian process with covariance function c. If $\{X(t) : t \in \mathbb{R}\}$ is a Markov process, then*

$$c(r,t)c(s,s) = c(r,s)c(s,t) \text{ for all } r \le s < t. \tag{2.5.4}$$

Conversely, if c satisfies (2.5.4) *and $t \in \mathbb{R} \rightsquigarrow c(t,t) \in \mathbb{R}$ is continuous, then $\{X(t) : t \in \mathbb{R}\}$ is a Markov process with transition probability function $P(s, x; t, \cdot\,) = \gamma_{b(s,t)x,a(s,t)}$, where*

$$b(s,t) = \begin{cases} \frac{c(s,t)}{c(s,s)} & \text{if } c(s,s) > 0 \\ 0 & \text{if } c(s,s) = 0, \end{cases}$$

$$a(s,t) = \begin{cases} c(t,t) - \frac{c(s,t)^2}{c(s,s)} & \text{if } c(s,s) > 0 \\ c(t,t) & \text{if } c(s,s) = 0, \end{cases}$$

and $\gamma_{b,0} = \delta_b$.

Proof. First suppose that $\{X(t) : t \in \mathbb{R}\}$ is a Markov process. Given $s < t$, either $c(s,s) = 0$, in which case $c(r,s) = 0$ and therefore (2.5.4) holds, or $c(s,s) > 0$, in which case $X(t) - \frac{c(s,t)}{c(s,s)}X(s)$ is independent of $X(s)$ and therefore $\mathbb{E}\big[X(t) \,\big|\, \mathcal{F}_s\big] = \frac{c(s,t)}{c(s,s)}X(s)$, which means that

$$c(r,t) = \mathbb{E}\big[X(t)X(r)\big] = \frac{c(s,t)}{c(s,s)}\mathbb{E}\big[X(s)X(r)\big] = \frac{c(r,s)c(s,t)}{c(s,s)}.$$

Conversely, assume $t \rightsquigarrow c(t,t)$ is continuous and that (2.5.4) holds. If $r \le s < t$ and $c(s,s) > 0$, then

$$\mathbb{E}\Big[\big(X(t) - b(s,t)X(s)\big)X(r)\Big] = c(r,t) - \frac{c(s,t)c(r,s)}{c(s,s)} = 0,$$

and so $X(t) - b(s,t)X(s)$ is a centered Gaussian with variance $a(s,t)$ that is independent of $\mathcal{F}_s$.

What remains is to handle the case when $c(s,s) = 0$ for some $s \in \mathbb{R}$. To this end, let $t \in \mathbb{R}$ be given. There is no problem if either $c(t,t) = 0$ or $c(s,s) > 0$ for all $s \le t$. Now set $s_0 = \sup\{s \le t : c(s,s) = 0\}$, and assume that $s_0 > -\infty$. By continuity, $c(s_0,s_0) = 0$, and so we are done if $s_0 = t$. If $s_0 < t$, then we have to show that $c(t,r) = 0$ for all $r < s_0$. But $r < s_0 \implies c(t,r)c(r,r) = c(r,s_0)c(s_0,t) = 0$, which, since $c(r,r) = 0 \implies c(r,t) = 0$, is possible only if $c(r,t) = 0$. □

2.5.3 Stationary Gaussian Processes

A stochastic process $\{X(\xi) : \xi \in V\}$, where V is a vector space, is said to be *stationary* if, for all $\eta \in V$, the distribution of $\{X(\xi+\eta) : \xi \in V\}$ is the same as that of $\{X(\xi) : \xi \in V\}$. For a centered Gaussian process, it is easy to show that stationarity is equivalent to its covariance function c being translation invariant. That is $c(\xi,\eta) = c(0,\eta-\xi)$ for all $\xi,\eta \in V$.

When $V = \mathbb{R}^N$, c is translation invariance if and only if there is an $\mathbb{R}$-valued, non-negative definite function f such that $c(\xi,\eta) = f(\eta-\xi)$, which means that $\frac{f}{f(0)}$ must be a characteristic function for a symmetric probability measure if c is continuous. Thus we have a ready source of translation invariant covariance functions. Notice that, by Corollary 2.5.5, a centered Gaussian process with covariance function $f(\eta-\xi)$ will admit a continuous version if f is Hölder continuous of some positive order at 0.

Theorem 2.5.9 *Suppose that $f : \mathbb{R} \longrightarrow \mathbb{R}$ is a continuous, non-negative definite function and that $\{X(t) : t \in \mathbb{R}\}$ is a continuous, centered Gaussian process with covariance function $f(t-s)$ on $(\Omega, \mathcal{F}, \mathbb{P})$. If $f(t) \longrightarrow 0$ as $t \to \infty$, then $\{X(t) : t \in \mathbb{R}\}$ is ergodic. Equivalently, for any bounded, Borel measurable $\varphi : \mathbb{R} \longrightarrow \mathbb{R}$,*

$$\lim_{T \nearrow \infty} \frac{1}{T} \int_0^T \varphi\big(X(t)\big)\, dt = \mathbb{E}\big[\varphi\big(X(0)\big)\big] = \int \varphi(y)\, \gamma_{0,f(0)}(dy) \quad \text{(a.s., } \mathbb{P}).$$

Proof. Without loss in generality, we will assume that $\Omega = C(\mathbb{R};\mathbb{R})$ and $\mathcal{F} = \sigma\big(\{\omega(t) : t \in \mathbb{R}\}\big)$.

For each $s \in \mathbb{R}$, define the time-shift map $T_s : \Omega \longrightarrow \Omega$ by $T_s\omega(t) = \omega(s+t)$. Clearly each T_s is a $\mathcal{F}$-measurable, $\mathbb{P}$-measure preserving map, and $T_s \circ T_t = T_{t+s}$ for all $s, t \in \mathbb{R}$.

By the Individual Ergodic Theorem, what we have to show is that if $A \in \mathcal{F}$ is time shift invariant (i.e., $A = T_t^{-1}A$ for all $t \geq 0$), then $\mathbb{P}(A) \in \{0,1\}$. For that purpose, we will show that, for any $A \in \mathcal{F}$,

$$\mathbb{P}\big(A \cap T_t^{-1}A\big) \longrightarrow \mathbb{P}(A)^2 \quad \text{as } t \to \infty, \tag{$*$}$$

from which it is clear that $\mathbb{P}(A) = \mathbb{P}(A)^2$ if A is time-shift invariant.

Let $\mathcal{A}$ be the collection of all subsets having the form

$$\big\{\omega : \big(\omega(t_1), \dots, \omega(t_n)\big)^\top \in \Gamma\big\}$$

for some $n \geq 1$, $t_1 < \cdots < t_n$, and $\Gamma \in \mathcal{B}_{\mathbb{R}^n}$, and let $\mathcal{C}$ be the set of $A \in \mathcal{F}$ for which $(*)$ holds. Clearly $\mathcal{A}$ is an algebra which generates $\mathcal{F}$ and $\mathcal{C}$ is closed under complementation. Thus, if we can show that $\mathcal{C} \supseteq \mathcal{A}$ and is closed under non-decreasing limits, then we will know that $\mathcal{C} = \mathcal{F}$.

To prove that $\mathcal{A} \subseteq \mathcal{C}$, let $A = \big\{\omega : \big(\omega(t_1), \dots \omega(t_n)\big)^\top \in \Gamma\big\}$, where $t_1 < \cdots < t_n$. By Lemma 2.4.1, we may assume that the matrix $B = \big(\big(f(t_j - t_k)\big)\big)_{1 \leq j,k \leq n}$ is non-degenerate. Next, for $t > t_n - t_1$, define $C(t) = \big(\big(f(t_j - t_k - t)\big)\big)_{1 \leq j,k \leq n}$ and

$$B(t) = \begin{pmatrix} B & C(t) \\ C(t)^\top & B \end{pmatrix}.$$

Then $\gamma_{0,B(t)}$ is the joint distribution of

$$\big(\omega(t_1), \dots \omega(t_n)\big)^\top \text{ and } \big(\omega(t_1 + t), \dots \omega(t_n + t)\big)^\top,$$

and so $\mathbb{P}(A \cap T_t^{-1}A) = \gamma_{0,B(t)}(\Gamma \times \Gamma)$. Since

$$B(t) \longrightarrow \begin{pmatrix} B & 0 \\ 0 & B \end{pmatrix} \quad \text{as } t \to \infty$$

and B is non-degenerate, $\gamma_{0,B(t)}$ converges in variation to $\gamma_{0,B}^2$, which means that $\mathbb{P}(A \cap T_t^{-1}A) \longrightarrow \gamma_{0,B}(\Gamma)^2 = \mathbb{P}(A)^2$.

Finally, suppose that $\{A_n : n \geq 1\} \subseteq \mathcal{C}$ and that $A_n \nearrow A$. Then

$$0 \leq \mathbb{P}\big(A \cap T_t^{-1}A\big) - \mathbb{P}\big(A_n \cap T_t^{-1}A_n\big) \leq \mathbb{P}(A \setminus A_n) + \mathbb{P}\big(T^{-1}A_n \setminus T_t^{-1}A_n\big) = 2\mathbb{P}(A \setminus A_n).$$

Hence, if $(*)$ holds for each A_n, then it also holds for A. □

We now have two examples of the influence that properties of f have on the properties of the associated Gaussian process: continuity properties of f are reflected in continuity properties of the paths, and decay properties of f at infinity influence long term properties, in particular ergodicity, of the process.

Periodicity provides a more dramatic example. To wit, take $f(t) = \cos(2\pi t)$. Of course one can use Corollary 2.5.5 to construct an associated Gaussian process, but there is a more revealing way to do so. Namely, take $\Omega = \mathbb{R}^2$ and $\mathbb{P} = \gamma_{0,I}$, and define

$$X(t, \xi) = \xi_1 \cos(2\pi t) + \xi_2 \sin(2\pi t) \text{ for } t \in \mathbb{R} \text{ and } \xi \in \Omega.$$

Clearly $\{X(t) : t \in \mathbb{R}\}$ is a centered Gaussian process under $\mathbb{P}$. Moreover,

$$\mathbb{E}\big[X(s)X(t)\big] = \cos(2\pi s)\cos(2\pi t) + \sin(2\pi s)\sin(2\pi t) = \cos\big(2\pi(t-s)\big).$$

Thus $\cos\big(2\pi(t-s)\big)$ is the covariance function for $\{X(t) : t \in \mathbb{R}\}$.

As J. Doob noticed, there are very few $\mathbb{R}$-valued characteristic functions $f : \mathbb{R} \longrightarrow \mathbb{R}$ for which the corresponding covariance function satisfies (2.5.4). Indeed, if (2.5.4) holds, $f(\xi + \eta) = f(\xi)f(\eta)$ for all $\xi, \eta \geq 0$. Thus, if $g(\xi) = \operatorname{sgn}(\xi)\log f(\xi)$, then $g(\xi + \eta) = g(\xi) + g(\eta)$ for all $\xi, \eta \in \mathbb{R}$, and so, by Theorem 2.2.4, $f(\xi) = e^{-\alpha|\xi|}$ for some $\alpha \geq 0$. As a consequence, if $\{X(t) : t \in \mathbb{R}\}$ is the centered Gaussian process with covariance function $e^{-\frac{|t|}{2}}$, then any other stationary centered Gaussian process that is Markov will have the same distribution as $\big\{aX(bt) : t \in \mathbb{R}\big\}$ for some $a, b \geq 0$. Because they were introduced by Ornstein and Uhlenbeck, these are called stationary *Ornstein–Uhlenbeck* processes.

There is an interesting way to construct a stationary Ornstein–Uhlenbeck process from a Brownian motion. Namely, let $\{B(t) : t \in \mathbb{R}\}$ be a Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ that supports a random variable $X_0 \in N(0,1)$ that is independent of $\{B(t) : t \in \mathbb{R}\}$, and define

$$X(t) = e^{-\frac{|t|}{2}}\Big(X_0 + B\big(\operatorname{sgn}(t)(e^{|t|} - 1)\big)\Big). \tag{2.5.5}$$

It is not hard to check that $\{X(t) : t \in \mathbb{R}\}$ is an stationary Ornstein–Uhlenbeck process with covariance function $e^{-\frac{|t|}{2}}$. For aficionados of Brownian motion, this representation facilitates computations. For example, it makes clear that, for $s < t$, $X(t) - e^{-\frac{t-s}{2}}X(s)$ is independent of $\mathcal{F}_s =$

$\sigma\big(\{X(\tau) : \tau \le s\}$ and has distribution $\gamma_{0,1-e^{-t}}$. Hence, $\{X(t) : t \in \mathbb{R}\}$ is a homogeneous Markov process, and the function $p(t,x,y)$ that appeared in (2.3.2) is the density of its transition probability function.

Exercise 2.5.2

(**i**) Assume that $f \in C^2(\mathbb{R};\mathbb{R})$ is a characteristic function whose second derivative is Hölder continuous of some positive order. Set $\mathcal{I} = \{(s,t) \in \mathbb{R}^2 : st = 0\}$, and define

$$c\big((s,0);(t,0)\big) = f(t-s),\ c\big((s,0);(0,t)\big) = c\big((0,t);(s,0)\big) = f'(t-s),\\ \text{and } c\big((0,s);(0,t)\big) = -f''(t-s).$$

Show that c is a covariance function on $\mathcal{I}$ and that there is a continuous centered Gaussian process $\{Z(s,t) : (s,t) \in \mathcal{I}\}$ for which c is the covariance function. Next, set $X(t) = Z(t,0)$ and $Y(t) = Z(0,t)$, and show that $\{X(t) : t \in \mathbb{R}\}$ is a centered Gaussian process with covariance function $f(t-s)$, that $\{Y(t) : t \in \mathbb{R}\}$ is a centered Gaussian process with covariance function $-f''(t-s)$, and that, almost surely,

$$X(t) - X(s) = \int_s^t Y(\tau)\,d\tau \text{ for all } s < t.$$

Equivalently, $X(\,\cdot\,) \in C^1(\mathbb{R};\mathbb{R})$ and $Y(\,\cdot\,) = \dot{X}(\,\cdot\,)$ almost surely.

(**ii**) Show that if f is an $\mathbb{R}$-valued characteristic function that is periodic and $\{X(t) : t \in \mathbb{R}\}$ is a continuous, centered Gaussian process for which $f(t-s)$ is the covariance function, show that almost surely $X(\,\cdot\,)$ is periodic with same period as f. In addition, if $f \in C^2(\mathbb{R};\mathbb{R})$ and f'' is Hölder continuous of some positive order, show that $X(t)$ is independent of $\dot{X}(t)$ for each $t \in \mathbb{R}$.

(**iii**) Let $\{B(t) : t \in \mathbb{R}\}$ be a Brownian motion. Using the representation of an Ornstein–Uhlenbeck process in terms of a Brownian motion, show that, for each $\alpha > 0$,

$$\lim_{|t|\to\infty} \frac{|B(t)|}{|t|^{\frac{1}{2}}(\log|t|)^{\alpha}} = 0$$

almost surely. This improves our earlier result but is still far short of the result coming from the law of the iterated logarithm (cf. § 3.5.2) which says that

$$\varlimsup_{|t|\to\infty} \frac{|B(t)|}{\sqrt{2t\log(\log t)}} = 1.$$

(**iv**) Let $\{B(t) : t \in \mathbb{R}\}$ be a Brownian motion, and define

$$B_0(t) = B(t) - \psi(t)B(1), \text{ where } \psi(t) = (t \wedge 1)^+.$$

Show that $B(1)$ is independent of $\sigma\big(\{B_0(t) : t \in \mathbb{R}\}\big)$. Next, given a non-negative, $\sigma\big(\{B(t) : t \in \mathbb{R}\}\big)$-measurable function F, set $\bar{F}(y) = \mathbb{E}\big[F \circ (B +$

$\psi y)\big]$ for $y \in \mathbb{R}$, and show that

$$\mathbb{E}\big[F \circ B,\, B(1) \in \Gamma\big] = \int_\Gamma \bar{F}(y)\, \gamma_{0,1}(dy)$$

for $\Gamma \in \mathcal{B}_{\mathbb{R}}$. Conclude that

$$\mathbb{E}\big[F \circ B \,\big|\, \sigma\big(\{B(1)\}\big)\big] = \bar{F}\big(B(1)\big).$$

In particular, the distribution of $\{B_0(t) : t \in \mathbb{R}\}\big)$ is that of a Brownian motion conditioned to be at 0 at time $t = 1$. For this reason, $\{B_0(t) : t \in \mathbb{R}\}$ is sometimes called a *pinned Brownian motion.*

(**v**) Referring to (**iv**), show that $\{B_0(t) : t \geq 0\}$ is a Markov process. Further, if $P(s, x; t, \cdot\,)$ is its transition probability function, show that

$$P(s, x; t, \cdot\,) = \gamma_{\frac{1-t}{1-s}x,\, \frac{(1-t)(t-s)}{1-s}} \text{ for } 0 < s < t < 1.$$

(**vi**) Let $\{X(t) : t \in \mathbb{R}\}$ be an Ornstein–Uhlenbeck process with coveriance function $e^{-\frac{|t|}{2}}$, and set $\mathcal{F}_s = \sigma\big(\{X(\tau) : \tau \in (-\infty, s]\}\big)$ for $s \in \mathbb{R}$. If H_n is the nth Hermite polynomial, show that, for each $n \in \mathbb{N}$,

$$\Big(e^{\frac{n|t|}{2}} H_n\big(X(t)\big), \mathcal{F}_s, \mathbb{P}\Big)$$

is a martingale.

(**vii**) Let $\{B(t) : t \in \mathbb{R}\}$ be a Brownian motion and set $\mathcal{F}_s = \sigma\big(\{B(\tau^+) : \tau \in (-\infty, s]\}\big)$ for $s \in \mathbb{R}$. Show that,

$$\left(\exp\left(\xi B(t^+) - \frac{t^+\xi^2}{2}\right), \mathcal{F}_s, \mathbb{P}\right)$$

is a martingale for each $\xi \in \mathbb{R}$. Next, for each $n \in \mathbb{N}$, check that $(t, x) \rightsquigarrow t^{\frac{n}{2}} H_n\big(t^{-\frac{1}{2}}x\big)$ is a polynomial in $(t, x) \in \mathbb{R}^2$ and that

$$\Big((t^+)^{\frac{n}{2}} H_n\big((t^+)^{-\frac{1}{2}} B(t^+)\big), \mathcal{F}_s, \mathbb{P}\Big)$$

is a martingale.

Hint: Use (2.3.18).

Chapter 3
Gaussian Measures on a Banach Space

3.1 Motivation

The theories of Gaussian measures and Hilbert spaces are inextricably related, and the contents of this chapter is to explain and explore that relationship.

To understand the relationship, it is best to begin in the finite dimensional setting. Let H be an N-dimensional Hilbert space. A Borel probability measure $\mathcal{W}$ on H is said to the standard Gauss measure there if $\{(\,\cdot\,,h)_H : h \in H\}$ is a centered Gaussian family under $\mathcal{W}$ with covariance function $(g,h)_H$. To see that such a $\mathcal{W}$ exists, let $\{h_m : 1 \le m \le N\}$ be an orthonormal basis in H, define $\Phi : \mathbb{R}^N \longrightarrow H$ by $\Phi(x) = \sum_{m=1}^N x_m h_m$, and take $\mathcal{W} = \Phi_* \gamma_{0,1}^N$. To see that there is only one such W, note that Φ is invertible and that $\Phi^{-1}(h) = \big((h,h_1)_H, \dots, (h,h_N)_H\big)^\top$, and check that $(\Phi^{-1})_* \mathcal{W} = \gamma_{0,1}^N$. Alternatively, set $\lambda_H = \Phi_* \lambda_{\mathbb{R}^N}$, check that λ_H is the unique translation invariant Borel measure on H that assigns measure 1 to the unit cube $\{h \in H : (h,h_m)_H \in [0,1] \text{ for } 1 \le m \le N\}$, and show that

$$\mathcal{W}(dh) = (2\pi)^{-\frac{N}{2}} e^{-\frac{\|h\|_H^2}{2}} \, \lambda_H(dh).$$

Given a finite dimensional Banach space B, let $\langle x, \xi\rangle$ denote the action of $\xi \in B^*$ on $x \in B$, and say that a $\mathcal{W} \in M_1(B)$ is a centered Gaussian measure on B if $\{\langle\,\cdot\,,\xi\rangle : \xi \in B^*\}$ is a centered Gaussian family under $\mathcal{W}$. Assuming that, in addition, $\mathcal{W}$ is non-degenerate in the sense that $\operatorname{var}\big(\langle\,\cdot\,,\xi\rangle\big) > 0$ if $\xi \ne 0$, there is a unique Hilbert structure on B such that $\mathcal{W}$ is the standard Gauss measure for B with that Hilbert stucture. Namely, let c be the covariance function for $\{\langle\,\cdot\,,\xi\rangle : \xi \in B^*\}$, and observe that c is a bilinear, symmetric form on B^* such that $c(\xi,\xi) > 0$ unless $\xi = 0$. Thus, for each $\xi \in B^*$ there is a unique $A\xi \in B$ such that $\langle A\xi, \eta\rangle = c(\xi,\eta)$ for all $\eta \in B^*$, and the map $A : B^* \longrightarrow B$ is linear isomorphism. Now define $(x,y)_H = \langle x, A^{-1}y\rangle$ for $x, y \in B$. Clearly $(x,y) \rightsquigarrow (x,y)_H$ is bilinear. In addition, if $x = A\xi$ and

$y = A\eta$, then $(x, y)_H = \langle A\xi, \eta\rangle = c(\xi, \eta)$, from which it follows that $(\,\cdot\,,\cdot\cdot\,)_H$ is an inner product on B. Finally, if $y_1, y_2 \in B$, then

$$\begin{aligned}\int (x, y_1)_H (x, y_2)_H \,\mathcal{W}(dx) &= \int \langle x, A^{-1}y_1\rangle\langle x, A^{-1}y_2\rangle\,\mathcal{W}(dx)\\ &= c\big(A^{-1}y_1, A^{-1}y_2\big) = (y_1, y_2)_H.\end{aligned}$$

Hence, $\mathcal{W}$ is the standard Gaussian measure for the Hilbert space B with inner product $(\,\cdot\,,\cdot\cdot)_H$.

The situation is much more complicated when B is infinite dimensional, and all the complications stem from the fact that there is no standard Gauss measure on an infinite dimensional Hilbert space. Indeed, suppose that $\mathcal{W}$ were the standard Gauss measure on the infinite dimensional Hilbert space H, and let $\{h_m : m \geq 0\}$ be an orthonormal sequence in H. Set $X_m(x) = (x, h_m)_H$, and observe that $\{X_m : m \geq 0\}$ would be a sequence of mutually independent $N(0, 1)$ random variables under $\mathcal{W}$. Thus

$$\int \exp\left(-\sum_{m=0}^{\infty} X_m(x)^2\right)\mathcal{W}(dx) = \lim_{n\to\infty}\left(\int e^{-t^2}\,\gamma_{0,1}(dt)\right)^n = \lim_{n\to\infty} 3^{-\frac{n}{2}} = 0,$$

which leads to the contradiction that $\|x\|_H^2 \geq \sum_{m=0}^{\infty} X_m(x)^2 = \infty$ for $\mathcal{W}$-almost every $x \in H$. When H is separable, and therefore $\dim(H)$ is countable, there is a more intuitive reason why $\mathcal{W}$ cannot exist. Namely, if $\mathcal{W}$ existed, then one would guess that

$$\mathcal{W}(dh) = \frac{1}{Z} e^{-\frac{\|h\|_H^2}{2}}\,\lambda_H(dh), \tag{3.1.1}$$

where $Z = (2\pi)^{\frac{\dim(H)}{2}}$ and λ_H is the translation invariant Borel measure on H with gives measure 1 to the unit cube $\{h \in H : (h, h_m)_H \in [0, 1] \text{ for } m \in \mathbb{N}\}$, where $\{h_m : m \geq 0\}$ is an orthonormal basis in H. However, even though R. Feynman was able to make remarkable computations using it, the formula in (3.1.1) defies mathematical rationalization: the measure λ_H does not exist and $Z = \infty$.

It was to overcome the problems caused by the preceding non-existence result that I. Sigal and L. Gross [6] developed the theory to which this chapter is devoted.

3.2 Some Background

Besides standard results about Hilbert and Banach spaces, I will need a few less familiar ones. Throughout, the Banach spaces here will be over $\mathbb{R}$, and I will use B^* to denote the dual of B and $\langle x, \xi\rangle$ to denote the action of $\xi \in B^*$ on $x \in B$.

Lemma 3.2.1 *If B is a separable Banach space, then, for each $R > 0$, the weak* topology on $\overline{B_{B^*}(0,R)}$ is compact and is metrizable. Hence the weak* topology on any bounded subset of B^* is second countable (i.e., is generated by a countable neighborhood basis).*

Proof. It suffices to show that the weak* topology on $\overline{B_{B^*}(0,1)}$ is metrizable and compact. To this end, choose a sequence $\{x_n : n \geq 1\} \subseteq B_B(0,1)$ of linearly independent elements whose span S is dense in B, and define

$$\rho(\xi,\nu) = \sum_{n=1}^{\infty} 2^{-n} |\langle x_n, \eta - \xi\rangle| \text{ for } \xi,\, \eta \in \overline{B_{B^*}(0,1)}.$$

Let $\{\xi_k : k \geq 1\} \subseteq \overline{B_{B^*}(0,1)}$ be given. Clearly, ρ is a metric, and $\rho(\xi_k, \xi) \longrightarrow 0$ if $\{\xi_k : k \geq 1\}$ is weak* convergent to ξ. Conversely, if $\rho(\xi_k,\xi)) \longrightarrow 0$, then $\lim_{k\to\infty}\langle x_n, \xi_k\rangle \longrightarrow \langle x_n, \xi\rangle$ for each $n \geq 1$ and therefore $\langle x, \xi_k\rangle \longrightarrow \langle x, \xi\rangle$ for all $x \in S$. Since S is dense in B and $\|\xi_k\|_{B^*} \leq 1$ for all $k \geq 1$, it follows that $\{\xi_k : k \geq 1\}$ is weak* convergent to ξ. Thus ρ is a metric for the weak* topology on $\overline{B_{B^*}(0,1)}$.

To prove compactness, use a diagonalization procedure to extract a subsequence $\{\xi_{k_j} : j \geq 1\}$ such that $a_m = \lim_{j\to\infty}\langle x_m, \xi_{k_j}\rangle$ exists for each $m \geq 1$, and set, for $x \in S$, $f(x) = \sum_{m=1}^n \alpha_m a_m$ if $x = \sum_{m=1}^n \alpha_m x_m$. Because the x_m's are linearly independent, f is a well defined linear function on S. Furthermore, $f(x) = \lim_{j\to\infty}\langle x, \xi_{k_j}\rangle$ and therefore $|f(x)| \leq \|x\|_B$ for $x \in S$. Hence there is a $\xi \in \overline{B_{B^*}(0,1)}$ such that

$$\lim_{j\to\infty}\langle x, \xi_{k_j}\rangle = f(x) = \langle x, \xi\rangle$$

for $x \in S$, and, because S is dense in B, this means that $\xi_{k_j} \longrightarrow \xi$ in the weak* topology. □

Given a Borel probability measure μ on a Banach space B, define its characteristic function $\hat{\mu} : B^* \longrightarrow \mathbb{C}$ by

$$\hat{\mu}(\xi) = \int e^{i\langle x,\xi\rangle}\, \mu(dx).$$

Lemma 3.2.2 *If B is a separable Banach space, then $\mathcal{B}_B$ is the smallest σ-algebra with respect to which $x \in B \rightsquigarrow \langle x, \xi\rangle \in \mathbb{R}$ is measurable for all $\xi \in B^*$. In particular, if μ, $\nu \in M_1(E)$, then $\mu = \nu$ if and only if $\hat{\mu} = \hat{\nu}$.*

Proof. Clearly $x \in B \rightsquigarrow \langle x, \xi\rangle \in \mathbb{R}$ is $\mathcal{B}_B$ measurable for all $\xi \in B^*$. To prove that sets in $\mathcal{B}_B$ are measurable with respect to the σ-algebra generated by these maps, it suffices to show that $\|\cdot\|_B$ is measurable with respect to them.

For that purpose, choose a sequence $\{\xi_n : n \geq 1\} \subseteq B^*$ which is weak* dense in $\overline{B_{B^*}(0,1)}$, and check that $\|x\|_B = \sup_{n\geq 1}\langle x, \xi_n\rangle$.

Given the preceding, we know that $\mu = \nu$ on $\mathcal{B}_B$ if, for all $\xi_1, \dots, \xi_n \in B^*$, the distribution of the map $x \in B \rightsquigarrow \big(\langle x, \xi_1\rangle, \cdots, \langle x, \xi_n\rangle\big)^\top \in \mathbb{R}^n$ is the same under μ as it is under ν, and, by Lemma 1.1.1, this will be true if and only if $\hat{\nu} = \hat{\mu}$. □

I will also be making use of Bochner's theory of integration for Banach space valued functions. What follows is a brief outline of his theory.

Let B be a separable, real Banach space and $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space. A function $X : \Omega \longrightarrow B$ is said to be simple if X is $\mathcal{F}$-measurable and X takes only finitely many values, in which case its expected value with respect to $\mathbb{P}$ is the element of B given by

$$\mathbb{E}[X] = \int_\Omega X(\omega)\, \mathbb{P}(d\omega) \equiv \sum_{x\in B} x\, \mathbb{P}(X = x).$$

Notice that an equivalent description of $\mathbb{E}[X]$ is as the unique element of B with the property that

$$\big\langle \mathbb{E}[X], \xi\big\rangle = \mathbb{E}\big[\langle X, \xi\rangle\big] \quad \text{for all } \xi \in B^*,$$

and therefore that the mapping taking simple X to $\mathbb{E}[X]$ is linear. Next, observe that, by the same argrument as was used in the proof of Lemma 3.2.2, $\omega \in \Omega \rightsquigarrow \|X(\omega)\|_B \in \mathbb{R}$ is $\mathcal{F}$-measurable if $X : \Omega \longrightarrow B$ is $\mathcal{F}$-measurable, and say that X is integrable if $\|X\|_B$ is. The space of B-valued integrable functions will be denoted by $L^1(\mathbb{P}; B)$, and, as usual, we will identify elements of $L^1(\mathbb{P}; B)$ that differ on a set of $\mathbb{P}$-measure 0 and will take $\|X\|_{L^1(\mathbb{P};B)} = \mathbb{E}\big[\|X\|_B\big]$.

Bochner's definition of the integral of an B-valued, integrable X is completed in the following theorem.

Theorem 3.2.3 *If $X : \Omega \longrightarrow B$ is integrable, then there is a unique element $\mathbb{E}[X] \in B$ satisfying $\langle \mathbb{E}[X], \xi\rangle = \mathbb{E}[\langle X, \xi\rangle]$ for all $\xi \in B^*$. In particular, the mapping $X \in L^1(\mathbb{P}; B) \longmapsto \mathbb{E}[X] \in B$ is linear and satisfies*

$$\big\|\mathbb{E}[X]\big\|_B \leq \mathbb{E}\big[\|X\|_B\big]. \tag{3.2.1}$$

Finally, if $X \in L^1(\mathbb{P}; B)$, then there is a sequence $\{X_n : n \geq 1\}$ of B-valued, simple functions with the property that $\|X_n - X\|_{L^1(\mathbb{P};B)} \longrightarrow 0$.

Proof. Clearly uniqueness, linearity, and (3.2.1) all follow immediately from the given characterization of $\mathbb{E}[X]$. Thus, what remains is to prove existence and the final approximation assertion. In fact, once the approximation assertion is proved, then existence will follow immediately from the observation that, by (3.2.1), $\mathbb{E}[X]$ can be taken equal to $\lim_{n\to\infty} \mathbb{E}[X_n]$ if $\|X - X_n\|_{L^1(\mathbb{P};B)} \longrightarrow 0$.

To prove the approximation assertion, begin with the case when $M = \sup_{\omega\in\Omega}\|X(\omega)\|_B < \infty$. Choose a dense sequence $\{x_\ell : \ell \ge 1\}$ in B, and set $A_{0,n} = \emptyset$ and

$$A_{\ell,n} = \left\{\omega : \|X(\omega) - x_\ell\|_E < \tfrac{1}{n}\right\} \quad \text{for } \ell,\, n \ge 1.$$

Then, for each $n \in \mathbb{Z}^+$ there exists an $L_n \in \mathbb{Z}^+$ with the property that

$$\mathbb{P}\left(\Omega \setminus \bigcup_{\ell=1}^{L_n} A_{\ell,n}\right) < \frac{1}{n}.$$

Hence, if $X_n : \Omega \longrightarrow B$ is defined so that

$$X_n(\omega) = x_\ell \quad \text{when } 1 \le \ell \le L_n \text{ and } \omega \in A_{\ell,n} \setminus \bigcup_{k=0}^{\ell-1} A_{k,n}$$

and $X_n(\omega) = 0$ when $\omega \notin \bigcup_1^{L_n} A_{\ell,n}$, then X_n is simple and

$$\mathbb{E}\big[\|X - X_n\|_B\big] \le \frac{M+1}{n}.$$

In order to handle the general case, let $X \in L^1(\mathbb{P}; B)$ and $n \in \mathbb{Z}$ be given. We can then find a sequence $r_n \nearrow \infty$ with the property that

$$\int\limits_{\Omega(r_n)\complement} \|X(\omega)\|_B \, \mu(d\omega) \le \frac{1}{2n},$$

where $\Omega(r) = \big\{\omega : \|X(\omega)\|_B \le r\big\}$ for $r \ge 1$. Since, $r\mathbb{P}\big(\Omega(r)\big) \le \mathbb{E}\big[\|X\|_B\big]$, we can apply the preceding to the restrictions of $\mathbb{P}$ and X to $\Omega(r_n)$ and thereby find a simple $X_n : \Omega(r_n) \longrightarrow B$ with the property

$$\mathbb{E}\big[\|X - X_n\|_B,\, \Omega(r_n)\big] \le \frac{1}{2n}.$$

Hence, after taking $X_n = 0$ off of $\Omega(r_n)$, we arrive at a simple X_n for which $\mathbb{E}\big[\|X - X_n\|_B\big] \le \frac{1}{n}$. $\square$

Once one has the results in Theorem 3.2.3, with little or no change in the arguments, one can prove for integrals of Banach space valued functions all the standard results for $\mathbb{R}$-valued ones that don't rely of special properties of $\mathbb{R}$. In particular, one can introduce the Lebesgue spaces $L^p(\mu; B)$ for $p \in [1, \infty)$ with norm $\|X\|_{L^p(\mu;B)} = \mathbb{E}\big[\|X\|_B^p\big]^{\frac{1}{p}}$ and can show that they are separable Banach spaces.

Having developed Bochner's integral for Banach space valued random variables, I will now introduce the corresponding notion of conditional expectation values.

Theorem 3.2.4 *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, B a separable Banach space, and $X \in L^1(\mathbb{P}; B)$. Then*

$$\mathbb{P}\big(X \neq 0\big) = 0 \iff \mathbb{E}\big[X,\, A\big] = 0 \text{ for all } A \in \mathcal{F}.$$

Next, if Σ is a sub-σ-algebra of $\mathcal{F}$, then there is a $\mathbb{P}$-almost everywhere unique Σ-measurable $X_\Sigma \in L^1(\mathbb{P}; B)$ such that

$$\mathbb{E}\big[X_\Sigma,\, A\big] = \mathbb{E}\big[X,\, A\big] \quad \text{for every } A \in \Sigma. \tag{3.2.2}$$

In particular, if Y is a second element of $L^1(\mathbb{P}; B)$, then, for all $\alpha,\, \beta \in \mathbb{R}$,

$$\big(\alpha X + \beta Y\big)_\Sigma = \alpha X_\Sigma + \beta Y_\Sigma \quad (\text{a.s.}, \mathbb{P}).$$

Finally,

$$\big\|X_\Sigma\big\|_B \leq \mathbb{E}\big[\|X\|_B \,\big|\, \Sigma\big] \quad (\text{a.s.}, \mathbb{P}). \tag{3.2.3}$$

Proof. Clearly, it is only necessary to prove the "$\Longleftarrow$" part of the first assertion. Thus, suppose that $\mathbb{P}(X \neq 0) > 0$. Then, because B is separable and therefore $B_{B^*}(0,1)$ with the weak* topology is also separable, there exists an $\epsilon > 0$ and a $\xi \in B_{B^*}(0,1)$ with the property that $\mu\big(\langle X, \xi\rangle \geq \epsilon\big) > 0$, from which it follows that there is an $A \in \mathcal{F}$ for which

$$\big\langle \mathbb{E}\big[X,\, A\big], \xi\big\rangle = \mathbb{E}\big[\langle X, \xi\rangle,\, A\big] \neq 0.$$

Turning to the uniqueness and other properties of X_Σ, it is obvious that uniqueness follows immediately from the preceding and that linearity follows from uniqueness. As for (3.2.3), notice that if $\xi \in B^*$ and $\|\xi\|_{B^*} \leq 1$, then

$$\mathbb{E}\big[\langle X_\Sigma, \xi\rangle,\, A\big] = \mathbb{E}\big[\langle X, \xi\rangle,\, A\big] \leq \mathbb{E}\big[\|X\|_B,\, A\big] = \mathbb{E}\Big[\mathbb{E}\big[\|X\|_B \,\big|\, \Sigma\big],\, A\Big]$$

for every $A \in \Sigma$. Hence, by the theory of conditional expectation for $\mathbb{R}$-valued random variables, $\langle X_\Sigma, \xi\rangle \leq \big(\|X\|_B\big)_\Sigma$ (a.s., $\mathbb{P}$) for each element ξ from the unit ball in B^*; and so, because $B_{B^*}(0,1)$ with the weak* topology is separable, (3.2.3) follows.

Finally, to prove the existence of X_Σ, suppose that X is simple, let R denote its range, and note that

$$X_\Sigma \equiv \sum_{x \in R} x\, \mathbb{P}\big(X = x \,\big|\, \Sigma\big)$$

has the required properties. In order to handle general $X \in L^1(\mathbb{P}; E)$, use the approximation result in Theorem 3.2.3 to find a sequence $\{X_n : n \geq 1\}$ of simple functions that tend to X in $L^1(\mathbb{P}; E)$. Then, since

$$(X_n)_\Sigma - (X_m)_\Sigma = \big(X_n - X_m\big)_\Sigma \quad (\text{a.s.}, \mathbb{P})$$

and therefore, by (3.2.3),

$$\mathbb{E}\big[\|(X_n)_\Sigma - (X_m)_\Sigma\big\|_B\big] \le \mathbb{E}\big[\|X_n - X_m\|_B\big],$$

we know that there exists a Σ-measurable $X_\Sigma \in L^1(\mathbb{P}; E)$ to which the sequence $\big\{(X_n)_\Sigma : n \ge 1\big\}$ converges in $L^1(\mathbb{P}; \mathbb{R})$; and clearly X_Σ has the required properties. □

In the future, I will call X_Σ the conditional expectation of X given Σ and will use $\mathbb{E}[X \mid \Sigma]$ to denote it. The following are essentially immediate consequences of uniqueness:

$$\mathbb{E}\big[YX\big|\,\Sigma\big] = Y\,\mathbb{E}\big[X\big|\Sigma\big] \quad (\text{a.s.}, \mathbb{P}) \quad \text{for bounded } \Sigma\text{-measurable } Y,$$

and

$$\mathbb{E}\big[X\big|\mathcal{T}\big] = \mathbb{E}\Big[\mathbb{E}\big[X\big|\Sigma\big]\Big|\mathcal{T}\Big] \quad (\text{a.s}, \mathbb{P})$$

whenever $\mathcal{T}$ is a sub-σ-algebra of Σ.

Once one knows how to take the conditional expectation of Banach space valued random variables, one should investigate what can be said about Banach space valued martingales. That is, given a non-decreasing sequence $\{\mathcal{F}_n : n \ge 0\}$ of sub-σ-algebras and an adapted sequence $\{X_n : n \ge 0\} \subseteq L^1(\mathbb{P}; B)$ (i.e., X_n is $\mathcal{F}_n$ measurable for each n) with the property that $X_n = \mathbb{E}[X_{n+1}|\mathcal{F}_n]$ for all $n \ge 0$, the triple $(X_n, \mathcal{F}_n, \mathbb{P})$ is called a *martingale*, and one should wonder how many of the properties of $\mathbb{R}$-valued martingales it has. As the following theorem demonstrates, many of the results for $\mathbb{R}$-valued martingales have easily proved analogs for Banach space valued ones. An exception is Doob's Martingale Convergence, which does not hold in general.

Theorem 3.2.5 *Let B be a separable Banach space and $\big(X_n, \mathcal{F}_n, \mathbb{P}\big)$ a B-valued martingale. Then $\big(\|X_n\|_B, \mathcal{F}_n, \mathbb{P}\big)$ is a non-negative submartingale and therefore, for each $N \in \mathbb{Z}^+$ and all $R \in (0, \infty)$,*

$$\mathbb{P}\left(\sup_{0\le n\le N} \|X_n\|_B \ge R\right) \le \frac{1}{R}\mathbb{E}\left[\|X_N\|_B,\ \sup_{0\le n\le N}\|X_n\|_B \ge R\right]. \qquad (3.2.4)$$

In particular, for each $p \in (1, \infty)$,

$$\mathbb{E}\left[\sup_{n\in\mathbb{N}} \|X_n\|_B^p\right]^{\frac{1}{p}} \le \frac{p}{p-1}\sup_{n\in\mathbb{N}}\mathbb{E}\big[\|X_n\|_B^p\big]^{\frac{1}{p}}. \qquad (3.2.5)$$

Finally, if $\mathcal{F} = \sigma\Big(\bigcup_{n\ge 0}\mathcal{F}_n\Big)$, $X \in L^1(\mathbb{P}; B)$, and $X_n = \mathbb{E}[X \mid \mathcal{F}_n]$, then $\|X_n - X\|_B \longrightarrow 0$ (a.s.,$\mathbb{P}$) and in $L^1(\mathbb{P}; \mathbb{R})$.

Proof. The fact that $\big(\|X_n\|_B, \mathcal{F}_n, \mu\big)$ is a submartingale is an easy application of the inequality in (3.2.3); and, given this fact, the inequalities in (3.2.4) and (3.2.5) follows from the result in the $\mathbb{R}$-valued case.

Let $X \in L^1(\mathbb{P}; B)$ be given, and set $X_n = \mathbb{E}[X|\mathcal{F}_n],\ n \in \mathbb{N}$. Because of (3.2.4), we know that the set of X for which $X_n \longrightarrow X$ (a.s.,$\mathbb{P}$) is a closed subset of $L^1(\mathbb{P}; B)$. Moreover, if X is simple, then the $\mathbb{P}$-almost everywhere convergence of X_n to X follows easily from the $\mathbb{R}$-valued result. Hence, we now know that $X_n \longrightarrow X$ (a.s,μ) for each $X \in L^1(\mu; B)$. To prove that the convergence is taking place in $L^1(\mathbb{P}; B)$, note that, by Fatou's Lemma,

$$\|X\|_{L^1(\mathbb{P};B)} \le \varliminf_{n\to\infty} \|X_n\|_{L^1(\mathbb{P};B)},$$

whereas (3.2.3) guarantees that

$$\|X\|_{L^1(\mathbb{P};B)} \ge \varlimsup_{n\to\infty} \|X_n\|_{L^1(\mathbb{P};B)}.$$

Hence, because

$$\Big|\|X_n\|_B - \|X\|_B - \|X_n - X\|_B\Big| \le 2\|X\|_B,$$

the convergence in $L^1(\mathbb{P}; B)$ is an application of Lebesgue's Dominated Convergence Theorem. □

3.2.1 Fernique's Theorem

A Borel probability measure $\mathcal{W}$ on separable Banach space B is said to a centered *Gaussian measure* if $\{\langle\,\cdot\,, \xi\rangle : \xi \in B^*\}$ is a centered Gaussian family under $\mathcal{W}$. Equivalently, if

$$c(\xi, \eta) = \int \langle x, \xi\rangle\langle x, \eta\rangle \text{ for } \xi, \eta \in B^*,$$

then c is the covariance function for this family and $\widehat{\mathcal{W}}(\xi) = e^{-\frac{c(\xi,\xi)}{2}}$ for $\xi \in B^*$.

The engine that drives much of what follows is the following remarkable result of X. Fernique. See Corollary 3.3.13 for a sharper statement.

Theorem 3.2.6 *Let $\mathcal{W}$ be a centered Gaussian measure on the separable Banach space B. If*

$$R = \inf\big\{r : \mathbb{P}(\|x\|_B \ge r) \le \tfrac{1}{10}\big\},$$

then

$$\int e^{\frac{\|x\|_B^2}{18R^2}}\,\mathcal{W}(dx) \le K \equiv e^{\frac{1}{2}} + \sum_{n=0}^{\infty}\left(\frac{e}{3}\right)^{2^n}. \tag{3.2.6}$$

Proof. Set $\mathbb{P} = \mathcal{W}^2$ on $\Omega = B^2$, and define $X_1\bigl(x_1,x_2\bigr) = x_1$ and $X_2\bigl(x_1,x_2\bigr) = x_2$. Then X_1 and X_2 are independent B-valued random variables under $\mathbb{P}$. Furthermore, if $Y_1 = \frac{x_1+x_2}{2^{\frac{1}{2}}}$ and $Y_2 = \frac{x_1-x_2}{2^{\frac{1}{2}}}$, then, for all $\xi_1,\xi_2 \in B^*$,

$$\begin{aligned}
&\mathbb{E}^{\mathbb{P}}\bigl[e^{i\langle Y_1,\xi_1\rangle}e^{i\langle Y_2,\xi_2\rangle}\bigr] = \widehat{\mathcal{W}}\bigl(2^{-\frac{1}{2}}(\xi_1+\xi_2)\bigr)\widehat{\mathcal{W}}\bigl(2^{-\frac{1}{2}}(\xi_1-\xi_2)\bigr)\\
&\quad = \exp\left(-\frac{c(\xi_1+\xi_2,\xi_1+\xi_2)}{4} - \frac{c(\xi_1-\xi_2,\xi_1-\xi_2)}{4}\right) = e^{-\frac{c(\xi_1,\xi_1)}{2}}e^{-\frac{c(\xi_2,\xi_2)}{2}}\\
&\quad = \mathbb{E}^{\mathbb{P}}\bigl[e^{i\langle X_1,\xi_1\rangle}e^{i\langle X_2,\xi_2\rangle}\bigr],
\end{aligned}$$

and so, by Lemma 3.2.2, (Y_1,Y_2) has the same distribution under $\mathbb{P}$ as (X_1,X_2).

Let $0 < s \le t$ be given, and use the preceding to justify

$$\begin{aligned}
&\mathbb{P}\bigl(\|X_1\|_B \le s\bigr)\mathbb{P}\bigl(\|X_1\|_B \ge t\bigr) = \mathbb{P}\bigl(\|X_1\|_B \le s \;\&\; \|X_2\|_B \ge t\bigr)\\
&\quad = \mathbb{P}\bigl(\|X_1 - X_2\|_B \le 2^{\frac{1}{2}}s \;\&\; \|X_1+X_2\|_B \ge 2^{\frac{1}{2}}t\bigr)\\
&\quad \le \mathbb{P}\bigl(\bigl|\|X_1\|_B - \|X_2\|_B\bigr| \le 2^{\frac{1}{2}}s \;\&\; \|X_1\|_E + \|X_2\|_B \ge 2^{\frac{1}{2}}t\bigr)\\
&\quad \le \mathbb{P}\bigl(\|X_1\|_B \wedge \|X_2\|_B \ge 2^{-\frac{1}{2}}(t-s)\bigr) = \mathbb{P}\bigl(\|X_1\|_B \ge 2^{-\frac{1}{2}}(t-s)\bigr)^2.
\end{aligned}$$

Now suppose that $\mathbb{P}\bigl(\|X_1\|_B \le R\bigr) \ge \frac{9}{10}$, and define $\{t_n : n \ge 0\}$ by $t_0 = R$ and $t_n = R + 2^{\frac{1}{2}}t_{n-1}$ for $n \ge 1$. Then

$$\mathbb{P}\bigl(\|X_1\|_B \le R\bigr)\mathbb{P}\bigl(\|X_1\|_B \ge t_n\bigr) \le \mathbb{P}\bigl(\|X_1\|_B \ge t_{n-1}\bigr)^2$$

and therefore

$$\frac{\mathbb{P}\bigl(\|X_1\|_B \ge t_n\bigr)}{\mathbb{P}\bigl(\|X_1\|_B \le R\bigr)} \le \left(\frac{\mathbb{P}\bigl(\|X_1\|_B \ge t_{n-1}\bigr)}{\mathbb{P}\bigl(\|X_1\|_B \le R\bigr)}\right)^2$$

for $n \ge 1$. Working by induction, one gets from this that

$$\frac{\mathbb{P}\bigl(\|X_1\|_B \ge t_n\bigr)}{\mathbb{P}\bigl(\|X_1\|_B \le R\bigr)} \le \left(\frac{\mathbb{P}\bigl(\|X_1\|_B \ge R\bigr)}{\mathbb{P}\bigl(\|X_1\|_B \le R\bigr)}\right)^{2^n}$$

and therefore, since $t_n = \frac{2^{\frac{n+1}{2}}-1}{2^{\frac{1}{2}}-1}R \le 3\cdot 2^{\frac{n+1}{2}}R$, that $\mathbb{P}\bigl(\|X_1\|_B \ge 3\cdot 2^{\frac{n}{2}}R\bigr) \le 3^{-2^n}$. Hence,

$$\mathbb{E}^{\mathbb{P}}\left[e^{\frac{\|X_1\|_E^2}{18R^2}}\right] \le e^{\frac{1}{2}}\mathbb{P}\big(\|X_1\|_B \le 3R\big) + \sum_{n=0}^{\infty} e^{2^n}\mathbb{P}\big(3\cdot 2^{\frac{n}{2}}R \le \|X_1\|_B \le 3\cdot 2^{\frac{n+1}{2}}R\big)$$
$$\le e^{\frac{1}{2}} + \sum_{n=0}^{\infty}\left(\frac{e}{3}\right)^{2^n} = K.$$

□

A B-valued random variable X on a probability space $(\Omega,\mathcal{F},\mathbb{P})$ is said to be a centered Gaussian random variable if, for all $\xi \in B^*$, $\langle X,\xi\rangle$ is a centered Gaussian random variable, which is equivalent to saying that $X_*\mathbb{P}$ is a centered Gaussian measure on B.

It should be obvious that Theorem 3.2.6 provides the following vast generalization of the estimate (2.4.3) in Lemma 2.4.2.

Corollary 3.2.7 *Let X be a B-valued, centered Gaussian random variable on $(\Omega,\mathcal{F},\mathbb{P})$. Then*

$$\mathbb{P}\big(\|X\|_B \ge R\big) \le \frac{1}{10} \implies \mathbb{E}\big[e^{\alpha\|X\|_B^2}\big] \le K,$$

where $\alpha = (18R^2)^{-1}$ and K is the one in Theorem 3.2.6. Hence, if $\{X_n : n \ge 1\}$ is a sequence of B-valued centered Gaussian random variable on $(\Omega,\mathcal{F},\mathbb{P})$ that converge in probability to a random variable X, then X is a centered Gaussian random variable and $\mathbb{E}\big[\|X_n - X\|_B^p\big] \longrightarrow 0$ for all $p \in [1,\infty)$.

Proof. The second assertion follows immediately from the first. To prove the first, simply set $\mathcal{W} = X_*\mathbb{P}$ and apply Theorem 3.2.6. □

Exercise 3.2.1

(**i**) Let $\{x_n : n \ge 0\}$ be a sequence in the separable Banach space B with the property that $\sum_{n=0}^{\infty}\|x_n\|_B < \infty$. Show that $\sum_{n=0}^{\infty}|\xi_n|\|x_n\|_B < \infty$ for $\gamma_{0,1}^{\mathbb{N}}$-almost every $\xi \in \mathbb{R}^{\mathbb{N}}$, and define $X : \mathbb{R}^{\mathbb{N}} \longrightarrow B$ so that $X(\xi) = \sum_{n=0}^{\infty}\xi_n x_n$ if $\sum_{n=0}^{\infty}|\xi_n|\|x_n\|_B < \infty$ and $X(\xi) = 0$ otherwise. Show that the distribution μ of X is a centered, Gaussian measure on B. In addition, show that μ is non-degenerate if and only if the span of $\{x_n : n \ge 0\}$ is dense in B.

(**ii**) Here an application of Fernique's Theorem and part (**i**) to functional analysis. Let E and F be a pair of separable Banach spaces and ψ a Borel measurable, linear map from E to F. Given a centered, Gaussian E-valued random variable X, use Lemma 2.2.3 to see that $\psi\circ X$ is an F-valued, centered Gaussian random variable, and apply Fernique's Theorem to conclude that $\psi\circ X$ is a square integrable and has mean value 0. Next, suppose that ψ were not continuous, and choose $\{x_n : n \ge 0\} \subseteq E$ and $\{\eta_n : n \ge 0\} \subseteq F^*$ so that $\|x_n\|_E = 1 = \|\eta_n\|_{F^*}$ and $\langle\psi(x_n),\eta_n\rangle \ge (n+1)^3$. Using part (**i**), show that there exist centered, Gaussian F-valued random variables $\{X_n : n \ge$

$0\} \cup \{X\}$, such that $X_n(\xi) = (n+1)^{-2} y_n x_n$ and $X(y) = \sum_{n=0}^{\infty} X_n(y)$ for $\gamma_{0,1}^{\mathbb{N}}$-almost every $y \in \mathbb{R}^{\mathbb{N}}$. Check that

$$\begin{aligned}\int \|\psi \circ X(y)\|_F^2 \, \gamma_{0,1}^{\mathbb{N}}(dy) &\geq \int \langle \psi \circ X(y), \eta_n \rangle^2 \, \gamma_{0,1}^{\mathbb{N}}(dy) \\ &\geq \int \langle \psi \circ X_n(y), \eta_n \rangle^2 \, \gamma_{0,1}^{\mathbb{N}}(dy) \geq (n+1),\end{aligned}$$

and thereby arrive at the contradiction that $\psi \circ X \notin L^2(\gamma_{0,1}^{\mathbb{N}}; F)$. Conclude that *every Borel measurable, linear map from E to F is continuous.*

3.2.2 Gaussian Measures on a Hilbert Space

Let H be a separable Hilbert space. If $\mathcal{W}$ is a centered Gaussian measure on H, then $\{(\,\cdot\,, h)_H : h \in H\}$ is a Gaussian family under $\mathcal{W}$ whose covariance function c is a non-negative definite, symmetric bilinear function on H^2. Moreover, by Theorem 3.2.6, $C \equiv \int \|x\|_H^2 \, \mathcal{W}(dx) < \infty$, and therefore

$$c(h,h) = \int (x,h)_H^2 \, \mathcal{W}(dx) \leq C\|h\|_H^2,$$

Hence there exists a non-negative definite, symmetric operator $A : H \longrightarrow H$ such that $c(g,h) = (g, Ah)_H$ for all $g, h \in H$. In particular, if $\{h_n : n \geq 1\}$ is an orthonormal basis in H, then

$$\mathrm{Trace}(A) = \sum_{n=1}^{\infty} (h_n, Ah_n) = \sum_{n=1}^{\infty} \int (x, h_n)_H^2 \, \mathcal{W}(dx) = \int \|x\|_H^2 \, \mathcal{W}(dx) = C < \infty,$$

which means that A is a symmetric trace class operator.

Conversely, if A is a symmetric, non-negative definite trace class operator on H, then there is a $\mathcal{W} \in M_1(H)$ under which $\{(\,\cdot\,, h)_H : h \in H\}$ is a Gaussian family with covariance function $c(g,h) = (g, Ah)_H$. To see this, let $\{h_n : n \geq 0\}$ be an orthonormal basis in H consisting of eigenfunctions for A, and set $\alpha_n = (h_n, Ah_n)_H$, the eigenvalue corresponding to h_n. Next, let $\{X_n : n \geq 0\}$ be a sequence of mutually independent $N(0,1)$ random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and define

$$S_n = \sum_{k=0}^{n} \alpha_k^{\frac{1}{2}} X_k h_k \text{ for } n \geq 0.$$

Then, for $m < n$,

$$\mathbb{E}\big[\|S_n - S_m\|_H^2\big] = \sum_{k=m+1}^{n} \alpha_k \le \sum_{k>m} \alpha_k,$$

and so there exists an $S : \Omega \longrightarrow H$ such that

$$\mathbb{E}\big[\|S - S_m\|_H^2\big] \le \sum_{k>m} \alpha_k.$$

Furthermore, for each $h \in H$, $(S,h)_H$ under $\mathbb{P}$ is a centered Gaussian random variable with variance

$$\sum_{k=0}^{\infty} \alpha_m (h, h_k)_H^2 = (h, Ah)_H.$$

Finally, take $\mathcal{W} = S_*\mathbb{P}$.

Combined with the preceding, we have now proved the following theorem.

Theorem 3.2.8 *Let A be a bounded, symmetric, non-negative definite operator of the separable Hilbert space H. Then there exists a $\mathcal{W} \in M_1(H)$ under which $\{(\,\cdot\,,h)_H : h \in H\}$ is a centered Gaussian process with covariance function $(g, Ah)_H$ if and only if A is trace class.*

3.3 Abstract Wiener Spaces

Let B be an infinite dimensional, separable Banach space and $\mathcal{W} \in M_1(B)$ a non-degenerate centered Gaussian measure with covariance c (i.e., the covariance function for the Gaussian family $\{\langle\,\cdot\,,\xi\rangle : \xi \in B^*\}$). Even though, as explained in § 3.1, we know that there is no Hilbert structure on B for which $\mathcal{W}$ is the standard Gauss measure, one can hope that there is a structure which provides some of the same advantages, and Gross's notion of an abstract Wiener space is one such structure. The following lemma is needed in order to explain Gross's idea.

Lemma 3.3.1 *Let B be a separable, real Banach space, and suppose that $H \subseteq B$ is a real Hilbert space that is continuously embedded as a dense subspace of B.*

(**i**) *For each $\xi \in B^*$ there is a unique $h_\xi \in H$ with the property that $\big(h, h_\xi\big)_H = \langle h, \xi\rangle$ for all $h \in H$. Moreover, the map $\xi \in B^* \longmapsto h_\xi \in H$ is linear, continuous, one-to-one, and onto a dense subspace of H. In fact, for any weak* dense subset S^* of B^*, $\{h_\xi : \xi \in S^*\}$ is dense in H.*

(**ii**) *If $x \in B$, then $x \in H$ if and only if there is a $K < \infty$ such that $|\langle x, \xi\rangle| \le K\|h_\xi\|_H$ for all $\xi \in B^*$. Moreover, for each $h \in H$, $\|h\|_H = \sup\{\langle h, \xi\rangle : \xi \in B^* \;\&\; \|h_\xi\|_H \le 1\}$.*

(**iii**) *If L^* is a weak* dense subspace of B^*, then there exists a sequence $\{\xi_n : n \geq 0\} \subseteq L^*$ such that $\{h_{\xi_n} : n \geq 0\}$ is an orthonormal basis for H. Moreover, if $x \in B$, then $x \in H$ if and only if $\sum_{n=0}^{\infty} \langle x, \xi_n \rangle^2 < \infty$. In particular, if $\|x\|_H \equiv \infty$ when $x \in B \setminus H$, then $\|\cdot\|_H : B \longrightarrow [0, \infty]$ is a lower semi-continuous function and so $\mathcal{B}_H \subseteq \mathcal{B}_B$. Finally,*

$$\big(h, h'\big)_H = \sum_{n=0}^{\infty} \langle h, \xi_n \rangle \langle h', \xi_n \rangle \quad \textit{for all } h, h' \in H.$$

Proof. Because H is continuously embedded in B, there exists a $C < \infty$ such that $\|h\|_B \leq C\|h\|_H$. Thus, if $\xi \in B^*$ and $f(h) = \langle h, \xi \rangle$, then f is linear and $|f(h)| \leq \|h\|_B \|\xi\|_{B^*} \leq C\|\xi\|_{B^*}\|h\|_H$, and so, by the Riesz Representation Theorem for Hilbert spaces, there exists a unique $h_\xi \in H$ such that $f(h) = \big(h, h_\xi\big)_H$. In fact, $\|h_\xi\|_H \leq C\|\xi\|_{B^*}$, and uniqueness can be used to check that $\xi \rightsquigarrow h_\xi$ is linear. To see that $\xi \rightsquigarrow h_\xi$ is one-to-one, it suffices to show that $\xi = 0$ if $h_\xi = 0$. But if $h_\xi = 0$, then $\langle h, \xi \rangle = 0$ for all $h \in H$, and therefore, because H is dense in B, $\xi = 0$. Finally, to complete the proof of (**i**), let S^* be a weak* dense subset of B^*, and suppose that $\{h_\xi : \xi \in S^*\}$ were not dense in H. Then there would exist an $h \in H \setminus \{0\}$ with the property that $\langle h, \xi \rangle = \big(h, h_\xi\big)_H = 0$ for all $\xi \in S^*$. But, since S^* is weak* dense in B^*, this would lead to the contradiction that $h = 0$.

Obviously, if $h \in H$, then $|\langle h, \xi \rangle| = |(h, h_\xi)_H| \leq \|h\|_H \|h_\xi\|_H$ for $\xi \in B^*$. Conversely, if $x \in B$ and $|\langle x, \xi \rangle| \leq K\|h_\xi\|_H$ for some $K < \infty$ and all $\xi \in B^*$, set $f(h_\xi) = \langle x, \xi \rangle$ for $\xi \in B^*$. Then, because $\xi \rightsquigarrow h_\xi$ is one-to-one, f is a well-defined, linear functional on $\{h_\xi : \xi \in B^*\}$. Moreover, $|f(\xi)| \leq K\|h_\xi\|_H$, and therefore, since $\{h_\xi : \xi \in B^*\}$ is dense in H, f admits a unique extension as a continuous, linear functional on H. Hence, by Riesz's theorem, there is an $h \in H$ such that

$$\langle x, \xi \rangle = f(h_\xi) = \big(h, h_\xi\big)_H = \langle h, \xi \rangle, \quad \xi \in B^*,$$

which means that $x = h \in H$. In addition, if $h \in H$, then $\|h\|_H = \sup\{\langle h, \xi \rangle : \|h_\xi\|_H \leq 1\}$ follows from the density of $\{h_\xi : \xi \in B^*\}$ in H, and this completes the proof of (**ii**).

Turning to (**iii**), note that, because B^* with the weak* topology is second countable , the weak* topology on L^* is also second countable. Thus, we can find a sequence in L^* that is weak* dense in B^*, and then extract a subsequence of linearly independent elements whose span S^* is weak* dense in B^*. Starting with this subsequence, apply the Grahm–Schmidt orthogonalization procedure to produce a sequence $\{\xi_n : n \geq 0\}$ whose span is S^* and for which $\{h_{\xi_n} : n \geq 0\}$ is orthonormal in H. Because the span of $\{h_{\xi_n} : n \geq 0\}$ is equal to $\{h_\xi : \xi \in S^*\}$, which, by what we proved earlier, is dense in H, $\{h_{\xi_n} : n \geq 0\}$ is an orthonormal basis in H. Knowing this, it is immediate that

$$\big(h, h'\big)_H = \sum_{n=0}^{\infty} \big(h, h_{\xi_n}\big)_H \big(h', h_{\xi_n}\big)_H = \sum_{n=0}^{\infty} \langle h, \xi_n\rangle\langle h', \xi_n\rangle.$$

In particular, $\|h\|_H^2 = \sum_{n=0}^{\infty}\langle h, \xi_n\rangle^2$. Finally, if $x \in B$ and $\sum_{n=0}^{\infty}\langle x, \xi_n\rangle^2 < \infty$, set $g = \sum_{m=0}^{\infty}\langle x, \xi_n\rangle h_{\xi_n}$. Then $g \in H$ and $\langle x - g, \xi\rangle = 0$ for all $\xi \in S^*$. Hence, since S^* is weak* dense in B^*, $x = g \in H$. Finally, because $x \rightsquigarrow \sum_{m=0}^{\infty}\langle x, \xi_m\rangle^2$ is lower semi-continuous, $\|\cdot\|_H$ is also. □

Given a separable Hilbert space H which is dense and continuously embedded in a separable Banach space B, and a $\mathcal{W} \in M_1(B)$, the triple $(H, B, \mathcal{W})$ is called an *abstract Wiener space* if

$$\widehat{\mathcal{W}}(\xi) = \exp\left(-\frac{\|h_\xi\|_H^2}{2}\right) \text{ for all } \xi \in B^*. \tag{3.3.1}$$

Equivalently, $(H, B, \mathcal{W})$ is an abstract Wiener space if and only if $\mathcal{W} \in M_1(B)$ and $\{\langle\,\cdot\,, \xi\rangle : \xi \in B^*\}$ is a centered Gaussian family under $\mathcal{W}$ for which $\operatorname{cov}(\xi, \eta) = (h_\xi, h_\eta)_H$.

As the following theorem shows, this notion applies to any centered Gaussian $\mathcal{W}$ on a Banach B that is *non-degenerate* in the sense that $\int \langle x, \xi\rangle^2\,\mathcal{W}(dx) > 0$ for all $\xi \in B^* \setminus \{0\}$.

Theorem 3.3.2 *If B is a separable Banach space and $\mathcal{W} \in M_1(B)$, then $\mathcal{W}$ is a non-degenerate, centered Gaussian measure if and only if there is separable Hilbert space for which $(H, B, \mathcal{W})$ is an abstract Wiener space, in which case there is only one such H. Conversely, if H is a separable Hilbert space, then there is a separable Banach space and a $\mathcal{W} \in M_1(B)$ for which $(H, B, \mathcal{W})$ is an abstract Wiener space.*

Proof. Keep in mind that if $\mathcal{W}$ is a centered Gaussian measure on B, then, by Theorem 3.2.6, $\|x\|_B^2 \in L^1(\mathcal{W}; \mathbb{R})$. Hence, for any $\xi \in B^*$, $\langle x, \xi\rangle \in L^2(\mathcal{W}; \mathbb{R})$ and $x\langle x, \xi\rangle \in L^1(\mathcal{W}; B)$.

Suppose that $(H, B, \mathcal{W})$ is an abstract Wiener space. Obviously, $\mathcal{W}$ is a centered Gaussian measure. To check that $\mathcal{W}$ is non-degenerate, let $\xi \in B^* \setminus \{0\}$ and observe that

$$\int \langle x, \xi\rangle^2\,\mathcal{W}(dx) = \|h_\xi\|_H^2 > 0.$$

Next, to see that H is uniquely determined, note that, for any ξ, $\eta \in B^*$,

$$\langle h_\xi, \eta\rangle = (h_\eta, h_\xi)_H = \int \langle x, \eta\rangle\langle x, \xi\rangle\,\mathcal{W}(dx) = \left\langle \int x\langle x, \xi\rangle\,\mathcal{W}(x), \eta\right\rangle,$$

and so

$$h_\xi = \int x\langle x, \xi\rangle\,\mathcal{W}(dx). \tag{$*$}$$

Since $\{h_\xi : \xi \in H\}$ is dense in H, this proves that H is uniquely determined.

Now assume that $\mathcal{W}$ is a centered Gaussian measure on B. Given $\xi \in B^*$, define h_ξ by $(*)$, and define the inner product $(\,\cdot\,,\cdot\cdot)_H$ by

$$(h_\eta, h_\xi)_H = \int \langle x, \eta\rangle\langle x, \xi\rangle\, \mathcal{W}(dx) \text{ for } \eta \in B^*.$$

Because $\mathcal{W}$ is non-degenerate, $\|h_\xi\|_H^2 \equiv (h_\xi, h_\xi)_H = 0 \implies \xi = 0 \implies h_\xi = 0$. Hence $(\,\cdot\,,\cdot\cdot)_H$ is a non-degenerate inner product. Now let H be the completion of $\{h_\xi : \xi \in B^*\}$ with respect to $\|\cdot\|_H$. Equivalently, let L to be the closure in $L^2(\mathcal{W};\mathbb{R})$ of $\{\langle\,\cdot\,,\xi\rangle : \xi \in B^*\}$, and take

$$H = \left\{\int x\psi(x)\,\mathcal{W}(dx) : \psi \in L\right\}.$$

It is clear that $\langle h, \xi\rangle = (h, h_\xi)_H$, first when $h = h_\eta$ for some $\eta \in B^*$ and then, by continuity, for all $h \in H$. In particular, for any $\xi \in B^*$ with $\|\xi\|_{B^*} = 1$,

$$\langle h, \xi\rangle \le \|h\|_H\|h_\xi\|_H = \|h\|_H \left(\int \langle x, \xi\rangle^2\, \mathcal{W}(dx)\right)^{\frac{1}{2}} \le C\|h\|_H$$

$$\text{where } C = \left(\int \|x\|_B^2\, \mathcal{W}(dx)\right)^{\frac{1}{2}},$$

and so $\|h\|_B \le C\|h\|_H$. From this it follows that H is continuously embedded in B as a subspace. In addition, because $\|h_\xi\|_H^2 = \int \langle x, \xi\rangle^2\, \mathcal{W}(dx)$, $\widehat{\mathcal{W}}(\xi) = \exp\left(-\frac{\|h_\xi\|_H^2}{2}\right)$. Finally, to see that H is dense in B, suppose that it were not. Then, by the Hahn–Banach Theorem, there would be a $\xi \in B^* \setminus \{0\}$ such that $\langle h, \xi\rangle = 0$ for all $h \in H$. But this would mean that $\|h_\xi\|_H^2 = \langle h_\xi, \xi\rangle = 0$, which contradicts $\xi \neq 0$. Thus, we have shown that $(H, B, \mathcal{W})$ is an abstract Wiener space.

Next let H be a separable Hilbert space. Choose an orthonormal basis $\{h_n : n \ge 0\}$ and, for $\alpha > 1$, define

$$\|h\|_{B^{(-\alpha)}} = \left(\sum_{n=0}^\infty (n+1)^{-\alpha}(h, h_n)_H^2\right)^{\frac{1}{2}} \in [0,\infty].$$

and let $B^{(-\alpha)}$ be the completion of H with respect to $\|\cdot\|_{B^{(-\alpha)}}$. Clearly $B^{(-\alpha)}$ is a separable Hilbert space with inner product

$$\big(x, y\big)_{B^{(-\alpha)}} = \sum_{n=0}^\infty (n+1)^{-\alpha}\big(x, h_n\big)_H\big(y, h_n\big)_H,$$

and H is dense in it. Further, if $h_n^{(-\alpha)} = (n+1)^{\frac{\alpha}{2}} h_n$, then $\{h_n^{(-\alpha)} : n \ge 0\}$ is an orthonormal basis in $B^{(-\alpha)}$. Now define $A : B^{(-\alpha)} \longrightarrow B^{(-\alpha)}$ by

$$Ax = \sum_{n=0}^{\infty} (n+1)^{-\frac{\alpha}{2}} \big(x, h_n^{(-\alpha)}\big)_{B^{(-\alpha)}} h_n.$$

Then

$$\sum_{m=0}^{\infty} \big(h_m^{(-\alpha)}, A h_m^{(-\alpha)}\big)_{B^{(-\alpha)}} = \sum_{m=0}^{\infty} (n+1)^{-\alpha} < \infty,$$

and so A is a trace class operator on $B^{(-\alpha)}$. Hence, by Theorem 3.2.8, there exists a non-degenerate, centered Gaussian measure $\mathcal{W} \in M_1(B^{(-\alpha)})$ for which

$$(x,y) \in B^{(-\alpha)} \times B^{(-\alpha)} \longmapsto (y, Ax)_{B^{(-\alpha)}} \in \mathbb{R}$$

is the covariance function. Moreover, if $x \in B^{(-\alpha)}$ and $h_x = Ax$, then $h_x \in H$, and $(h, h_x)_H = \big(h, x\big)_{B^{(-\alpha)}}$ for all $h \in H$. Finally,

$$\|h_x\|_H^2 = \sum_{n=0}^{\infty} (n+1)^{-\alpha} \big(x, h_n^{(-\alpha)}\big)^2_{B^{(-\alpha)}} = \big(x, Ax\big)_{B^{(-\alpha)}},$$

and so $\widehat{\mathcal{W}}(\xi) = \exp\Big(-\frac{\|h_x\|_H^2}{2}\Big)$. $\square$

As is evident from the proof of the second part of Theorem 3.3.2, there are many choices of the Banach space for a given Hilbert space even though there is only one Hilbert space for a given Banach space. Thus, when thinking in terms of abstract Wiener spaces, the canonical object is the Hilbert space. Metaphorically speaking, the Hilbert space is the skeleton, whereas the Banach space is simply a flesh coating that provides a comfy place for the measure to live, and there are a multitude of ways in which the same skeleton can be coated. A challenging problem is that of determining when one can find a better coating than the one produced in the preceding proof.

Related to the construction of abstract Wiener spaces for a given Hilbert space is the following. When $H = \mathbb{R}^N$, then there is a unique symmetric, positive definite transformation on $\mathbb{R}^N$ such $(g,h)_H = (g, Ah)_{\mathbb{R}^N}$, and the $(H, \mathbb{R}^N, \mathcal{W})$ with $\mathcal{W} = \gamma_{0,A^{-1}}$ is an abstract Wiener space. In particular, one can construct all these abstract Wiener spaces from the standard one, the one corresponding to I using the transformation $A^{-\frac{1}{2}}$, and one should suspect that there is a general principle of which this is a special case. The following theorem provides such a principle.

Theorem 3.3.3 *Let $(H, B, \mathcal{W})$ be an abstract Wiener space and F an isometric, isomorphism from H to the Hilbert space G. If C is a separable Banach space in which G is continuously embedded as a dense subspace and for which there exists an extension of F as a continuous, linear map $\tilde{F}$ from B to C, then $(G, C, \tilde{F}_*\mathcal{W})$ is an abstract Wiener space. In particular, one can take C to be the completion of G with respect to the norm $\|\cdot\|_C$ on G given by $\|g\|_C = \|F^{-1}g\|_H$ and $\tilde{F}$ to be the continuous extention of F as a map from B onto C.*

Proof. Let $\tilde{F}^\top : C^* \longrightarrow B^*$ be the adjoint map for $\tilde{F}$. If $\eta \in C^*$, then, for all $g \in G$,

$$\begin{aligned}\big(g, g_\eta\big)_G &= \langle g, \eta\rangle = \langle F^{-1}g, \tilde{F}^\top\eta\rangle \\ &= \big(F^{-1}g, h_{\tilde{F}^\top\eta}\big)_B = \big(g, Fh_{\tilde{F}^\top\eta}\big)_G,\end{aligned}$$

and so $g_\eta = Fh_{\tilde{F}^\top\eta}$. Hence,

$$\begin{aligned}\widehat{\tilde{F}_*\mathcal{W}}(\eta) &= \int e^{i\langle \tilde{F}x,\eta\rangle}\,\mathcal{W}(dx) = \hat{\mathcal{W}}\big(\tilde{F}^\top\eta\big) \\ &= \exp\left(-\frac{\|h_{\tilde{F}^\top\eta}\|_H^2}{2}\right) = \exp\left(-\frac{\|g_\eta\|_G^2}{2}\right),\end{aligned}$$

which completes the proof that $\big(G, C, \tilde{F}_*\mathcal{W}\big)$ is an abstract Wiener space.

Turning to the concluding assertion, it is clear that $\|\cdot\|_C$ is a norm on G, and the completion C of G with respect to it is a separable Banach space in which G is continuously embedded as a dense subspace. Moreover, F has a unique extension $\tilde{F}$ as an isometric, isomorphism from B onto C. □

Theorem 3.3.3 is the abstract Wiener space analog of the classical isomorphism theorem for separable Hilbert space of the same dimension. Namely, it says that all abstract Wiener spaces corresponding to Hilbert spaces of the same dimension are, from a sufficiently abstract perspective, the same. Of course, just as in practice it is foolish to think of all Hilbert spaces of the same dimension as being the same, one should not think that the associated abstract Wiener spaces are all the same.

3.3.1 The Cameron–Martin Subspace and Formula

Given a centered, non-degenerate Gaussian measure $\mathcal{W}$ on B, the Hilbert space H for which $(H, B, \mathcal{W})$ is an abstract Wiener space is called its *Cameron–Martin subspace*.

It is important to understand the relationship between the Cameron–Martin subspace for $\mathcal{W}$ and the covariance function c for the Gaussian family $\{\langle\,\cdot\,,\xi\rangle : \xi \in B^*\}$. Indeed, since $\langle h_\xi, \eta\rangle = (h_\eta, h_\xi)_H = c(\xi,\eta)$ for $\xi, \eta \in B^*$, the map $\xi \rightsquigarrow h_\xi$ is the map $A : B^* \longrightarrow B$ described in § 3.1, and therefore H is the completion of $\{A\xi : \xi \in B^*\}$ with respect to the Hilbert norm $\sqrt{c(\xi,\xi)}$. When B is finite dimensional, $\{A\xi : \xi \in B^*\} = B$, and therefore $H = B$, but, as is shown below, this is far from the case when B is infinite dimensional.

Theorem 3.3.4 *If $(H, B, \mathcal{W})$ is an abstract Wiener space, then the map $\xi \in B^* \longmapsto h_\xi \in H$ is continuous from the weak* topology on B^* into the* *strong topology on H. In particular, for each $R > 0$, $\{h_\xi : \xi \in \overline{B_{B^*}(0,R)}\}$ is a compact subset of H, $\overline{B_H(0,R)}$ is a compact subset of B, $H \in \mathcal{B}_B$, and, when H is infinite dimensional, $\mathcal{W}(H) = 0$. Finally, there is a unique linear, isometric map $\mathcal{I} : H \longrightarrow L^2(\mathcal{W};\mathbb{R})$ such that $\mathcal{I}(h_\xi) = \langle\,\cdot\,,\xi\rangle$ for all $\xi \in B^*$, and $\{\mathcal{I}(h) : h \in H\}$ is a Gaussian family in $L^2(\mathcal{W};\mathbb{R})$ for which $(\,\cdot\,,\cdot\cdot)_H$ is the covariance function.*

Proof. To prove the initial assertion, suppose that $\{\xi_k : k \geq 1\} \subseteq B^*$ is weak* convergent to ξ. Then $\langle x,\xi_k\rangle \longrightarrow \langle x,\xi\rangle$ for all $x \in B$, and therefore, by Lemma 2.1.1, $\langle\,\cdot\,,\xi_k\rangle \longrightarrow \langle\,\cdot\,,\xi\rangle$ in $L^2(\mathcal{W};\mathbb{R})$. Hence

$$\|h_{\xi_k} - h_\xi\|_H^2 = \big\|\langle\,\cdot\,,\xi_k\rangle - \langle\,\cdot\,,\xi\rangle\big\|^2_{L^2(\mathcal{W};\mathbb{R})} \longrightarrow 0.$$

Given the first assertion, the compactness of $\{h_\xi : \xi \in \overline{B_{B^*}(0,R)}\}$ in H follows from the compactness of $\overline{B_{B^*}(0,R)}$ in the weak* topology. To see that $\overline{B_H(0,R)}$ is compact in B, first note that $\overline{B_H(0,R)}$ is compact in the weak topology on H. Therefore, all that we have to show is that the embedding map $h \in \overline{B_H(0,R)} \longmapsto h \in B$ is continuous from the weak topology on H into the strong topology on B. Thus, suppose that $h_k \longrightarrow h$ weakly in H. Because $\big\{h_\xi : \xi \in \overline{B_{B^*}(0,1)}\big\}$ is compact in H, for each $\epsilon > 0$ there exist an $n \in \mathbb{Z}^+$ and $\{\xi_1,\dots,\xi_n\} \subseteq \overline{B_{B^*}(0,1)}$ such that

$$\{h_\xi : \xi \in \overline{B_{B^*}(0,1)}\} \subseteq \bigcup_1^n B_H(h_{\xi_m},\epsilon).$$

Now choose ℓ so that $\max_{1\leq m\leq n} |\langle h_k - h, \xi_m\rangle| < \epsilon$ for all $k \geq \ell$. Then, for any $\xi \in \overline{B_{B^*}(0,1)}$ and all $k \geq \ell$,

$$|\langle h_k - h,\xi\rangle| \leq \epsilon + \min_{1\leq m\leq n}\big|\big(h_k - h, h_\xi - h_{\xi_m}\big)_H\big| \leq (2R+1)\epsilon,$$

which proves that $\|h_k - h\|_H \leq (2R+1)\epsilon$ for all $k \geq \ell$.

Because $H = \{x \in B : \|x\|_H < \infty\}$ and $\|\cdot\|_H$ is lower semi-continuous on B, $H \in \mathcal{B}_B$. To see that $\mathcal{W}(H) = 0$ when H is infinite dimensional, choose $\{\xi_n : n \geq 0\}$ as (**iii**) of Lemma3.3.1, and set $X_n(x) = \langle x,\xi_n\rangle$. Then $\{X_n : n \geq 0\}$ is an infinite sequence of mutually independent $N(0,1)$-random variables, and so, since

$$H = \left\{x \in B : \sum_{n=0}^\infty X_n^2 < \infty\right\} \text{ and } \int \exp\left(-\sum_{n=0}^\infty X_n^2\right) \mathcal{W}(dx) = 0,$$

$\mathcal{W}(H) = 0$.

Turning to the map $\mathcal{I}$, define $\mathcal{I}(h_\xi) = \langle\,\cdot\,,\xi\rangle$. Then, for each ξ, $\mathcal{I}(h_\xi)$ is a centered Gaussian with variance $\|h_\xi\|_H^2$, and so $\mathcal{I}$ is a linear isometry from $\{h_\xi : \xi \in B^*\}$ into $L^2(\mathcal{W};\mathbb{R})$. Hence, since $\{h_\xi : \xi \in B^*\}$ is dense in H, $\mathcal{I}$ admits a unique extension as a linear isometry from H into $L^2(\mathcal{W};\mathbb{R})$.

Moreover, as the $L^2(\mathcal{W};\mathbb{R})$-limit of centered Gaussians, $\mathcal{I}(h)$ is a centered Gaussian with variance $\|h\|_H^2$ for each $h \in H$. $\square$

The map $\mathcal{I}$ in Theorem 3.3.4 was introduced for the classical Wiener space by R.E.A.C. Paley and Wiener, and so I will call it the *Paley–Wiener map.* To appreciate its importance here, observe that $\{h_\xi : \xi \in B^*\}$ is the subspace of those $g \in H$ with the property that $h \in H \longmapsto (h,g)_H \in \mathbb{R}$ admits a continuous extension to B. Even though, when $\dim(H) = \infty$, no such continuous extension exists for general $g \in H$, $\mathcal{I}(g)$ can be thought of as an extension of $h \rightsquigarrow (h,g)_H$, albeit one that is defined only up to a $\mathcal{W}$-null set. Of course, one has to be careful when using this interpretation, since, when H is infinite dimensional, $\mathcal{I}(g)(x)$ for a given $x \in E$ is not well-defined simultaneously for all $g \in H$. Nonetheless, by adopting it, one gets further evidence for the idea that $\mathcal{W}$ wants to be the standard Gaussian measure on H. Namely, because

$$\int e^{i\mathcal{I}(h)}\, d\mathcal{W} = e^{-\frac{\|h\|_H^2}{2}}, \quad h \in H,$$

if $\mathcal{W}$ lived on H, then it would certainly be the standard Gauss measure there.

The Paley–Wiener map also provides strong support for the idea that it is the Hilbert space H that is the canonical component in the triple $(H, B, \mathcal{W})$. Indeed suppose that $(H, B_1, \mathcal{W}_1)$ and $(H, B_2, \mathcal{W}_2)$ are both abstract Wiener spaces. Because there may be no obvious correspondence between elements of B_1^* and B_2^*, in general there will be no obvious relationship between the Gaussian families $\{\langle\,\cdot\,, \xi_1\rangle : \xi_1 \in B_1^*\}$ and $\{\langle\,\cdot\,, \xi_1\rangle : \xi_2 \in B_2^*\}$, and so it will be hard to understand the connection between the measures $\mathcal{W}_1$ and $\mathcal{W}_2$ in terms of their Fourier transforms. On the other hand, if $\mathcal{I}_1$ and $\mathcal{I}_2$ are the Paley–Wiener maps for $(H, B_1, \mathcal{W}_1)$ and $(H, B_2, \mathcal{W}_2)$, then the Gaussian family $\{\mathcal{I}_1(h) : h \in H\}$ will have the same distribution under $\mathcal{W}_1$ as $\{\mathcal{I}_2 : h \in H\}$ has under $\mathcal{W}_2$. In particular, if $\xi_1 \in B_1^*$ then the $\mathcal{W}_2$-random variable that should be associated with $\langle\,\cdot\,, \xi_1\rangle$ is $\mathcal{I}_2(h_{\xi_1})$, which need not equal $\langle\,\cdot\,, \xi_2\rangle$ for any $\xi_2 \in B_2^*$.

Among the most important applications of the Paley–Wiener map is the following theorem about the behavior of Gaussian measures under translation. That is, if $y \in B$ and $T_y : B \longrightarrow B$ is given by $T_y(x) = x + y$, we will be looking at the measure $(T_y)_*\mathcal{W}$ and its relationship to $\mathcal{W}$. Using the reasoning suggested above, the result is easy to guess. Namely, if $\mathcal{W}$ really lived on H and were given by a Feynman-type representation (cf. § 3.1)

$$\mathcal{W}(dh) = \frac{1}{Z} e^{-\frac{\|h\|_H^2}{2}}\, \lambda_H(dh),$$

then $(T_g)_*\mathcal{W}$ should have the Feynman representation

$$\frac{1}{Z}e^{-\frac{\|h-g\|_H^2}{2}}\,\lambda_H(dh),$$

which could be rewritten as

$$\big[(T_g)_*\mathcal{W}\big](dh) = \exp\big[\big(h,g\big)_H - \tfrac{1}{2}\|g\|_H^2\big]\,\mathcal{W}(dh).$$

Hence, if we are correct to interpret $\mathcal{I}(g)$ as $(\,\cdot\,,g)_H$, we are led to guess that, at least for $g \in H$,

$$\big[(T_g)_*\mathcal{W}(dx)\big](dx) = R_g(x)\,\mathcal{W}(dx), \text{ where } R_g = \exp\big[\mathcal{I}(g) - \tfrac{1}{2}\|g\|_H^2\big]. \quad (3.3.2)$$

That (3.3.2) is correct was proved for the classical Wiener space by R. Cameron and T. Martin, and for this reason it is called the *Cameron–Martin formula*. In fact, one has the following result, the second half of which is due to I. Segal.

Theorem 3.3.5 *If $(H,B,\mathcal{W})$ is an abstract Wiener space, then, for each $g \in H$, $(T_g)_*\mathcal{W} \ll \mathcal{W}$ and the R_g in* (3.3.2) *is the corresponding Radon–Nikodym derivative. Conversely, if $y \in B \setminus H$, then $(T_y)_*\mathcal{W}$ is singular with respect to $\mathcal{W}$.*

Proof. Let $g \in H$, and set $\mu = (T_g)_*\mathcal{W}$. Then

$$\hat{\mu}(\xi) = \int e^{i\langle x+g,\xi\rangle}\,\mathcal{W}(dx) = \exp\big(i\langle g,\xi\rangle - \tfrac{1}{2}\|h_\xi\|_H^2\big). \quad (*)$$

Now define ν by the right-hand side of (3.3.2). Clearly $\nu \in M_1(E)$. Thus, we will have proved the first part of this theorem once we show that $\hat{\nu}$ is given by the right-hand side of $(*)$. To this end, observe that, for any $h_1, h_2 \in H$,

$$\int e^{\alpha_1\mathcal{I}(h_1)+\alpha_2\mathcal{I}(h_2)}\,d\mathcal{W} = \exp\left(\frac{\alpha_1^2}{2}\|h_1\|_H^2 + \alpha_1\alpha_2\big(h_1,h_2\big)_H + \frac{\alpha_2^2}{2}\|h_2\|_H^2\right)$$

for all $\alpha_1, \alpha_2 \in \mathbb{C}$. Indeed, this is obvious when α_1 and α_2 are pure imaginary, and, since both sides are entire functions of $(\alpha_1,\alpha_2) \in \mathbb{C}^2$, it follows in general by analytic continuation. In particular, by taking $h_1 = g$, $\alpha_1 = 1$, $h_2 = h_\xi$, and $\alpha_2 = i$, it is easy to check that the right-hand side of $(*)$ is equal to $\hat{\nu}(\xi)$.

To prove the second assertion, begin by recalling from Lemma 3.3.1 that for any $y \in B$, $y \in H$ if and only if there is a $K < \infty$ with the property that $|\langle y,\xi\rangle| \le K$ for all $\xi \in B^*$ with $\|h_\xi\|_H = 1$. Now suppose that $(T_y)_*\mathcal{W} \not\perp \mathcal{W}$, and let R be the Radon–Nikodym derivative of its absolutely continuous part. Given $\xi \in B^*$ with $\|h_\xi\|_H = 1$, let $\mathcal{F}_\xi$ be the σ-algebra generated by $x \rightsquigarrow \langle x,\xi\rangle$, and check that $(T_y)_*\mathcal{W} \restriction \mathcal{F}_\xi \ll \mathcal{W} \restriction \mathcal{F}_\xi$ with Radon–Nikodym derivative

$$Y(x) = \exp\left(\langle y,\xi\rangle\langle x,\xi\rangle - \frac{\langle y,\xi\rangle^2}{2}\right).$$

Hence,

$$Y \geq \mathbb{E}^{\mathcal{W}}\big[R \,\big|\, \mathcal{F}_\xi\big] \geq \mathbb{E}^{\mathcal{W}}\big[R^{\frac{1}{2}} \,\big|\, \mathcal{F}_\xi\big]^2,$$

and so

$$\exp\left(-\frac{\langle y, \xi\rangle^2}{8}\right) = \langle Y^{\frac{1}{2}}, \mathcal{W}\rangle \geq \alpha \equiv \langle R^{\frac{1}{2}}, \mathcal{W}\rangle \in (0,1].$$

Since this means that $\langle y, \xi\rangle^2 \leq 8 \log \frac{1}{\alpha}$, $y \in H$, we have shown that $(T_y)_*\mathcal{W} \perp \mathcal{W}$ unless $y \in H$. □

To appreciate that Gaussian measures are as translation invariant as any probability measures on an infinite dimensional space, one should know about the theorem of V. Sudakov below.

Recall that a subset of a metric space is said to be *meager* if it can be written as the countable union of nowhere dense (i.e., with empty interior) closed sets. If the metric space is complete, then the Baire Category Theorem says that meager subsets are nowhere dense.

Lemma 3.3.6 *If $\{K_n : n \geq\}$ is a sequence of compact subsets of an infinite dimensional Banach space E, then the interior of $\bigcup_{n=1}^\infty K_n$ is empty.*

Proof. By the Baire Category Theorem, all that we have to show is that compact subsets of E are nowhere dense. Thus, suppose that $K \subseteq E$ is compact. To show that its interior is empty, it suffices to prove that $K - K = \{y - x : x, y \in K\}$ contains no neighborhood of 0. Indeed, if we knew that and if $B_E(x,r) \subseteq K$ for some $x \in E$ and $r > 0$, then we would have the contradiction that $B_E(0,r) \subseteq B_E(x,r) - B(x,r) \subseteq K - K$. Thus, suppose that $r > 0$ is given, and choose $x_1, \dots, x_n \in E$ so that $K \subseteq \bigcup_{m=1}^n B_E\left(x_m, \frac{r}{4}\right)$. By the Hahn–Banach Theorem, there exists a $\xi \in E^*$ such that $\|\xi\|_{E^*} = 1$ and $\langle x_m, \xi\rangle = 0$ for $1 \leq m \leq n$. Now choose $y \in B_E(0,r)$ so that $\langle y, \xi\rangle \geq \frac{3r}{4}$. Then, $\langle x, \xi\rangle \leq \frac{r}{4}$ and $\langle x + y, \xi\rangle \geq \frac{r}{2}$ for all $x \in K$. Hence $x + y \notin K$ for any $x \in K$, and so $y \in B_E(0,r) \setminus (K - K)$. □

Theorem 3.3.7 *If μ is a Borel probability measure on an infinite dimensional, separable Banach space E, then the set of $x \in E$ for which $(T_x)_*\mu \not\perp \mu$ is contained in a meager set S, and therefore the set of $x \in E$ for which $(T_x)_* \perp \mu$ is dense in E.*

Proof. Using Ulam's Lemma (cf. Lemma 9.1.7 in [10]), for each $n \geq 1$ choose a compact $K_m \subseteq E$ for which $\mu(K_n) \geq 1 - \frac{1}{n}$, and set $\mathit{\Gamma} = \bigcup_{n=1}^\infty K_n$ and $\tilde{\mathit{\Gamma}} = \mathit{\Gamma} - \mathit{\Gamma} = \bigcup_{m,n=1}^\infty (K_m - K_n)$. Clearly $x \in \tilde{\mathit{\Gamma}} \iff -x \in \tilde{\mathit{\Gamma}}$ and $\mu(\mathit{\Gamma}) = 1$, and, by Lemma 3.3.6, the interior of $\tilde{\mathit{\Gamma}}$ is empty, which means its complement is dense. Finally, if $-x \notin \tilde{\mathit{\Gamma}}$, then $(-x + \mathit{\Gamma}) \cap \mathit{\Gamma} = \emptyset$, and so $(T_x)_*\mu(\mathit{\Gamma}) = \mu(-x + \mathit{\Gamma}) = 0$, which means that $(T_x)_*\mu \perp \mu$. □

In view of Sudakov's result, one sees that centered Gaussian measures provide as good a substitute for Lebesgue measure as one can hope to find: not only does one know the directions in which the are quasi-invariant under translation, one has a simple formula for the associated Radon–Nikodym derivatives. In particular, (3.3.5) provides the basis for an integration by parts formula that plays a central in the development of the Sobolev type theory used in what is called *Malliavin's calculus*. The integration by parts formula is based on the observation that if $h \in H$, then, for any non-negative, Borel measurable functions f and g

$$\mathbb{E}^{\mathcal{W}}\big[(f \circ T_h)g\big] = \mathbb{E}^{\mathcal{W}}\big[R_h f(g \circ T_{-h})\big]. \tag{$*$}$$

Hence, if $f \in L^2(\mathcal{W};\mathbb{R})$ and $g \in L^p(\mathcal{W};\mathbb{R})$ for some $p > 2$, then $(f \circ T_h)g \in L^1(\mathcal{W};\mathbb{R})$ and $(*)$ again holds. Now choose $\{\xi_j : j \geq 1\} \subseteq B^*$ so that $\{h_j : j \geq 1\}$ is an orthonormal basis in H when $h_j = h_{\xi j}$, and let $\mathcal{P}$ be the set of functions φ of the form $\varphi(x) = p\big(\langle x, \xi_1\rangle, \dots, \langle x, \xi_N\rangle\big)$ for some $N \geq 1$ and real polynomial p on $\mathbb{R}^N$. Then, for any $f \in L^2(\mathcal{W};\mathbb{R})$ and $\psi \in \mathcal{P}$,

$$\frac{d}{dt}\mathbb{E}^{\mathcal{W}}\big[(f \circ T_{th})\varphi\big] = \frac{d}{dt}\mathbb{E}^{\mathcal{W}}\big[R_{th} f(\varphi \circ T_{-th})\big] = \mathbb{E}^{\mathcal{W}}\big[R_{th} f\big((\partial_h^\top \psi) \circ T_{th}\big)\big],$$

where $\partial_h^\top \psi(x) = \langle x, h\rangle \psi(x) - \partial_h \psi$ and

$$\partial_h \psi = \frac{d}{ds}\psi(x + sh)\Big|_{s=0} = \sum_{j=1}^N (h, h_j)_H \partial_j p\big(\langle x, \xi_1\rangle, \dots, \langle x, \xi_N\rangle\big),$$

$\partial_j p$ being the partial derivative of p with respect to the j coordinate. In particular, when $t = 0$, we have

$$\frac{d}{dt}\mathbb{E}^{\mathcal{W}}\big[(f \circ T_{th})\psi\big]\Big|_{t=0} = \mathbb{E}^{W}\big[f \partial_h^\top \psi\big],$$

and so, if $\varphi \in \mathcal{P}$,

$$\mathbb{E}^{\mathcal{W}}\big[\partial_h \varphi \psi\big] = \mathbb{E}^{\mathcal{W}}\big[\varphi \partial_h^\top \psi\big]. \tag{$**$}$$

The equation in $(**)$ is the basic result, but there is a better way to formulate it. Namely, for $\varphi \in \mathcal{P}$, define $\nabla \varphi = \sum_{j=1}^N \partial_{h_j} \varphi h_j$. Next, let $\mathcal{P}H$ be the set of $\Psi : B \longrightarrow H$ of the form $\Psi = \sum_{j=1}^N \psi_j h_j$ for some $N \geq 1$ and $\{\psi_j : 1 \leq j \leq N\} \subseteq \mathcal{P}$, and define

$$\nabla^\top \Psi(x) = \sum_{j=1}^N \big(\langle x, \xi_j\rangle \psi_j(x) - \partial_{h_j} \psi_j(x)\big).$$

Then, after applying $(**)$ to each term in $\big(\nabla \varphi, \Psi\big)_H$, one sees that

$$\big(\nabla \varphi, \Psi\big)_{L^2(\mathcal{W};H)} = \big(\varphi \nabla^\top \Psi\big)_{L^2(\mathcal{W};\mathbb{R})}. \tag{3.3.3}$$

This equation allows us to extend the definition of ∇ well beyond $\mathcal{P}$. Namely, it shows that ∇ on $\mathcal{P}$ is a closeable operator. That is, the closure of its graph in $L^2(\mathcal{W};\mathbb{R}) \times L^2(\mathcal{W};H)$ is the graph of an operator. To see this, suppose that $\varphi_k \longrightarrow f$ in $L^2(\mathcal{W};\mathbb{R})$ and that $\nabla\varphi_k \longrightarrow F$ in $L^2(\mathcal{W};H)$. What we have to check is that F is uniquely determined by f. But, by (3.3.3),

$$\begin{aligned}\big(F,\Psi\big)_{L^2(\mathcal{W};H)} &= \lim_{k\to\infty}\big(\nabla\varphi_k,\Psi\big)_{L^2(\mathcal{W};H)}\\ &= \lim_{k\to\infty}\big(\varphi_k,\nabla^\top\Psi\big)_{L^2(\mathcal{W};\mathbb{R})} = \big(f,\nabla^\top\Psi\big)_{L^2(\mathcal{W};\mathbb{R})},\end{aligned}$$

which, since (cf. part (**viii**) of Exercise 3.7.1) $\mathcal{P}$ is dense in $L^2(\mathcal{W};\mathbb{R})$ and therefore $\mathcal{P}H$ is dense in $L^2(\mathcal{W};H)$, means that F is determined by f. Hence, if $W_1^2(\mathcal{W};\mathbb{R})$ is the set of $f \in L^2(\mathcal{W};\mathbb{R})$ for which there is an $F \in L^2(\mathcal{W};H)$ such that (f,F) is in the closure of the graph of ∇ on $\mathcal{P}$, then we can extend ∇ to $W_1^2(\mathcal{W};\mathbb{R})$ by taking $\nabla f = F$.

In the language of operator theory, the preceding can be summarized by saying that, because the domain of $\nabla^\top$ is dense, it has a well defined adjoint $(\nabla^\top)^\top$. Further, (3.3.3) says that $\mathcal{P}$ is contained in the domain of $(\nabla^\top)^\top$ and that $(\nabla^\top)^\top\varphi = \nabla\varphi$. Thus, since the graph of an adjoint is always closed and the graph of ∇ is contained in that of $(\nabla^\top)^\top$, the closure of the graph of ∇ is the graph of an operator, namely it is graph of the restriction of $(\nabla^\top)^\top$ to $W_1^2(\mathcal{W};\mathbb{R})$. What remains open is the question whether $W_1^2(\mathcal{W};\mathbb{R})$ is equal to the domain of $(\nabla^\top)^\top$. The reason why this is an important question is that, in applications, checking whether a function is in the domain of $(\nabla^\top)^\top$ is often much easier than showing directly that it is in $W_1^2(\mathcal{W};\mathbb{R})$. Indeed, $f \in L^2(\mathcal{W};\mathbb{R})$ will be in the domain of $(\nabla^\top)^\top$ if and only if there is a $C < \infty$ such that

$$\left|\big(f,\nabla^\top\Psi\big)_{L^2(\mathcal{W};\mathbb{R})}\right| \le C\|\Psi\|_{L^2(\mathcal{W};H)} \text{ for } \Psi \in \mathcal{P}H.$$

A proof that $W_1^2(\mathcal{W};\mathbb{R})$ is in fact equal to the domain of $(\nabla^\top)^\top$ is outlined in part (**viii**) of Exercise 3.3.1.

3.3.2 Some Examples of Abstract Wiener Spaces

The Gauss measures discussed in § 3.2.2 provide a ready source of examples. Given a positive definite, symmtric trace class operator A on a separable Hilbert space B, let $\mathcal{W}$ be the centered Gaussian measure on B with covariance $(x,Ay)_B$. Using the reasoning in § 3.1, one sees that B itself with the inner product $(x,y)_H = (x,A^{-1}y)_B$ is the Cameron–Martin subspace for $\mathcal{W}$. When B is infinite dimensional, A^{-1} will be an unbounded operator and its domain $\{Ax : x \in B\}$ will not be complete with respect to the Hilbert norm $\|\cdot\|_H$. To see what its completion is, choose an orthonormal basis $\{e_m : m \ge 0\}$ for B, where e_m is an eigenvectors for A with eigenvalue α_m. Then $(x,A^{-1}y)_B = \sum_{m=0}^\infty \alpha_m^{-1}(x,e_m)_B^2$, and so the $\|\cdot\|_H$-completion of $\{Ax : x \in B\}$ is the Hilbert space

$$H = \left\{ h \in B : \sum_{m=0}^{\infty} \alpha_m^{-1} (h, e_m)_B^2 < \infty \right\}$$

with inner product

$$(g, h)_H = \sum_{m=0}^{\infty} \alpha_m^{-1} (g, e_m)_B (h, e_m)_B.$$

It should now be evident that $(B, H, \mathcal{W})$ is an abstract Wiener space.

Although the construction described in Theorem 3.2.8 always works, it is far too crude to yield a refined result of the sort proved by N. Wiener when he constructed Brownian motion. To formulate Wiener's result in terms of an abstract Wiener spaces, take Ω_0 to be the space of all paths $\omega : \mathbb{R} \longrightarrow \mathbb{R}$ with the properties that $\omega(0) = 0$ and $\lim_{|t|\to\infty} \frac{|\omega(t)|}{|t|} = 0$. Then Ω_0 becomes a separable Banach space when one takes $\|\omega\|_{\Omega_0} = \sup_{t\in\mathbb{R}} \frac{|\omega(t)|}{1+|t|}$, and, using Riesz's Representation Theorem, it is easy to identify Ω_0^* with the space of signed Borel measures μ on $\mathbb{R}$ for which $\mu(\{0\}) = 0$ and $\|\mu\|_{\Omega_0^*} \equiv \int (1 + |t|)\,|\mu|(dt) < \infty$, where $|\mu|$ denotes the variation measure for μ (i.e., if $\mu = \mu_+ - \mu_-$ is the Hahn decomposition of μ, then $|\mu| = \mu_+ + \mu_-$).

It was shown in Theorem 2.5.7 that almost all Brownian paths are in Ω_0, and so their distribution induces a centered Gaussian measure $\mathcal{W}_0$ on Ω_0. The most intuitive way to guess the Cameron–Martin subspace for $\mathcal{W}_0$ is to think about its Feynman type representation. Namely, we are looking for a separable Hilbert space $H \subseteq \Omega_0$ with the property that, for any $n \geq 1$, the $\mathcal{W}_0$-distribution of $\left\{h\left(\frac{m}{n}\right) : -n^2 \leq m \leq n^2\right\}$ is

$$(2\pi)^{-n^2} \exp\left(-\frac{1}{2} \sum_{m=-n^2+1}^{n^2} n\Big(-\big(h\big(\tfrac{m}{n}\big) - h\big(\tfrac{m-1}{n}\big)^2\Big)\right) \prod_{\substack{-n^2\leq m\leq n^2 \\ m\neq 0}} dh\big(\tfrac{m}{n}\big).$$

Ignoring everything except the exponential in the density, and rewriting the sum in the exponent as

$$-\frac{1}{n} \sum_{m=-n^2+1}^{n^2} \left(\frac{h\big(\frac{m}{n}\big) - h\big(\frac{m-1}{n}\big)}{\frac{1}{n}}\right)^2,$$

one is led to the Feynman representation

$$\mathcal{W}_0(dh) = \frac{1}{Z} \exp\left(-\frac{1}{2} \int_{\mathbb{R}} \dot{h}(t)^2\, dt\right) \lambda_H(dh).$$

With this in mind, take $H_0^1(\mathbb{R};\mathbb{R})$ to be the set of absolutely continuous $h \in \Omega_0$ whose derivative is in $L^2(\lambda_{\mathbb{R}};\mathbb{R})$, and turn $H_0^1(\mathbb{R};\mathbb{R})$ into a separable Hilbert space with $\|h\|_{H_0^1(\mathbb{R};\mathbb{R})} = \|\dot{h}\|_{L^2(\lambda_{\mathbb{R}};\mathbb{R})}$. Using a standard mollification procedure, one sees that $H^1_0(\mathbb{R};\mathbb{R})$ is dense in Ω_0. Furthermore, $|h(t)| \le |t|^{\frac{1}{2}}\|h\|_{H^1_0(\mathbb{R};\mathbb{R})}$, and so $\|h\|_{\Omega_0} \le \|h\|_{H^1_0(\mathbb{R};\mathbb{R})}$, which means that $H^1_0(\mathbb{R};\mathbb{R})$ is continuously embedded in Ω_0.

Theorem 3.3.8 *If $\mathcal{W}_0 \in M_1(\Omega_0)$ is the distribution of a Brownian motion, then $\bigl(H^1_0(\mathbb{R};\mathbb{R}), \Omega_0, \mathcal{W}_0\bigr)$ is an abstract Wiener space.*

Proof. $\{\omega(t) : t \in \mathbb{R}\}$ is centered Gaussian process with covariance $w(s,t) = \mathbf{1}_{[0,\infty)}(st)|s| \wedge |t|$ under $\mathcal{W}_0$. Thus, if $\mu \in \Omega_0^*$ and one uses the formula

$$\langle \omega, \mu\rangle = \lim_{n\to\infty} \sum_{m=-n}^{n} \omega\bigl(\tfrac{m}{n}\bigr)\mu\Bigl(\bigl(\tfrac{m}{n}, \tfrac{m+1}{n}\bigr)\Bigr),$$

it is easy to check that $\omega \rightsquigarrow \langle \omega, \mu\rangle$ is a centered Gaussian random variable with variance $\iint w(s,t)\,\mu(ds)\mu(dt)$, and therefore $\{\langle\,\cdot\,,\mu\rangle : \mu \in \Omega_0^*\}$ is a centered Gaussian family with covariance function

$$c(\mu,\nu) = \iint w(s,t)\,\mu(ds)\mu(dt).$$

Next, for $\mu \in \Omega_0^*$, let h_μ be the function given by $h_\mu(t) = \int w(s,t)\,\mu(ds)$. Because

$$w(s,t) = \mathbf{1}_{[0,\infty)}(st)\int_0^\infty \mathbf{1}_{[0,|s|]}(\tau)\mathbf{1}_{[0,|t|]}(\tau)\,d\tau,$$

one sees that

$$h_\mu(t) = \begin{cases} \int_0^t \mu\bigl((\tau,\infty)\bigr)\,d\tau & \text{if } t \ge 0\\ \int_t^0 \mu\bigl((-\infty,\tau]\bigr)\,d\tau & \text{if } t < 0.\end{cases}$$

Thus h_μ is absolutely continuous, and

$$\dot h_\mu(t) = \begin{cases} \mu\bigl((t,\infty)\bigr) & \text{if } t \ge 0\\ -\mu\bigl((-\infty,t]\bigr) & \text{if } t < 0.\end{cases}$$

Since $|\mu|\bigl((-t,t)\complement\bigr) \le (1+t)^{-1}\|\mu\|_{\Omega_0^*}$, it follows that $h_\mu \in H^1_0(\mathbb{R};\mathbb{R})$. Furthermore, if $h \in H^1_0(\mathbb{R};\mathbb{R})$, then

$$(h,h_\mu)_{H^1_0} = \int_0^\infty \dot h(\tau)\mu\bigl([\tau,\infty)\bigr)\,d\tau - \int_{-\infty}^0 \dot h(\tau)\mu\bigl((-\infty,\tau]\bigr)\,d\tau = \langle h,\mu\rangle.$$

In particular,

$$\|h_\mu\|^2_{H^1_0(\mathbb{R};\mathbb{R})} = \langle h_\mu,\mu\rangle = \iint w(s,t)\,\mu(ds)\mu(dt) = c(\mu,\mu),$$

and so $H^1_0(\mathbb{R};\mathbb{R})$ is the Cameron–Martin space for $\mathcal{W}_0$. $\square$

One should understand how the preceding result can be interpreted in terms of the reasoning used for Hilbert spaces. Namely, the analog here of the operator A there is the operator $A : \Omega_0^* \longrightarrow \Omega_0$ given by $A\mu(t) = \int w(s,t)\,\mu(dt)$. Because

$$\partial_t^2 \int w(s,t)\varphi(s)\,ds = -\varphi(t) \text{ for } \varphi \in C^2_c(\mathbb{R};\mathbb{R}),$$

μ is minus the second distributional derivative of $h_\mu = A\mu$, and so $A^{-1}h_\mu = -\partial^2 h_\mu$. Therefore, if H is the Cameron–Martin space for $\mathcal{W}_0$, we should expect that

$$\big(h, h_\mu\big)_H = \big(-\partial^2 h, h_\mu\big)_{L^2(\lambda_{\mathbb{R}};\mathbb{R})} = \big(\dot{h}, \dot{h}_\mu\big)_{L^2(\lambda_R;\mathbb{R})},$$

at least for $h \in C^2_c(\mathbb{R};\mathbb{R})$. Starting from this, it is evident why H must be $H^1_0(\mathbb{R};\mathbb{R})$.

There is an important interpretation of the Payley–Wiener map $h \in H^1_0(\mathbb{R};\mathbb{R}) \longmapsto \mathcal{I}(h) \in L^2(\mathcal{W}_0;\mathbb{R})$. Namely, for $\mu \in \Omega_0^*$, $\langle \omega, \mu\rangle$ is the Lebesgue integral

$$\int \omega(t)\,\mu(dt) = (\mathrm{R})\int_0^\infty \omega(t)\,d\mu\big((t,\infty)\big) - (\mathrm{R})\int_{-\infty}^0 \omega(t)\,d\mu\big((-\infty,t)\big),$$

where, taking advantage of the facts that ω is continuous and $t \rightsquigarrow \mu\big((t,\infty)\big)$ and $t \rightsquigarrow \mu\big((-\infty,t)\big)$ are functions of bounded variation, the integrals on the right are taken in the sense of Riemann–Stieltjes. A fundamental property (cf. Theorem 1.2.1 in [11]) of Riemann–Stieltjes integration is that if φ is Riemann integrable with respect to ψ on an interval $[a,b]$, then ψ is Riemann integrable with respect to φ and

$$(\mathrm{R})\int_a^b \varphi\,d\psi = \varphi(b)\psi(b) - \varphi(a)\psi(a) - (\mathrm{R})\int_a^b \psi\,d\varphi.$$

Hence, since $\omega(0) = 0$ and $|\mu|\big((t,\infty)\big) \vee |\mu|\big((-\infty,-t)\big) \le \|\mu\|_{\Omega_0^*}(1+t)^{-1}$ for $t > 0$,

$$\langle \omega, \mu\rangle = (\mathrm{R})\int_0^\infty \mu\big((t,\infty)\,d\omega(t) - (\mathrm{R})\int_{-\infty}^0 \mu\big((-\infty,t)\,d\omega(t). \tag{3.3.4}$$

This means that $\mathcal{I}(h_\mu)$ is a bonafide Riemann–Stieltjes integral. Even though $\mathcal{I}(h)$ will be neither a Lebesgue or Riemann–Stieltjes for general $h \in H^1_0(\mathbb{R};\mathbb{R})$ and is only defined up to a set of $\mathcal{W}_0$-measure 0, it is the limit in $L^2(\mathcal{W}_0;\mathbb{R})$ of Riemann integrals, and that accounts for the use by many authors for the term Paley–Wiener integral.

Paley and Wiener did not interpret their integral in terms of abstract Wiener spaces. Instead, they developed their theory from the observation that if $f \in C^1_c(\mathbb{R};\mathbb{R})$, then $(\mathrm{R})\int f(t)\,d\omega(t)$ is a well defined Riemann integral which, under $\mathcal{W}_0$, is a centered Gaussian random variable with variance

$\|f\|^2_{L^2(\lambda_{\mathbb{R}};\mathbb{R})}$. They then used this observation and a mollification procedure to give meaning to $\int f(t)\,d\omega(t)$ for general $f \in L^2(\lambda_{\mathbb{R}};\mathbb{R})$. That is, given $f \in L^2(\lambda_{\mathbb{R}};\mathbb{R})$, they chose a sequence $\{f_n : n \geq 1\} \subseteq C^1_c(\mathbb{R};\mathbb{R})$ that converges to f in $L^2(\lambda_{\mathbb{R}};\mathbb{R})$ and realized that the integrals (R) $\int f_n(t)\,d\omega(t)$ would converge in $L^2(\mathcal{W}_0;\mathbb{R})$ to a centered Gaussian random variable, which they denoted by $\int f(t)\,d\omega(t)$, with variance $\|f\|^2_{L^2(\lambda_{\mathbb{R}};\mathbb{R})}$. From the abstract Wiener space perspective, what they did is start with $\mu \in \Omega_0^*$ which are absolutely continuous with respect to $\lambda_{\mathbb{R}}$ and have a continuous, compactly supported Radon–Nikodym derivative and then use the same extention procedure as we used in the proof of Theorem 3.3.4.

The stationary Gaussian processes discussed in § 2.6.3 are another source of abstract Wiener spaces. For example, recall the Ornstein–Uhlenbeck process discussed at the end of that section, only here rescale time and space its covariance function is $2^{-1}e^{-|t|}$ instead of $e^{-\frac{|t|}{2}}$. Again the paths of this process are continuous and have sublinear growth at infinity. Hence its distribution is a Borel measure on the space Ω of continuous functions $\omega : \mathbb{R} \longrightarrow \mathbb{R}$ satisfying $\lim_{|t|\to\infty} \frac{|\omega(t)}{|t|} = 0$, which becomes a separable Banach space when one uses the norm $\|\omega\|_\Omega = \sup_{t\in\mathbb{R}} \frac{|\omega(t)}{1+|t|}$. Just as before, Ω^* can be identified as the space of signed Borel measures on $\mathbb{R}$ for which $\int (1+|t|)\,|\mu|(dt) < \infty$, only now $\mu(\{0\})$ need not be 0. Using the fact that this Ornstien–Uhlenbeck process is a homogeneous Markov process with transition probability density

$$\big(\pi(1-e^{-2t})\big)^{-\frac{1}{2}} \exp\left(-\frac{\big(y-e^{-t}x\big)^2}{1-e^{-2t}}\right),$$

one can use the Feynman formalism to see that its Cameron–Martin subspace must consist of absolutely continuous $h \in \Omega$ for which $(h+\dot h) \in L^2(\lambda_{\mathbb{R}};\mathbb{R})$. Since elements of Ω are tempered distributions and therefore have Fourier transforms, the Fourier transform $\hat h$ of such an h will have the property that $(1-i\xi)\hat h \in L^2(\lambda_{\mathbb{R}};\mathbb{R})$ and therefore both h and $\dot h$ are in $L^2(\lambda_{\mathbb{R}};\mathbb{R})$ and $\|h+\dot h\|^2_{L^2(\lambda_{\mathbb{R}};\mathbb{R})} = \|h\|^2_{L^2(\lambda_{\mathbb{R}};\mathbb{R})} + \|\dot h\|^2_{L^2(\lambda_{\mathbb{R}};\mathbb{R})}$. This line of reasoning is made precise in the following theorem.

Theorem 3.3.9 *Let $\mathcal{U} \in M_1(\Omega)$ be the distribution of the Ornstein–Uhlenbeck process with covariance function $\frac{1}{2}e^{-|t|}$, and let $H^1(\mathbb{R};\mathbb{R})$ be the space of absolutely continuous $h : \mathbb{R} \longrightarrow \mathbb{R}$ for which both h and $\dot h$ are in $L^2(\lambda_{\mathbb{R}};\mathbb{R})$. Then $H^1(\mathbb{R};\mathbb{R})$ is a separable Hilbert space with norm*

$$\|h\|_{H^1(\mathbb{R};\mathbb{R})} = \Big(\|h\|^2_{L^2(\lambda_{\mathbb{R}};\mathbb{R})} + \|\dot h\|^2_{L^2(\lambda_{\mathbb{R}};\mathbb{R})}\Big)^{\frac{1}{2}},$$

$H^1(\mathbb{R};\mathbb{R})$ is dense in Ω, and $(H^1(\mathbb{R};\mathbb{R}), \Omega, \mathcal{U})$ is an abstract Wiener space.

Proof. Everything but the final statement is left as an exercise.

Proceeding as in the proof of Theorem 3.3.8, one sees that

$$\int \langle \omega, \mu \rangle \langle \omega, \nu \rangle \, \mathcal{U}(d\omega) = \frac{1}{2} \iint e^{-|t-s|} \, \mu(ds)\nu(dt) \text{ for } \mu, \nu \in \Omega^*.$$

Given $\mu \in \Omega^*$, define $h_\mu(t) = \frac{1}{2} \int e^{-|t-s|} \, \mu(ds)$. Then

$$\dot{h}_\mu(t) = -\frac{1}{2} \int \operatorname{sgn}(t-s) e^{-|t-s|} \, \mu(ds),$$

and so, for $h \in H^1(\mathbb{R}; \mathbb{R})$,

$$\begin{aligned} \big(\dot{h}, \dot{h}_\mu\big)_{L^2(\lambda_{\mathbb{R}}; \mathbb{R})} &= -\frac{1}{2} \int \left(\int \dot{h}(t) \operatorname{sgn}(t-s) e^{-|t-s|} \, dt \right) \mu(ds) \\ &= \int h(s) \, \mu(ds) - \int h(t) h_\mu(t) \, dt. \end{aligned}$$

Hence $\langle h, \mu \rangle = \big(h, h_\mu\big)_{H^1(\mathbb{R};\mathbb{R})}$. In addition,

$$\int \langle \omega, \mu \rangle^2 \, \mathcal{U}(d\omega) = \frac{1}{2} \iint e^{-|t-s|} \, \mu(ds)\mu(dt) = \langle h_\mu, \mu \rangle = \|h_\mu\|^2_{H^1(\mathbb{R};\mathbb{R})},$$

which completes the proof that $(H^1(\mathbb{R}; \mathbb{R}), \Omega, \mathcal{U})$ is an abstract Wiener space. □

Just as the Cameron–Martin subspace for the distribution of Brownian motion can be guessed using the ideas in § 3.1, so too Cameron–Martin subspace for the distribution of the Ornstein–Uhlenbeck process. The operator $A : \Omega^* \longrightarrow \Omega$ here is given by $A\mu(t) = \frac{1}{2} \int e^{-|t-s|} \, \mu(ds)$, and one can easily check that $A\mu - \partial^2 A\mu = \mu$ in the sense distributions. Using integration by parts and proceeding in the same way as we did before, one concludes that Cameron–Martin space for $\mathcal{U}$ must be $H^1(\mathbb{R}; \mathbb{R})$.

There is an interesting connection between stationary Gaussian processes and Paley–Wiener integrals. Namely, suppose that $f \in L^2(\lambda_{\mathbb{R}}; \mathbb{R})$ is an $\mathbb{R}$-valued characteristic function. Then $(2\pi)^{-1} \check{f}$ is the density of a even Borel probability measure on $\mathbb{R}$. Hence $\check{f}$ is a non-negative, even $\lambda_{\mathbb{R}}$-integrable function. Take ψ to be the Fourier transform of $\big((2\pi)^{-1} \check{f}\big)^{\frac{1}{2}}$, and observe that ψ is an even element of $L^2(\mathbb{R}; \mathbb{R})$ and that $f(t) = \int \psi(t-s)\psi(s) \, ds$. Using Paley–Wiener integration, now define

$$X(t, \omega) = \int \psi(t-\tau) \, d\omega(\tau) \in L^2(\mathcal{W}_0; \mathbb{R}) \text{ for } t \in \mathbb{R}.$$

Then

$$\begin{aligned} \int X(s, \omega) X(t, \omega) \, \mathcal{W}_0(d\omega) &= \int \psi(t-\tau)\psi(s-\tau) \, d\tau \\ &= \int \psi(t-s-\tau)\psi(\tau) \, d\tau = f(t-s), \end{aligned}$$

and so $\{X(t) : t \in \mathbb{R}\}$ under $\mathcal{W}_0$ is a centered stationary Gaussian process with covariance $f(t-s)$. This representation of stationary Gaussian processes is a key ingredient in much of prediction theory.

3.3.3 Wiener Series

In this subsection I will give a modern version of Wiener's approach to providing a mathematically satisfactory treatment of Brownian motion.

Theorem 3.3.10 *Let H be an infinite dimensional, separable, real Hilbert space and B a Banach space into which H is continuously embedded as a dense subspace. If for some orthonormal basis $\{h_m : m \geq 0\}$ in H the series*

$$\begin{gathered}\sum_{m=0}^{\infty} y_m h_m \text{ converges in } B \\ \text{for } \gamma_{0,1}^{\mathbb{N}}\text{-almost every } y = (y_0, \dots, y_m, \dots) \in \mathbb{R}^{\mathbb{N}}\end{gathered} \tag{3.3.5}$$

and if $S : \mathbb{R}^{\mathbb{N}} \longrightarrow B$ is given by

$$S(y) = \begin{cases} \sum_{m=0}^{\infty} y_m h_m & \text{when the series converges in } B \\ 0 & \text{otherwise,} \end{cases}$$

then $\big(H, B, \mathcal{W}\big)$ with $\mathcal{W} = S_*\gamma_{0,1}^{\mathbb{N}}$ is an abstract Wiener space. Conversely, if $(H, B, \mathcal{W})$ is an abstract Wiener space and $\{h_m : m \geq 0\}$ is an orthogonal sequence in H such that, for each $m \in \mathbb{N}$, either $h_m = 0$ or $\|h_m\|_H = 1$, then

$$\mathbb{E}^{\mathcal{W}}\left[\sup_{n\geq 0}\left\|\sum_{m=0}^{n} \mathcal{I}(h_m) h_m\right\|_B^p\right] < \infty \quad \text{for all } p \in [1, \infty), \tag{3.3.6}$$

and, for $\mathcal{W}$-almost every $x \in B$, $\sum_{m=0}^{\infty}[\mathcal{I}(h_m)](x)h_m$ converges in B to the $\mathcal{W}$-conditional expectation value of x given $\sigma\big(\{\mathcal{I}(h_m) : m \geq 0\}\big)$. Moreover,

$$\sum_{m=0}^{\infty}[\mathcal{I}(h_m)](x)h_m \text{ is } \mathcal{W}\text{-independent of } x - \sum_{m=0}^{\infty}[\mathcal{I}(h_m)](x)h_m.$$

Finally, if $\{h_m : m \geq 0\}$ is an orthonormal basis in H, then, for $\mathcal{W}$-almost every $x \in B$, $\sum_{m=0}^{\infty}[\mathcal{I}(h_m)](x)h_m$ converges in B to x, and the convergence is also in $L^p(\mathcal{W}; B)$ for every $p \in [1, \infty)$.

Proof. First assume that (3.3.5) holds for some orthonormal basis, and set $S_n(y) = \sum_{m=0}^{n} y_m h_m$ and $\mathcal{W} = S_*\gamma_{0,1}^{\mathbb{N}}$. Then, because $S_n(y) \longrightarrow S(y)$ in B for $\gamma_{0,1}^{\mathbb{N}}$-almost every $y \in \mathbb{R}^{\mathbb{N}}$,

$$\widehat{\mathcal{W}}(\xi) = \lim_{n\to\infty} \mathbb{E}^{\gamma_{0,1}^{\mathbb{N}}}\left[e^{i\langle S_n,\xi\rangle}\right] = \lim_{n\to\infty}\prod_{m=0}^{n} e^{-\frac{1}{2}(h_\xi,h_m)_H^2} = e^{-\frac{1}{2}\|h_\xi\|_H^2},$$

which proves that $(H,B,\mathcal{W})$ is an abstract Wiener space.

Next suppose that $(H,B,\mathcal{W})$ is an abstract Wiener space and that $\{h_m : m\ge 0\}$ is an orthogonal sequence with $\|h_m\|_H\in\{0,1\}$ for each $m\ge 0$. For each $n\in\mathbb{N}$, set $\mathcal{F}_n=\sigma\big(\{\mathcal{I}(h_m) : 0\le m\le n\}\big)$. Clearly, $\mathcal{F}_n\subseteq\mathcal{F}_{n+1}$ and $\mathcal{F}\equiv\sigma\left(\bigcup_{n=0}^{\infty}\mathcal{F}_n\right)$ is the σ-algebra generated by $\{\mathcal{I}(h_m) : m\ge 0\}$. Moreover, if $S_n=\sum_{m=0}^{n}\mathcal{I}(h_m)h_m$, then, since $\{\mathcal{I}(h_m) : m\ge 0\}$ is a Gaussian family and $\langle x-S_n(x),\xi\rangle$ is perpendicular in $L^2(\mathcal{W};\mathbb{R})$ to $\mathcal{I}(h_m)$ for all $\xi\in B^*$ and $0\le m\le n$, $x-S_n(x)$ is $\mathcal{W}$-independent of $\mathcal{F}_n$. Thus $S_n=\mathbb{E}^{\mathcal{W}}[x\,|\,\mathcal{F}_n]$, and so, by Theorem 3.2.5, we know both that (3.3.6) holds and that $S_n\longrightarrow\mathbb{E}^{\mathcal{W}}[x\,|\,\mathcal{F}]$ $\mathcal{W}$-almost surely and in $L^p(\mathcal{W};B)$ for all $p\in[1,\infty)$. In addition, the $\mathcal{W}$-independence of $S_n(x)$ from $x-S_n(x)$ implies that the limit quantities possess the same independence property.

In order to complete the proof at this point, all that we have to do is show that $x=\mathbb{E}^{\mathcal{W}}[x\,|\,\mathcal{F}]$ $\mathcal{W}$-almost surely when $\{h_m : m\ge 0\}$ is an orthonormal basis. Equivalently, we must check that $\mathcal{B}_B$ is contained in the $\mathcal{W}$-completion $\overline{\mathcal{F}}^{\mathcal{W}}$ of $\mathcal{F}$. To this end, note that, for each $h\in H$, because $\sum_{m=0}^{n}(h,h_m)_H h_m$ converges in H to h,

$$\sum_{m=0}^{n}\big(h,h_m\big)_H\mathcal{I}(h_m)=\mathcal{I}\left(\sum_{m=0}^{n}\big(h,h_m\big)_H h_m\right)\longrightarrow\mathcal{I}(h)\quad\text{in } L^2(\mathcal{W};\mathbb{R}).$$

Hence, $\mathcal{I}(h)$ is $\overline{\mathcal{F}}^{\mathcal{W}}$-measurable for every $h\in H$. In particular, this means that $x\rightsquigarrow\langle x,\xi\rangle$ is $\overline{\mathcal{F}}^{\mathcal{W}}$-measurable for every $\xi\in B^*$, and so, since $\mathcal{B}_B$ is generated by $\{\langle\,\cdot\,,\xi\rangle : \xi\in B^*\}$, $\mathcal{B}_B\subseteq\overline{\mathcal{F}}^{\mathcal{W}}$. □

It is important to acknowledge that the preceding theorem would not have made Wiener's life easier. Indeed, Wiener knew full well that what he had to do is prove a series of the sort described in the first part of Theorem 3.3.10 converges in the space of continuous paths. Being an expert in harmonic analysis, he chose an orthonormal basis consisting of trigonometric functions and used non-trivial results from harmonic analysis to prove the required convergence.

Corollary 3.3.11 *If $\mathcal{W}$ is a non-degenerate, centered Gaussian measure on the separable Banach space B, then $\mathcal{W}(G)>0$ for all non-empty open subsets $G\subseteq B$.*

Proof. Let H be the Cameron–Martin subspace for $\mathcal{W}$ on B, and assume that $\|\cdot\|_B\le\|\cdot\|_H$.

Since H is dense in B, it suffices to show that $\mathcal{W}\big(B_B(h,r)\big)>0$ for all $h\in H$ and $r>0$. In addition, because

$$\mathcal{W}\big(B_B(0,r)\big) = \int_{B_B(h,r)} R_{-h}\, d\mathcal{W} \le \|R_{-h}\|_{L^2(\mathcal{W};\mathbb{R})}\Big(\mathcal{W}\big(B_B(h,r)\big)\Big)^{\frac{1}{2}}$$

and

$$\|R_{-h}\|^2_{L^2(\mathcal{W};\mathbb{R})} = e^{-\|h\|_H^2}\int e^{-2\mathcal{I}(h)}\, d\mathcal{W} = e^{\|h\|_H^2},$$

it is enough to prove that $\mathcal{W}\big(B_B(0,r)\big) > 0$ for all $r > 0$. To this end, let $\{h_n : n \ge 0\}$ be an orthonormal basis in H, and set $S_n = \sum_{m=0}^n \mathcal{I}(h_m)h_m$. By Theorem 3.3.10, for each $r > 0$ there is an n such that $\mathcal{W}\big(\|x - S_n\|_B \le \frac{r}{2}\big) \ge \frac{1}{2}$. At the same time, $\|S_n\|_B^2 \le \|S_n\|_H^2 = \sum_{m=0}^n \mathcal{I}(h_m)^2$, and therefore, since $x - S_n$ is independent of S_n,

$$\mathcal{W}\big(B_B(0,r)\big) \ge \tfrac{1}{2}\gamma_{0,1}^{n+1}\Big(B_{\mathbb{R}^{n+1}}\big(t,\tfrac{r}{2}\big)\Big) > 0$$

for all $r > 0$. □

3.3.4 Isoperimetric Inequality for Abstract Wiener Space

By taking advantage of the dimension independence of (2.4.6), we will prove here the analog of (2.4.6) for abstract Wiener spaces.

Let $(H, B, \mathcal{W})$ be an infinite dimensional, non-degenerate abstract Wiener space. Given $A \in \mathcal{B}_B$ and $t \ge 0$, define

$$A^{(t)} = \big\{x + h : x \in A \;\&\; h \in \overline{B_H(0,t)}\big\}.$$

Equivalently, $x \in A^{(t)}$ if and only if there is a $y \in A$ such that $\|y - x\|_H \le t$. Notice that $A^{(t)}$ is significantly smaller than it would have been had we used $\|\cdot\|_B$ rather than $\|\cdot\|_H$ to define it. Also, observe that, because, by (**iii**) in Lemma 3.3.1, $\|\cdot\|_H$ is a lower semi-continuous function on B, $A^{(t)}$ is a closed subset of B and therefore Borel measurable.

Theorem 3.3.12 *For any $A \in \mathcal{B}_B$ and $t \ge 0$,*

$$\Phi^{-1}\big(\mathcal{W}(A^{(t)})\big) \ge \Phi^{-1}\big(\Phi^{-1}\big(\mathcal{W}(A)\big) + t\big). \tag{3.3.7}$$

In particular, if $A \in \mathcal{B}_B$ and $x \in A \implies x + h \in A$ for all $h \in H$, then $\mathcal{W}(A) \in \{0,1\}$.

Proof. Choose a sequence $\{\xi_n : n \ge 0\} \subseteq B^*\}$ so that $\{h_n : n \ge 0\}$ is an orthonormal basis in H when $h_n = h_{\xi_n}$. By Theorem 3.3.10, we know that $\mathcal{B}_B$ is contained in the $\mathcal{W}$-completion of $\sigma\big(\{\langle\,\cdot\,,\xi_n\rangle : n \ge 0\}\big)$. Thus, since the set of A for which (3.3.7) holds is closed under monotone limits, it suffices for us to show that it holds when $A \in \mathcal{F}_N \equiv \sigma\big(\{\langle\,\cdot\,,\xi_n\rangle : 0 \le n \le N-1\}\big)$

for some $N \geq 1$. To this end, define the projection map $\pi_N : B \longrightarrow \mathbb{R}^N$ by $\pi_N(x) = \big(\langle x, \xi_0\rangle, \dots, \langle x, \xi_{N-1}\rangle\big)^\top$. Then, $A \in \mathcal{F}_N$ if and only if $A = \pi_N^{-1}(\Gamma)$ for some $\Gamma \in \mathcal{B}_{\mathbb{R}^N}$, in which case $A^{(t)} = \pi_N^{-1}(\Gamma^{(t)})$, where, as in (2.4.6), $\Gamma^{(t)} = \{y \in \mathbb{R}^N : |y - \Gamma| \leq t\}$. Further, $\mathcal{W}(A) = \gamma_{01}^N(\Gamma)$ and $\mathcal{W}(A^{(t)}) = \gamma_{0,1}^N(\Gamma^{(t)})$, and so (3.3.7) for A follows from (2.4.6) for Γ.

Given the first part, the second part follows from the fact that, under the stated condition, $A = A^{(t)}$ for all $t > 0$, and therefore $\mathcal{W}(A) > 0 \implies \mathcal{W}(A) = \mathcal{W}(A^{(t)}) \geq \Phi\big(\Phi^{-1}\big(\mathcal{W}(A)\big) + t\big) \longrightarrow 1$ as $t \to \infty$. □

An interesting consequence of (3.3.7) is the fact that if $L \in \mathcal{B}_B$ is a subspace of B containing H, then $\mathcal{W}(L) \in \{0,1\}$. Thus, for example, with probability 0 or 1, a Brownian path will be in any Banach space of paths that contains $H_0^1(\mathbb{R};\mathbb{R})$.

Corollary 3.3.13 *Let $f : B \longrightarrow [-\infty, \infty]$ be a Borel measurable function satisfying $|f(y) - f(x)| \leq \lambda\|y - x\|_H$ for some $\lambda \in [0,\infty)$ and all $x, y \in B$. If $|f(x)| < \infty$ for $\mathcal{W}$-almost every $x \in B$ and m is a median of f under $\mathcal{W}$, then*

$$\mathcal{W}\big(\{|f(x) - m| \geq \lambda t\}\big) \leq 2\big(1 - \Phi(t)\big) \leq 2e^{-\frac{t^2}{2}}, \tag{3.3.8}$$

and so, for any $g \in C\big(\mathbb{R};[0,\infty)\big)$ satisfying $f(x) \leq f(y)$ for $|x| \leq |y|$,

$$\int_B g\big(|f(x) - m|\big)\,\mathcal{W}(dx) \leq \int_{\mathbb{R}} g(\lambda|t|)\,\gamma_{0,1}(dt). \tag{3.3.9}$$

In particular, if $\Sigma = \sup\big\{\|h\|_B : h \in \overline{B_H(0,1)}\big\}$ and m is the smallest median of $x \rightsquigarrow \|x\|_B$ under $\mathcal{W}$, then

$$\begin{aligned} 2\big(1 - \Phi(\Sigma^{-1}t)\big) &\leq \mathcal{W}\big(\{x : \|x\|_B \geq t\}\big) \\ &\leq 2\big(1 - \Phi(\Sigma^{-1}(t-m)^+\big) \leq 2e^{-\frac{\left((t-m)^+\right)^2}{2\Sigma^2}}, \end{aligned} \tag{3.3.10}$$

and so

$$\int \exp\left(\frac{\alpha(\|x\|_B - m)^2}{2\Sigma^2}\right)\mathcal{W}(dx) \leq \frac{1}{(1-\alpha)^{\frac{1}{2}}} \leq \int \exp\left(\frac{\alpha\|x\|_B^2}{2\Sigma^2}\right)\mathcal{W}(dx)$$

for $\alpha \in [0,1)$.

Proof. The estimate in (3.3.8) is proved from (3.3.7) in exactly the same way as the one in (2.4.13) was derived from (2.4.6). To prove the upper bounds in (3.3.10), simply observe that

$$\big|\|y\|_B - \|x\|_B\big| \leq \|x - y\|_B \leq \Sigma\|y - x\|_H.$$

The first step in proving the lower bound in (3.3.10) is to show that

$$\Sigma = \sup\{\|h_\xi\|_H : \|\xi\|_{B^*} = 1\}. \tag{$*$}$$

To check this, begin with the observation that

$$\Sigma = \sup\bigl\{\|h\|_H^{-1} : \|h\|_B = 1\bigr\}.$$

Now suppose that $\xi \in B^*$ with $\|\xi\|_{B^*} = 1$, and set $g = \frac{h_\xi}{\|h_\xi\|_B}$. Note that, since $\|g\|_B = 1$, $\|g\|_H^{-1} \le \Sigma$. Hence, because $1 \ge \langle g, \xi\rangle = \bigl(g, h_\xi\bigr)_H = \|g\|_H \|h_\xi\|_H$, $\|h_\xi\|_H \le \|g\|_H^{-1} \le \Sigma$, and so left hand side of $(*)$ dominates the right hand side. To prove the opposite inequality, suppose that $h \in H$ with $\|h\|_B = 1$. By the Hahn–Banach Theorem, there exists a $\xi \in B^*$ with $\|\xi\|_{B^*} = 1$ and $\langle h, \xi\rangle = 1$. In particular, $\|h\|_H \|h_\xi\|_H \ge \bigl(h, h_\xi\bigr)_H = \langle h, \xi\rangle = 1$, and therefore $\|h\|_H^{-1} \le \|h_\xi\|_H$, which, together with the preceding, completes the verification.

The next step is to show that there exists an $\xi \in B^*$ with $\|\xi\|_{B^*} = 1$ such that $\|h_\xi\|_H = \Sigma$. To this end, choose $\{\xi_k : k \ge 1\} \subseteq B^*$ with $\|\xi_k\|_{B^*} = 1$ so that $\|h_{\xi_k}\|_H \longrightarrow \Sigma$. Because $\overline{B_{B^*}(0,1)}$ is compact in the weak* topology and, by Theorem 3.3.4, $\xi \in \overline{B_{B^*}(0,1)} \longmapsto h_\xi \in H$ is continuous from the weak* topology into the strong topology, we can assume that $\{\xi_k : k \ge 1\}$ is weak* convergent to some $\xi \in \overline{B_{B^*}(0,1)}$ and that $\|h_\xi\|_H = \Sigma$, which is possible only if $\|\xi\|_{B^*} = 1$. Finally, knowing that this ξ exists, note that $\langle\,\cdot\,, \xi\rangle$ is a centered Gaussian under $\mathcal{W}$ with variance Σ^2. Hence, since $\|x\|_B \ge |\langle x, \xi\rangle|$,

$$\mathcal{W}\bigl(\|x\|_B \ge t\bigr) \ge \mathcal{W}\bigl(|\langle x, \xi\rangle| \ge t\bigr) = \gamma_{0,1}\bigl(\mathbb{R}\backslash(-\Sigma^{-1}t, \Sigma^{-1}t)\bigr) = 2\bigl(1 - \Phi(\Sigma^{-1}t)\bigr).$$

□

The assumption in Corollary 3.3.13 that f is $\mathcal{W}$-almost surely finite is vital. Indeed, otherwise we would get a contradiction by taking $f(x) = \|x\|_H$.

The estimates in (3.3.10) and its application represent a significant sharpening of Fernique's result. An asymptotic version of (3.3.10) was proved by M. Donsker and S.R.S. Varadhan using the result in Theorem 3.4.1 below. They showed that

$$\lim_{R\to\infty} R^{-2} \log \mathcal{W}\bigl(\|x\|_B \ge R) = -\frac{1}{2\Sigma^2},$$

and from this they concluded that $e^{\frac{\alpha\|x\|_B^2}{2\Sigma^2}}$ is $\mathcal{W}$-integrable if and only if $\alpha < 1$. The proof given here of the lower bound in (3.3.10) is an adaptation of their argument.

3.3.5 Rademacher's Theorem for Abstract Wiener Space

An important theorem proved by H. Rademacher states that if $f : \mathbb{R}^N \longrightarrow \mathbb{R}$ is a uniformly Lipschitz continuous function, then it has a gradient at $\lambda_{\mathbb{R}^N}$-almost every point. More precisely, there is a Borel measurable function $\nabla f : \mathbb{R}^N \longrightarrow \mathbb{R}^N$ such that

$$\lim_{t\searrow 0} \sup_{e\in\mathbb{S}^{N-1}} \frac{f(x+te) - f(x) - t\big(\nabla f(x), e\big)_{\mathbb{R}^N}}{t} = 0 \text{ for } \lambda_{\mathbb{R}^N} - \text{a.e. } x \in \mathbb{R}^N.$$

In particular, ∇f can be chosen so that $\big\| |\nabla f| \big\|_{\mathrm{c}}$ is the Lipschitz constant for f.

The analog for abstract Wiener spaces given below appeared originally in [3].

Theorem 3.3.14 *Let $(H, B, \mathcal{W})$ be an abstract Wiener space and $f : B \longrightarrow \mathbb{R}$ is Borel measurable function for which there exists a $\lambda < \infty$ such that $|f(y) - f(x)| \le \lambda \|y - x\|_H$ for all $x, y \in B$. Then there exists an $F \in L^\infty(\mathcal{W}; H)$ and an $A \in \mathcal{B}_B$ such that $\mathcal{W}(A) = 1$ and, for $x \in A$*

$$\frac{f(s+th) - f(x)}{t} - \big(F(x), h\big)_H \longrightarrow 0 \text{ as } t \to 0$$

uniformly for h in compact subsets of H.

Proof. First observe that, by Corollary 3.3.13, $f \in L^p(\mathcal{W}; \mathbb{R})$ for all $p \in [1, \infty)$.

Given $\xi \in B^*$ with $\|h_\xi\|_H = 1$, define

$$\varGamma_\xi = \left\{ x \in B : \lim_{t\searrow 0} \frac{f(x+th_\xi) - f(x)}{t} \text{ exists in } \mathbb{R} \right\}$$
$$\text{and } \varDelta_\xi = \big\{ (x, s) \in B \times \mathbb{R} : x + sh_\xi \in \varGamma_\xi \big\}.$$

Since the function $t \rightsquigarrow f(x + th_\xi)$ is Lipschitz continuous and therefore absolutely continuous, Lebesgue's Differentiation Theorem guarantees that, for each $x \in B$, $t \rightsquigarrow f(x + th_\xi)$ is $\lambda_\mathbb{R}$-almost everywhere differentiable, and therefore, for each $x \in B$, $\gamma_{0,1}\big(\{s : (x, s) \in \varDelta_\xi\}\big) = 1$. Next observe that

$$\varGamma_\xi = \big\{ x : \big(x - \langle x, \xi\rangle h_\xi, \langle x, \xi\rangle\big) \in \varDelta_\xi \big\},$$

and therefore, since $x - \langle x, \xi\rangle h_\xi$ is $\mathcal{W}$-independent of $\langle x, \xi\rangle \in N(0, 1)$,

$$\mathcal{W}(\varGamma_\xi) = \int \gamma_{0,1}\Big(\big\{ s : \big(x - \langle x, \xi\rangle, s\big) \in \varDelta_\xi \big\}\Big)\, \mathcal{W}(dx) = 1.$$

Define

$$F_\xi(x)=\begin{cases}\lim_{t\searrow 0}\frac{f(x+th_\xi)-f(x)}{t} & \text{if } x\in\Gamma_\xi\\ 0 & \text{if } x\notin\Gamma_\xi.\end{cases}$$

Now choose $\{\xi_m : m\ge 0\}\subseteq B^*$ so that $\{h_m : m\ge 0\}$ is an orthonormal basis for H when $h_m=h_{\xi_m}$, and take D be the set of $\xi\in\operatorname{span}\bigl(\{\xi_m : m\ge 0\}\bigr)$ for which $\{(h_\xi,h_m)_H : m\ge 0\}\subseteq\mathbb{Q}$. Clearly $\{h_\xi : \xi\in D\}$ is a dense subset of H, and since D is countable, $\Gamma\equiv\bigcap_{\xi\in D}\Gamma_\xi$ has $\mathcal{W}$-measure 1.

Given $\xi\in D$ and $\psi\in\mathcal{P}$ (cf. the notation in § 3.3.1),

$$\begin{aligned}\bigl(F_\xi,\psi\bigr)_{L^2(\mathcal{W};\mathbb{R})}&=\lim_{t\searrow 0}\int\frac{f(x+th_\xi)-f(x)}{t}\psi(x)\,\mathcal{W}(dx)\\&=\lim_{t\searrow 0}\int f(x)\frac{R_{th_\xi}\psi\circ T_{-th_\xi}(x)-\psi(x)}{t}\,\mathcal{W}(dx)\\&=\int f(x)\partial_{h_\xi}^{\top}\psi(x)\,\mathcal{W}(dx)=\sum_{m=0}^{\infty}(h,h_m)_H\int f(x)\partial_{h_m}^{\top}\psi(x)\,\mathcal{W}(dx).\end{aligned}$$

Since, for each m, by reversing the preceding one has

$$\begin{aligned}\int f(x)\partial_{h_m}^{\top}\psi(x)\,\mathcal{W}(dx)&=\lim_{t\searrow 0}\int\frac{f(x+th_m)-f(x)}{t}\psi(x)\,\mathcal{W}(dx)\\&=\int F_m(x)\psi(x)\,\mathcal{W}(dx),\end{aligned}$$

it follows that

$$(F_\xi,\psi)_{L^2(\mathcal{W};\mathbb{R})}=\sum_{m=0}^{\infty}(h_\xi,h_m)_H(F_m,\psi)_{L^2(\mathcal{W};\mathbb{R})}$$

for all $\psi\in\mathcal{P}$ and therefore that $F_\xi=\sum_{m=0}^{\infty}(h_\xi,h_m)_H F_m$ (a.s., $\mathcal{W}$) for each $\xi\in D$.

Let A be the set of $x\in\Gamma$ for which $F_\xi(x)=\sum_{m=0}^{\infty}(h_\xi,h_m)_H F_m(x)$ for all $\xi\in D$. Then $\mathcal{W}(A)=1$, and, for $x\in A$,

$$\left|\sum_{m=0}^{\infty}(h_\xi,h_m)_H F_m(x)\right|\le\lambda\|h_\xi\|_H\ \text{ for all }\xi\in D.$$

Thus $\sum_{m=0}^{\infty}F_m(x)^2\le\lambda^2$ for $x\in A$, and so, if we define the Borel measurable function $F:B\longrightarrow H$ by

$$F(x)=\begin{cases}\sum_{m=0}^{\infty}F_m(x)h_m & \text{if } x\in A\\ 0 & \text{if } x\notin A,\end{cases}$$

then $F_\xi(x)=\bigl(F(x),h_\xi\bigr)_H$ for $\xi\in D$ and $x\in A$. Finally, for $x\in A$ and $h\in H$, define $t\in[0,1]\longrightarrow[\rho(x,h)](t)\in\mathbb{R}$ so that $[\rho(x,h)](0)=0$ and

$$[\rho(x,h)](t) = \frac{f(x+th)-f(x)}{t} - \big(F(x),h\big)_H \text{ for } t \in (0,1]\setminus\{0\}.$$

Then,

$$\sup_{t\in(0,1]} \big|[\rho(x,g)](t) - [\rho(x,h)](t)\big| \le 2\lambda\|g-h\|_H \text{ for } g,\, h \in H,$$

and, for $x \in A$ and $\xi \in D$, $\rho(x,h_\xi) \in C\big([0,1];\mathbb{R}\big)$. Therefore $h \rightsquigarrow \rho(x,h) \restriction (0,1]$ is the continuous extention of its restriction to $\{h_\xi : \xi \in D\}$, and so $h \in H \longmapsto \rho(x,h) \in C\big([0,1];\mathbb{R}\big)$ is a continuous map. Hence, if $x \in A$, then, as $t \searrow 0$,

$$\frac{f(x+th)-f(x)}{t} - \big(F(x),h\big)_H = [\rho(x,h)](t) \longrightarrow 0$$

uniformly for h's in compact subsets of H. □

As a corollary of Theorem 3.3.14, one knows that a function f satisfying its hypothesis is an element of the space $W_1^2(\mathcal{W};\mathbb{R})$ described in § 3.3.1 and that $F = \nabla f$. Indeed, for any $\psi \in \mathcal{P}$ and $m \ge 0$,

$$\big(F_m,\psi\big)_{L^2(\mathcal{W};\mathbb{R})} = \big(f,\partial_{h_m}^\top\psi\big)_{L^2(\mathcal{W};\mathbb{R})},$$

and so

$$\big(F,\Psi\big)_{L^2(\mathcal{W};H)} = \big(f,\nabla^\top\Psi\big)_{L^2(\mathcal{W};\mathbb{R})}$$

for all $\Psi \in \mathcal{P}H$. Hence, f is in the domain of $(\nabla^\top)^\top$ and therefore, by (**viii**) of Exercise 3.3.1, it is an element of $W_1^2(\mathcal{W};\mathbb{R})$. Conversely, one can show that if $f \in W_1^2(\mathcal{W};\mathbb{R})$ and $\|\nabla f\|_{L^\infty(\mathcal{W};H)} < \infty$, then f is $\mathcal{W}$-almost everywhere equal to a Borel function $\tilde f$ satisfying the condition in Rademacher's Theorem.

Here is an amusing application of Theorem 3.3.14. Given a weak* compact subset K of B^*, set $f(x) = \max_{\xi\in K}\langle x,\xi\rangle$. Obviously, f satisfies the hypothesis in the theorem. Let A be the set of x with the property that

$$\big(F(x),h\big)_H = \lim_{t\searrow 0}\frac{f(x+th)-f(x)}{t} \text{ for all } h \in H,$$

and define $g : B \times H \longrightarrow \mathbb{R}$ by

$$g(x,h) = \max\big\{\langle h,\xi\rangle : \xi \in \Gamma_x\big\} \text{ where } \Gamma_x = \big\{\xi \in K : \langle x,\xi\rangle = f(x)\big\}.$$

If $x \in A$ and $\xi \in \Gamma_x$, then $\frac{f(x+th)-f(x)}{t} \ge \langle h,\xi\rangle$ for $t > 0$, and so $\big(F(x),h\big)_H \ge \langle h,\xi\rangle$ for all $h \in H$. Since this means that $\big(F(x),h\big)_H = (h,h_\xi)_H$ for all $h \in H$, it follows that $h_\xi = F(x)$ and therefore that there is precisely one element of Γ_x for each $x \in A$. In other words, for each $x \in A$, there is a unique $\xi_x \in K$ such that $\langle x,\xi_x\rangle = \max\big\{\langle x,\xi\rangle : \xi \in K\big\}$. When applied to

Brownian motion, this result says that, on each compact set of times, almost every Brownian path achieves its maximum value precisely once.

3.3.6 Gross's Operator Extention Procedure

A major goal of both Segal and Gross was to develop a procedure for extending operators defined on the Cameron–Martin space to the Banach space in which it is embedded. This subsection contains a couple of Gross's fundamental results in that direction. Throughout, H will be an infinite dimensional, separable Hilbert space over $\mathbb{R}$.

Lemma 3.3.15 *Let $(H,B,\mathcal{W})$ be an abstract Wiener space and $\{h_m : m \geq 0\}$ an orthonormal basis in H. Then, for each $h \in H$, $\sum_{m=0}^{\infty}(h,h_m)_H \mathcal{I}(h_m)$ converges to $\mathcal{I}(h)$ $\mathcal{W}$-almost surely and in $L^p(\mathcal{W};\mathbb{R})$ for every $p \in [1,\infty)$.*

Proof. Define the σ-algebras $\mathcal{F}_n$ and $\mathcal{F}$ as in the proof of Theorem 3.3.10. Then, by the same argument as was used there, one can identify $\sum_{m=0}^{n}(h,h_m)_H \mathcal{I}(h_m)$ as $\mathbb{E}^{\mathcal{W}}[\mathcal{I}(h)\,|\,\mathcal{F}_n]$. Thus, since $\overline{\mathcal{F}}^{\mathcal{W}} \supseteq \mathcal{B}_B$, the required convergence statement is an immediate consequence of Theorem 3.2.5. □

Theorem 3.3.16 *Let $(H,B,\mathcal{W})$ be an abstract Wiener space. For each finite dimensional subspace L of H there is a $\mathcal{W}$-almost surely unique map $P_L : E \longrightarrow H$ such that, for every $h \in H$ and $\mathcal{W}$-almost every $x \in B$, $\big(h, P_L x\big)_H = \mathcal{I}(\Pi_L h)(x)$, where Π_L denotes the orthogonal projection map from H onto L. In fact, if $\{g_1,\dots,g_{\dim(L)}\}$ is an orthonormal basis for L, then $P_L x = \sum_1^{\dim(L)}[\mathcal{I}(g_i)](x) g_i$, and so $P_L x \in L$ for $\mathcal{W}$-almost every $x \in B$. In particular, the distribution of $x \in B \longmapsto P_L x \in L$ under $\mathcal{W}$ is the same as that of $(y_1,\dots,y_{\dim(L)}) \in \mathbb{R}^{\dim(L)} \longmapsto \sum_1^{\dim(L)} y_j g_j \in L$ under $\gamma_{0,1}^{\dim(L)}$. Finally, $x \rightsquigarrow P_L x$ is $\mathcal{W}$-independent of $x \rightsquigarrow x - P_L x$.*

Proof. Set $\ell = \dim(L)$. It suffices to note that

$$\mathcal{I}(\Pi_L h) = \mathcal{I}\left(\sum_{k=1}^{\ell}(h,g_k)_H g_k\right) = \sum_{k=1}^{\ell}(h,g_k)_H \mathcal{I}(g_k) = \left(\sum_{k=1}^{\ell}\mathcal{I}(g_k)g_k, h\right)_H$$

for all $h \in H$ □

The definition of an abstract Wiener space that we have been using is not the same as Gross's. The difference is that his definition included the property that is derived in the following theorem.

Theorem 3.3.17 *Let $(H,B,\mathcal{W})$ be an abstract Wiener space and $\{h_n : n \geq 0\}$ an orthonormal basis for H. If $L_n = \operatorname{span}\big(\{h_0,\dots,h_n\}\big)$, then, for all $\epsilon > 0$ there exists an $n \in \mathbb{N}$ such that $\mathbb{E}^{\mathcal{W}}\big[\|P_L x\|_B^2\big] \leq \epsilon^2$ whenever L is a finite dimensional subspace of H that is perpendicular to L_n.*

Proof. Without loss in generality, we will assume that $\|\cdot\|_B \le \|\cdot\|_H$.

Arguing by contradiction, we will show that if the asserted property did not hold, then there would exist an orthonormal basis $\{f_n : n \ge 0\}$ for H such that $\sum_0^\infty \mathcal{I}(f_n)f_n$ fails to converge in $L^2(\mathcal{W};B)$.

Suppose that there is an $\epsilon > 0$ such that for all $n \in \mathbb{N}$ there exists a finite dimensional $L \perp L_n$ with $\mathbb{E}^{\mathcal{W}}\big[\|P_L x\|_B^2\big] \ge \epsilon^2$. Under this assumption, define $\{n_m : m \ge 0\} \subseteq \mathbb{N}$, $\{\ell_m : m \ge 0\} \subseteq \mathbb{N}$, and $\big\{\{f_0,\dots,f_{n_m}\} : m \ge 0\big\} \subseteq L_{n_m}$ inductively by the following prescription. First, take $n_0 = 0 = \ell_0$ and $f_0 = h_0$. Next, knowing n_m and $\{f_0,\dots,f_{n_m}\}$, choose a finite dimensional subspace $L \perp L_{n_m}$ so that $\mathbb{E}^{\mathcal{W}}\big[\|P_L x\|_B^2\big] \ge \epsilon^2$, set $\ell_m = \dim(L)$, and let $\{g_{m,1},\dots,g_{m,\ell_m}\}$ be an orthonormal basis for L. For any $\delta > 0$ there exists an $n \ge n_m + \ell_m$ such that

$$\sum_{j,k=1}^{\ell_m} \big|\big(\Pi_{L_n} g_{m,j}, \Pi_{L_n} g_{m,k}\big)_H - \delta_{j,k}\big| \le \delta.$$

In particular, if $\delta \in (0,1)$, then the elements of $\{\Pi_{L_n} g_{m,i} : 1 \le i \le \ell_m\}$ are linearly independent and the orthonormal set $\{\tilde{g}_{m,j} : 1 \le j \le \ell_m\}$ obtained from them via the Gram–Schmidt orthogonalization procedure satisfies

$$\sum_{j=1}^{\ell_m} \|\tilde{g}_{m,j} - \Pi_{L_n} g_{m,j}\|_H \le K_m \sum_{j,k=1}^{\ell_m} \big|\big(\Pi_{L_n} g_{m,j}, \Pi_{L_n} g_{m,k}\big) - \delta_{i,j}\big|$$

for some $K_m < \infty$ which depends only on ℓ_m. Moreover, because $L \perp L_{n_m}$, $\tilde{g}_{m,j} \perp L_{n_m}$ for all $1 \le j \le \ell_m$. Hence, we can find an $n_{m+1} \ge n_m + \ell_m$ for which $\operatorname{span}\big(\{h_n : n_m < n \le n_{m+1}\}\big)$ admits an orthonormal basis $\{f_{n_m+1},\dots,f_{n_{m+1}}\} \perp L_{n_m}$ with the property that $\sum_1^{\ell_m} \|g_{m,j} - f_{n_m+j}\|_H \le \frac{\epsilon}{4}$.

Clearly $\{f_n : n \ge 0\}$ is an orthonormal basis for H. On the other hand,

$$\begin{aligned}
&\mathbb{E}^{\mathcal{W}}\left[\left\|\sum_{n=n_m+1}^{n_m+\ell_m} \mathcal{I}(f_n)f_n\right\|_B^2\right]^{\frac{1}{2}} \\
&\qquad \ge \epsilon - \mathbb{E}^{\mathcal{W}}\left[\left\|\sum_1^{\ell_m}\big(\mathcal{I}(g_{m,j})g_{m,j} - \mathcal{I}(f_{n_m+j})f_{n_m+j}\big)\right\|_B^2\right]^{\frac{1}{2}} \\
&\qquad \ge \epsilon - \sum_1^{\ell_m} \mathbb{E}^{\mathcal{W}}\Big[\big\|\mathcal{I}(g_{m,j})g_{m,j} - \mathcal{I}(f_{n_m+j})f_{n_m+j}\big\|_H^2\Big]^{\frac{1}{2}},
\end{aligned}$$

and so, since $\mathbb{E}^{\mathcal{W}}\Big[\big\|\mathcal{I}(g_{i,m})g_{m,j} - \mathcal{I}(f_{n_m+i})f_{n_m+i}\big\|_H^2\Big]^{\frac{1}{2}}$ is dominated by

$$\mathbb{E}^{\mathcal{W}}\big[\big\|\big(\mathcal{I}(g_{m,j}) - \mathcal{I}(f_{n_m+j})\big)g_{m,j}\big\|_H^2\big]^{\frac{1}{2}} + \mathbb{E}^{\mathcal{W}}\big[\mathcal{I}(f_{n_m+j})^2\big]^{\frac{1}{2}}\|g_{m,j} - f_{n_m+j}\|_H$$
$$\leq 2\|g_{m,j} - f_{n_m+j}\|_H,$$

we have that

$$\mathbb{E}^{\mathcal{W}}\left[\left\|\sum_{n_m+1}^{n_m+\ell_m}\mathcal{I}(f_n)f_n\right\|_B^2\right]^{\frac{1}{2}} \geq \frac{\epsilon}{2} \quad \text{for all } m \geq 0,$$

and this means that $\sum_0^\infty \mathcal{I}(f_n)f_n$ cannot be converging in $L^2(\mathcal{W}; B)$. □

Besides showing that my definition of an abstract Wiener space is the same as Gross's, Theorem 3.3.17 allows us to prove a very convincing statement, again due to Gross, of just how non-unique is the Banach space for which a given Hilbert space is the Cameron–Martin subspace.

Corollary 3.3.18 *If $(H, B, \mathcal{W})$ is an abstract Wiener space, then there exists a separable Banach space B_0 that is continuously embedded in B as a measurable subset and has the properties that $\mathcal{W}(B_0) = 1$, bounded subsets of B_0 are relatively compact in B, and $\big(H, B_0, \mathcal{W} \upharpoonright B_0\big)$ is again an abstract Wiener space.*

Proof. Again we will assume that $\|\cdot\|_E \leq \|\cdot\|_H$.

Choose $\{\xi_n : n \geq 0\} \subseteq B^*$ so that $\{h_n : n \geq 0\}$ is an orthonormal basis in H when $h_n = h_{\xi_n}$, and set $L_n = \operatorname{span}\big(\{h_0, \dots, h_n\}\big)$. Next, using Theorem 3.3.17, choose an increasing sequence $\{n_m : m \geq 0\}$ so that $n_0 = 0$ and $\mathbb{E}^{\mathcal{W}}\big[\|P_L x\|_B^2\big]^{\frac{1}{2}} \leq 2^{-m}$ for $m \geq 1$ and finite dimensional $L \perp L_{n_m}$, and define Q_ℓ for $\ell \geq 0$ on B into H so that

$$Q_0 x = \langle x, \xi_0\rangle h_0 \quad \text{and} \quad Q_\ell x = \sum_{n=n_{\ell-1}+1}^{n_\ell}\langle x, \xi_n\rangle h_n \quad \text{when } \ell \geq 1.$$

Finally, set $S_m = P_{L_{n_m}} = \sum_{\ell=0}^m Q_\ell$, and define B_0 to be the set of $x \in B$ such that

$$\|x\|_{B_0} \equiv \|Q_0 x\|_E + \sum_{\ell=1}^\infty \ell^2\|Q_\ell x\|_B < \infty \quad \text{and} \quad \|S_m x - x\|_B \longrightarrow 0.$$

To show that $\|\cdot\|_{B_0}$ is a norm on B_0 and that B_0 with norm $\|\cdot\|_{B_0}$ is a Banach space, first note that if $x \in B_0$, then

$$\|x\|_B = \lim_{m\to\infty}\|S_m x\|_B \leq \|Q_0 x\|_B + \varliminf_{m\to\infty}\sum_{\ell=1}^m\|Q_\ell x\|_B \leq \|x\|_{B_0},$$

and therefore $\|\cdot\|_{B_0}$ is certainly a norm on B_0. Next, suppose that the sequence $\{x_k : k \geq 1\} \subseteq B_0$ is a Cauchy sequence with respect to $\|\cdot\|_{B_0}$. By the preceding, we know that $\{x_k : k \geq 1\}$ is also Cauchy convergent with respect to $\|\cdot\|_B$, and so there exists an $x \in B$ such that $x_k \longrightarrow x$ in B. We need to show that $x \in B_0$ and that $\|x_k - x\|_{B_0} \longrightarrow 0$. Because $\{x_k : k \geq 1\}$ is bounded in B_0, it is clear that $\|x\|_{B_0} < \infty$. In addition, for any $m \geq 0$ and $k \geq 1$,

$$\begin{aligned}\|x - S_m x\|_B &= \lim_{\ell\to\infty} \|x_\ell - S_m x_\ell\|_B \leq \varliminf_{\ell\to\infty} \|x_\ell - S_m x_\ell\|_{B_0} \\ &= \varliminf_{\ell\to\infty} \sum_{n>m} n^2 \|Q_n x_\ell\|_B \leq \sum_{n>m} n^2 \|Q_n x_k\|_B + \sup_{\ell>k} \|x_\ell - x_k\|_{B_0}.\end{aligned}$$

Thus, by choosing k for a given $\epsilon > 0$ so that $\sup_{\ell>k} \|x_\ell - x_k\|_{B_0} < \epsilon$, we conclude that $\varlimsup_{m\to\infty} \|x - S_m x\|_B < \epsilon$ and therefore that $S_m x \longrightarrow x$ in B. Hence, $x \in B_0$. Finally, to see that $x_k \longrightarrow x$ in B_0, simply note that

$$\begin{aligned}&\|x - x_k\|_{B_0} = \|Q_0(x - x_k)\|_B + \sum_{m=1}^{\infty} m^2 \|Q_m(x - x_k)\|_B \\ &\leq \varliminf_{\ell\to\infty} \left(\|Q_0(x_\ell - x_k)\|_B + \sum_{m=1}^{\infty} m^2 \|Q_m(x_\ell - x_k)\|_B \right) \leq \sup_{\ell>k} \|x_\ell - x_k\|_{B_0},\end{aligned}$$

which tends to 0 as $k \to \infty$.

To show that bounded subsets of B_0 are relatively compact in B, it suffices to show that for any sequence $\{x_\ell : \ell \geq 1\} \subseteq \overline{B_{B_0}(0,R)}$, then there is an $x \in B$ to which a subsequence converges in B. For this purpose, observe that, for each $m \geq 0$, there is a subsequence $\{x_{\ell_k} : k \geq 1\}$ along which $\{S_m x_{\ell_k} : k \geq 1\}$ converges in L_{n_m}. Hence, by a diagonalization argument, $\{x_{\ell_k} : k \geq 1\}$ can be chosen so that $\{S_m x_{\ell_k} : k \geq 1\}$ converges in L_{n_m} for all $m \geq 0$. Since, for $1 \leq j < k$,

$$\begin{aligned}\|x_{\ell_k} - x_{\ell_j}\|_B &\leq \|S_m x_{\ell_k} - S_m x_{\ell_j}\|_B + \sum_{n>m} \|Q_n(x_{\ell_k} - x_{\ell_j})\|_B \\ &\leq \|S_m x_{\ell_k} - S_m x_{\ell_j}\|_B + 2R \sum_{n>m} \frac{1}{n^2},\end{aligned}$$

it follows that $\{x_{\ell_k} : k \geq 1\}$ is Cauchy convergent in B and therefore that it converges in B.

We must still show that $B_0 \in \mathcal{B}_B$ and that $\big(H, B_0, \mathcal{W} \restriction B_0\big)$ is an abstract Wiener space. To see the first of these, observe that $x \in B \longmapsto \|x\|_{B_0} \in [0,\infty]$ is lower semicontinuous and that $\{x : \|S_m x - x\|_B \longrightarrow 0\} \in \mathcal{B}_B$. In addition, because, by Theorem 3.3.10, $\|S_m x - x\|_B \longrightarrow 0$ for $\mathcal{W}$-almost every $x \in B$, we will know that $\mathcal{W}(B_0) = 1$ once we show that $\mathcal{W}\big(\|x\|_{B_0} < \infty\big) = 1$, which follows immediately from

$$\begin{aligned}
\mathbb{E}^{\mathcal{W}}\big[\|x\|_{B_0}\big] &= \mathbb{E}^{\mathcal{W}}\big[\|Q_0 x\|_B\big] + \sum_1^\infty m^2 \mathbb{E}^{\mathcal{W}}\big[\|Q_m x\|_B\big] \\
&\le \mathbb{E}^{\mathcal{W}}\big[\|Q_0 x\|_B\big] + \sum_1^\infty m^2 \mathbb{E}^{\mathcal{W}}\big[\|Q_m x\|_B^2\big]^{\frac{1}{2}} \\
&\le \mathbb{E}^{\mathcal{W}}\big[\|Q_0 x\|_B\big] + \sum_{m=1}^\infty m^2 2^{-m} < \infty.
\end{aligned}$$

The next step is to check that H is continuously embedded in B_0. Certainly $h \in H \implies \|S_m h - h\|_B \le \|S_m h - h\|_H \longrightarrow 0$. Next suppose that $h \in H \setminus \{0\}$ and that $h \perp L_{n_m}$, and let L be the line in H spanned by h. Then $P_L x = \|h\|_H^{-2}[\mathcal{I}(h)](x)h$, and so, because $L \perp L_{n_m}$,

$$\frac{1}{2^m} \ge \mathbb{E}^{\mathcal{W}}\big[\mathcal{I}(h)^2\big]^{\frac{1}{2}} \frac{\|h\|_B}{\|h\|_H^2} = \frac{\|h\|_B}{\|h\|_H}.$$

Hence, we now know that $h \perp L_{n_m} \implies \|h\|_B \le 2^{-m}\|h\|_H$. In particular, $\|Q_{m+1}h\|_B \le 2^{-m}\|Q_{m+1}h\|_H \le 2^{-m}\|h\|_H$ for all $m \ge 0$ and $h \in H$, and so

$$\|h\|_{B_0} = \|Q_0 h\|_B + \sum_{m=1}^\infty m^2 \|Q_m h\|_B \le \left(1 + 2\sum_{m=1}^\infty \frac{m^2}{2^m}\right)\|h\|_H = 25\|h\|_H.$$

To complete the proof, we must show that H is dense in B_0 and that, for each $\xi \in B_0^*$, $\widehat{\mathcal{W}}(\xi) = e^{-\frac{1}{2}\|h_\xi\|_H^2}$, where $h_\xi \in H$ is determined by $\big(h, h_\xi\big)_H = \langle h, \xi\rangle$ for $h \in H$. Both these facts rely on the observation that

$$\|x - S_m x\|_{B_0} = \sum_{n>m} n^2 \|Q_n x\|_B \longrightarrow 0 \quad \text{for all } x \in B_0.$$

Knowing this, the density of H in B_0 is obvious. Finally, if $\xi \in B_0^*$, then, by the preceding and Lemma 3.3.15,

$$\begin{aligned}
\langle x, \xi\rangle &= \lim_{m\to\infty} \langle S_m x, \xi\rangle = \lim_{m\to\infty} \sum_{n=0}^{n_m} \langle x, \xi_n\rangle \langle h_n, \xi\rangle \\
&= \lim_{m\to\infty} \sum_{n=0}^{n_m} \big(h_\xi, h_n\big)_H \big[\mathcal{I}(h_n)\big](x) = \big[\mathcal{I}(h_\xi)\big](x)
\end{aligned}$$

for $\mathcal{W}$-almost every $x \in B_0$. Hence $\langle\,\cdot\,, \xi\rangle$ under $\mathcal{W}$ is a centered Gaussian with variance $\|h_\xi\|_H^2$. $\square$

Using the ideas in the preceding proof, one can show that, given a separable Hilbert space H, H itself equals the intersection of all the Banach spaces in which H is the Cameron–Martin subspace. See [2] for details.

3.3.7 Orthogonal Invariance

Consider the standard Gauss distribution $\gamma_{0,I}$ on $\mathbb{R}^N$. Obviously, $\gamma_{0,I}$ is orthogonal invariant. That is, if $\mathcal{O}$ is an orthogonal transformation of $\mathbb{R}^N$, then $\gamma_{0,I}$ is invariant under the transformation $T_{\mathcal{O}} : \mathbb{R}^N \longrightarrow \mathbb{R}^N$ given by $T_{\mathcal{O}}x = \mathcal{O}x$. On the other hand, none of these transformations can be ergodic, since any radial function on $\mathbb{R}^N$ is invariant under $T_{\mathcal{O}}$ for every $\mathcal{O}$.

Now think about the analogous situation when $\mathbb{R}^N$ is replaced by an infinite dimensional Hilbert space H and $(H, B, \mathcal{W})$ is an associated abstract Wiener space. As we are about to show, $\mathcal{W}$ is invariant under orthogonal transformations on H. On the other hand, because $\|x\|_H = \infty$ for $\mathcal{W}$-almost every $x \in B$, there are no non-trivial radial functions now, a fact that leaves open the possibility that some orthogonal transformation of H give rise to ergodic transformations for $\mathcal{W}$. The purpose of this subsection is to investigate these matters, and I begin with the following formulation of the orthogonal invariance of $\mathcal{W}$.

Theorem 3.3.19 *Let $(H, B, \mathcal{W})$ be an abstract Wiener space and $\mathcal{O}$ an orthogonal transformation on H. Then there is a $\mathcal{W}$-almost surely unique, Borel measurable map $T_{\mathcal{O}} : B \longrightarrow B$ such that $\mathcal{I}(h) \circ T_{\mathcal{O}} = \mathcal{I}(\mathcal{O}^\top h)$ $\mathcal{W}$-almost surely for each $h \in H$. Moreover, $\mathcal{W} = (T_{\mathcal{O}})_* \mathcal{W}$.*

Proof. To prove uniqueness, note that if T and T' both satisfy the defining property for $T_{\mathcal{O}}$, then, for each $\xi \in B^*$,

$$\langle Tx, \xi \rangle = \mathcal{I}(h_\xi)(Tx) = \mathcal{I}(\mathcal{O}^\top h_\xi) = \mathcal{I}(h_\xi)(T'x) = \langle T'x, \xi \rangle$$

for $\mathcal{W}$-almost every $x \in B$. Hence, since $B_{B^*}(0,1)$ is separable in the weak* topology, $Tx = T'x$ for $\mathcal{W}$-almost every $x \in B$.

To prove existence, choose an orthonormal basis $\{h_m : m \geq 0\}$ for H, and let C be the set of $x \in B$ for which both $\sum_{m=0}^\infty [\mathcal{I}(h_m)](x) h_m$ and $\sum_{m=0}^\infty [\mathcal{I}(h_m)](x) \mathcal{O} h_m$ converge in B. By Theorem 3.3.10, we know that $\mathcal{W}(C) = 1$ and that

$$x \rightsquigarrow T_{\mathcal{O}}x \equiv \begin{cases} \sum_{m=0}^\infty [\mathcal{I}(h_m)](x) \mathcal{O} h_m & \text{if } x \in C \\ 0 & \text{if } x \notin C \end{cases}$$

has distribution $\mathcal{W}$. Hence, all that remains is to check that $\mathcal{I}(h) \circ T_{\mathcal{O}} = \mathcal{I}(\mathcal{O}^\top h)$ $\mathcal{W}$-almost surely for each $h \in H$. To this end, let $\xi \in B^*$, and observe that

$$\begin{aligned} [\mathcal{I}(h_\xi)](T_{\mathcal{O}}x) = \langle T_{\mathcal{O}}x, \xi \rangle &= \sum_{m=0}^\infty \big(h_\xi, \mathcal{O} h_m\big)_H [\mathcal{I}(h_m)](x) \\ &= \sum_{m=0}^\infty \big(\mathcal{O}^\top h_\xi, h_m\big)_H [\mathcal{I}(h_m)](x) \end{aligned}$$

for $\mathcal{W}$-almost every $x \in B$. Thus, since, by Lemma 3.3.15, the last of these series convergences $\mathcal{W}$-almost surely to $\mathcal{I}(\mathcal{O}^\top h_\xi)$, we have that $\mathcal{I}(h_\xi) \circ T_{\mathcal{O}} = \mathcal{I}(\mathcal{O}^\top h_\xi)$ $\mathcal{W}$-almost surely. To handle general $h \in H$, simply note that both $h \in H \longmapsto \mathcal{I}(h) \circ T_{\mathcal{O}} \in L^2(\mathcal{W};\mathbb{R})$ and $h \in H \longmapsto \mathcal{I}(\mathcal{O}^\top h) \in L^2(\mathcal{W};\mathbb{R})$ are isometric, and remember that $\{h_\xi : \xi \in B^*\}$ is dense in H. □

I discuss next the possibility of $T_{\mathcal{O}}$ being ergodic for some orthogonal transformations $\mathcal{O}$. First notice that $T_{\mathcal{O}}$ cannot be ergodic if $\mathcal{O}$ has a non-trivial, finite dimensional invariant subspace L, since if $\{h_1, \dots, h_n\}$ were an orthonormal basis for L, then $\sum_{m=1}^n \mathcal{I}(h_m)^2$ would be a non-constant, $T_{\mathcal{O}}$-invariant function. Thus, the only candidates for ergodicity are $\mathcal{O}$'s that have no non-trivial, finite dimensional, invariant subspaces. In a more general and highly abstract context, Segal [7] showed that the existence of a non-trivial, finite dimensional subspace for $\mathcal{O}$ is the only obstruction to $T_{\mathcal{O}}$ being ergodic. Here I will show less.

Theorem 3.3.20 *Let $(H, B, \mathcal{W})$ be an abstract Wiener space. If $\mathcal{O}$ is an orthogonal transformation on H with the property that, for every $g, h \in H$, $\varliminf_{n\to\infty} \big(\mathcal{O}^n g, h\big)_H = 0$, then $T_{\mathcal{O}}$ is ergodic.*

Proof. What we have to show is that any $T_{\mathcal{O}}$-invariant element $\Phi \in L^2(\mathcal{W};\mathbb{R})$ is $\mathcal{W}$-almost surely constant, and for this purpose it suffices to check that

$$\varliminf_{n\to\infty} \left|\mathbb{E}^{\mathcal{W}}\big[(\Phi \circ T_{\mathcal{O}}^n)\Phi\big]\right| = 0 \tag{$*$}$$

for all $\Phi \in L^2(\mathcal{W};\mathbb{R})$ with mean value 0. In fact, if $\{h_m : m \geq 1\}$ is an orthonormal basis for H, then it suffices to check $(*)$ when

$$\Phi(x) = F\Big([\mathcal{I}(h_1)](x), \dots, [\mathcal{I}(h_N)](x)\Big)$$

for some $N \in \mathbb{Z}^+$ and bounded, Borel measurable $F : \mathbb{R}^N \longrightarrow \mathbb{R}$. The reason why it is sufficient to check it for such Φ's is that, because $T_{\mathcal{O}}$ is $\mathcal{W}$-measure preserving, the set of Φ's for which $(*)$ holds is closed in $L^2(\mathcal{W};\mathbb{R})$. Hence, if we start with any $\Phi \in L^2(\mathcal{W};\mathbb{R})$ with mean value 0, we can first approximate it in $L^2(\mathcal{W};\mathbb{R})$ by bounded functions with mean value 0 and then condition these bounded approximates with respect to $\sigma\big(\{\mathcal{I}(h_1), \dots, \mathcal{I}(h_N)\}\big)$ to give them the required form.

Now suppose that $\Phi = F\big(\mathcal{I}(h_1), \dots, \mathcal{I}(h_N)\big)$ for some N and bounded, measurable F. Then

$$\mathbb{E}^{\mathcal{W}}\big[(\Phi \circ T_{\mathcal{O}}^n)\Phi\big] = \iint_{\mathbb{R}^N \times \mathbb{R}^N} F(\xi)F(\eta)\, \gamma_{0,A_n}(d\xi \times d\eta),$$

where

$$A_n = \begin{pmatrix} I & B_n \\ B_n^\top & I \end{pmatrix} \quad \text{with } B_n = \Big(\Big(\big(h_k, \mathcal{O}^n h_\ell\big)_H\Big)\Big)_{1\le k,\ell\le N}$$

and the block structure corresponds to $\mathbb{R}^N \times \mathbb{R}^N$. Finally, by our hypothesis about $\mathcal{O}$, we can find a subsequence $\{n_m : m \ge 0\}$ such that $\lim_{m\to\infty} B_{n_m} = 0$, from which it is clear that $\gamma_{0,C_{n_m}}$ tends to $\gamma_{0,I}\times\gamma_{0,I}$ in variation and therefore

$$\lim_{m\to\infty} \mathbb{E}^{\mathcal{W}}\big[(\Phi\circ T_{\mathcal{O}}^{n_m})\Phi\big] = \mathbb{E}^{\mathcal{W}}[\Phi]^2 = 0.$$

□

Perhaps the best tests for whether an orthogonal transformation satisfies the hypothesis in Theorem 3.3.20 come from spectral theory. To be more precise, if H_c and $\mathcal{O}_c$ are the space and operator obtained by complexifying H and $\mathcal{O}$, the Spectral Theorem for normal operators allows one to write

$$\mathcal{O}_c = \int_0^{2\pi} e^{i\alpha}\, dE_\alpha,$$

where $\{E_\alpha : \alpha \in [0,2\pi)\}$ is a resolution of the identity in H_c by orthogonal projection operators. The spectrum of $\mathcal{O}_c$ is said to be *absolutely continuous* if, for each $h \in H_c$, the non-decreasing function $\alpha \rightsquigarrow \big(E_\alpha h, h\big)_{H_c}$ is absolutely continuous, which, by polarization, means that $\alpha \rightsquigarrow \big(E_\alpha h, h'\big)_{H_c}$ is absolutely continuous for all $h,\, h' \in H_c$. The reason for introducing this concept here is that, by combining the Riemann–Lebesgue Lemma with Theorem 3.3.20, one can prove that $T_{\mathcal{O}}$ *is ergodic if the spectrum of* $\mathcal{O}_c$ *is absolutely continuous.* Indeed, given $h,\, h' \in H$, let f be the Radon–Nikodym derivative of $\alpha \rightsquigarrow \big(E_\alpha h, h'\big)_{H_c}$, and apply the Riemann–Lebesgue Lemma to see that

$$\big(\mathcal{O}^n h, h'\big)_H = \int_0^{2\pi} e^{in\alpha} f(\alpha)\, d\alpha \longrightarrow 0 \quad \text{as } n\to\infty.$$

This conclusion is substantially weaker than one proved by Segal. He proved that $T_{\mathcal{O}}$ will be ergodic if and only if $\mathcal{O}$ has no non-trivial eigenfunctions. See [12] for another proof of this result.

Exercise 3.3.1

(**i**) Let μ be a Borel probability measure on the separable Banach space E, set $A = \{x \in E : (T_x)_*\mu \ll \mu\}$, and, for $x \in A$, let R_x be the Radon-Nikodym derivative of $(T_x)_*\mu$ with respect to μ. Show that, for $x, y \in A \implies x+y \in A$ and $R_{x+y} = R_y \circ T_{-x} R_x$. Use this to show that if $(H, B, \mathcal{W})$ be an abstract Wiener space and $g, h \in H$, then $\mathcal{I}(g)\circ T_h = \mathcal{I}(g) + (g,h)_H$ (a.s., $\mathcal{W}$).

(**ii**) Let $\{X(t) : t \in \mathbb{R}\}$ be a centered Gaussian process with covaiance function $\cos\big(2\pi(t-s)\big)$. Describe the distribution of this process in terms of an abstract Wiener space.

(**iii**) Let $(H, B, \mathcal{W})$ be an abstract Wiener space, and suppose that $\{\xi_n : n \geq 0\} \subseteq B^*$ has the property that $\{h_n : n \geq 0\}$ is an orthonormal sequence when $h_h = h_{\xi_n}$. Set $S_n(x) = \sum_{m=0}^{n} \langle x, \xi_m \rangle h_m$, and show that there exists an $\alpha > 0$ such that

$$\mathbb{E}\left[\exp\left(\alpha \sup_{n \geq 0} |S_n|\right)\right] < \infty.$$

(**iv**) Let $\bigl(H_0^1(\mathbb{R};\mathbb{R}), \Omega_0, \mathcal{W}_0\bigr)$ be the abstract Wiener space for Brownian motion, and let $0 = s_0 < s_1 < \cdots < s_n$ be given. Set $L = \operatorname{span}\bigl(\{h_m : 0 \leq m \leq n\}\bigr)$ where $h_m(t) = (t \wedge s_m)^+$ for $0 \leq m \leq n$, and show that

$$P_L\omega(t) = \sum_{m=1}^{n} \frac{\bigl(\omega(s_m) - \omega(s_{m-1})\bigr)\bigl(h_m(t) - h_{m-1}(t)\bigr)}{s_m - s_{m-1}}.$$

Next, for $y \in \mathbb{R}^n$, define

$$\psi(t, y) = \sum_{m=1}^{n} \frac{(y_m - y_{m-1})\bigl(h_m(t) - h_{m-1}(t)\bigr)}{s_m - s_{m-1}}$$

where $y_0 = 0$. Show that if $F : \Omega_0 \longrightarrow [0, \infty)$ is $\mathcal{B}_{\Omega_0}$-measurable, then, for all $\Gamma \in \mathcal{B}_{\mathbb{R}^n}$,

$$\mathbb{E}^{\mathcal{W}_0}\bigl[F,\ \bigl(\omega(s_1), \ldots, \omega(s_n)\bigr) \in \Gamma\bigr] = \int_{\Gamma} \mathbb{E}^{\mathcal{W}_0}\bigl[F\bigl(P_L\omega + \psi(\,\cdot\,, y)\bigr)\bigr]\, \gamma_{0,A}(dy),$$

where $A = \bigl((s_k \wedge s_\ell)\bigr)_{1 \leq k, \ell \leq n}$, and conclude that the distribution of $P_L\omega$ is that of Brownian motion conditioned to return to 0 at times $s_1 < \cdots < s_n$.

(**v**) Let $\bigl(\Omega, H^1(\mathbb{R};\mathbb{R}), \mathcal{U}\bigr)$ be the abstract Wiener space in Theorem 3.3.9. Given $s_1 < s_2$, set $L = \operatorname{span}\bigl(\{h_1, h_2\}\bigr)$ where $h_k(t) = \frac{1}{2}e^{-|t - s_k|}$, and show that

$$P_L\omega(t) = 2\omega(s_1)h_1(t) + \frac{2\bigl(\omega(s_2) - 2h_1(s_2)\omega(s_1)\bigr)\bigl(h_2(t) - 2h_1(s_2)h_1(t)\bigr)}{1 - 4h_1(s_2)^2}.$$

Conclude that the distribution of $\omega - P_L\omega$ is that of the Ornstein–Uhlenbeck process conditioned to be at 0 at times s_1 and s_2.

(**vi**) Show that if $\mathcal{O} : H_0^1(\mathbb{R};\mathbb{R}) \longrightarrow H_0^1(\mathbb{R};\mathbb{R})$ is an orthogonal transformation and $\lim_{n\to\infty}\bigl(\mathcal{O}^n g, h\bigr)_{H_0^1(\mathbb{R};\mathbb{R})} = 0$ for all $g, h \in H_0^1(\mathbb{R};\mathbb{R}) \cap C_c^\infty(\mathbb{R};\mathbb{R})$, then $\lim_{n\to\infty}\bigl(\mathcal{O}^n g, h\bigr)_{H_0^1(\mathbb{R};\mathbb{R})} = 0$ for all $g, h \in H_0^1(\mathbb{R};\mathbb{R})$.

(**vii**) Given $\alpha \in (0, \infty)$, define $\mathcal{O}_\alpha : H_0^1(\mathbb{R};\mathbb{R}) \longrightarrow H_0^1(\mathbb{R};\mathbb{R})$ by $\mathcal{O}_\alpha h(t) = \alpha^{-\frac{1}{2}}h(\alpha t)$, and check that $\mathcal{O}_\alpha$ is an orthogonal transformation and that it extends to Ω_0 as the scaling map in (**ii**) of Theorem 2.5.7. Next, use (**vi**) above and Theorem 3.3.20 to show that $\mathcal{O}_\alpha$ is ergodic under $\mathcal{W}_0$ if $\alpha \neq 1$.

(**viii**) Let H be an infinite dimensional, separable Hilbert space and $(H, B, \mathcal{W})$ an abstract Wiener space. Refering to the discussion at the end of § 3.3.1, choose $\{\xi_j : j \geq 1\} \subseteq B^*$ so that $\{h_j : j \geq 1\}$ is an orthonormal basis for H when $h_j = h_{\xi_j}$, and define the spaces $\mathcal{P}$, $\mathcal{P}H$ and $W_1^2(\mathcal{W};\mathbb{R})$ as they were defined there. Show that $\mathcal{P}$ are dense in $L^2(\mathcal{W};\mathbb{R})$ and $L^2(\mathcal{W};H)$, respectively. Next, recall the operators $\nabla : \mathcal{P} \longrightarrow \mathcal{P}H$ and $\nabla^\top : \mathcal{P}H \longrightarrow \mathcal{P}$, and note that $\mathcal{P}$ and therefore $W_1^2(\mathcal{W};\mathbb{R})$ are contained in the domain $\mathrm{Dom}\big((\nabla^\top)^\top\big)$ of the adjoint $(\nabla^\top)^\top$ of $\nabla^\top$. You are to show that in fact $W_1^2(\mathcal{W};\mathbb{R}) = \mathrm{Dom}\big((\nabla^\top)^\top\big)$.

Hint: Here are steps that you might want to take. Given $f \in \mathrm{Dom}\big((\nabla^\top)^\top\big)$, set $F = (\nabla^\top)^\top f$. For $N \geq 1$, let $f_N = \mathbb{E}^{\mathcal{W}}\big[f \,\big|\, \mathcal{F}_N\big]$ and $F_N = \mathbb{E}^{\mathcal{W}}\big[F \,\big|\, \mathcal{F}_N\big]$, where $\mathcal{F}_N = \sigma\big(\{\langle\,\cdot\,,\xi_j\rangle : 1 \leq j \leq N\}\big)$, and show that $f_N \in \mathrm{Dom}\big((\nabla^\top)^\top\big)$ and $F_N = (\nabla^\top)^\top f_N$. Since $f_N \longrightarrow f$ in $L^2(\mathcal{W};\mathbb{R})$ and $F_N \longrightarrow F$ in $L^2(\mathcal{W};H)$ as $N \to \infty$, conclude that it suffices to prove the result when $H = B = \mathbb{R}^N$ and $\mathcal{W} = \gamma_{0,1}^N$. That is, let ∂_j on $C^1(\mathbb{R}^N;\mathbb{R})$ denote partial differential with respect to the jth coordinate and $\partial_j^\top \psi(x) = (x_j\psi(x) - \partial_j\psi(x)$, and define $\nabla^\top \Psi = \sum_{j=1}^N \partial_j^\top \psi_j$ for $\Psi = \big(\psi_1,\dots,\psi_N\big)^\top \in C^1(\mathbb{R}^N;\mathbb{R}^N)$. What one needs to show is that if $(f,F) \in L^2\big(\gamma_{0,1}^N;\mathbb{R}\big) \times L^2(\gamma_{0,1}^N;\mathbb{R}^N)$ satifies

$$\big(f,\nabla^\top\Psi\big)_{L^2(\gamma_{0,1};\mathbb{R})} = \big(F,\Psi\big)_{L^2(\gamma_{0,1};\mathbb{R}^N)}$$

for all Ψ's with polynomial coordinates, then there exists a sequence $\{\varphi_k : k \geq 1\}$ of polynomials on $\mathbb{R}^N$ such that $\varphi_k \longrightarrow f$ in $L^2(\gamma_{0,1}^N;\mathbb{R})$ and $\big(\partial_1\varphi,\dots,\partial_N\varphi\big)^\top \longrightarrow F$ in $L^2(\gamma_{0,1}^N;\mathbb{R}^N)$. Perhaps the best way to do this is to introduce the functions $H_{\mathbf{k}}(x) = \prod_{j=1}^N H_{k_j}(x_j)$ for $\mathbf{k} = (k_1,\dots,k_N) \in \mathbb{N}^N$, where H_k is the kth Hermite polynomial, set $\tilde{H}_{\mathbf{k}} = (\mathbf{k}!)^{-\frac{1}{2}}$ where $\mathbf{k}! = \prod_{j=1}^N k_j!$, check that $\{\tilde{H}_{\mathbf{k}} : \mathbf{k} \in \mathbb{N}^N\}$ is an orthonomal basis in $L^2(\gamma_{0,1}^N;\mathbb{R})$, and show that

$$\|F\|^2_{L^2(\gamma_{0,1}^N;\mathbb{R}^N)} = \sum_{\mathbf{k}\in\mathbb{N}^N} \|\mathbf{k}\|_1 \big(f,\tilde{H}_{\mathbf{k}}\big)^2_{L^2(\gamma_{0,1}^N;\mathbb{R})},$$

where $\|\mathbf{k}\|_1 = \sum_{j=1}^N k_j$.

3.4 Asymptotic Properties of Abstract Wiener Spaces

3.4.1 Large Deviaviations in Abstract Wiener Spaces

The goal of this subsection is to derive the following result.

Theorem 3.4.1 *Let $(H,B,\mathcal{W})$ be an abstract Wiener space, and, for $\epsilon > 0$, denote by $\mathcal{W}_\epsilon$ the $\mathcal{W}$-distribution of $x \rightsquigarrow \epsilon^{\frac{1}{2}}x$. Then, for each $\Gamma \in \mathcal{B}_B$,*

$$
\begin{aligned}
-\inf_{h\in\mathring{\Gamma}} \frac{\|h\|_H^2}{2} &\le \varliminf_{\epsilon\searrow 0} \epsilon \log \mathcal{W}_\epsilon(\Gamma) \\
&\le \varlimsup_{\epsilon\searrow 0} \epsilon \log \mathcal{W}_\epsilon(\Gamma) \le -\inf_{h\in\overline{\Gamma}} \frac{\|h\|_H^2}{2}.
\end{aligned} \tag{3.4.1}
$$

The original version of Theorem 3.4.1 was proved by M. Schilder for the classical Wiener measure using a method that does not extend easily to the general case. The statement that I have given is due to Donsker and Varadhan, and my proof derives from an approach that was introduced into this context by Varadhan.

The lower bound is an easy application of the Cameron–Martin formula. Indeed, all that one has to do is show that if $h \in H$ and $r > 0$, then

$$
\varliminf_{\epsilon\searrow 0} \epsilon \log \mathcal{W}_\epsilon\big(B_B(h,r)\big) \ge -\frac{\|h\|_H^2}{2}. \tag{$*$}
$$

To this end, note that, for any $\xi \in B^*$ and $\delta > 0$,

$$
\begin{aligned}
\mathcal{W}_\epsilon\big(B_B(h_\xi,\delta)\big) &= \mathcal{W}\big(B_B(\epsilon^{-\frac{1}{2}}h_\xi, \epsilon^{-\frac{1}{2}}\delta)\big) \\
&= \mathbb{E}^{\mathcal{W}}\Big[e^{-\epsilon^{-\frac{1}{2}}\langle x,\xi\rangle - \frac{1}{2\epsilon}\|h_\xi\|_H^2}, B_B(0,\epsilon^{-\frac{1}{2}}\delta)\Big] \\
&\ge e^{-\delta\epsilon^{-1}\|\xi\|_{B^*} - \frac{1}{2\epsilon}\|h_\xi\|_H^2}\mathcal{W}\big(B_B(0,\epsilon^{-\frac{1}{2}}\delta)\big),
\end{aligned}
$$

which means that

$$
B_B(h_\xi,\delta) \subseteq B_B(h,r) \implies \varlimsup_{\epsilon\searrow 0} \epsilon \log \mathcal{W}_\epsilon\big(B_B(h,r)\big) \ge -\delta\|\xi\|_{B^*} - \frac{\|h_\xi\|_H^2}{2},
$$

and therefore, after letting $\delta \searrow 0$ and remembering that $\{h_\xi : x \in B^*\}$ is dense in H, that $(*)$ holds.

The proof of the upper bound in (3.4.1) is a little more involved. The first step is to show that it suffices to treat the case when Γ is relatively compact. To this end, refer to Corollary 3.3.18, and set C_R equal to the closure in B of $B_{B_0}(0,R)$. By Fernique's Theorem applied to $\mathcal{W}$ on B_0, we know that $\mathbb{E}^{\mathcal{W}}\big[e^{\alpha\|x\|_{B_0}^2}\big] \le K$ for some $\alpha > 0$ and $K < \infty$. Hence

$$
\mathcal{W}_\epsilon\big(B\setminus C_R\big) = \mathcal{W}\big(B\setminus C_{\epsilon^{-\frac{1}{2}}R}\big) \le Ke^{-\alpha\frac{R^2}{\epsilon}},
$$

and so, for any $\Gamma \in \mathcal{B}_B$ and $R > 0$,

$$
\mathcal{W}_\epsilon\big(\Gamma\big) \le 2\mathcal{W}_\epsilon(\Gamma\cap C_R) \vee \big(Ke^{-\alpha\frac{R^2}{\epsilon}}\big).
$$

Thus, if we can prove the upper bound for relatively compact Γ's, then, because $\Gamma \cap C_R$ is relatively compact, we will know that, for all $R > 0$,

$$\varlimsup_{\epsilon\searrow 0}\epsilon\log\mathcal{W}_\epsilon(\Gamma)\le -\left[\left(\inf_{h\in\overline{\Gamma}}\frac{\|h\|_H^2}{2}\right)\wedge\left(\alpha R^2\right)\right],$$

from which the general result is immediate.

To prove the upper bound when Γ is relatively compact, we will show that, for any $y\in B$,

$$\varlimsup_{r\searrow 0}\varlimsup_{\epsilon\searrow 0}\epsilon\log\mathcal{W}_\epsilon\big(B_B(y,r)\big)\le\begin{cases}-\frac{\|y\|_H^2}{2} & \text{if } y\in H\\ -\infty & \text{if } y\notin H.\end{cases}\tag{$**$}$$

To see that $(**)$ is enough, assume that it is true, and let $\Gamma\in\mathcal{B}_B\setminus\{\emptyset\}$ be relatively compact. Given $\beta\in(0,1)$, for each $y\in\overline{\Gamma}$ choose $r(y)>0$ and $\epsilon(y)>0$ so that

$$\mathcal{W}_\epsilon\big(B_B(y,r(y))\big)\le\begin{cases}e^{-\frac{(1-\beta)}{2\epsilon}\|y\|_H^2} & \text{if } y\in H\\ e^{-\frac{1}{\beta\epsilon}} & \text{if } y\notin H\end{cases}$$

for all $0<\epsilon\le\epsilon(y)$. Because Γ is relatively compact, we can find $N\in\mathbb{Z}^+$ and $\{y_1,\dots,y_N\}\subseteq\overline{\Gamma}$ such that $\Gamma\subseteq\bigcup_1^N B_B(y_n,r_n)$, where $r_n=r(y_n)$. Thus, for sufficiently small $\epsilon>0$,

$$\mathcal{W}_\epsilon(\Gamma)\le N\exp\left(-\left[\left(\frac{1-\beta}{2\epsilon}\inf_{h\in\overline{\Gamma}}\|h\|_H^2\right)\wedge\frac{1}{\epsilon\beta}\right]\right),$$

and so

$$\varlimsup_{\epsilon\searrow 0}\epsilon\log\mathcal{W}_\epsilon(\Gamma)\le -\left[\left(\frac{1-\beta}{2}\inf_{h\in\overline{\Gamma}}\|h\|_H^2\right)\wedge\frac{1}{\beta}\right].$$

Now let $\beta\searrow 0$.

Finally, to prove $(**)$, observe that

$$\begin{aligned}\mathcal{W}_\epsilon\big(B_B(y,r)\big)&=\mathcal{W}\Big(B_B(\tfrac{y}{\sqrt{\epsilon}},\tfrac{r}{\sqrt{\epsilon}})\Big)=\mathbb{E}^{\mathcal{W}}\Big[e^{-\epsilon^{-\frac12}\langle x,\xi\rangle}e^{\epsilon^{-\frac12}\langle x,\xi\rangle},\,B_B(\tfrac{y}{\sqrt{\epsilon}},\tfrac{r}{\sqrt{\epsilon}})\Big]\\ &\le e^{-\epsilon^{-1}(\langle y,\xi\rangle-r\|\xi\|_{B^*})}\mathbb{E}^{\mathcal{W}}\Big[e^{\epsilon^{-\frac12}\langle x,\xi\rangle}\Big]=e^{-\epsilon^{-1}\left(\langle y,\xi\rangle-\frac{\|h_\xi\|_H^2}{2}-r\|\xi\|_{B^*}\right)},\end{aligned}$$

for all $\xi\in B^*$. Hence,

$$\varlimsup_{r\searrow 0}\varlimsup_{\epsilon\searrow 0}\epsilon\log\mathcal{W}_\epsilon\big(B_B(y,r)\big)\le -\sup_{\xi\in B^*}\Big(\langle y,\xi\rangle-\tfrac12\|h_\xi\|_H^2\Big).$$

Note that the preceding supremum is the same as half the supremum of $\langle y,\xi\rangle$ over ξ with $\|h_\xi\|_H=1$, which, by Lemma 3.3.1, is equal to $\frac{\|y\|_H^2}{2}$ if $y\in H$ and to ∞ if $y\notin H$.

3.5 Brownian Motion on a Banach Space

This section is devoted to the study of Brownian motion on a Banach space. To be precise, given a non-degenerate, centered, Gaussian measure $\mathcal{W}$ on a separable Banach space B, what we will show is that there exists an B-valued stochastic process $\{B(t) : t \in \mathbb{R}\}$ with the properties that: $B(0) = 0$, $t \rightsquigarrow B(t)$ is continuous, and, for each $\xi \in B^*$ with $\|h_\xi\|_H = 1$, $\{\langle B(t), \xi\rangle : t \in \mathbb{R}\}$ is an $\mathbb{R}$-valued Brownian motion. Equivalently, for all $n \geq 1$ and $t_0 < \cdots < t_n$, $B(t_1) - B(t_0), \cdots, B(t_n) - B(t_{n-1})$ are mutually independent random variables, the mth one of which has distribution (cf. the notation in Theorem 3.4.1) $\mathcal{W}_{t_m - t_{m-1}}$.

3.5.1 Abstract Wiener Formulation

Let $\mathcal{W}$ on B be as above, use H to denote its Cameron–Martin subspace, and take $H^1_0(\mathbb{R}; H)$ to be the Hilbert space of absolutely continuous $h : \mathbb{R} \longrightarrow H$ such that $h(0) = 0$ and $\|h\|_{H^1_0(\mathbb{R};H)} = \|\dot{h}\|_{L^2(\mathbb{R};H)} < \infty$. Finally, let $\Omega_0(B)$ be the space of continuous $\omega : \mathbb{R} \longrightarrow B$ satisfying $\omega(0) = 0$ and $\lim_{t\to\infty} \frac{\|\omega(t)\|_B}{t} = 0$, and turn $\Omega_0(B)$ into a Banach space with norm $\|\omega\|_{\Omega_0(B)} = \sup_{t\geq 0}(1+t)^{-1}\|\omega(t)\|_E$. Exactly the same line of reasoning used when $B = \mathbb{R}$ applies here and shows that $\Omega_0(B)$ is a separable Banach space in which $H^1_0(\mathbb{R}; H)$ is continuously embedded as a dense subspace.

Theorem 3.5.1 *With $H^1_0(\mathbb{R}; H)$ and $\Omega_0(B)$ as above, there is a unique $\mathcal{W}^{(B)}_0 \in M_1\big(\Omega_0(B)\big)$ such that $\big(H^1_0(\mathbb{R}; H), \Omega_0(B), \mathcal{W}^{(B)}_0\big)$ is an abstract Wiener space. Moreover, $\{\omega(t) : t \in \mathbb{R}\}$ is a B-valued Brownian motion under $\mathcal{W}^{(B)}_0$.*

We will use an approach based on an analog Theorem 3.3.10. Begin by choosing an orthonormal basis $\{g_m : m \geq 0\}$ in $H^1_0(\mathbb{R}; \mathbb{R})$, and, for $n \geq 0$, $t \in \mathbb{R}$, and $y = (y_0, \dots, y_m, \dots) \in B^{\mathbb{N}}$, set $S_n(t, y) = \sum_{m=0}^{n} g_m(t) y_m$.

Lemma 3.5.2 *For $\mathcal{W}^{\mathbb{N}}$-almost every $y \in B^{\mathbb{N}}$, $\{S_n : n \geq 1\}$ is relatively compact in $\Omega_0(B)$.*

Proof. Choose $B_0 \subseteq B$, as in Corollary 3.3.18, so that bounded subsets of B_0 are relatively compact in B and $\big(H, B_0, \mathcal{W} \restriction B_0\big)$ is again an abstract Wiener space. Without loss in generality, assume that $\|\cdot\|_B \leq \|\cdot\|_{B_0}$. By Fernique's Theorem, we know that $C \equiv \mathbb{E}^{\mathcal{W}}\big[\|x\|^4_{B_0}\big] < \infty$.

Next, for $s \in \mathbb{R}$, define

$$f_s(\tau) = \begin{cases} |s| \wedge |\tau| & \text{if } s\tau \geq 0 \\ 0 & \text{if } s\tau < 0 \end{cases},$$

observe that

$$S_n(t,y) - S_n(s,y) = \sum_{m=0}^{n} \big(f_t - f_s, g_m\big)_{H^1_0} y_m,$$

and conclude that the $\mathcal{W}^{\mathbb{N}}$-distribution of $S_n(t) - S_n(s)$ is $\mathcal{W}_{\epsilon_n}$, where $\epsilon_n = \sum_{m=0}^{n}\big(f_t - f_s, g_m\big)^2_{H^1(\mathbb{R})} \le t-s$. Hence, $\mathbb{E}^{\mathcal{W}^{\mathbb{N}}}\big[\|S_n(t)-S_n(s)\|^4_{B_0}\big] \le C(t-s)^2$. In addition, $\{\|S_n(t)-S_n(s)\|_{B_0} : n \ge 1\}$ is a submartingale, and so, by Doob's Inequality plus Kolmogorov's Continuity Criterion, there exists a $K < \infty$ such that, for each $T > 0$,

$$\mathbb{E}\left[\sup_{n\ge 0}\sup_{-T\le s<t\le T}\frac{\|S_n(t)-S_n(s)\|_{B_0}}{(t-s)^{\frac{1}{8}}}\right] \le KT^{\frac{3}{4}}. \tag{$*$}$$

From $(*)$ and $S_n(0) = 0$, we know that for $\mathcal{W}^{\mathbb{N}}$-almost every $y \in B^{\mathbb{N}}$, $\{S_n(\,\cdot\,,y) : n \ge 0\}$ is uniformly bounded and uniformly equicontinuous on each interval $[-T,T]$ with respect to $\|\cdot\|_{B_0}$. Since this means that, for every $T > 0$ and $\mathcal{W}^{\mathbb{N}}$-almost every $y \in B^{\mathbb{N}}$, $\{S_n(t,y) : n \ge 0 \;\&\; t \in [-T,T]\}$ is relatively compact in B and $\{S_n(\,\cdot\,,y) \restriction [-T,T] : n \ge 0\}$ is uniformly $\|\cdot\|_B$-equicontinuous, the Ascoli–Arzela Theorem guarantees that, $\mathcal{W}^{\mathbb{N}}$-almost surely, $\{S_n(\,\cdot\,,y) : n \ge 0\}$ is relatively compact in $C(\mathbb{R};B)$ with the topology of uniform convergence on compacts. Thus, what remains to be shown is that $\mathcal{W}^{\mathbb{N}}$-almost surely,

$$\lim_{T\to\infty}\sup_{n\ge 0}\sup_{|t|\ge T}\frac{\|S_n(t,y)\|_B}{|t|} = 0.$$

But,

$$\begin{aligned}\sup_{|t|\ge 2^k}\frac{\|S_n(t,y)\|_B}{|t|} &\le \sum_{\ell\ge k}\sup_{2^\ell\le|t|\le 2^{\ell+1}}\frac{\|S_n(t,y)\|_B}{|t|}\\ &\le \sum_{\ell\ge k}2^{-\frac{7\ell}{8}}\sup_{0\le|t|\le 2^{\ell+1}}\frac{\|S_n(t,y)\|_B}{|t|^{\frac{1}{8}}},\end{aligned}$$

and therefore, by $(*)$,

$$\mathbb{E}\left[\sup_{n\ge 0}\sup_{|t|\ge 2^k}\frac{\|S_n(t)\|_B}{|t|}\right] \le \frac{2^{\frac{3}{4}}K}{2^{\frac{1}{8}}-1}2^{-\frac{k}{8}}.$$

□

In addition to the preceding compactness result, we need the following simple criterion for checking when a relatively compact sequence in $\Omega_0(B)$ converges.

Lemma 3.5.3 *Suppose that $\{\omega_n : n \ge 0\}$ is a relatively compact sequence in $\Omega_0(B)$. If $\lim_{n\to\infty}\langle \omega_n(t), \xi\rangle$ exists for each t in a dense subset of $\mathbb{R}$ and ξ in a weak* dense subset of B^*, then $\{\omega_n : n \ge 0\}$ converges in $\Omega_0(B)$.*

Proof. For a relatively compact sequence to be convergent, it is necessary and sufficient that every convergent subsequence have the same limit. Thus, suppose that ω and ω' are limit points of $\{\omega_n : n \ge 0\}$. Then, by hypothesis, $\langle \omega(t), \xi\rangle = \langle \omega'(t), \xi\rangle$ for t in a dense subset of $\mathbb{R}$ and ξ in a weak* dense subset of B^*. But this means that the same equality holds for all $(t, \xi) \in \mathbb{R} \times B^*$ and therefore that $\omega = \omega'$. $\square$

Proof of Theorem 3.5.1. In view of Lemmas 3.5.2 and 3.5.3 and the separability of $B_{B^*}(0,1)$ in the weak* topology, we will know that $\{S_n(\,\cdot\,, y) : n \ge 0\}$ converges in $\Omega_0(B)$ for $\mathcal{W}^{\mathbb{N}}$-almost every $y \in B^{\mathbb{N}}$ once we show that, for each $(t, \xi) \in [0,\infty) \times B^*$, $\{\langle S_n(t,y), \xi\rangle : n \ge 0\}$ converges in $\mathbb{R}$ for $\mathcal{W}^{\mathbb{N}}$-almost every $y \in B^{\mathbb{N}}$. But if $\xi \in B^*$, then $\langle S_n(t,y), \xi\rangle = \sum_0^n \langle y_m, \xi\rangle g_m(t)$, the random variables $y \rightsquigarrow \langle y_m, \xi\rangle g_m(t)$ are mutually independent, centered Gaussians under $\mathcal{W}^{\mathbb{N}}$ with variance $\|h_\xi\|_H^2 g_m(t)^2$, and $\sum_0^\infty g_m(t)^2 = \|f_t\|^2_{H^1(\mathbb{R};\mathbb{R})} = t$. Thus, by Kolmogorov's convergence theorem for sums of mutually independent, square integrable random variables, we have the required convergence.

Next, define $S : [0,\infty) \times B^{\mathbb{N}} \longrightarrow B$ so that

$$S(t,y) = \begin{cases} \lim_{n\to\infty} S_n(t,y) & \text{if } \{S_n(\,\cdot\,,y) : n \ge 0\} \text{ converges in } \Omega_0(B) \\ 0 & \text{otherwise.} \end{cases}$$

Given $\Xi \in \Omega_0(B)^*$, determine $h_\Xi \in H^1_0(\mathbb{R}; H)$ by $\big(h, h_\Xi\big)_{H^1_0(\mathbb{R};H)} = \langle h, \Xi\rangle$ for all $h \in H^1_0(\mathbb{R}; H)$. We must show that, under $\mathcal{W}^{\mathbb{N}}$, $y \rightsquigarrow \langle S(\,\cdot\,, y), \Xi\rangle$ is a centered Gaussian with variance $\|h_\Xi\|^2_{H^1_0(\mathbb{R};H)}$. To this end, define $\xi_m \in B^*$ so that ${}_B\langle x, \xi_m\rangle_{B^*} = {}_{\Omega_0(B)}\langle g_m x, \Xi\rangle_{\Omega_0(B)^*}$ for $x \in B$, where $g_m x$ is the element of $\Omega_0(B)$ such that $g_m x(t) = g_m(t)x$. Then,

$$\langle S(\,\cdot\,, y), \Xi\rangle = \lim_{n\to\infty} \langle S_n(\,\cdot\,, y), \Xi\rangle = \lim_{n\to\infty} \sum_0^n \langle y_m, \xi_m\rangle \quad \mathcal{W}^{\mathbb{N}}\text{-almost surely.}$$

Hence, $\langle S(\,\cdot\,, y), \Xi\rangle$ is certainly a centered Gaussian under $\mathcal{W}^{\mathbb{N}}$, and, because we are dealing with Gaussian random variables, almost sure convergence implies L^2-convergence. To compute its variance, choose an orthonormal basis $\{h_k : k \ge 0\}$ for H, and note that, for each $m \ge 0$,

$$\int \langle y_m, \xi_m\rangle^2 \, \mathcal{W}^{\mathbb{N}}(dy) = \|h_{\xi_m}\|_H^2 = \sum_{k=0}^\infty \langle g_m h_k, \Xi\rangle^2.$$

Thus, since $\{g_m h_k : (m,k) \in \mathbb{N}^2\}$ is an orthonormal basis in $H^1_0(\mathbb{R}; H)$,

$$\mathbb{E}\big[\langle S(\,\cdot\,),\Xi\rangle^2\big]=\sum_{m,k=0}^{\infty}\langle g_m h_k,\Xi\rangle^2=\sum_{m,k=0}^{\infty}\big(g_m h_k,h_\Xi\big)^2_{H_0^1(\mathbb{R};H)}=\|h_\Xi\|^2_{H_0^1(\mathbb{R};H)}.$$

Finally, if $\mathcal{W}_0^{(B)}$ is the $\mathcal{W}^{\mathbb{N}}$-distribution of $y \rightsquigarrow S(\,\cdot\,,y)$, then the preceding shows that $\big(H_0^1(B),\Omega_0(B),\mathcal{W}_0^{(B)}\big)$ is an abstract Wiener space. To see that $\{\omega(t) : t\in\mathbb{R}\}$ is a Brownian motion under $\mathcal{W}_0^{(B)}$, it suffices to show that, for all $s<t$ and $\xi\in B^*$,

$$\mathbb{E}\big[\langle\omega(s),\xi\rangle\langle\omega(s),\xi\rangle\big]=w(s,t)\|h_\xi\|_H^2,$$

where $w(s,t)=\mathbf{1}_{[0,\infty)}(st)|s|\wedge|t|$. To this end, define $\Xi_s\in\Omega_0(B)$ by

$$_{\Omega_0(B)}\langle\omega,\Xi_s\rangle_{\Omega_0(B)^*}={}_B\langle\omega(s),\xi\rangle_{B^*}\ \text{for}\ \omega\in\Omega_0(B).$$

Then $h_{\Xi_s}=f_s h_\xi$, where $f_s\in H_0^1$ is defined as in the proof of Lemma 3.5.2. Hence

$$\mathbb{E}\big[\langle\omega(s),\xi\rangle\langle\omega(s),\xi\rangle\big]=\big(h_{\Xi_t},h_{\Xi_s}\big)_{H_0^1(\mathbb{R};B)}=(f_t,f_s)_{H_0^1(\mathbb{R};\mathbb{R})}\|h_\xi\|_H^2=w(s,t)\|h_\xi\|_H^2.$$

□

3.5.2 Strassen's Theorem

This subsection is devoted to a beautiful version of the law of the iterated logarithm, proved originally by V. Strassen for classical Brownian motion.

The original statement of the law of the iterated logarithm says that if $\{X_m : m\geq 1\}$ is a sequence of independent, identically distributed square integrable $\mathbb{R}$-valued random variables with mean 0 and variance 1 and $\tilde{S}_n=\frac{S_n}{\Lambda_n}$, where $S_n=\sum_{m=1}^n X_m$ and $\Lambda_n=\sqrt{2n\log(\log n\vee 3)}$, then, with probability 1, the sequence $\{\tilde{S}_n : n\geq 1\}$ is compact and the set of its limit points coincides with the interval $[-1,1]$. A. Khinchine was the first to prove such a result when he did so for Bernoulli random variables. Over time, various people, including Kolmogorov, extended the result to more general random variables, and it was finally proved in full generality by P. Hartman and A. Wintner. Strassen's version introduced an innovation that was not anticipated.

Theorem 3.5.4 *Given $\omega\in\Omega_0(B)$, define $\tilde{\omega}_n(t)=\frac{\omega(nt)}{\Lambda_n}$ for $n\geq 1$ and $t\in\mathbb{R}$, where $\Lambda_n=\sqrt{2n\log_{(2)}(n\vee 3)}$. Then, for $\mathcal{W}_0^{(B)}$-almost every ω, the sequence $\{\tilde{\omega}_n : n\geq 1\}$ is relatively compact in $\Omega_0(B)$ and its set of limit points coincides with the closure $\overline{B_{H_0^1(\mathbb{R};H)}(0,1)}$ in B of $B_H(0,1)$. Equivalently, for $\mathcal{W}_0^{(B)}$-almost every ω,*

$$\overline{\lim_{n\to\infty}} \|\tilde{\omega}_n - \overline{B_{H^1_0(\mathbb{R};H)}(0,1)}\|_{\Omega_0(B)} = 0$$

and, for each $y \in \overline{B_{H^1_0(\mathbb{R};H)}(0,1)}$, $\underline{\lim}_{n\to\infty} \|\tilde{\omega}_n - y\|_{\Omega_0(B)} = 0$.

Proof. Without loss in generality, we will assume that $\|\cdot\|_B \le \|\cdot\|_H$.

The proof relies on the Brownian scaling invariance property (cf. (**v**) in Exercise 3.3.1, which says that $\mathcal{W}_0^{(B)}$ is invariant under the scaling maps $S_\alpha : \Omega_0(B) \longrightarrow \Omega_0(B)$ given by $S_\alpha\omega = \alpha^{-\frac{1}{2}}\omega(\alpha\,\cdot)$ for $\alpha > 0$ and is easily proved as a consequence of the fact that these maps are isometric from $H^1_0(\mathbb{R};H)$ onto itself. In addition, we will use the fact that, for $R > 0$, $r \in (0,1]$, and $\omega \in \Omega_0(B)$, $\|\omega(r\,\cdot) - B_{H^1_0(\mathbb{R};H)}(0,R)\|_{\Omega_0(B)} \le \|\omega - B_{H^1_0(\mathbb{R};H)}(0,R)\|_{\Omega_0(B)}$. To see this, let $h \in B_{H^1_0(\mathbb{R};H)}(0,R)$ be given, and check that $h(r\,\cdot)$ is again in $B_H(0,R)$ and that $\|\omega(r\,\cdot) - h(r\,\cdot)\|_{\Omega_0(B)} \le \|\omega - h\|_{\Omega_0(B)}$.

To prove that $\tilde{w}_n$ tends to $\overline{B_H(0,1)}$, begin by observing that, for any $\beta \in (1,2)$,

$$\overline{\lim_{n\to\infty}} \|\tilde{\omega}_n - \overline{B_H(0,1)}\|_B \le \overline{\lim_{m\to\infty}} \max_{\beta^{m-1}\le n\le\beta^m} \|\tilde{\omega}_n - \overline{B_H(0,1)}\|_B.$$

Taking the preceding comments into account and applying the upper bound in (3.4.1), one can justify

$$\begin{aligned}
&\mathcal{W}_0^{(B)}\left(\max_{\beta^{m-1}\le n\le\beta^m} \big\|\tilde{\omega}_n - \overline{B_{H^1_0(\mathbb{R};H)}(0,1)}\big\|_{\Omega_0(B)} \ge \delta\right)\\
&= \mathcal{W}_0^{(B)}\left(\max_{\beta^{m-1}\le n\le\beta^m} \left\|\frac{\beta^{\frac{m}{2}}\omega(n\beta^{-m}\,\cdot)}{\Lambda_n} - \overline{B_{H^1_0(\mathbb{R};H)}(0,1)}\right\|_{\Omega_0(B)} \ge \delta\right)\\
&\le \mathcal{W}_0^{(B)}\left(\max_{\beta^{m-1}\le n\le\beta^m} \left\|\omega\big(n\beta^{-m}\,\cdot\big) - \overline{B_{H^1_0(\mathbb{R};H)}\left(0, \frac{\Lambda_{\lfloor\beta^{m-1}\rfloor}}{\beta^{\frac{m}{2}}}\right)}\right\|_{\Omega(B)} \ge \frac{\delta\Lambda_{\lfloor\beta^{m-1}\rfloor}}{\beta^{\frac{m}{2}}}\right)\\
&\le \mathcal{W}_0^{(B)}\left(\left\|\omega - \overline{B_{H^1_0(\mathbb{R};H)}\left(0, \frac{\Lambda_{\lfloor\beta^{m-1}\rfloor}}{\beta^{\frac{m}{2}}}\right)}\right\|_{\Omega_0(B)} \ge \frac{\delta\Lambda_{\lfloor\beta^{m-1}\rfloor}}{\beta^{\frac{m}{2}}}\right)\\
&= \mathcal{W}_0^{(B)}\Big(\big\|\beta^{\frac{m}{2}}\Lambda^{-1}_{\lfloor\beta^{m-1}\rfloor}\omega - \overline{B_{H^1_0(\mathbb{R};H)}(0,1)}\big\|_{\Omega(B)} \ge \delta\Big)\\
&= \mathcal{W}^{(B)}_{\beta^m\Lambda^{-2}_{\lfloor\beta^{m-1}\rfloor}}\big(\|\omega - \overline{B_{H^1_0(\mathbb{R};H)}(0,1)}\|_{\Omega_0(B)} \ge \delta\big)\\
&\le \exp\left(-\frac{R^2\lfloor\beta^{m-1}\rfloor}{\beta^m}\log_{(2)}\lfloor\beta^{m-1}\rfloor\right)
\end{aligned}$$

for all $\beta \in (1,2)$ and $R < 1+\delta$. Now, take $R = \frac{1+\delta}{2}$ choose $\beta \in (1,2)$ so that $\frac{R^2}{\beta} > 1$, and conclude that

$$\sum_{m=1}^{\infty} \mathcal{W}_0^{(B)}\left(\max_{\beta^{m-1}\le n\le\beta^m} \|\tilde{\omega}_n - \overline{B_H(0,1)}\|_{\Omega_0(B)} \ge \delta\right) < \infty$$

and therefore that

$$\mathcal{W}_0^{(B)}\Big(\varlimsup_{n\to\infty}\|\tilde{\omega}_n-\overline{B_H(0,1)}\|_{\Omega_0(B)}\geq\delta\Big)=0.$$

Because B is separable, to prove that, $\mathcal{W}_0^{(B)}$-almost surely, $\varliminf_{n\to\infty}\|\tilde{\omega}_n-\eta\|_B=0$ for every $\eta\in\overline{B_{H_0^1(\mathbb{R};H)}(0,1)}^B$, it suffices to prove that $\varliminf_{n\to\infty}\|\tilde{\omega}_n-h\|_B=0$ (a.s.,$\mathcal{W}_0^{(B)}$) for each $h\in B_{H_0^1(\mathbb{R};H)}(0,1)$. In addition, because

$$\lim_{T\to\infty}\sup_{\omega\in A}\sup_{|t|\notin[T^{-1},T]}\frac{\|\omega(t)\|_B}{1+|t|}=0,$$

for any relatively compact $A\subseteq\Omega_0(B)$, and, by the preceding, for $\mathcal{W}_0^{(B)}$-almost every ω, the union of $\{\omega_n:\,n\geq1\}$ and $\overline{B_{H_0^1(\mathbb{R};H)}(0,1)}$ is relatively compact in $\Omega_0(B)$, it suffices to prove that

$$\varliminf_{n\to\infty}\sup_{t\in[k^{-1},k]}\frac{\big\|\big(\tilde{\omega}_n(\pm t)-\tilde{\omega}_n(\pm k^{-1})\big)-\big(h(\pm t)-h(\pm k^{-1})\big)\|_B}{1+t}=0\ (\text{a.s.},\mathcal{W}_0^{(B)})$$

for each $h\in B_{H_0^1(\mathbb{R};H)}(0,1)$ and $k\geq2$. Since, for a fixed $k\geq2$, the random variables

$$\big(\tilde{\omega}_{k^{2m}}-\tilde{\omega}_{k^{2m}}(\pm k^{-1})\big)\restriction\{t:\,\pm t\in[k^{-1},k]\},\quad m\geq1,$$

are $\mathcal{W}_0^{(B)}$-independent random variables, we can use the Borel–Cantelli Lemma to reduce the problem to showing that, if

$$\check{\omega}_{k^m}(t)=\begin{cases}\tilde{\omega}_{k^m}(t+k^{-1})-\tilde{\omega}_{k^m}(k^{-1}) & \text{if } t\geq0\\ \tilde{\omega}_{k^m}(t-k^{-1})-\tilde{\omega}_{k^m}(-k^{-1}) & \text{if } t<0,\end{cases}$$

then

$$\sum_{m=1}^{\infty}\mathcal{W}_0^{(B)}\big(\|\check{\omega}_{k^{2m}}-h\|_{\Omega_0(B)}\leq\delta\big)=\infty$$

for each $\delta>0$, $k\geq2$, and $h\in B_{H_0^1(\mathbb{R};H)}(0,1)$. Finally, since $(\mathcal{W}_0^{(B)})_{(k^m\Lambda_{k^{2m}}^{-1})^2}$ is the $\mathcal{W}_0^{(B)}$ distribution of $\omega\rightsquigarrow\check{\omega}_{k^{2m}}$, then by the lower bound in (3.4.1) with $R=\frac{1+\delta}{2}\in(0,1)$, one sees that

$$\mathcal{W}_0^{(B)}\big(\|\check{\omega}_{k^{2m}}-h\|_{\Omega_0(B)}\leq\delta\big)\geq e^{-R^2\log(\log k^m)}=(m\log k)^{-R^2}$$

for sufficiently large m. □

An essentially trivial corollary of Theorem 3.5.4 is the law of the law of iterated for centered Gaussian random variables with values in a separable Banach space B. Indeed, if $(H,B,\mathcal{W})$ is an associated abstract Wiener space and S_n is the nth partial sum of such random variables, then S_n has the same distribution as $\omega(n)$ under $\mathcal{W}_0^{(B)}$, and therefore, almost surely,

$$\overline{\lim_{n\to\infty}} \left\| \frac{S_n}{\Lambda_n} - \{h(1) : h \in B_{H_0^1(\mathbb{R};H)}(0,1)\} \right\|_B = 0$$

and

$$\underline{\lim_{n\to\infty}} \left\| \frac{S_n}{\Lambda_n} - h(1) \right\|_B = 0 \text{ for all } h \in B_{H_0^1(\mathbb{R};H)}(0,1).$$

Further, by taking $h(t) = (1 \wedge t)^+ g$ for $g \in H$, one sees that $B_H(0,1) = \{h(1) : h \in B_{H_0^1(\mathbb{R};H)}(0,1)\}$.

The preceding corollary is far less interesting than the one Strassen had in mind. Namely, by combining his theorem when $H = \mathbb{R}$ with a beautiful idea of A. Skorokhod (cf. Chapter 7 in [9]), he realized that it provides an elegant proof of the Hartman–Wintner law of the iterated logarithm. What Skorokhod had shown is that if X is an $\mathbb{R}$-valued random variable with mean value 0 and variance 1 and if $\{B(t) : t \in \mathbb{R}\}$ is a Brownian motion, then there is a stopping time σ for $\{B(t) : t \geq 0\}$ such that $B(\sigma)$ has the same distribution as X. For example, if $\mathbb{P}(X = \pm 1) = \frac{1}{2}$, then one can take σ to be the first time $t \geq 0$ that $|B(t)| = 1$. In any case, the expected value of σ is 1, and $\{B(t+\sigma) - B(\sigma) : t \geq 0\}$ is independent of $B(\sigma)$ and has the same distribution as $\{B(t) : t \geq 0\}$. Next, use induction to construct a sequence of stopping times $\{\zeta_n : n \geq 0\}$ so that $\zeta_0 = 0$ and, for $n \geq 1$, $\tau_n = \zeta_n - \zeta_{n-1}$ is the stopping time σ relative to $\{B(t) - B(\zeta_{n-1}) : t \geq 0\}$. Then $\{\tau_n : n \geq 1\}$ is a sequence of mutually independent, identically distributed random variables with mean value 1, and $\big\{B(\zeta_n) - B(\zeta_{n-1}) : n \geq 1\big\}$ is a sequence of mutually idependent random variables each of which has the same distribution as X. Therefore proving the law of the iterated logarithm for X comes down to understanding the bavior of $\frac{B(\zeta_n)}{\Lambda_n}$ as $n \to \infty$. To this end, set $\tilde{B}_n(t) = \frac{B(nt)}{\Lambda_n}$, and observe that, by Theorem 3.5.4 applied when $H = H_0^1(\mathbb{R};\mathbb{R})$, $\{\tilde{B}_n : n \geq 1\}$ is almost surely relatively compact in $\Omega_0(\mathbb{R})$, and so, since, by the law of large numbers, $\frac{\zeta_n}{n} \longrightarrow 1$ almost surely,

$$\left| \frac{B(\zeta_n)}{\Lambda_n} - \tilde{B}_n(1) \right| = \left| \tilde{B}_n\big(n^{-1}\zeta_n\big) - \tilde{B}_n(1) \right| \longrightarrow 0$$

almost surely. Finally, we know that $\{\tilde{B}_n(1) : n \geq 1\}$ is almost surely compact and that its set of limit points is the closure of $\{h(1) : \|h\|_{H_0^1(\mathbb{R};\mathbb{R})} \leq 1\}$ in $\mathbb{R}$ and is therefore the interval $[-1, 1]$.

Exercise 3.5.1 For $\omega \in \Omega_0(B)$, define

$$\omega^*(t) = \begin{cases} |t|\omega\left(\frac{1}{t}\right) & \text{if } t \neq 0 \\ 0 & \text{if } t = 0. \end{cases}$$

(**i**) Show that $\omega^* \in \Omega_0(B)$ and that $\omega \rightsquigarrow \omega*$ is an isometric, linear map of $\Omega_0(B)$ onto itself. Further show that $h^* \in H^1_0(\mathbb{R};B)$ if $h \in H^1_0(\mathbb{R};B)$ and that $h \rightsquigarrow h^*$ is an orthognal transformation on $H^1_0(\mathbb{R};B)$. In particular, conclude that $\omega*$ has the same distribution under $\mathcal{W}_0^{(B)}$ as ω.

(**ii**) Set

$$X_n(t,\omega) = \frac{n\omega\left(\frac{t}{n}\right)}{\sqrt{2\log_{(2)}(n \vee 3)}},$$

and show that, for $\mathcal{W}_0^{(B)}$-almost every $\omega \in \Omega_0(B)$, $\{X_n(\,\cdot\,,\omega) : n \geq 1\}$ is a relatively compact sequence for which $\overline{B_{H^1(B)}(0,1)}$ is the set of limit points.
Hint: Note that

$$\|X_n(\,\cdot\,,\omega) - h\|_{\Omega_0(B)} = \big\|X_n(\,\cdot\,,\omega)^* - h^*)\|_{\Omega_0(B)}$$

and that $X_n(\,\cdot\,,\omega) = \Lambda_n^{-1}|t|\omega^*\left(\frac{n}{t}\right)$. Conclude that

$$\|X_n(\,\cdot\,,\omega) - h\|_{\Omega_0(B)} = \big\|\widetilde{\omega_n^*} - h\big\|_{\Omega_0(B)}.$$

3.6 One Dimensional Euclidean Fields

In this section we will be studying Hilbert spaces for which the only Banach spaces that can be used to construct an associated abstract Wiener space consist of generalized functions (i.e., distributions) because evaluation at a point cannot be an element of its dual. Such a abstract Wiener spaces describe what physicists call a *field*, and when the, like $L^2(\lambda_{\mathbb{R}^N};\mathbb{R})$, the Hilbert space is invariant under the Euclidean group, it is called a *Euclidean field*.

3.6.1 Some Background

Recall the normalized Hermite functions $\{\tilde{h}_n : n \geq 0\}$ introduced in § 2.3.3 They form an orthonormal basis not only in $L^2(\lambda_{\mathbb{R}};\mathbb{R})$, they also form a basis in L. Schwartz's test function space $\mathscr{S}(\mathbb{R};\mathbb{R})$ of smooth functions φ all of whose derivatives are *rapidly decreasing* in the sense that

$$\lim_{|x|\to\infty} |x|^k \big|\partial^\ell \varphi(x)\big| = 0 \quad \text{for all } k,\ell \in \mathbb{N}.$$

To be precise, using (cf. § 7.3.4 in [11] for more details) the results in § 2.3.3, especially (2.3.20) and (2.3.22), one can show that $\varphi \in \mathscr{S}(\mathbb{R};\mathbb{R})$ if and only if

$$\|\varphi\|_{\mathscr{S}^{(m)}(\mathbb{R};\mathbb{R})} \equiv \left(\sum_{n=0}^{\infty}\bigl(n+\tfrac{1}{2}\bigr)^m\bigl(\varphi,\tilde{h}_n\bigr)^2_{L^2(\lambda_{\mathbb{R}};\mathbb{R})}\right) < \infty \quad \text{for all } m \in \mathbb{N},$$

and that $\mathscr{S}(\mathbb{R};\mathbb{R})$ becomes a complete, separable metric space when one uses the metric

$$\rho(\varphi,\psi) = \sum_{m=0}^{\infty} 2^{-m}\frac{\|\varphi-\psi\|_{\mathscr{S}^{(m)}(\mathbb{R};\mathbb{R})}}{1+\|\varphi-\psi\|_{\mathscr{S}^{(m)}(\mathbb{R};\mathbb{R})}}.$$

As a consequence, the dual space $\mathscr{S}^*(\mathbb{R};\mathbb{R})$ of tempered distributions can be described as the set of linear functionals u on $\mathscr{S}(\mathbb{R};\mathbb{R})$ for which the sequence $\bigl\{\langle\tilde{h}_n,u\rangle : n\geq 0\bigr\}$ has at most polynomial growth, and the action of $u \in \mathscr{S}^*(\mathbb{R};\mathbb{R})$ on $\varphi \in \mathscr{S}(\mathbb{R};\mathbb{R})$ is given by

$$\langle\varphi,u\rangle = \sum_{n=0}^{\infty}(\varphi,\tilde{h}_h)_{L^2(\lambda_{\mathbb{R}};\mathbb{R})}\langle\tilde{h}_n,u\rangle. \tag{3.6.1}$$

The preceding leads to a natural decomposition of $\mathscr{S}^*(\mathbb{R};\mathbb{R})$ into subspaces corresponding to their elements relationship to $L^2(\lambda_{\mathbb{R}};\mathbb{R})$. Namely, set $\lambda_n = n+\frac{1}{2}$, and, for $m\geq 0$, define $\mathscr{S}^{(-m)}(\mathbb{R};\mathbb{R})$ to be the space of $u\in\mathscr{S}^*(\mathbb{R};\mathbb{R})$ for which

$$\|u\|_{\mathscr{S}^{(-m)}(\mathbb{R};\mathbb{R})} \equiv \left(\sum_{n=0}^{\infty}\lambda_n^{-m}\langle\tilde{h}_n,u\rangle^2\right)^{\frac{1}{2}} < \infty.$$

Clearly $\mathscr{S}^{(-m)}(\mathbb{R};\mathbb{R})$ is a separable Hilbert space with inner product

$$\bigl(u,v\bigr)_{\mathscr{S}^{(-m)}(\mathbb{R};\mathbb{R})} = \sum_{n=0}^{\infty}\lambda_n^{-m}\langle\tilde{h}_n,u\rangle\langle\tilde{h}_n,v\rangle$$

and, if $h_n^{(-m)} = \lambda_n^{\frac{m}{2}}\tilde{h}_n$, then $\{h_n^{(-m)} : n\geq 0\}$ is an orthonormal basis for it. Alternatively, let $\mathcal{H}$ be the Hermite operator in (2.3.21), and define the operators $(-\mathcal{H})^{\frac{m}{2}}$ for $m\in\mathbb{Z}$ by

$$(-\mathcal{H})^{\frac{m}{2}}\varphi = \sum_{n=0}^{\infty}\lambda_n^{\frac{m}{2}}\bigl(\varphi,\tilde{h}_n\bigr)_{L^2(\lambda_{\mathbb{R}};\mathbb{R})}\tilde{h}_n \text{ for } \varphi\in\mathscr{S}(\mathbb{R};\mathbb{R}).$$

These operators are self-adjoint and therefore extend to $\mathscr{S}^*(\mathbb{R};\mathbb{R})$ by defining $(-\mathcal{H})^{\frac{m}{2}}u$ so that

$$\langle\varphi,(-\mathcal{H})^{\frac{m}{2}}u\rangle = \langle(-\mathcal{H})^{\frac{m}{2}}\varphi,u\rangle.$$

By (2.3.21), we know that $(-\mathcal{H})^{\frac{m}{2}}\tilde{h}_n = \lambda_n^{\frac{m}{2}}\tilde{h}_n$, and so it should now be clear that $u\in\mathscr{S}^{(-m)}(\mathbb{R};\mathbb{R}) \iff (-\mathcal{H})^{-\frac{m}{2}}u\in L^2(\lambda_{\mathbb{R}};\mathbb{R})$ and that $(-\mathcal{H})^{\frac{m}{2}}$ maps $L^2(\lambda_{\mathbb{R}};\mathbb{R})$ isometrically onto $\mathscr{S}^{(-m)}(\mathbb{R};\mathbb{R})$. Finally, if Q_t is the operator in (2.3.24), then, by (2.3.25),

$$\langle\tilde{h}_n,Q_tu\rangle = \langle Q_t\tilde{h}_n,u\rangle = e^{-\lambda_n t}\langle h_n,u\rangle,$$

and so Q_t maps $\mathscr{S}^*(\mathbb{R};\mathbb{R})$ into $\mathscr{S}(\mathbb{R};\mathbb{R})$ and, as $t \searrow 0$, $Q_t u \longrightarrow u$ in $\mathscr{S}^{(-m)}(\mathbb{R};\mathbb{R})$ if $u \in \mathscr{S}^{(-m)}(\mathbb{R};\mathbb{R})$.

3.6.2 An Abstract Wiener Space for $L^2(\lambda_{\mathbb{R}};\mathbb{R})$

The Hilbert space $L^2(\lambda_{\mathbb{R}};\mathbb{R})$ is an example of a Hilbert space that requires the preceding considerations. Indeed, if B were a Banach space of distributions for which δ_0 was represented by an element of B^*, then there would have to exist a $f_0 \in L^2(\lambda_{\mathbb{R}};\mathbb{R})$ such that $(\varphi, f_0)_{L^2(\lambda_{\mathbb{R}};\mathbb{R})} = \varphi(0)$ when $\varphi \in \mathscr{S}(\mathbb{R};\mathbb{R})$. But this would mean that $|\varphi(0)| \leq \|f_0\|_{L^2(\lambda_{\mathbb{R}};\mathbb{R})}\|\varphi\|_{L^2(\lambda_{\mathbb{R}};\mathbb{R})}$, and so $\|f_0\|_{L^2(\lambda_{\mathbb{R}};\mathbb{R})}$ would be infinite.

Proceeding as in the proof of Theorem 3.3.2, define $A : \mathscr{S}^{(-2)}(\mathbb{R};\mathbb{R}) \longrightarrow \mathscr{S}^{(-2)}(\mathbb{R};\mathbb{R})$ by

$$Au = \sum_{n=0}^{\infty} \lambda_n^{-1} \big(h_n^{(-2)}, u\big)_{\mathscr{S}^{(-2)}(\mathbb{R};\mathbb{R})} \tilde{h}_n,$$

and observe that

$$\sum_{n=0}^{\infty} \big(h_n^{(-2)}, Ah_n^{(-2)}\big)_{\mathscr{S}^{(-2)}(\mathbb{R};\mathbb{R})} = \sum_{n=0}^{\infty} \lambda_n^{-2} < \infty.$$

Hence, by Theorem 3.2.8, there is a centered, Gaussian measure $\mathcal{W}_L \in M_1\big(\mathscr{S}^{(-2)}(\mathbb{R};\mathbb{R})\big)$ for which $\big(u, Av\big)_{\mathscr{S}^{(-2)}(\mathbb{R};\mathbb{R})}$ is the covariance function. Moreover, given $u \in \mathscr{S}^{(-2)}(\mathbb{R};\mathbb{R})$, set $f_u = Au$, note that $f_u \in L^2(\lambda_{\mathbb{R}};\mathbb{R})$, and check that, for $g \in L^2(\lambda_{\mathbb{R}};\mathbb{R})$,

$$(g, u)_{\mathscr{S}^{(-2)}(\mathbb{R};\mathbb{R})}\rangle = \sum_{n=0}^{\infty} \big(g, h_n^{(-2)}\big)_{\mathscr{S}^{(-2)}(\mathbb{R};\mathbb{R})} \big(h_n^{(-2)}, u\big)_{\mathscr{S}^{(-2)}(\mathbb{R};\mathbb{R})} = \big(g, f_u\big)_{L^2(\lambda_{\mathbb{R}};\mathbb{R})}$$

and that $\|f_u\|^2_{L^2(\lambda_{\mathbb{R}};\mathbb{R})} = \big(u, Au\big)_{\mathscr{S}^{(-2)}(\mathbb{R};\mathbb{R})}$. Therefore we have proved the following theorem.

Theorem 3.6.1 $\big(L^2(\lambda_{\mathbb{R}};\mathbb{R}), \mathscr{S}^{(-2)}(\mathbb{R};\mathbb{R}), \mathcal{W}_L\big)$ *is an abstract Wiener space.*

There is an interesting connection between this abstract Wiener space and Brownian motion. Namely, define

$$B(t) = \begin{cases} \mathcal{I}\big(\mathbf{1}_{[0,t]}\big) & \text{for } t \geq 0 \\ -\mathcal{I}\big(\mathbf{1}_{[t,0]}\big) & \text{for } t < 0. \end{cases}$$

Then $\{B(t) : t \in \mathbb{R}\}$ is a centered Gaussian process under $\mathcal{W}_L$ with the covariance function w in (2.5.3). Hence there is a continuous version of $\{B(t) : t \geq 0\}$, and this version will be a Brownian motion.

If elements of $\mathscr{S}^{(-2)}(\mathbb{R};\mathbb{R})$ were bonafide functions and we pretended that $\mathcal{I}(f)(u) = (f,u)_{L^2(\lambda_{\mathbb{R}};\mathbb{R})}$, then the preceding would be saying that

$$B(t,u) = \begin{cases} \int_0^t u(\tau)\,d\tau & \text{if } t \geq 0 \\ -\int_t^0 u(\tau)\,d\tau & \text{if } t < 0. \end{cases}$$

That is, "$u(t) = \dot{B}(t,u)$". To make this mathematically kosher, we have to formulate it in the language of distribution theory. That is, given $v \in \mathscr{S}^*(\mathbb{R};\mathbb{R})$, ∂v is the element of $\mathscr{S}^*(\mathbb{R};\mathbb{R})$ satisfying

$$\langle \varphi, \partial v\rangle = -\langle \varphi', v\rangle \quad \text{for all } \varphi \in \mathscr{S}(\mathbb{R};\mathbb{R}).$$

To see that $u = \partial B(\,\cdot\,,u)$ in this sense, set

$$f_t = \begin{cases} \mathbf{1}_{[0,t]} & \text{if } t \geq 0 \\ -\mathbf{1}_{[t,0]} & \text{if } t < 0 \end{cases}$$

and $B_s(t,u) = \langle Q_s f_t, u\rangle$ for $s > 0$. Then, as $s \searrow 0$,

$$\big\|B_s(t,\,\cdot\,) - B(t,\,\cdot\,)\big\|_{L^2(\mathcal{W}_L;\mathbb{R})} = \|Q_s f_t - f_t\|_{L^2(\lambda_{\mathbb{R}};\mathbb{R})} \longrightarrow 0 \qquad (*)$$

uniformly for t in compacts. At the same time,

$$\begin{aligned} \langle Q_s f_t, u\rangle &= \lim_{\sigma\searrow 0}\big(Q_s f_t, Q_\sigma u\big)_{L^2(\lambda_{\mathbb{R}};\mathbb{R})} = \lim_{\sigma\searrow 0}\big(f_t, Q_{s+\sigma} u\big)_{L^2(\lambda_{\mathbb{R}};\mathbb{R})} \\ &= \lim_{\sigma\searrow 0}\big(Q_\sigma f_t, Q_s u\big)_{L^2(\lambda_{\mathbb{R}};\mathbb{R})} = \big(f_t, u_s\big)_{L^2(\lambda_{\mathbb{R}};\mathbb{R})}, \end{aligned}$$

where $u_s = Q_s u$, and so

$$B_s(t,u) = \begin{cases} \int_0^t u_s(\tau)\,d\tau & \text{if } t \geq 0 \\ -\int_t^0 u_s(\tau)\,d\tau & \text{if } t < 0. \end{cases}$$

Hence, for any $\varphi \in \mathscr{S}(\mathbb{R};\mathbb{R})$,

$$\int \varphi'(t) B_s(t,u)\,dt = -\int \varphi(t) u_s(t)\,dt,$$

and therefore

$$
\begin{aligned}
&\left(\int \left|\int \varphi'(t)B(t,u)\,dt + \langle \varphi, u\rangle\right|^2 \mathcal{W}_L(du)\right)^{\frac{1}{2}} \\
&\le \left(\int \left|\int \varphi'(t)\big(B(t,u) - B_s(t,u)\big)\,dt\right|^2 \mathcal{W}_L(du)\right)^{\frac{1}{2}} \\
&\qquad + \left(\int \langle \varphi, u - u_s\rangle^2 \,\mathcal{W}_L(du)\right)^{\frac{1}{2}}.
\end{aligned}
$$

By Minkowski's inequality and $(*)$, the first term on the right tends to 0 as $s \searrow 0$, and, because $u_s \longrightarrow u$ in $\mathscr{S}^{(-2)}(\mathbb{R};\mathbb{R})$ and $\|u_s\|_{\mathscr{S}^{(-2)}(\mathbb{R};\mathbb{R})} \le \|u\|_{\mathscr{S}^{(-2)}(\mathbb{R};\mathbb{R})} \in L^2(\mathcal{W}_L;R)$, the second term does also. Finally, because $\mathscr{S}(\mathbb{R};\mathbb{R})$ is separable, having proved that $\langle \varphi', B(\,\cdot\,,u)\rangle = -\langle \varphi, u\rangle$ (a.s.,$\mathcal{W}_L$) for each $\varphi \in \mathscr{S}(\mathbb{R};\mathbb{R})$, it follows that $u = \partial B(\,\cdot\,,u)$ for $\mathcal{W}_L$-almost every $u \in \mathscr{S}^{(-2)}(\mathbb{R};\mathbb{R})$. As a dividend of these considerations, we see that $\mathcal{W}_L$ is supported on a much smaller class of distributions than $\mathscr{S}^{(-2)}(\mathbb{R};\mathbb{R})$. Namely, we now know that $\mathcal{W}_L$ is supported on the space of tempered distributions that are the first derivative of functions that are Hölder continuous of every order less than $\frac{1}{2}$ and grow at infinity slower than every power greater than $\frac{1}{2}$.

In the engineering literature, the derivative of Brownian motion is known as *white noise*. That is because, if one pretends the $\dot{B}(t)$ exists in a classical sense, then the process $\{\dot{B}(t) : t \in \mathbb{R}\}$ would be totally uncorrelated Gaussian process. In fact, its covariance function $c(s,t)$ would be ∞ when $s = t$ and 0 when $s \ne t$. Hence, for each $t \in \mathbb{R}$, $\dot{B}(t)$ would be a centered Gaussian random variable with infinite variance which is independent of $\{\dot{B}(s) : s \ne t\}$. Admittedly, this picture is much more intuitively appealing than the more mathematically correct one given above, but loss of intuition is a price that mathematicians are accustomed to paying.

The Ornstein–Uhlenbeck process is another process that can be constructed starting from

$$\big(L^2(\lambda_{\mathbb{R}};\mathbb{R}), \mathscr{S}^{(-2)}(\mathbb{R};\mathbb{R}), \mathcal{W}_L\big)$$

by the procedure we just used to construct Brownian motion. The reason why the Gaussian process $\{\mathcal{I}(\mathbf{1}_{[0,t]}) : t \in \mathbb{R}\}$ under $\mathcal{W}_L$ has the distribution of a Brownian motion is that $\big(\mathbf{1}_{[0,s]}, \mathbf{1}_{[0,t]}\big)_{L^2(\lambda_{\mathbb{R}};\mathbb{R})} = s \wedge t$. Thus, what is needed to construct an Ornstein–Uhlenbeck process is a family $\{\psi_t : t \in \mathbb{R}\} \subseteq L^2(\lambda_{\mathbb{R}};\mathbb{R})$ for which $\big(\psi_s, \psi_t\big)_{L^2(\lambda_{\mathbb{R}};\mathbb{R})} = \frac{1}{2}e^{-|t-s|}$. To find such a family, remember that $e^{-|t|}$ is the characteristic function of the Cauchy distribution $\frac{1}{\pi(1+\xi^2)}$, and conclude that we can take $\psi_t \in L^2(\lambda_{\mathbb{R}};\mathbb{R})$ to be the function whose Fourier transform is $e^{it\xi}(1+\xi^2)^{-\frac{1}{2}}$. Another expression of ψ_t is contained in (**ii**) of Exercise 3.7.1.

3.7 An Abstract Wiener Space for $L^2(\mathbb{R}^N;\mathbb{R})$

As we have seen, the abstract Wiener space for $L^2(\lambda_{\mathbb{R}};\mathbb{R})$ already requires the introduction of distributions, albeit distributions of a mild order.

Although the order of the distributions required goes up with dimension, only a few new ideas are required. The first step is to introduce the Hermite functions for $\mathbb{R}^N$. Given $\mathbf{n}=(n_1,\dots,n_N)\in\mathbb{N}^N$, define

$$\tilde{h}_{\mathbf{n}}(x)=\prod_{j=1}^{N}\tilde{h}_{n_j}(x_j)\text{ for }x\in\mathbb{R}^N.$$

Then $\{\tilde{h}_{\mathbf{n}}:\mathbf{n}\in\mathbb{N}^N\}$ forms an orthonormal basis in $L^2(\mathbb{R}^N;\mathbb{R})$. In addition, for each $\mathbf{n}$

$$H\tilde{h}_{\mathbf{n}}=-\lambda_{\mathbf{n}}\tilde{h}_{\mathbf{n}}\text{ where }H=\tfrac{1}{2}\big(\Delta-|x|^2\big),\ \lambda_{\mathbf{n}}=\Big(\|\mathbf{n}\|_1+\frac{N}{2}\Big),\text{ and }\|\mathbf{n}\|_1=\sum_{j=1}^{N}n_j.$$

The first of these is just the standard construction of bases on product spaces from bases on the factors, and the second is an easy consequence of (2.3.21). Further, these Hermite functions play the same role in Schwartz's theory of tempered distributions on $\mathbb{R}^N$ as their antecedents do for his theory of tempered distributions on $\mathbb{R}$. That is, the test function space $\mathscr{S}(\mathbb{R}^N;\mathbb{R})$ consists of $\varphi\in L^2(\lambda_{\mathbb{R}^N};\mathbb{R})$ with the property that

$$\sum_{\mathbf{n}\in\mathbb{N}^N}\lambda_{\mathbf{n}}^m\big(\varphi,\tilde{h}_{\mathbf{n}}\big)^2_{L^2(\lambda_{\mathbb{R}^N};\mathbb{R})}<\infty\text{ for all }m\ge 0,$$

and the space $\mathscr{S}^*(\mathbb{R}^N;\mathbb{R})$ of tempered distributions is the set of linear functionals u on $\mathscr{S}(\mathbb{R}^N;\mathbb{R})$ with the property that

$$\sum_{\mathbf{n}\in\mathbb{N}^N}\lambda_{\mathbf{n}}^{-m}\langle\tilde{h}_n,u\rangle^2<\infty\quad\text{for some }m\ge 0.$$

Finally, for each $m\ge 0$, one takes $\mathscr{S}^{(-m)}(\mathbb{R}^N;\mathbb{R})$ to be the separable Hilbert space $u\in\mathscr{S}^*(\mathbb{R}^N;\mathbb{R})$ for which

$$\|u\|_{\mathscr{S}^{(-m)}(\mathbb{R}^N;\mathbb{R})}=\Bigg(\sum_{\mathbf{n}\in\mathbb{N}^N}\lambda_{\mathbf{n}}^{-m}\langle\tilde{h}_{\mathbf{n}},u\rangle^2\Bigg)^{\frac{1}{2}}<\infty.$$

Obviously, if $h_{\mathbf{n}}^{(-m)}=\lambda_{\mathbf{n}}^{\frac{m}{2}}\tilde{h}_{\mathbf{n}}$, then $\{h_{\mathbf{n}}^{(-m)}:\mathbf{n}\in\mathbb{N}^N\}$ is an orthonormal basis for $\mathscr{S}^{(-m)}(\mathbb{R}^N;\mathbb{R})$.

In view of the preceding preparations, it should be clear how to go about constructing an abstract Wiener space for $L^2(\lambda_{\mathbb{R}^N};\mathbb{R})$. Indeed, define A :

$\mathscr{S}^{(-N-1)}(\mathbb{R}^N;\mathbb{R}) \longrightarrow \mathscr{S}^{(-N-1)}(\mathbb{R}^N;\mathbb{R})$ by

$$Au = \sum_{\mathbf{n}\in\mathbb{N}^N} \lambda_{\mathbf{n}}^{-\frac{N+1}{2}} \big(h_{\mathbf{n}}^{(-N-1)}, u\big)_{\mathscr{S}^{(-N-1)}(\mathbb{R}^N;\mathbb{R})} \tilde{h}_{\mathbf{n}}.$$

Then

$$\begin{aligned}
\sum_{\mathbf{n}\in\mathbb{N}^N} \big(h_{\mathbf{n}}^{(-N-1)}, Ah_{\mathbf{n}}^{(-N-1)}\big)_{\mathscr{S}^{(-N-1)}(\mathbb{R}^N;\mathbb{R})} &= \sum_{\ell=0}^{\infty} \lambda_{\mathbf{n}}^{-N-1} \mathrm{card}\big\{\mathbf{n} : \|\mathbf{n}\|_1 = \ell\big\} \\
&\le \left(\frac{2}{N}\right)^{N+1} + \sum_{\ell=1}^{\infty} \big(\ell + \tfrac{N}{2}\big)^{-N-2} \mathrm{card}\big\{\mathbf{n} : \|\mathbf{n}\|_1 \le \ell\big\} \\
&\le \left(\frac{2}{N}\right)^{N+1} + \sum_{\ell=1}^{\infty} \frac{N(\ell+1)^N}{\big(\ell + \frac{N}{2}\big)^{N+2}} < \infty.
\end{aligned}$$

Hence, by Theorem 3.2.8, there is a centered Gaussian measure

$$\mathcal{W}_{LN} \in M_1\big(\mathscr{S}^{(-N-1)}(\mathbb{R}^N;\mathbb{R})\big)$$

for which $\big(v, Au\big)_{\mathscr{S}^{(-N-1)}(\mathbb{R}^N;\mathbb{R})}$ is the covariance function. Next, for a given $u \in \mathscr{S}^{(-N-1)}(\mathbb{R}^N;\mathbb{R})$, set $f_u = Au$, and check that $\big(g, u\big)_{\mathscr{S}^{(-N-1)}(\mathbb{R}^N;\mathbb{R})} = (g, f_u)_{L^2(\lambda_{\mathbb{R}^N};\mathbb{R})}$ for $g \in L^2(\lambda_{\mathbb{R}^N};\mathbb{R})$ and $\|f_u\|^2_{L^2(\lambda_{\mathbb{R}^N};\mathbb{R})} = \big(u, Au\big)_{\mathscr{S}^{(-N-1)}}$. Therefore $\big(L^2(\lambda_{\mathbb{R}^N};\mathbb{R}), \mathscr{S}^{(-N-1)}(\mathbb{R}^N;\mathbb{R}), \mathcal{W}_{LN}\big)$ is an abstract Wiener space.

When $N \ge 2$, there is no true analog of Brownian motion because there is no way to interpret the Hilbert space $H_0^1(\mathbb{R}^N;\mathbb{R})$ as a space of tempered distributions. One might naïvely guess that it should be a space consisting of $u \in \mathscr{S}^*(\mathbb{R}^N;\mathbb{R})$ for which $\nabla u \in L^2(\lambda_{\mathbb{R}^N};\mathbb{R}^N)$, but, because, when $N \ge 2$ such elements of $L^2(\lambda_{\mathbb{R}^N};\mathbb{R})$ are defined only a set of measure 0, the condition $u(0) = 0$ makes no sense, one cannot use that condition to determine u uniquely. See part (**i**) of Exercise 3.7.1 for more details.

In spite of preceding, there is an approximate analog of Brownian motion, known as the *Brownian sheet*, in higher dimensions. To describe it, for $\mathbf{t} = (t_1, \dots, t_N) \in \mathbb{R}^N$ define

$$\varPhi_{\mathbf{t}}(x) = \prod_{j=1}^{N} \big(\mathbf{1}_{[0,t_j^+]}(x_j) - \mathbf{1}_{[t_j^-,0)}(x_j)\big) \text{ for } x = (x_1, \dots, x_N) \in \mathbb{R}^N,$$

and set $B(\mathbf{t}) = \mathcal{I}\big(\varPhi_{\mathbf{t}}\big)$. Then $\{B(\mathbf{t}) : \mathbf{t} \in \mathbb{R}^N\}$ is a centered Gaussian family under $\mathcal{W}_{LN}$ with covariance function

$$\mathbb{E}^{\mathcal{W}_{LN}}\big[B(\mathbf{s})B(\mathbf{t})\big] = \begin{cases} \prod_{j=1}^N |s_j| \wedge |t_j| & \text{if } s_j t_j \ge 0 \text{ for all } 1 \le j \le N \\ 0 & \text{otherwise.} \end{cases}$$

As a consequence, one finds that $B(\mathbf{t})-B(\mathbf{s})$ is independent of $B(\mathbf{s})$ if $s_j \le t_j$ for all $1 \le j \le N$ and that there is a $C<\infty$ such that

$$\mathbb{E}^{\mathcal{W}_{L^N}}\left[\big(B(\mathbf{t})-B(\mathbf{s})\big)^2\right] \le CT^{N-1}|\mathbf{t}-\mathbf{s}| \text{ for } T \ge 1 \text{ and } \mathbf{s},\mathbf{t} \in [-T,T]^N.$$

Hence, by Corollary 2.5.5, there is a version of $\{B(\mathbf{t}) : \mathbf{t} \in \mathbb{R}^N\}$ which is Hölder continuous of every order less than $\frac{1}{2}$. Finally, by essentially the same argument as we used when $N=1$, one can show that

$$\frac{\partial^N B(t,u)}{\partial_{t_1}\cdots\partial_{t_N}} = u \text{ for } \mathcal{W}_{L^N}\text{-almost every } u \in \mathscr{S}^{(-N-1)}(\mathbb{R}^N;\mathbb{R}).$$

3.7.1 The Ornstein–Uhlenbeck Free Field in Higher Dimensions

As distinguished from $H^1_0(\mathbb{R};\mathbb{R})$, the analog $H^1(\mathbb{R}^N;\mathbb{R})$ of $H^1(\mathbb{R};\mathbb{R})$ is a function space: one simply has to complete $\mathscr{S}(\mathbb{R}^N;\mathbb{R})$ with respect to the Hilbert norm corresponding to the inner product

$$\big(g,h\big)_{H^1(\mathbb{R}^N;\mathbb{R})} = \big(g,h\big)_{L^2(\lambda_{\mathbb{R}^N};\mathbb{R})} + \big(\nabla g,\nabla h\big)_{L^2(\lambda_{\mathbb{R}^N};\mathbb{R}^N)}.$$

However, when $N \ge 2$, it will be shown that $H^1(\mathbb{R}^N;\mathbb{R})$ is not the Cameron–Martin space for a centered Gaussian measure on a Banach space of distributions on which evaluation at a point is given by an element of its dual. Thus, like the one for $L^2(\mathbb{R}^N;\mathbb{R})$, a Gaussian measure for which $H^1(\mathbb{R}^N;\mathbb{R})$ is the Cameron–Martin space lives on a space of distributions that cannot be evaluated at a point, and so the associated abstract Wiener space describes a Euclidean field.

To understand what follows, it is helpful to re-interpret the construction that we made at the end of § 3.6.2 of the Ornstein–Uhlenbeck process starting from $\big(L^2(\lambda_{\mathbb{R}};\mathbb{R}),\mathscr{S}^{(-2)}(\mathbb{R};\mathbb{R}),\mathcal{W}_L\big)$. If $\varphi \in \mathscr{S}(\mathbb{R};\mathbb{R})$, then $(h,\varphi)_{H^1(\mathbb{R};\mathbb{R})} = \big(h,Lg\big)_{L^2(\lambda_{\mathbb{R}};\mathbb{R})}$ for $h \in H^1(\mathbb{R};\mathbb{R})$, where L is the Bessel operator $\mathbf{1}-\partial^2$. Clearly L maps $\mathscr{S}(\mathbb{R};\mathbb{R})$ continuously into itself, and, in terms of the Fourier transform, the action of L is given by $\widehat{L\varphi} = (1+\xi^2)\hat{\varphi}(\xi)$. Further, if, for any $\alpha \in \mathbb{R}$ and $\varphi \in \mathscr{S}(\mathbb{R};\mathbb{R})$, $L^\alpha\varphi$ is given by $\widehat{L^\alpha\varphi}(\xi) = (1+\xi^2)^\alpha\hat{\varphi}(\xi)$, then L^α also maps $\mathscr{S}(\mathbb{R};\mathbb{R})$ continuously into itself and $L^{\alpha+\beta} = L^\alpha \circ L^\beta$. Since $\big(\psi,L^\alpha\varphi\big)_{L^2(\lambda_{\mathbb{R}};\mathbb{R})} = \big(L^\alpha\varphi,\psi\big)_{L^2(\lambda_{\mathbb{R}};\mathbb{R})}$ for $\varphi,\psi \in \mathscr{S}(\mathbb{R};\mathbb{R})$, this definition of L^α on $\mathscr{S}^*(\mathbb{R};\mathbb{R})$ is consistent with the one on $\mathscr{S}(\mathbb{R};\mathbb{R})$, and the function ψ_t in § 3.6.2 is $L^{-\frac{1}{2}}\delta_t$.

When $\alpha \le 0$, it is easy to check that L^α is a continuous, linear map of $\mathscr{S}^{(-m)}(\mathbb{R};\mathbb{R})$ into itself for every $m \in \mathbb{N}$ and that $L^{-\frac{1}{2}}$ is an isometric isomorphism of $L^2(\lambda_{\mathbb{R}};\mathbb{R})$ onto $H^1(\mathbb{R};\mathbb{R})$. In particular, we can apply

Theorem 3.3.3 to see that $\big(H^1(\mathbb{R};\mathbb{R}), B, (L^{-\frac{1}{2}})_*\mathcal{W}_L\big)$ is an abstract Wiener space when B is the Banach space $\big\{L^{-\frac{1}{2}}u : u \in \mathscr{S}^{(-2)}(\mathbb{R};\mathbb{R})\big\}$ with norm $\|x\|_B = \|L^{\frac{1}{2}}x\|_{\mathscr{S}^{(-2)}(\mathbb{R};\mathbb{R})}$. As we know, there is a more pleasing choice, namely the one in Theorem 3.3.9, but the construction used here is too crude to arrive at that choice.

With the preceding in mind, it should be clear how to construct an abstract Wiener space whose Cameron–Martin space is $H^1(\mathbb{R}^N;\mathbb{R})$ from $\big(L^2(\lambda_{\mathbb{R}^N};\mathbb{R}), \mathscr{S}^{(-N-1)}(\mathbb{R}^N;\mathbb{R}), \mathcal{W}_{LN}\big)$. Namely, one first observes that, for $\varphi \in \mathscr{S}(\mathbb{R}^N;\mathbb{R})$, $(h,\varphi)_{H^1(\mathbb{R}^N;\mathbb{R})} = \big(h, L\varphi\big)_{L^2(\lambda_{\mathbb{R}^N};\mathbb{R})}$, where L is the Bessel operator $\mathbf{1}-\Delta$. Proceeding as in the case when $N=1$, one uses the Fourier transform to define the operators L^α for $\alpha \in \mathbb{R}$ first on $\mathscr{S}(\mathbb{R}^N;\mathbb{R})$ and then on $\mathscr{S}^*(\mathbb{R}^N;\mathbb{R})$ and checks that $L^{-\frac{1}{2}}$ is a bounded linear map of $\mathscr{S}^{(-N-1)}(\mathbb{R}^N;\mathbb{R})$ into itself and an isometric isomorphism from $L^2(\lambda_{\mathbb{R}^N};\mathbb{R})$ onto $H^1(\mathbb{R}^N;\mathbb{R})$. Hence, by Theorem 3.3.3, $\big(H^1(\mathbb{R}^N;\mathbb{R}), B, (L^{-\frac{1}{2}})_*\mathcal{W}_{LN}\big)$ is an abstract Wiener space when B is the Banach space $\big\{L^{-\frac{1}{2}}u : u \in \mathscr{S}^{(-N-1)}(\mathbb{R}^N;\mathbb{R})\big\}$ with norm $\|x\|_B = \|L^{\frac{1}{2}}x\|_{\mathscr{S}^{(-N-1)}(\mathbb{R}^N;\mathbb{R})}$.

The critical difference between the cases $N=1$ and $N \geq 2$ comes from the fact that $(1+|\xi|^2)^{-\frac{1}{2}}$ is an element of $L^2(\lambda_{\mathbb{R}^N};\mathbb{R})$ if and only if $N=1$, and therefore $L^{-\frac{1}{2}}\delta_x \in L^2(\lambda_{\mathbb{R}^N};\mathbb{R})$ if and only if $N=1$. Thus, when $N \geq 2$, there are no functions ψ_x, $x \in \mathbb{R}^N$, for which $\big(\psi_x, \psi_y\big)_{L^2(\lambda_{\mathbb{R}^N};\mathbb{R})} = k(y-x)$, where $\hat{k}(\xi) = (1+|\xi|^2)^{-1}$. As a consequence, when $N \geq 2$, there is no centered Gaussian process $\{X(x) : x \in \mathbb{R}^N\}$ that plays the role for $\big(H^1(\mathbb{R}^N;\mathbb{R}), \mathscr{S}^{(-N-1)}(\mathbb{R}^N;\mathbb{R}), \mathcal{W}_{FN}\big)$ that the Ornstein–Uhlenbeck plays when $N=1$. Alternatively, this difference can be seen in terms of L^{-1}. For δ_0 to be an element of B^*, it is necessary that there exist an $h_0 \in H^1(\mathbb{R}^N;\mathbb{R})$ such that

$$\langle \varphi, \delta_0\rangle = (\varphi, h_0)_{H^1(\mathbb{R}^N;\mathbb{R})} = \big(L\varphi, h_0\big)_{L^2(\lambda_{\mathbb{R}^N};\mathbb{R})} = \langle \varphi, Lh_0\rangle$$

for all $\varphi \in \mathscr{S}(\mathbb{R}^N;\mathbb{R})$. This means that $Lh_0 = \delta_0$ and therefore $\widehat{h_0}(\xi) = (1+|\xi|^2)^{-1}$. But $(1+|\xi|^2)^{-1} \in L^2(\lambda_{\mathbb{R}^N};\mathbb{R})$ if and only if $N \in \{1,2\}$ and $\widehat{\nabla h_0}(\xi) = (1+|\xi|^2)^{-1}\xi \in L^2(\lambda_{\mathbb{R}^N};\mathbb{R}^N)$ if and only if $N=1$. Thus h_0 exists if and only if $N=1$.

3.7.2 Is There any Physics Here?

There are many reasons why the answer is a resounding **NO**. Physicists want non-trivial *quantum fields*, and all we have done is produce Euclidean fields, and not even particularly interesting ones.

Feynman's path-intregal formalism provides a way of understanding why we have not been doing physics. If Feynman were asked to describe the measure $\mathcal{W}$ in an abstract Wiener space $(H, B, \mathcal{W})$, he would say that $\mathcal{W}$ is given by the formula in (3.1.1). Of course, unless H is finite dimensional, from a mathematical standpoint, (3.1.1) is irreparably flawed when H is infinite dimensional: as we have seen, the measure $\mathcal{W}$ lives on the Banach space B and does not even see H. Nonetheless, as we saw when we derived (3.3.2), one can make accurate predictions based on expressions like (3.1.1), and so I will ignore its flaws in the following.

The Feynman representation of measure $\mathcal{W}_{F4}$ in

$$\big(H^1(\mathbb{R}^4;\mathbb{R}), \mathscr{S}^{(-5)}(\mathbb{R}^4;\mathbb{R}), \mathcal{W}_{F4}\big)$$

is

$$\mathcal{W}_{FN}(dh) = \frac{1}{Z}\exp\left(-\frac{\|h\|^2_{L^2(\mathbb{R}^4;\mathbb{R})} + \|\nabla h\|^2_{L^2(\mathbb{R}^4;\mathbb{R}^4)}}{2}\right)\lambda_{H^1(\mathbb{R}^4;\mathbb{R})}(dh).$$

Although we have given a mathematically satisfactory description of $\mathcal{W}_{F4}$, that is only the first step in producing a physically satisfactory quantum field. A basic physical requirement is that a quantum field be invariant under Lorentz transformations (i.e., ones that preserve the quadratic form $-x_1^2 + \sum_{j=2}^4 x_j^2$) of coordinates, whereas $\mathcal{W}_{F4}$ is invariant under Euclidean (i.e., ones that preserve $|x|^2$) transformations of coordinates. The obvious way to convert a Euclidean invariant field into Lorentz invariant one is to replace x_1 by ix_1, a step that is easier to describe than it is to carry out.

Besides the issue raised in the preceding paragraph, there is another serious problem to be confronted even in the Euclidean setting. The problem is that the free field is a Euclidean model of a system of free particles, particles which do not interact. The simplest Euclidean model of interacting particles would be one in which the density $\exp\left(-\frac{\|h\|^2_{H^1(\mathbb{R}^4;\mathbb{R})}}{2}\right)$ is replaced by

$$\exp\left(-\frac{\|h\|^2_{H^1(\mathbb{R}^4;\mathbb{R})} + \|h\|^4_{L^4(\mathbb{R}^4;\mathbb{R})}}{2}\right),$$

and the problem is that, although this would make perfectly good sense if $\mathcal{W}_{F4}$ lived on $H^1(\mathbb{R}^4;\mathbb{R})$, no naïve interpretation is available when one takes into account of the fact that $\mathcal{W}_{F4}$ lives on a space of distributions. Indeed, one knows how to apply lots of linear operations to distributions, but one doesn't know how to apply non-linear ones to them. Nelson was able to handle this problem for a two dimensional Euclidean field and showed that it could be transformed into a Lorentz invariant field. Using different techniques, in a sequence of articles J. Glimm and A. Jaffe carried out the same program for three dimensional fields. Seeing as the book [4] that they subsequently wrote is 535 pages long, I think that I can be forgiven, maybe even thanked, for not attempting to summarize their work here. As far as I know, to date, nobody has succeeded in constructing a non-trivial four dimensional quantum field, and there are results that indicate that nobody ever will.

Exercise 3.7.1

(**i**) In connection with the problems with the space $H^1_0(\mathbb{R}^N;\mathbb{R})$, let $\mathcal{C}$ be the space of $\varphi \in C^1(\mathbb{R}^2;\mathbb{R})$ for which $\varphi_0(0) = 0$ and $\nabla\varphi \in L^2(\mathbb{R}^2;\mathbb{R})$. Then $H^1_0(\mathbb{R}^2;\mathbb{R})$ is the completion of $\mathcal{C}$ with respect to the Hilbert norm $\|\varphi\|_{H^1_0(\mathbb{R}^2;\mathbb{R})} = \|\nabla\varphi\|_{L^2(\mathbb{R}^2;\mathbb{R}^2)}$. Now define

$$\psi_n(x) = \begin{cases} 1 & \text{if } |x| \geq 1 \\ \frac{log(nx)}{\log n} & \text{if } \frac{1}{n} \leq |x| < 1 \\ 0 & \text{if } |x| < \frac{1}{n}. \end{cases}$$

Choose a $\rho \in C^\infty\big(\mathbb{R}^2;;[0,\infty)\big)$ which vanishes off of $B_{\mathbb{R}^2}(0,1)$ and has integral 1, set $\rho_n(x) = n^2\rho(nx)$, and take $\varphi_n = \rho_n * \psi_n$. Clearly, $\varphi_n \in \mathcal{C}$. Show that $\varphi_n \longrightarrow 1$ in $\mathscr{S}^*(\mathbb{R}^2;\mathbb{R})$ and $\|\nabla\varphi_n\|_{L^2(\mathbb{R}^2;\mathbb{R}^2)} \longrightarrow 0$. Thus, just because the $\|\cdot\|_{H^1_0(\mathbb{R}^2;\mathbb{R})}$ norms of a sequence in $\mathcal{C}$ tends to 0, the sequence need not converge to 0 in $\mathscr{S}^*(\mathbb{R}^2;\mathbb{R})$.

(**ii**) Our representation in § 3.7.1 of powers of the Bessel operator L was as Fourier multipliers. The goal in here is to describe the distributions of which the Fourier mulipliers for L^{-1} and $L^{-\frac{1}{2}}$ are the Fourier transform. To do so, define

$$r_N(\lambda) = (4\pi)^{-\frac{N}{2}} \int_0^\infty t^{\frac{N}{2}-2} e^{-\lambda t - \frac{1}{t}}\, dt \text{ for } \lambda > 0,$$

and set

$$k_N(x) = r_N\left(\frac{|x|^2}{4}\right) \text{ and } \psi_N(x) = r_{N+1}\left(\frac{|x|^2}{4}\right) \text{ for } x \in \mathbb{R}^N \setminus \{0\}.$$

Show that $\widehat{k_N}(\xi) = (1+|\xi|^2)^{-1}$ and $\widehat{\psi_N}(\xi) = (1+|\xi|^2)^{-\frac{1}{2}}$.

Hint: Observe that

$$r_N\left(\frac{|x|^2}{4}\right) = \int_0^\infty e^{-t} g_t(x)\, dt \text{ where } g_t(x) = (4\pi t)^{-\frac{N}{2}} e^{-\frac{|x|^2}{4t}}.$$

(**iii**) Continuing (**ii**), show that $r_1(\lambda) = 2^{-1}e^{-\sqrt{\lambda}}$, and use Abelian assymtotics for Laplace transforms to show that

$$r_N(\lambda) \sim \begin{cases} \log\frac{1}{\lambda} & \text{if } N = 2 \\ \frac{2\Gamma\left(\frac{N}{2}\right)\lambda^{1-\frac{N}{2}}}{N-2} & \text{if } N \geq 3 \end{cases} \quad \text{as } \lambda \searrow 0.$$

References

[1] Bryc, Wlodzimierz, *Normal distribution: characterizations with applications*, Lecture Notes in Statistics vol 100, Springer (1995)

[2] Chen, L. & Stroock, D.,*Additive Functions and Gaussian Measures.* Prokhorov and Contemporary Probability Theory, Springer Proceedings in Mathematics & Statistics, vol 33. Springer, Berlin, Heidelberg (1993)

[3] Enchev, O. & Stroock, D.,*Rademacher's Theorem for Wiener Functionals*, Ann. Probab. 21, pp. 25–33 (1993)

[4] Glimm, J. & Jaffe, A.,*Quantum Physics, 2nd ed.*, Springer (1987)

[5] Gross, L.,*Logarithmic Sobolev inequalities*, Am. J. Math. 97, pp. 1061-1083 (1975)

[6] ——— *Abstract Wiener spaces.* Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66), Vol. II: Contributions to Probability Theory, Part 1. Berkeley, Calif.: Univ. California Press., pp. 31–42 (1965/66)

[7] Segal, I.,*Ergodic Subgroups of the Orthogonal Group on a Real Hilbert Space*, Ann. of Math., vol. 66 #2 (1957)

[8] Simon, B., *The $P(\phi)2$ Euclidean* (*Quantum*) *Field Theory*, Princeton Series in Physics, Princeton Univ. Press, Princeton, N.J., (1974)

[9] Skorohod, A.V.,*Studies in the Theory of Random Processes*, Dover Books on Mathematics, (1982)

[10] Stroock, D.W., *Probability Theory, An Analytic View, 2nd Ed.*, Cambridge Univ. Press (2010)

[11] ———, *Essentials of Integration Theory for Analysis, 2nd ed.*, Springer GTM # 262. (2020)

[12] ———,*Some Thoughts about Segal's Ergodic Theorem*, Colloq. Math. 118(1), pp. 89-105 (2010)