Supersymmetric Theory of Stochastic Dynamics

Sourangshu Ghosh*

*sourangshug@iisc.ac.in, Department of Civil Engineering,
Indian Institute of Science Bangalore

Contents

Preface

This monograph develops a fully rigorous mathematical foundation for the supersymmetric structure underlying stochastic dynamics. We present a synthesis of measure theory, differential geometry, and quantum field theory with complete proofs, functional analytic rigor, and geometric intuition. Our goal is to place the Parisi–Sourlas supersymmetry and its generalizations on the same level of precision as stochastic analysis in the sense of Itô, Stroock, and Varadhan.

Chapter 1

Probability Measures on Infinite-Dimensional Function Spaces

1.1 Introduction

This chapter develops the measure-theoretic and functional-analytic foundations required for the rigorous treatment of stochastic processes and their supersymmetric extensions in subsequent chapters. We focus on Gaussian measures on separable Banach and Hilbert spaces, the construction of the Wiener measure, the Cameron–Martin theorem, Fernique's integrability theorem, and the basic change-of-measure result (Girsanov). The presentation aims for maximum mathematical precision: proofs are given in full detail and the necessary functional-analytic background is stated when used.

1.2 Notation and standing hypotheses

Throughout this chapter:

- $\mathbb{N} = \{1, 2, 3, \dots\}$. For $T > 0$ we denote by $C([0,T]; \mathbb{R}^d)$ the Banach space of continuous functions $x : [0,T] \to \mathbb{R}^d$ endowed with the supremum norm $\|x\|_\infty = \sup_{t\in[0,T]} |x(t)|$.
- If E is a separable Banach space we write $\mathcal{B}(E)$ for its Borel sigma-algebra.
- For a probability measure μ on $(E, \mathcal{B}(E))$ we denote expectation by $\mathbb{E}_\mu[\cdot]$ or simply $\mathbb{E}[\cdot]$ when μ is understood.
- For a separable Hilbert space H we denote the inner product by $\langle \cdot, \cdot \rangle_H$ and the induced norm by $\| \cdot \|_H$.
- We use the standard notation $L^p(\mu; E)$ for Bochner spaces when necessary.

1.3 Cylindrical measures and projective limits

We begin with finite-dimensional projections and the general projective construction that produces measures on infinite-dimensional spaces from consistent finite-dimensional distributions. This material is standard but we give full proofs to ensure later arguments are self-contained.

Definition 1.1 (Cylindrical sigma-algebra)**.** Let E be a separable Banach space and let E^* denote its continuous dual. For finite collections $\ell_1, \dots, \ell_n \in E^*$ define the projection $\pi_{\ell_1,\dots,\ell_n} : E \to \mathbb{R}^n$ by

$$\pi_{\ell_1,\dots,\ell_n}(x) = (\ell_1(x), \dots, \ell_n(x)) \tag{1.1}$$

The *cylindrical sigma-algebra* $\mathscr{C}(E)$ is the sigma-algebra generated by all such finite-dimensional projections.

The following lemma identifies when a collection of finite-dimensional measures arises from a bona fide measure on $(E, \mathcal{B}(E))$.

Theorem 1.2 (Kolmogorov–Prokhorov extension theorem, version for Banach spaces)**.** *Let $\{\mu_I\}$ be a projective family of probability measures indexed by finite-dimensional subspaces $I \subset E^*$ (i.e. laws for the projections π_I), satisfying the natural consistency conditions. Suppose further that for every $\varepsilon > 0$ there exists a compact $K \subset E$ with $\sup_I \mu_I(\pi_I(K)) \geq 1 - \varepsilon$. Then there exists a unique probability measure μ on $(E, \mathcal{B}(E))$ whose finite-dimensional projections coincide with $\{\mu_I\}$.*

Proof. The proof breaks into three parts:

- Construction of a probability measure on an ambient product space having the μ_I as finite-dimensional marginals.
- Verification that the constructed measure is carried by the subset of the product that equals the image of E under the evaluation map.
- Pushing the product measure forward to E and checking uniqueness and Radon property.

A. *Projective family to a measure on a product space.*

For each finite-dimensional $I \subset E^*$ identify I' (the algebraic dual of I) with $\mathbb{R}^{\dim I}$ in any linear coordinate system; endow I' with the Euclidean topology and Borel sigma-algebra. Consider the product space

$$X := \prod_{I \in \mathscr{I}} I', \tag{1.2}$$

equipped with the product topology and the product Borel sigma-algebra $\mathcal{B}(X)$. (Each factor I' is Polish, hence X with the product topology is Polish when the index set is countable; for possibly uncountable $\mathscr{I}$ we still use the product sigma-algebra and Tychonoff topology — the arguments that follow use tightness and Prokhorov and so work in the standard projective-limit framework.)

For each finite subcollection $I_1, \dots, I_m \in \mathscr{I}$ define a probability measure $\nu_{I_1,\dots,I_m}$ on $I_1' \times \cdots \times I_m'$ by taking it to be the pushforward of μ_I under the projection $I' \to I_1' \times \cdots \times I_m'$ for any I containing $I_1, \dots, I_m$ (such an I exists since the set of finite-dimensional subspaces is directed). By the projective consistency hypothesis the definition does not depend on the chosen $I \supset \{I_1, \dots, I_m\}$. The family $\{\nu_{I_1,\dots,I_m}\}$ is a consistent family of finite-dimensional distributions on the index set $\mathscr{I}$ (consistency under further marginalization holds by construction).

By the classical Kolmogorov (or Daniell–Kolmogorov) extension principle for finite-dimensional consistent marginals there exists a unique probability measure Π on the measurable space $(X, \mathcal{C})$ where $\mathcal{C}$ is the cylinder sigma-algebra generated by coordinate projections $\mathrm{pr}_I : X \to I'$. At this stage Π is a probability measure defined on the cylinder sigma-algebra. We will now upgrade Π to a Borel measure on X by using tightness (Prokhorov).

B. *Tightness and extension to the product Borel sigma-algebra; concentration on the evaluation image.*

Define the evaluation (or embedding) map

$$\iota : E \longrightarrow X, \qquad \iota(x) := \big(\pi_I(x)\big)_{I \in \mathscr{I}}, \tag{1.3}$$

where $\pi_I(x) = (\ell(x))_{\ell \in I} \in I'$. The map ι is linear and measurable (coordinate maps are continuous linear functionals on E).

By hypothesis $(*)$ there is uniform tightness with respect to compact sets in E: for each $\varepsilon > 0$ there exists a compact $K_\varepsilon \subset E$ with

$$\sup_{I \in \mathscr{I}} \mu_I(\pi_I(K_\varepsilon)) \geq 1 - \varepsilon. \tag{1.4}$$

Fix $\varepsilon > 0$ and such a compact K_ε. The set $\iota(K_\varepsilon) \subset X$ is compact in the product topology: indeed, each coordinate projection pr_I restricted to $\iota(K_\varepsilon)$ equals π_I applied to the compact set K_ε, hence has relatively compact image in the finite-dimensional Euclidean space I'; Tychonoff's theorem (and finite-dimensionality of each coordinate) shows $\iota(K_\varepsilon)$ is relatively compact, and closedness (hold because ι is continuous and K_ε compact) implies compactness. Therefore $\iota(K_\varepsilon)$ is a compact subset of X.

Now, for any finite-dimensional index I the marginal of the cylinder measure Π on the coordinate I' equals μ_I by construction; therefore

$$\Pi\big(\iota(K_\varepsilon)\big) \;=\; \lim_{F \nearrow \mathscr{I}} \Pi\big(\{x \in X : \mathrm{pr}_I(x) \in \pi_I(K_\varepsilon) \text{ for each } I \in F\}\big) \tag{1.5}$$

where the limit is over finite $F \subset \mathscr{I}$ directed by inclusion. Each cylinder event $\{x : \mathrm{pr}_I(x) \in \pi_I(K_\varepsilon)\ \forall I \in F\}$ has Π-measure at least $1 - \varepsilon$ because for each coordinate $I \in F$ the marginal

$$\mu_I(\pi_I(K_\varepsilon)) \geq 1 - \varepsilon \tag{1.6}$$

and by the union bound one gets a measure lower bound $1 - |F|\varepsilon$; letting ε be smaller appropriately and refining K_ε one obtains

$$\Pi(\iota(K_\varepsilon)) \geq 1 - \varepsilon \tag{1.7}$$

More directly: the projective consistency produces for each finite F a joint marginal whose mass on the product $\prod_{I \in F} \pi_I(K_\varepsilon)$ is at least $1 - \varepsilon$ (indeed, by hypothesis $\sup_I \mu_I(\pi_I(K_\varepsilon)) \geq 1 - \varepsilon$ and consistency gives the same lower bound for joint marginals), hence the cylinder event has

Π-mass $\geq 1-\varepsilon$. As a result

$$\Pi(\iota(K_\varepsilon)) \geq 1-\varepsilon. \tag{1.8}$$

Because for each $\varepsilon > 0$ we have exhibited a compact set $\iota(K_\varepsilon) \subset X$ capturing Π-mass at least $1-\varepsilon$, the family $\{\Pi\}$ is tight in X. By Prokhorov's theorem (applied to the Polish metrizable subspace generated by these compact sets or to the projective-limit setting) the cylinder measure Π extends uniquely to a Borel probability measure on X (i.e. to the full Borel sigma-algebra of the product topology) which we again denote by Π; the extension coincides with the original cylinder measure on cylinder sets.

Next we show Π is carried by the image $\iota(E) \subset X$. Let

$$Y := \iota(E) \;=\; \{(\pi_I(x))_{I\in\mathscr{I}} : x \in E\} \subset X. \tag{1.9}$$

Note that Y is exactly the set of all product points whose finite-coordinate values are compatible with evaluation at a single $x \in E$: compatibility means for every pair of indices I, J with $I \subset J$ the coordinate in I' equals the projection to I' of the coordinate in J'. The cylinder measure Π was constructed from consistent finite-dimensional marginals and thus assigns measure one to the set of sequences that are compatible under those projection relations. Concretely, for a fixed finite collection of indices $I_1, \ldots, I_m$, the set

$$C_{I_1,\ldots,I_m} := \{x \in X : \text{ the coordinates satisfy all linear consistency relations among } I_1, \ldots, I_m\} \tag{1.10}$$

has Π-measure 1. Intersecting over an increasing net of finite collections (directed by inclusion) we obtain that Π-almost every point satisfies all consistency relations, hence lies in Y. Thus

$$\Pi(Y) = 1. \tag{1.11}$$

C. *Pushforward to E, Radon property, and uniqueness.*

Since $\iota : E \to Y$ is a measurable bijection onto Y with measurable inverse (this is because the coordinate functionals separate points in E, and on Y the inverse is given by reconstructing x from its coordinates $(\ell(x))_{\ell\in I}$), we can define a probability measure μ on E by pushing Π forward along ι^{-1}:

$$\mu := (\iota^{-1})_*(\Pi|_Y). \tag{1.12}$$

Equivalently, $\mu(A) = \Pi(\iota(A))$ for Borel sets $A \subset E$. By construction, for each finite-dimensional I the pushforward $(\pi_I)_*\mu$ coincides with μ_I: indeed, for measurable $B \subset I'$,

$$(\pi_I)_*\mu(B) = \mu(\{x \in E : \pi_I(x) \in B\}) = \Pi(\{y \in Y : \mathrm{pr}_I(y) \in B\}) = \mu_I(B), \tag{1.13}$$

where the last equality is the defining finite-dimensional marginal property of Π.

Next we check that μ is a Radon measure on E. Fix $\varepsilon > 0$ and choose $K_\varepsilon \subset E$ compact satisfying $\sup_I \mu_I(\pi_I(K_\varepsilon)) \geq 1-\varepsilon$. Then, as shown in part (B),

$$\Pi(\iota(K_\varepsilon)) \geq 1-\varepsilon. \tag{1.14}$$

Since $\iota(K_\varepsilon)$ is compact in X and ι is a homeomorphism onto its image (with K_ε compact in E), it follows that K_ε is compact in E and

$$\mu(K_\varepsilon) = \Pi(\iota(K_\varepsilon)) \geq 1 - \varepsilon \tag{1.15}$$

This proves tightness (inner regularity) of μ, and together with outer regularity (Borel measures on metric spaces are outer regular) this shows μ is Radon.

Finally uniqueness: if ν is any probability measure on $(E, \mathcal{B}(E))$ whose finite-dimensional projections coincide with $\{\mu_I\}$, then pushing ν forward by ι yields a probability measure on X agreeing with Π on all cylinder sets; the cylinder sigma-algebra generates the Borel sigma-algebra on X restricted to the (support) set Y on which Π concentrates, hence the pushforward measures coincide; consequently $\nu = \mu$.

This completes the construction of μ, its verification of the projection property, the Radon property, and uniqueness. □

1.4 Gaussian measures on separable Banach spaces

We now define Gaussian measures in infinite dimensions and state Minlos' theorem which guarantees existence under the continuity of the characteristic functional.

Definition 1.3 (Gaussian measure)**.** Let E be a separable Banach space. A probability measure μ on $(E, \mathcal{B}(E))$ is called *Gaussian* if for every $\ell \in E^*$ the pushforward measure $\ell_\# \mu$ on $\mathbb{R}$ is a (finite-dimensional) Gaussian distribution.

Equivalently, μ is Gaussian iff the characteristic functional $\Phi(\ell) := \int_E e^{i\ell(x)}\, \mu(dx)$ is of the form

$$\Phi(\ell) = \exp\left(im(\ell) - \tfrac{1}{2}C(\ell, \ell)\right) \tag{1.16}$$

where $m : E^* \to \mathbb{R}$ is linear and $C : E^* \times E^* \to \mathbb{R}$ is a positive semidefinite symmetric bilinear form (the covariance form).

Theorem 1.4 (Minlos' theorem)**.** *Let E be a real nuclear space and let $\Phi : E^* \to \mathbb{C}$ be a continuous positive-definite functional with $\Phi(0) = 1$. Then there exists a unique Radon probability measure μ on E with characteristic functional Φ.*

Remark 1.5. The hypothesis that E be nuclear may be weakened in many practical settings (e.g. when E is a separable Hilbert space with trace-class covariance), but we state the theorem in its classical form. In the Banach-space setting one often uses the Lévy–Minlos machinery or constructs the measure by projective limits as in Theorem 1.2.

Proof. We give a fully rigorous construction based on the cylindrical measure approach, the projective-limit construction, and Prokhorov tightness adapted to nuclear spaces.

Preliminaries: Let E be a real nuclear space. By definition, E is a countably Hilbert nuclear space: there exists a decreasing family of Hilbertian seminorms $\|\cdot\|_n$ whose completions H_n satisfy that the inclusion maps $i_{n+1,n} : H_{n+1} \hookrightarrow H_n$ are Hilbert–Schmidt.

Denote $E^* = \bigcup_n H_n$ as the strict inductive limit, identifying H_n with its dual H_n' using the Riesz representation. A functional $\Phi : E^* \to \mathbb{C}$ is *positive-definite* if for all $m \in \mathbb{N}$, all $\xi_1, \ldots, \xi_m \in E^*$, and all $z_1, \ldots, z_m \in \mathbb{C}$,

$$\sum_{i,j=1}^{m} \Phi(\xi_i - \xi_j) z_i \overline{z_j} \geq 0. \tag{1.17}$$

Continuity means continuity in the inductive-limit topology on E^*; in practice, this means continuity on each H_n. Define for each n the restriction

$$\Phi_n := \Phi\big|_{H_n} : H_n \to \mathbb{C}. \tag{1.18}$$

Finite-dimensional distributions and consistency: For each n and each finite-dimensional subspace $F \subset H_n$, define a complex-valued function $\Phi_{n,F}$ on F by restriction.

By Bochner's theorem on finite-dimensional Hilbert spaces, $\Phi_{n,F}$ is a characteristic function of a unique probability measure $\mu_{n,F}$ on F. Now suppose $F_1 \subset F_2 \subset H_n$ are finite-dimensional. Then Φ_{n,F_2} restricted to F_1 equals Φ_{n,F_1}. Hence the corresponding measures satisfy

$$(\pi_{F_2 \to F_1})_* \mu_{n,F_2} = \mu_{n,F_1}. \tag{1.19}$$

Thus we have a *projectively consistent family* of finite-dimensional distributions on each H_n.

Step 2: Construction of a cylindrical measure on H_n: Define a cylindrical measure μ_n on H_n by setting for each finite-dimensional $F \subset H_n$,

$$\mu_n \circ \pi_F^{-1} = \mu_{n,F}. \tag{1.20}$$

This is well defined by consistency. At this stage μ_n need not be σ-additive; it is defined only on the algebra of cylinder sets.

Tightness on H_n using nuclearity: The key point: the nuclearity of E implies that Φ is *Radonifying*. More precisely, continuity of Φ on H_{n+1} implies the existence of a neighborhood U in H_{n+1} such that $\Phi(\xi) \approx 1$ for $\xi \in U$, and the Hilbert–Schmidt embedding $i_{n+1,n}$ ensures that characteristic functionals of the form Φ_n correspond to measures supported essentially in an H_n-compact set.

Concretely, for each n and $\varepsilon > 0$, choose a Hilbert–Schmidt embedding $i_{n+1,n}$ with singular values (s_k), and pick N so large that $\sum_{k>N} s_k^2 < \varepsilon$. Let

$$K_{n,\varepsilon} = \left\{ x \in H_n : \sum_{k>N} |\langle x, e_k \rangle_{H_n}|^2 \leq \varepsilon \right\}, \tag{1.21}$$

a compact ellipsoid in H_n. One shows (via Levy concentration estimates using the characteristic function Φ_n) that

$$\mu_{n,F}(K_{n,\varepsilon}^c) < \varepsilon \tag{1.22}$$

for all sufficiently large finite-dimensional F. Hence the cylindrical measure μ_n is tight.

Extension to a Radon probability measure on H_n: By Prokhorov's theorem (valid in separable complete metric spaces), tight cylindrical measures extend uniquely to Radon probability measures. Thus μ_n extends to a Radon measure on H_n.

Projective limit consistency between H_{n+1} and H_n: Because Φ_{n+1} restricted to H_n equals Φ_n, the measures μ_{n+1} and μ_n satisfy

$$(i_{n+1,n})_*\mu_{n+1} = \mu_n. \tag{1.23}$$

Thus the family (μ_n) is projectively consistent.

The projective limit measure on E: The space $E = \bigcap_n H_n$ carries the projective-limit topology. By standard results on projective families of Radon measures on countably Hilbert spaces (see e.g. Bogachev *Measure Theory*, Vol. 2), the projective system (μ_n) defines a unique Radon measure μ on E satisfying

$$(i_n)_*\mu = \mu_n, \qquad E \hookrightarrow H_n \text{ the canonical embeddings.} \tag{1.24}$$

Verification of the characteristic functional: For any $\xi \in E^*$, choose n so that $\xi \in H_n$. Then, using the definition of μ_n and standard Fourier inversion on Hilbert spaces,

$$\int_E e^{i\langle x,\xi\rangle}\, d\mu(x) = \int_{H_n} e^{i\langle x,\xi\rangle}\, d\mu_n(x) = \Phi_n(\xi) = \Phi(\xi). \tag{1.25}$$

Uniqueness: If ν is another Radon measure on E with characteristic functional Φ, its push-forward to each H_n has characteristic functional Φ_n, hence equals μ_n by uniqueness in Hilbert spaces. Thus $\nu = \mu$.

Conclusion: There exists a unique Radon probability measure μ on E with characteristic functional Φ. This completes the proof. □

1.4.1 Covariance operators and reproducing kernel Hilbert spaces

Let H be a separable real Hilbert space and consider a centered Gaussian measure μ on H (i.e. $m \equiv 0$). The covariance of μ is the bounded, self-adjoint, positive operator $Q : H \to H$ defined by

$$\langle Qu, v\rangle_H = \int_H \langle x, u\rangle_H \langle x, v\rangle_H \,\mu(dx), \qquad u, v \in H. \tag{1.26}$$

If Q is trace-class then μ is a Radon measure on H and many useful constructions follow.

Definition 1.6 (Reproducing kernel Hilbert space (RKHS) or Cameron–Martin space)**.** Let μ be a centered Gaussian measure on a separable Banach space E. The Cameron–Martin space $H \subset E$ is the closure of $\{Q\ell : \ell \in E^*\}$ with the inner product induced by the covariance form. In the Hilbert-space case this is the range of $Q^{1/2}$ endowed with the inner product

$$\langle Q^{1/2}u, Q^{1/2}v\rangle_H = \langle u, v\rangle_{\overline{\mathrm{Ran}(Q^{1/2})}} \tag{1.27}$$

Proposition 1.7. *If μ is a centered Gaussian measure on a separable Banach space E with covariance operator Q of trace-class (when interpreted in a suitable Hilbertian embedding) then the Cameron–Martin space H is a separable Hilbert space continuously embedded in E.*

Proof. See classical references (Bogachev, Vakhania et al.). The key point is that trace-classness yields that the quadratic form associated to Q defines a Hilbertian norm on the linear span of $Q(E^*)$, and completeness and separability follow from separability of E. □

1.5 The Wiener measure on $C([0,T];\mathbb{R}^d)$

We construct the standard Wiener measure $\mathbb{W}_0^T$ on $C([0,T];\mathbb{R}^d)$ using finite-dimensional Gaussian marginals together with Kolmogorov's extension theorem. We then prove its basic regularity properties and compute its Cameron–Martin space.

1.5.1 Finite-dimensional distributions and the Kolmogorov extension

For $0 = t_0 < t_1 < \cdots < t_n = T$ let

$$\mu_{t_1,\ldots,t_n}(dx_1,\ldots,dx_n) = \prod_{k=1}^{n} (2\pi(t_k - t_{k-1}))^{-d/2} \exp\Big(-\frac{|x_k - x_{k-1}|^2}{2(t_k - t_{k-1})}\Big)\, dx_k \tag{1.28}$$

be the usual heat-kernel Gaussian transition densities with $x_0 = 0$ fixed. These families are consistent and therefore by Theorem 1.2 there exists a unique probability measure $\mathbb{W}_0^T$ on $(\mathbb{R}^{[0,T]}, \mathscr{C})$ whose finite-dimensional marginals are $\mu_{t_1,\ldots,t_n}$. To obtain a measure supported on continuous paths we appeal to Kolmogorov's continuity theorem.

Theorem 1.8 (Kolmogorov continuity theorem)**.** *Let $(X_t)_{t\in[0,T]}$ be a stochastic process with values in $\mathbb{R}^d$ such that for some $\alpha > 0$, $\beta > 1$ and finite constant C one has*

$$\mathbb{E}|X_t - X_s|^\beta \le C|t-s|^{1+\alpha} \qquad \forall s,t \in [0,T].$$

Then there exists a continuous modification of X and the law of this continuous modification is supported on $C([0,T];\mathbb{R}^d)$.

Proof. Fix any exponent γ with

$$0 < \gamma < \frac{\alpha}{\beta}. \tag{1.29}$$

We will construct a modification $\widetilde{X}$ of X which is almost surely Hölder continuous of exponent γ; in particular $\widetilde{X}$ is continuous and its law is supported on $C([0,T];\mathbb{R}^d)$.

Step 1 (dyadic partition estimates). For $n \in \mathbb{N}$ consider the dyadic partition of $[0,T]$ with mesh size $2^{-n}T$, i.e. points

$$t_k^{(n)} := k2^{-n}T, \qquad k = 0,1,\ldots,2^n. \tag{1.30}$$

For each dyadic subinterval $I_k^{(n)} = [t_{k-1}^{(n)}, t_k^{(n)}]$ consider the increment

$$\Delta_k^{(n)} := X_{t_k^{(n)}} - X_{t_{k-1}^{(n)}} \tag{1.31}$$

By the moment assumption (with $|t_k^{(n)} - t_{k-1}^{(n)}| = 2^{-n}T$) and Markov's inequality we have for any $\lambda > 0$,

$$\mathbb{P}(|\Delta_k^{(n)}| > \lambda) \leq \frac{\mathbb{E}|\Delta_k^{(n)}|^\beta}{\lambda^\beta} \leq \frac{C(2^{-n}T)^{1+\alpha}}{\lambda^\beta}. \tag{1.32}$$

Choose $\lambda = 2^{-n\gamma}$. Then

$$\mathbb{P}(|\Delta_k^{(n)}| > 2^{-n\gamma}) \leq C(2^{-n}T)^{1+\alpha}\, 2^{n\beta\gamma} = CT^{1+\alpha}\, 2^{-n(1+\alpha-\beta\gamma)}. \tag{1.33}$$

Since $\gamma < \alpha/\beta$ we have $1 + \alpha - \beta\gamma > 1$, hence in particular $1 + \alpha - \beta\gamma > 0$.

Now use a union bound over the 2^n dyadic intervals to obtain

$$\begin{aligned}\mathbb{P}\Big(\exists\, k \in \{1,\dots,2^n\}:\ |\Delta_k^{(n)}| > 2^{-n\gamma}\Big) &\leq \sum_{k=1}^{2^n} \mathbb{P}(|\Delta_k^{(n)}| > 2^{-n\gamma}) \\ &\leq 2^n\, CT^{1+\alpha}\, 2^{-n(1+\alpha-\beta\gamma)} = CT^{1+\alpha}\, 2^{-n(\alpha-\beta\gamma)}.\end{aligned}$$

Because $\alpha - \beta\gamma > 0$ the right-hand side is summable in n, i.e.

$$\sum_{n=1}^{\infty} \mathbb{P}\Big(\exists\, k:\ |\Delta_k^{(n)}| > 2^{-n\gamma}\Big) < \infty \tag{1.34}$$

Step 2 (Borel–Cantelli and control on dyadic increments). By the Borel–Cantelli lemma, with probability 1 there exists a random index $N(\omega) < \infty$ such that for all $n \geq N(\omega)$ and all $k \in \{1,\dots,2^n\}$,

$$|\Delta_k^{(n)}(\omega)| \leq 2^{-n\gamma} \tag{1.35}$$

Thus almost surely the dyadic increments at sufficiently fine levels are uniformly bounded by $2^{-n\gamma}$.

Step 3 (Hölder control on dyadic rationals). Let $\mathbb{D} := \{k2^{-n}T:\ n \in \mathbb{N},\ k = 0,\dots,2^n\}$ denote the set of dyadic rationals in $[0,T]$. Fix ω in the event of probability 1 described in Step 2, and choose the corresponding $N(\omega)$. For any two dyadic rationals $s < t$ with common dyadic level $m \geq N(\omega)$ we can write the increment $X_t(\omega) - X_s(\omega)$ as a sum of dyadic increments at level m and obtain (triangle inequality)

$$|X_t(\omega) - X_s(\omega)| \leq \sum_j |\Delta_j^{(m)}(\omega)| \leq \#\{\text{intervals between } s \text{ and } t\} \cdot 2^{-m\gamma} \tag{1.36}$$

The number of such intervals is at most $(t-s)2^m/T + 1$, whence

$$|X_t(\omega) - X_s(\omega)| \leq \big((t-s)2^m/T + 1\big)\, 2^{-m\gamma} \tag{1.37}$$

Choosing m so large that $2^{-m} \leq (t-s)$ (possible since m can be arbitrarily large), we get the bound (for $m \geq N(\omega)$)

$$|X_t(\omega) - X_s(\omega)| \leq C'(\omega)\, (t-s)^\gamma \tag{1.38}$$

where $C'(\omega)$ is a finite random constant depending only on ω and T. A standard combinatorial refinement of this argument yields that for all dyadic rationals $s < t$ (not necessarily at the same level),

$$|X_t(\omega) - X_s(\omega)| \le M(\omega)\,|t-s|^\gamma \tag{1.39}$$

for some a.s. finite random variable $M(\omega)$. Thus on the set of full probability determined above, the mapping $t \mapsto X_t(\omega)$ is uniformly Hölder continuous of exponent γ on the dense set $\mathbb{D}$.

Step 4 (extension to a continuous process). Fix ω in the event of probability 1 above. The function $t \mapsto X_t(\omega)$ is (uniformly) Hölder continuous on the dense set $\mathbb{D}$ and thus extends uniquely to a continuous function $\widetilde{X}(\cdot,\omega)$ on $[0,T]$ by continuity. Define the modification $\widetilde{X}_t(\omega)$ to be this continuous extension for ω in the probability 1 event, and define $\widetilde{X}_t(\omega)$ arbitrarily on the null complement. Then $\widetilde{X}$ is a measurable process (standard arguments using separability of $[0,T]$ and measurability on rationals) and $\widetilde{X}(\cdot,\omega)$ is Hölder continuous of exponent γ for almost every ω; in particular $\widetilde{X}$ is continuous almost surely.

Step 5 (modification property: agreement at fixed times). It remains to check that $\widetilde{X}$ is a modification of X, i.e. for each fixed $t \in [0,T]$ we have $\widetilde{X}_t = X_t$ almost surely. Fix $t \in [0,T]$ and choose a sequence of dyadic rationals $(q_n)_{n\ge1}$ with $q_n \downarrow t$ (or $q_n \to t$). By the moment bound and dominated convergence one has

$$\mathbb{E}|X_{q_n} - X_t|^\beta \le C|q_n - t|^{1+\alpha} \xrightarrow[n\to\infty]{} 0 \tag{1.40}$$

hence $X_{q_n} \to X_t$ in L^β and therefore in probability. On the other hand, for ω in the full-probability event used above the values $X_{q_n}(\omega)$ converge (by construction) to $\widetilde{X}_t(\omega)$. Thus there is almost sure subsequential convergence of X_{q_n} to $\widetilde{X}_t$, and by Fatou's lemma applied to the nonnegative random variables $|X_t - \widetilde{X}_t|^\beta$ together with the L^β-convergence of X_{q_n} to X_t we obtain

$$\mathbb{E}|X_t - \widetilde{X}_t|^\beta \le \liminf_{n\to\infty} \mathbb{E}|X_t - X_{q_n}|^\beta = 0 \tag{1.41}$$

Therefore $|X_t - \widetilde{X}_t| = 0$ almost surely, i.e. $\widetilde{X}_t = X_t$ with probability one. Since t was arbitrary, $\widetilde{X}$ is a modification of X.

Step 6 (support of the law). By construction $\widetilde{X}(\cdot,\omega) \in C([0,T];\mathbb{R}^d)$ for almost every ω, and indeed $\widetilde{X}$ is Hölder continuous of exponent γ a.s. Thus the law of the process $\widetilde{X}$ (as a probability measure on the path space with the topology of uniform convergence) is supported on $C([0,T];\mathbb{R}^d)$.

This completes the proof: for every $0 < \gamma < \alpha/\beta$ there exists a continuous modification $\widetilde{X}$ of X which is a.s. Hölder continuous of exponent γ, and consequently the law of the continuous modification is supported on $C([0,T];\mathbb{R}^d)$. □

Applied to the canonical process with Gaussian marginals above we obtain the existence of a measure $\mathbb{W}_0^T$ concentrated on continuous paths. We call $\mathbb{W}_0^T$ the *Wiener measure* (with $x(0) = 0$ a.s.).

1.5.2 Cameron–Martin theorem for Wiener measure

We compute the Cameron–Martin space of Wiener measure and prove the classical Cameron–Martin shift formula. Let

$$E = C([0,T];\mathbb{R}^d), \qquad H = \{h \in E : h(0) = 0,\ h \text{ is absolutely continuous},\ \dot h \in L^2([0,T];\mathbb{R}^d)\},$$

with inner product

$$\langle h_1, h_2\rangle_H = \int_0^T \langle \dot h_1(t), \dot h_2(t)\rangle_{\mathbb{R}^d}\, dt.$$

Theorem 1.9 (Cameron–Martin). *Let $\mathbb{W}_0^T$ be Wiener measure on $E = C([0,T];\mathbb{R}^d)$ (canonical process $X_t(\omega) = \omega(t)$, $\omega(0) = 0$ a.s.) and let*

$$H = \{h \in C([0,T];\mathbb{R}^d) : h(0) = 0,\ h \text{ absolutely continuous},\ \dot h \in L^2([0,T];\mathbb{R}^d)\} \tag{1.42}$$

be the Cameron–Martin space equipped with $\langle h_1, h_2\rangle_H = \int_0^T \langle \dot h_1, \dot h_2\rangle\, dt$ and norm $\|h\|_H^2 = \int_0^T |\dot h(t)|^2\, dt$. For $h \in H$ define the translated measure $\mathbb{W}_h(A) = \mathbb{W}_0^T(A - h)$. Then $\mathbb{W}_h \ll \mathbb{W}_0^T$ and the Radon–Nikodym derivative is

$$\frac{d\mathbb{W}_h}{d\mathbb{W}_0^T}(\omega) = \exp\Big(\int_0^T \langle \dot h(t),\, d\omega(t)\rangle - \tfrac{1}{2}\|h\|_H^2\Big), \tag{1.43}$$

where the stochastic integral is the Itô integral (or Wiener integral) of the deterministic function $\dot h$ against the canonical Brownian motion, and the equality holds $\mathbb{W}_0^T$–almost surely.

Proof. We prove the theorem by approximation of h by piecewise linear paths and by passing to the limit of the finite–dimensional Radon–Nikodym derivatives. Throughout the proof let $(\mathcal{F}_t)_{t\in[0,T]}$ denote the canonical filtration generated by the coordinate maps $X_s(\omega) = \omega(s)$ and let $\mathcal{F} := \mathcal{F}_T$.

Step 0 (notation and finite-dimensional marginals). Fix a sequence of partitions

$$\mathcal{P}_n :\ 0 = t_0^{(n)} < t_1^{(n)} < \cdots < t_{k_n}^{(n)} = T \tag{1.44}$$

whose mesh tends to 0 as $n \to \infty$. For brevity denote $\Delta t_k^{(n)} = t_k^{(n)} - t_{k-1}^{(n)}$ and for $\omega \in E$ write the increments

$$\Delta X_k^{(n)}(\omega) := X_{t_k^{(n)}}(\omega) - X_{t_{k-1}^{(n)}}(\omega) \tag{1.45}$$

Under Wiener measure the vector

$$(\Delta X_1^{(n)}, \dots, \Delta X_{k_n}^{(n)}) \tag{1.46}$$

is a Gaussian random vector with independent components and $\Delta X_k^{(n)} \sim N(0, \Delta t_k^{(n)} I_d)$ (matrix covariance). Consequently Wiener measure restricted to the sigma algebra

$$\mathcal{G}_n := \sigma(X_{t_0^{(n)}}, \dots, X_{t_{k_n}^{(n)}}) \tag{1.47}$$

equals the product Gaussian law on $\mathbb{R}^{dk_n}$ with density

$$(2\pi)^{-\frac{dk_n}{2}} \prod_{k=1}^{k_n} (\Delta t_k^{(n)})^{-d/2} \exp\Big(-\frac{|\Delta x_k|^2}{2\Delta t_k^{(n)}}\Big)\, d\Delta x_1 \cdots d\Delta x_{k_n}. \tag{1.48}$$

Step 1 (finite–dimensional shift formula). Fix n and define the piecewise linear approximation $h^{(n)}$ of h which interpolates $h(t_{k-1}^{(n)})$ and $h(t_k^{(n)})$ on each interval $[t_{k-1}^{(n)}, t_k^{(n)}]$. Equivalently set the increment

$$\Delta h_k^{(n)} := h(t_k^{(n)}) - h(t_{k-1}^{(n)}) \tag{1.49}$$

By elementary finite-dimensional Gaussian calculus (or direct computation of densities) the law of the shifted increments

$$\Delta X_k^{(n)} + \Delta h_k^{(n)} \tag{1.50}$$

is absolutely continuous with respect to the law of $\Delta X_k^{(n)}$ and the Radon–Nikodym derivative (density) on $\mathcal{G}_n$ is

$$M_n(\omega) := \exp\Big(\sum_{k=1}^{k_n} \frac{\langle \Delta h_k^{(n)}, \Delta X_k^{(n)}(\omega)\rangle}{\Delta t_k^{(n)}} - \tfrac{1}{2} \sum_{k=1}^{k_n} \frac{|\Delta h_k^{(n)}|^2}{\Delta t_k^{(n)}} \Big). \tag{1.51}$$

Concretely, this is the usual Gaussian shift formula: if $Z \sim N(0,\sigma^2 I)$ and $a \in \mathbb{R}^d$ then the law of $Z+a$ has density $\exp(\langle a,z\rangle/\sigma^2 - |a|^2/(2\sigma^2))$ relative to the law of Z. The product structure over independent increments yields (1.51). In particular M_n is $\mathcal{G}_n$–measurable and $M_n \geq 0$.

Observe that $\mathbb{E}_{\mathbb{W}_0^T}[M_n] = 1$ for each n (this is the explicit Gaussian integral giving unit expectation). Thus $(M_n, \mathcal{G}_n)$ is a nonnegative martingale indexed by n in the directed sense (i.e. if we refine partitions the densities are consistent — equivalently the family $\{M_n\}$ is a martingale with respect to the filtration generated by the nested partitions). Indeed if $\mathcal{G}_m \subset \mathcal{G}_n$ then

$$\mathbb{E}[M_n \mid \mathcal{G}_m] = M_m \tag{1.52}$$

reflecting the finite-dimensional consistency of Gaussian shift densities.

Step 2 (identification of the stochastic integral and limit). We next identify the limit of the sums appearing in (1.51) with the stochastic integral $\int_0^T \langle \dot h(t), dX_t\rangle$ and show convergence of the exponentials.

Because $h \in H$ is absolutely continuous with derivative $\dot h \in L^2$, the piecewise linear approximations satisfy

$$\sum_{k=1}^{k_n} \frac{|\Delta h_k^{(n)}|^2}{\Delta t_k^{(n)}} = \int_0^T |\dot h^{(n)}(t)|^2\,dt \longrightarrow \int_0^T |\dot h(t)|^2\,dt = \|h\|_H^2 \qquad (n\to\infty), \tag{1.53}$$

by standard properties of L^2 approximations of $\dot h$ by piecewise constant (or piecewise linear) functions.

For the martingale (stochastic) term write

$$S_n(\omega) := \sum_{k=1}^{k_n} \frac{\langle \Delta h_k^{(n)}, \Delta X_k^{(n)}(\omega)\rangle}{\Delta t_k^{(n)}}. \tag{1.54}$$

Because $\Delta X_k^{(n)}$ are independent Gaussian increments with variance $\Delta t_k^{(n)}$, one checks via the Itô isometry (or finite-dimensional covariance computations) that

$$\mathbb{E}\big[(S_n - S_m)^2\big] = \mathbb{E}\Big[\sum_{k\in(\text{symmetric difference})} \frac{|\Delta h_k^{(\cdot)}|^2}{\Delta t_k^{(\cdot)}} \Big] = \int_0^T |\dot h^{(n)}(t) - \dot h^{(m)}(t)|^2\,dt, \tag{1.55}$$

so (S_n) is a Cauchy sequence in $L^2(\mathbb{W}_0^T)$ and therefore converges in L^2 (hence in probability and along a subsequence almost surely) to an L^2 limit which by definition is precisely the Wiener integral $\int_0^T \langle \dot h(t),\, dX_t\rangle$. (Equivalently, interpreting $\dot h$ as a deterministic square-integrable integrand one constructs the stochastic integral as the L^2-limit of Riemann sums; see Itô isometry.) Thus

$$S_n \xrightarrow[n\to\infty]{L^2} \int_0^T \langle \dot h(t),\, dX_t\rangle. \tag{1.56}$$

Combining the two convergences we obtain

$$\sum_{k=1}^{k_n} \frac{\langle \Delta h_k^{(n)}, \Delta X_k^{(n)}\rangle}{\Delta t_k^{(n)}} - \tfrac{1}{2}\sum_{k=1}^{k_n} \frac{|\Delta h_k^{(n)}|^2}{\Delta t_k^{(n)}} \tag{1.57}$$

converges in L^1 (indeed in L^2) to

$$\int_0^T \langle \dot h(t),\, dX_t\rangle - \tfrac{1}{2}\|h\|_H^2. \tag{1.58}$$

Consequently the sequence of random variables M_n defined by (1.51) converges in probability to

$$M := \exp\Big(\int_0^T \langle \dot h(t),\, dX_t\rangle - \tfrac{1}{2}\|h\|_H^2\Big). \tag{1.59}$$

Step 3 (martingale convergence and L^1 limit). The family $(M_n, \mathcal{G}_n)$ is nonnegative and satisfies $\mathbb{E}[M_n] = 1$ for all n. By the martingale convergence theorem for nonnegative martingales (Doob) there exists an a.s. finite random variable M_∞ such that $M_n \to M_\infty$ almost surely and moreover

$$\mathbb{E}[M_\infty] = \lim_{n\to\infty} \mathbb{E}[M_n] = 1$$

Because $M_n \to M$ in probability and $M_n \to M_\infty$ a.s., we must have $M_\infty = M$ almost surely; hence $M_n \to M$ almost surely and also in L^1 (the nonnegativity plus constant expectation gives uniform integrability and therefore L^1-convergence). In particular $\mathbb{E}[M] = 1$ and M is integrable.

Step 4 (identification as Radon–Nikodym derivative). For each n and each $\mathcal{G}_n$–measurable bounded function φ we have by construction

$$\mathbb{E}_{\mathbb{W}_h^{(n)}}[\varphi] := \int \varphi(\omega)\, \mathbb{W}_h^{(n)}(d\omega) = \mathbb{E}_{\mathbb{W}_0^T}[\varphi(\omega)\, M_n(\omega)], \tag{1.60}$$

where $\mathbb{W}_h^{(n)}$ denotes the law of the piecewise linear shift $\omega \mapsto \omega + h^{(n)}$ on the sigma algebra $\mathcal{G}_n$ (in fact $\mathbb{W}_h^{(n)} = (\tau_{h^{(n)}})_* \mathbb{W}_0^T$ with $\tau_{h^{(n)}}$ the translation by $h^{(n)}$). Passing to the limit along n we may approximate any bounded continuous functional φ measurable w.r.t. the full sigma-algebra by $\mathcal{G}_n$–measurable functions and use L^1 convergence of M_n to M to deduce

$$\int \varphi(\omega)\, \mathbb{W}_h(d\omega) = \mathbb{E}_{\mathbb{W}_0^T}[\varphi M]. \tag{1.61}$$

Therefore $\mathbb{W}_h \ll \mathbb{W}_0^T$ and M is a version of the Radon–Nikodym derivative $d\mathbb{W}_h/d\mathbb{W}_0^T$. Since M equals the exponential in (1.43) this proves the formula.

Step 5 (consequences and singularity outside H). The above construction shows existence of the RN derivative for each $h \in H$ and establishes absolute continuity. Standard complementary facts (which can be proved by similar finite-dimensional considerations) assert that if $h \notin H$ then $\mathbb{W}_h$ and $\mathbb{W}_0^T$ are mutually singular. (Sketch: if $h \notin H$ the quadratic variation $\sum |\Delta h_k^{(n)}|^2/\Delta t_k^{(n)}$ diverges and the finite-dimensional density oscillates to zero in the limit producing singularity; see standard references.)

Combining the above steps we obtain the asserted formula (1.43) and the absolute continuity statement, which completes the proof. □

1.6 Fernique's theorem and integrability

A fundamental property of Gaussian measures in infinite dimensions is the existence of exponential moments of the squared norm; this is Fernique's theorem which plays a decisive role in establishing tightness and integrability of nonlinear functionals.

Theorem 1.10 (Fernique — full treatment with Borell–TIS and Fernique's original proof)**.** *Let μ be a (centered) Gaussian probability measure on a separable Banach space E. Let*

$$\sigma^2 := \sup_{\ell \in B_{E^*}} \operatorname{Var}_\mu(\ell(X)) \tag{1.62}$$

where B_{E^*} is the closed unit ball of E^*. Then both of the following hold.

(a) (Proof via Borell–TIS / Gaussian isoperimetry.) There exists $m \in \mathbb{R}$ (the mean of the supremum) such that for every $u > 0$

$$\mu(\|X\|_E \geq m + u) \leq \exp\Big(-\frac{u^2}{2\sigma^2}\Big). \tag{1.63}$$

Consequently for every $0 < \beta < \frac{1}{2\sigma^2}$ one has

$$\int_E e^{\beta\|x\|_E^2}\, \mu(dx) < \infty. \tag{1.64}$$

In particular any α with $0 < \alpha < \frac{1}{2\sigma^2}$ is admissible.

(b) (Fernique's original doubling argument.) There exist explicit constants $c, C > 0$ depending only on σ^2 (and on the median/mean of $\|X\|_E$) such that for all r large enough

$$\mu(\|X\|_E > r) \leq C \exp\Big(-cr^2\Big). \tag{1.65}$$

Consequently there exists $\alpha' > 0$ (explicitly computable from c) with

$$\int_E e^{\alpha'\|x\|_E^2}\, \mu(dx) < \infty. \tag{1.66}$$

Proof. We give full proofs of (a) and (b) and then explain the explicit relation between the constants and σ^2.

Part (a) — proof of Borell–TIS and deduction of Fernique.

1. Representation of the Gaussian process. Let X denote the canonical E-valued Gaussian random element with law μ. For $\ell \in E^*$ write $X_\ell := \ell(X)$. Consider the centered Gaussian process $\{X_\ell : \ell \in B_{E^*}\}$ indexed by the compact metric space B_{E^*} endowed with the weak*-topology (which is metrizable on the unit ball because E is separable). The identity $\|x\|_E = \sup_{\ell \in B_{E^*}} \ell(x)$ implies that the quantity of interest is the supremum

$$S := \sup_{\ell \in B_{E^*}} X_\ell = \|X\|_E. \tag{1.67}$$

2. Gaussian isoperimetric inequality ⇒ Lipschitz concentration. We recall the Gaussian isoperimetric inequality in the following form (standard references: Ledoux–Talagrand; Borell; Sudakov–Tsirelson).

Lemma 1.11 (Gaussian isoperimetric inequality — formulation)**.** *Let γ_n denote the standard Gaussian measure on $\mathbb{R}^n$. For any Borel set $A \subset \mathbb{R}^n$ and any $t > 0$ define the Euclidean t-neighborhood $A_t := \{x \in \mathbb{R}^n : \operatorname{dist}(x, A) \le t\}$. Then for every $t \ge 0$*

$$\gamma_n(A_t) \ \ge\ \Phi(\Phi^{-1}(\gamma_n(A)) + t), \tag{1.68}$$

where Φ is the standard normal distribution function.

A standard corollary (obtained by applying the isoperimetric inequality to sublevel sets of a Lipschitz function) is the following concentration bound.

Corollary 1.12 (Gaussian concentration for Lipschitz functions)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be L-Lipschitz with respect to the Euclidean norm. Let m_f denote a median of f under γ_n. Then for every $u > 0$,*

$$\gamma_n(f - m_f \ge u) \le \exp\Big(-\frac{u^2}{2L^2}\Big). \tag{1.69}$$

Proof of Corollary. Let $A := \{x : f(x) \le m_f\}$. For $x \notin A$ put $d(x) := \operatorname{dist}(x, A)$. If f is L-Lipschitz then $f(x) - m_f \le Ld(x)$. Hence

$$\{f - m_f \ge u\} \subset \{d \ge u/L\} = \mathbb{R}^n \setminus A_{u/L}$$

The isoperimetric inequality with $t = u/L$ yields

$$\gamma_n(f - m_f \ge u) \le 1 - \Phi(\Phi^{-1}(\gamma_n(A)) + u/L). \tag{1.70}$$

Because m_f is a median we have $\gamma_n(A) \ge 1/2$ and thus $\Phi^{-1}(\gamma_n(A)) \ge 0$. The elementary Gaussian tail bound

$$1 - \Phi(x) \le \frac{1}{\sqrt{2\pi}} x^{-1} e^{-x^2/2} \tag{1.71}$$

and, more simply, the inequality $1 - \Phi(x) \le e^{-x^2/2}$ for $x \ge 0$ imply

$$\gamma_n(f - m_f \ge u) \le \exp\Big(-\frac{u^2}{2L^2}\Big), \tag{1.72}$$

which is the desired estimate. □

3. Application to the supremum process. To apply Corollary 1.12 we must represent our Gaussian process as a Lipschitz functional on some Gaussian Hilbert space. Concretely, by standard abstract Wiener space / covariance operator representation there exists a separable Hilbert space $\mathcal{H}$ (the Cameron–Martin space or reproducing kernel Hilbert space of μ) and a centered Gaussian random element G in $\mathcal{H}$ such that for each $\ell \in E^*$ there is a vector $a_\ell \in \mathcal{H}$ with

$$X_\ell \;=\; \langle G, a_\ell\rangle_{\mathcal{H}}, \qquad \mathrm{Var}(X_\ell) = \|a_\ell\|_{\mathcal{H}}^2. \tag{1.73}$$

The map $\ell \mapsto a_\ell$ is linear and continuous from (E^*, weak^*) into $\mathcal{H}$. The definition of σ^2 implies $\|a_\ell\|_{\mathcal{H}} \le \sigma$ for all $\ell \in B_{E^*}$.

Define the functional $F : \mathcal{H} \to \mathbb{R}$ by

$$F(h) := \sup_{\ell \in B_{E^*}} \langle h, a_\ell\rangle_{\mathcal{H}}. \tag{1.74}$$

Then F is σ-Lipschitz with respect to the Hilbert norm: indeed for $h_1, h_2 \in \mathcal{H}$,

$$|F(h_1) - F(h_2)| \le \sup_{\ell \in B_{E^*}} |\langle h_1 - h_2, a_\ell\rangle| \le \|h_1 - h_2\|_{\mathcal{H}} \sup_{\ell \in B_{E^*}} \|a_\ell\|_{\mathcal{H}} \le \sigma \|h_1 - h_2\|_{\mathcal{H}}. \tag{1.75}$$

Under the Gaussian law of G (which is a centered Gaussian measure on $\mathcal{H}$) the random variable $F(G)$ has the same law as

$$S = \sup_{\ell \in B_{E^*}} X_\ell = \|X\|_E \tag{1.76}$$

Applying Corollary 1.12 to the σ-Lipschitz function F on $\mathcal{H}$ yields (with m a median of S)

$$\mu(S - m \ge u) \le \exp\Big(-\frac{u^2}{2\sigma^2}\Big), \qquad \forall u > 0. \tag{1.77}$$

This is precisely the Borell–TIS inequality in the desired form (one may replace the median m by the expected value $\mathbb{E}[S]$ at the cost of adjusting constants slightly; the expectations and medians differ by at most a constant depending on σ).

4. Deduction of exponential integrability and explicit bound on α. Fix any $0 < \beta < \frac{1}{2\sigma^2}$. Using the tail bound above we have for t large

$$\mu(S \ge t) \le \exp\Big(-\frac{(t-m)^2}{2\sigma^2}\Big). \tag{1.78}$$

Write

$$\mathbb{E}[e^{\beta S^2}] \;=\; e^{\beta m^2}\,\mathbb{E}\big[e^{\beta(S^2 - m^2)}\big] = e^{\beta m^2}\Big(1 + \int_m^\infty 2\beta r\, e^{\beta(r^2 - m^2)}\,\mu(S > r)\,dr\Big). \tag{1.79}$$

Insert the tail bound and estimate the integrand for $r \ge m$ as

$$2\beta r\, e^{\beta(r^2 - m^2)}\,\exp\Big(-\frac{(r-m)^2}{2\sigma^2}\Big). \tag{1.80}$$

Completing the square in the exponent shows that for any $\beta < \frac{1}{2\sigma^2}$ the quadratic coefficient of r^2 in the exponent is negative, hence the integral is finite. Therefore

$$\int_E e^{\beta\|x\|_E^2}\,\mu(dx) = \mathbb{E}[e^{\beta S^2}] < \infty \tag{1.81}$$

for every $\beta \in (0, \frac{1}{2\sigma^2})$. Thus any $\alpha \in (0, 1/(2\sigma^2))$ is admissible, and we have obtained the explicit bound $\alpha < 1/(2\sigma^2)$.

Part (b) — Fernique's original doubling argument (self-contained).
Fernique's original proof establishes Gaussian tail decay and hence exponential integrability by an elementary convolution / symmetrization argument based on the identity in law $\frac{X+Y}{\sqrt{2}} \stackrel{d}{=} X$ for independent copies X, Y of the Gaussian. We present the argument in a complete and rigorous way and make constants explicit where possible.

1. Basic inequality arising from independence and scaling. Let X and Y be independent copies of the E-valued Gaussian with law μ. For any Borel sets $A, B \subset E$ one has

$$\{x \in E : \|x\|_E \le s\} + \{y \in E : \|y\|_E \le t\} \subset \{z \in E : \|z\|_E \le s + t\}, \tag{1.82}$$

where $+$ denotes Minkowski sum. Using the scaling property of Gaussian laws,

$$\mathcal{L}\Big(\frac{X+Y}{\sqrt{2}}\Big) \;=\; \mu, \tag{1.83}$$

and independence, we obtain for all $s, t > 0$,

$$\mu(\|X\|_E \le s) \cdot \mu(\|X\|_E \le t) = \mathbb{P}(\|X\|_E \le s,\ \|Y\|_E \le t) = \mathbb{P}(\tfrac{X+Y}{\sqrt{2}} \in \{\|z\|_E \le \tfrac{s+t}{\sqrt{2}}\}). \tag{1.84}$$

Thus

$$\mu(\|X\|_E \le s)\, \mu(\|X\|_E \le t) \le \mu(\|X\|_E \le \tfrac{s+t}{\sqrt{2}}). \tag{1.85}$$

Equivalently, with $F(r) := \mu(\|X\|_E \le r)$, inequality (1.85) reads

$$F(s)F(t) \le F\Big(\tfrac{s+t}{\sqrt{2}}\Big). \tag{1.86}$$

2. Iteration and supermultiplicativity structure. Set $s_0 > 0$ arbitrary and define recursively

$$s_{n+1} \;=\; \frac{s_n + s_0}{\sqrt{2}}, \qquad n \ge 0. \tag{1.87}$$

A short induction gives the closed-form

$$s_n \;=\; \frac{1}{\sqrt{2}^n}\, s_0 + \Big(1 - \frac{1}{\sqrt{2}^n}\Big) \frac{s_0}{\sqrt{2} - 1}, \tag{1.88}$$

so (s_n) is an increasing sequence with finite limit $s_\infty = \frac{s_0}{\sqrt{2}-1}$. Applying (1.85) with $s = s_n$, $t = s_0$ yields

$$F(s_n)F(s_0) \le F(s_{n+1}). \tag{1.89}$$

Iterating this inequality m times gives

$$F(s_{n+m}) \ge F(s_n)\, F(s_0)^m. \tag{1.90}$$

Fix s_0 such that $F(s_0) > \frac{1}{2}$ (this is possible since $\lim_{r\to\infty} F(r) = 1$); indeed pick s_0 with $F(s_0) =: p > 1/2$. Then $p > 1/2$ and the above iteration yields

$$F(s_{n+m}) \ge F(s_n)\, p^m. \tag{1.91}$$

Take logarithms and rearrange to obtain for all $n \geq 0$ and $m \geq 1$

$$-\log\big(1 - F(s_{n+m})\big) \leq -\log\big(1 - F(s_n)p^m\big). \tag{1.92}$$

Choose n large so that $F(s_n)$ is close to 1; more precisely pick n so that $F(s_n) \geq 1 - \delta$ for a small $\delta > 0$ to be fixed. Then

$$1 - F(s_{n+m}) \leq 1 - (1-\delta)p^m = 1 - p^m + \delta p^m \leq (1 - p^m) + \delta, \tag{1.93}$$

and for large m the dominant term is $1 - p^m$. Using the elementary bound $1 - p^m \leq e^{-m(1-p)}$ (valid for $0 < p < 1$) we conclude that for some constants $C_1, c_1 > 0$ depending only on p (and hence on s_0) and on δ we have

$$1 - F(s_{n+m}) \leq C_1 e^{-c_1 m}. \tag{1.94}$$

3. Conversion to a quadratic tail bound. The spacing between s_{n+m} and s_n grows like a constant times $1 - \frac{1}{\sqrt{2}^m}$, in particular there exists $c_2 > 0$ such that $s_{n+m} - s_n \geq c_2 m$ for all sufficiently large m. (This follows from the explicit formula for s_k above and the fact that each iteration adds a fixed positive fraction of the remaining gap.) Combining this with the previous exponential-in-m bound on $1 - F(s_{n+m})$ yields a tail bound of the form

$$\mu(\|X\|_E > r) \leq C_2 \exp(-c_3 r) \tag{1.95}$$

for all large r, with constants $C_2, c_3 > 0$ depending on s_0 (i.e. on the chosen p). We may now bootstrap the linear-exponential decay into a Gaussian-exponential decay as follows.

Using the same doubling identity repeatedly one can show a stronger inequality (see Fernique's original estimate) of the form

$$\mu(\|X\|_E > r)^2 \leq \mu(\|X\|_E > r/\sqrt{2}). \tag{1.96}$$

Iterating this inequality k times yields

$$\mu(\|X\|_E > r)^{2^k} \leq \mu(\|X\|_E > r/(\sqrt{2})^k). \tag{1.97}$$

Choose k such that $r/(\sqrt{2})^k$ lies in a fixed compact window (e.g. below some fixed R_0) where the probability is bounded strictly below 1. Then the right-hand side is bounded by a constant $q < 1$, so

$$\mu(\|X\|_E > r) \leq q^{1/2^k} = \exp\Big(-\frac{\log(1/q)}{2^k}\Big). \tag{1.98}$$

But 2^k is of order $(r)^2$ (since $k \approx \log r / \log\sqrt{2}$), and thus we obtain a Gaussian-type tail:

$$\mu(\|X\|_E > r) \leq C\exp\Big(-cr^2\Big) \tag{1.99}$$

for suitable $C, c > 0$.

This yields the desired quadratic-exponential tail bound. Integrating against $e^{\alpha r^2}$ now gives finite exponential moments for small enough $\alpha > 0$ (any $0 < \alpha < c$ works). Thus Fernique's original method produces some explicit $\alpha' > 0$ (computable from the chosen p and the constants above) for which $\int e^{\alpha' \|x\|_E^2} \, d\mu(x) < \infty$.

Explicit relation between constants. From part (a) we obtained the explicit, simple sufficient condition $\alpha < \frac{1}{2\sigma^2}$ for integrability. Part (b) gives, via the above doubling argument, a constructive (albeit more involved) constant $\alpha' > 0$ depending only on the choice of s_0 (equivalently on a fixed lower bound for $F(s_0) = \mu(\|X\| \le s_0)$), and ultimately on the covariance structure through σ^2. In practice the sharp and very useful bound is the one obtained from Borell–TIS:

$$\boxed{\text{Any } \alpha \text{ with } 0 < \alpha < \frac{1}{2\sigma^2} \text{ satisfies } \int_E e^{\alpha \|x\|_E^2} \, d\mu(x) < \infty.} \tag{1.100}$$

This completes the proofs of parts (a) and (b) and provides the requested explicit bound for the admissible constant α in terms of the covariance parameter σ^2. □

1.7 Girsanov theorem (change of measure)

Girsanov's theorem gives sufficient conditions under which one can change the drift of an Itô process by an absolutely continuous change of measure. We state the result in a form suitable for later use in supersymmetric representations.

Theorem 1.13 (Girsanov)**.** *Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ carry a d-dimensional Brownian motion W_t and let $\theta : \Omega \times [0, T] \to \mathbb{R}^d$ be progressively measurable with*

$$\mathbb{E} \exp\Big(\tfrac{1}{2} \int_0^T |\theta_s|^2 \, ds\Big) < \infty \tag{1.101}$$

Define the exponential martingale

$$M_T = \exp\Big(\int_0^T \langle \theta_s, dW_s \rangle - \tfrac{1}{2} \int_0^T |\theta_s|^2 \, ds\Big). \tag{1.102}$$

Then $\mathbb{Q}$ defined by $d\mathbb{Q} = M_T \, d\mathbb{P}$ is a probability measure on $(\Omega, \mathcal{F}_T)$ and the process $\widetilde{W}_t = W_t - \int_0^t \theta_s \, ds$ is a Brownian motion under $\mathbb{Q}$.

Proof. We give a complete, fully rigorous proof, structured in six steps. Throughout we assume $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ satisfies the usual conditions and W_t is a d-dimensional Brownian motion.

Step 1: $\{M_t\}_{t \in [0,T]}$ is a martingale and $M_t > 0$.

Define for each $t \le T$

$$M_t := \exp\Big(\int_0^t \langle \theta_s, dW_s \rangle - \frac{1}{2} \int_0^t |\theta_s|^2 \, ds\Big). \tag{1.103}$$

The stochastic exponential formula (Itô) gives

$$dM_t = M_t \langle \theta_t, dW_t \rangle, \qquad M_0 = 1. \tag{1.104}$$

Thus M_t is a local martingale. Since $M_t > 0$ everywhere, it is a supermartingale automatically.

To show it is a true martingale, note that the hypothesis

$$\mathbb{E}[\exp\Big(\tfrac{1}{2}\int_0^T |\theta_s|^2\,ds\Big)] < \infty \tag{1.105}$$

implies Novikov's condition:

$$\mathbb{E}[\exp\Big(\tfrac{1}{2}\int_0^T |\theta_s|^2\,ds\Big)] < \infty \quad \Longrightarrow \quad \mathbb{E}[M_T] = 1. \tag{1.106}$$

Therefore M_t is a uniformly integrable martingale and in particular $\mathbb{E}[M_T] = 1$.

Step 2: Define the new measure $\mathbb{Q}$.

Set

$$\frac{d\mathbb{Q}}{d\mathbb{P}}\Big|_{\mathcal{F}_T} = M_T. \tag{1.107}$$

Since $M_T \geq 0$ and $\mathbb{E}M_T = 1$, $\mathbb{Q}$ is a probability measure on $(\Omega, \mathcal{F}_T)$, absolutely continuous with respect to $\mathbb{P}$.

Step 3: Girsanov formula for stochastic integrals.

For any bounded progressively measurable vector process ϕ_t, we show

$$\mathbb{E}_{\mathbb{Q}}\Big[\int_0^T \langle \phi_s,\, dW_s\rangle\Big] = \mathbb{E}_{\mathbb{Q}}\Big[\int_0^T \langle \phi_s,\, d\widetilde{W}_s\rangle\Big], \tag{1.108}$$

where

$$\widetilde{W}_t := W_t - \int_0^t \theta_s\,ds. \tag{1.109}$$

Using $d\mathbb{Q} = M_T\,d\mathbb{P}$ and the product rule,

$$\int_0^T \phi_s \cdot dW_s = \int_0^T \phi_s \cdot d\widetilde{W}_s + \int_0^T \phi_s \cdot \theta_s\,ds. \tag{1.110}$$

Under $\mathbb{Q}$,

$$\mathbb{E}_{\mathbb{Q}}\left[\int_0^T \phi_s \cdot dW_s\right] = \mathbb{E}_{\mathbb{P}}\left[M_T\Big(\int_0^T \phi_s \cdot d\widetilde{W}_s + \int_0^T \phi_s \cdot \theta_s\,ds\Big)\right]. \tag{1.111}$$

Since $dM_t = M_t\theta_t \cdot dW_t$, applying integration by parts to $M_t \int_0^t \phi_s \cdot d\widetilde{W}_s$ and using that $\langle \widetilde{W}\rangle = \langle W\rangle$ because adding a drift does not change quadratic variation, we obtain that

$$\mathbb{E}_{\mathbb{P}}\Big[M_T \int_0^T \phi_s \cdot d\widetilde{W}_s\Big] = 0. \tag{1.112}$$

The remaining term yields

$$\mathbb{E}_{\mathbb{P}}\Big[M_T \int_0^T \phi_s \cdot \theta_s\,ds\Big] = \mathbb{E}_{\mathbb{Q}}\Big[\int_0^T \phi_s \cdot \theta_s\,ds\Big]. \tag{1.113}$$

Thus the identity above follows.

Step 4: Martingale property of $\widetilde{W}_t$ under $\mathbb{Q}$.

Let $0 \le s < t \le T$ and $B \in \mathcal{F}_s$. For any coordinate vector $e_i \in \mathbb{R}^d$ pick $\phi_u = 1_{(s,t]}(u)e_i$. Then

$$\int_0^T \phi_u \cdot d\widetilde{W}_u = \widetilde{W}_t^{(i)} - \widetilde{W}_s^{(i)}. \tag{1.114}$$

Using the identity of Step 3,

$$\mathbb{E}_{\mathbb{Q}}[(\widetilde{W}_t^{(i)} - \widetilde{W}_s^{(i)})1_B] = \mathbb{E}_{\mathbb{Q}}\Big[\int_s^t \phi_u \cdot dW_u\, 1_B\Big]. \tag{1.115}$$

But under $\mathbb{Q}$, the right-hand side equals

$$\mathbb{E}_{\mathbb{P}}\Big[M_T 1_B \int_s^t dW_u^{(i)}\Big] = \mathbb{E}_{\mathbb{P}}\Big[1_B \int_s^t M_u dW_u^{(i)}\Big] = 0, \tag{1.116}$$

because $\int_0^{\cdot} M_u dW_u$ is a $\mathbb{P}$-martingale and $B \in \mathcal{F}_s$. Hence

$$\mathbb{E}_{\mathbb{Q}}\Big[\widetilde{W}_t^{(i)} \mid \mathcal{F}_s\Big] = \widetilde{W}_s^{(i)}. \tag{1.117}$$

Thus $\widetilde{W}_t$ is a $\mathbb{Q}$-martingale.

Step 5: Quadratic variation under $\mathbb{Q}$.

Since $\widetilde{W}_t = W_t - \int_0^t \theta_s\, ds$ and the drift term has zero quadratic variation,

$$\langle \widetilde{W}^{(i)} \rangle_t = \langle W^{(i)} \rangle_t = t. \tag{1.118}$$

The same holds for cross-variations:

$$\langle \widetilde{W}^{(i)}, \widetilde{W}^{(j)} \rangle_t = \langle W^{(i)}, W^{(j)} \rangle_t = \delta_{ij}\, t. \tag{1.119}$$

Step 6: Lévy characterization.

Lévy's characterization theorem states:

A continuous $\mathbb{R}^d$-valued martingale with quadratic variation $\langle M^{(i)}, M^{(j)} \rangle_t = \delta_{ij}\, t$ is a d-dimensional Brownian motion.

Steps 4 and 5 verify all hypotheses. Hence $\widetilde{W}_t$ is a d-dimensional Brownian motion under $\mathbb{Q}$.

Conclusion: (1) M_T is a true martingale with $\mathbb{E}[M_T] = 1$, so

$$d\mathbb{Q} = M_T\, d\mathbb{P} \tag{1.120}$$

defines a probability measure.

(2) The process

$$\widetilde{W}_t = W_t - \int_0^t \theta_s \, ds \tag{1.121}$$

is a Brownian motion under $\mathbb{Q}$. This completes the proof. □

1.8 Remarks and further directions

This chapter provided the probabilistic and analytic underpinnings used in the construction of supersymmetric path integrals. In the next chapter we will introduce stochastic calculus on manifolds, the Malliavin calculus, and the functional-analytic spectral theory necessary to formulate the supersymmetric Hamiltonians.

1.9 References

Bogachev (1998) [1] is the definitive reference for sections 1.3–1.7, Vakhania et. al. (2012) [2] for Classical construction via cylindrical measures, Kuo (2006) [3] for early rigorous Wiener measure theory, Da Prato & Zabczyk (2014) [4] for Modern Banach/Hilbert framework for Wiener processes, Fernique (1974) [5], Cameron & Martin (1944) [6], Girsanov (1960) [7], Stroock (2010) [8], Albeverio (1998) [9], Minlos (1959) [10], Ghosh (2025) [11].

Chapter 2

Fréchet Differentiability on Banach Spaces

2.1 Introduction

The purpose of this chapter is to develop the theory of Fréchet differentiable maps between Banach spaces at a level of rigor suitable for applications in stochastic analysis, infinite-dimensional geometry, and supersymmetric stochastic dynamics. Unlike the finite-dimensional setting, differentiability in Banach spaces involves subtle issues related to boundedness of linear approximations, the geometry of the ambient space, the existence of higher order derivatives, and the relationship between Gâteaux and Fréchet differentiability. These issues become particularly delicate in infinite dimensions, where compactness arguments generally fail and reflexivity or other geometric hypotheses play nontrivial roles.

We present a systematic and rigorous treatment of these topics, beginning with basic definitions and concluding with deep results including the mean value theorem, Taylor expansions, and the inverse and implicit function theorems in Banach spaces.

Throughout this chapter, all vector spaces are assumed to be real (unless explicitly stated otherwise) and endowed with norms making them Banach spaces. We write $L(E, F)$ for the Banach space of bounded linear operators between Banach spaces E and F.

2.2 Preliminaries on Banach Spaces

We record notational and foundational material needed for this chapter.

Definition 2.1. A *Banach space* is a pair $(E, \|\cdot\|_E)$ consisting of a real (or complex) vector space E and a complete norm $\|\cdot\|_E$.

Definition 2.2. Let E, F be Banach spaces. A linear operator $A : E \to F$ is *bounded* if there exists $C \geq 0$ such that $\|Ax\|_F \leq C\|x\|_E$ for all $x \in E$. The least such constant is

$$\|A\|_{L(E,F)} = \sup_{\|x\|_E=1} \|Ax\|_F. \tag{2.1}$$

We denote the space of bounded linear operators by $L(E, F)$; it is itself a Banach space.

A normed vector space is constructed from a real or complex vector space E together with a function $|\cdot|_E : E \to [0, \infty)$ satisfying the three structural properties that turn E into a metric space with rich geometric structure. First, the norm must satisfy positive definiteness, meaning that

$$|x|_E = 0 \quad \Longleftrightarrow \quad x = 0, \tag{2.2}$$

which guarantees that the algebraic identity element is metrically unique. Second, it must respect absolute homogeneity, so that for all scalars λ and all $x \in E$,

$$|\lambda x|_E = |\lambda|, |x|_E, \tag{2.3}$$

ensuring compatibility between scalar multiplication and the induced metric. Third, it must satisfy the triangle inequality,

$$|x + y|_E \leq |x|_E + |y|_E, \tag{2.4}$$

which ensures that the metric derived from the norm has the geometric structure expected of linear spaces. When these conditions hold, the pair $(E, |\cdot|_E)$ is called a normed space.

Completeness enters the picture when one considers sequences. A sequence (x_n) in a normed space is called Cauchy if for every $\varepsilon > 0$ there exists N such that $|x_m - x_n|_E < \varepsilon$ for all $m, n \geq N$. Completeness is then the requirement that every Cauchy sequence converges to a limit inside the space. If the normed space possesses this property, then it is called a Banach space. This definition is equivalent to the requirement that every absolutely convergent series converges in E, because if $\sum |x_n|_E < \infty$, then partial sums form a Cauchy sequence, and completeness ensures convergence. Thus a Banach space is, in effect, a normed vector space endowed with a metric that is "closed" under Cauchy limits.

Linear operators between normed spaces are maps $A : E \to F$ satisfying

$$A(\lambda x + \mu y) = \lambda Ax + \mu Ay \tag{2.5}$$

for all $x, y \in E$ and all scalars λ, μ. Such a linear operator is called bounded if there exists a constant $C \geq 0$ such that

$$|Ax|_F \leq C|x|_E \quad \text{for all } x \in E. \tag{2.6}$$

This condition implies uniform control of the operator across the entire unit sphere in E. An equivalent formulation, using the unit ball rather than arbitrary vectors, is obtained by noting that the inequality above holds for all $x \neq 0$ if and only if it holds for all x with $|x|_E = 1$; indeed, if $x \neq 0$, then

$$|Ax|_F = |x|_E \left| A\left(\frac{x}{|x|_E}\right) \right|_F \leq |x|_E \sup_{|y|_E=1} |Ay|_F. \tag{2.7}$$

This supremum defines the operator norm,

$$|A|_{L(E,F)} = \sup_{|x|_E=1} |Ax|_F, \tag{2.8}$$

and the reasoning above shows that the constant C in the boundedness condition can always be taken to be this supremum. Moreover, this operator norm is the minimal constant satisfying the inequality: if $|Ax|_F \leq C|x|_E$ for all x, then necessarily

$$\sup_{|x|_E=1} |Ax|_F \leq C, \tag{2.9}$$

implying that $|A|_{L(E,F)} \leq C$. Thus the operator norm furnishes a canonical measurement of the strength of a linear operator between normed spaces.

When E and F are Banach spaces, the collection $L(E,F)$ of all bounded linear operators can itself be made into a normed vector space using the operator norm. The fact that it is complete under this norm—that is, that $L(E,F)$ is again a Banach space—requires a careful argument. Consider a Cauchy sequence (A_n) in $L(E,F)$. By Cauchyness, for every $\varepsilon > 0$ there exists N such that $|A_n - A_m|_{L(E,F)} < \varepsilon$ whenever $n,m \geq N$. Fix an arbitrary $x \in E$ with $|x|_E = 1$. The defining property of the operator norm ensures that

$$|A_nx - A_mx|_F \leq |A_n - A_m|_{L(E,F)} < \varepsilon, \tag{2.10}$$

so for each fixed x, the sequence (A_nx) is Cauchy in F. Because F is complete, there exists a limit, denoted (Ax). This procedure defines a map $A : E \to F$. Linearity of A follows from the fact that limits respect addition and scalar multiplication: if $x, y \in E$, then

$$A(\lambda x+\mu y) = \lim *n \to \infty A_n(\lambda x+\mu y) = \lambda \lim *n \to \infty A_nx+\mu \lim_{n\to\infty} A_ny = \lambda Ax+\mu Ay. \tag{2.11}$$

To see that A is bounded, consider any x with $|x|_E = 1$. The sequence $(|A_nx|_F)$ is bounded above by $\sup_n |A_n|_{L(E,F)}$, which is finite because (A_n) is Cauchy. Passing to the limit yields

$$|Ax| * F = \lim *n \to \infty |A_nx| * F \leq \limsup *n \to \infty |A_n| * L(E,F) < \infty. \tag{2.12}$$

Hence A is a bounded linear operator. Finally, convergence in operator norm follows from the fact that for any $\varepsilon > 0$, there exists N such that for all $n \geq N$,

$$|A_n - A_m|_{L(E,F)} < \varepsilon \quad \text{for all sufficiently large } m, \tag{2.13}$$

and letting $m \to \infty$ gives

$$|A_n - A| * L(E,F) = \sup_{|x|_E=1} |A_nx - Ax|_F \leq \varepsilon. \tag{2.14}$$

Thus (A_n) converges to A in operator norm, and $A \in L(E,F)$. Because every Cauchy sequence in $L(E,F)$ converges to a limit inside $L(E,F)$, this space is complete. Therefore $L(E,F)$ is a Banach space whenever both E and F are Banach spaces.
We will frequently use the uniform boundedness principle.

Theorem 2.3 (Uniform Boundedness Principle)**.** *Let E be a Banach space, F a normed space, and $\{A_\alpha\}_{\alpha\in\mathcal{A}} \subset L(E,F)$ a family of bounded linear operators. Assume that for every $x \in E$ the set $\{\|A_\alpha x\|_F : \alpha \in \mathcal{A}\}$ is bounded (i.e. each $x \in E$ is sent to a uniformly bounded set of values). Then the operator norms are uniformly bounded:*

$$\sup_{\alpha\in\mathcal{A}} \|A_\alpha\|_{L(E,F)} < \infty. \tag{2.15}$$

Proof. We give the classical proof based on Baire's category theorem, written with full detail.

For $n \in \mathbb{N}$ define the subset

$$E_n \ := \ \{x \in E : \sup_{\alpha \in \mathcal{A}} \|A_\alpha x\|_F \leq n\}. \tag{2.16}$$

Each set E_n is the intersection of the closed sets $\{x : \|A_\alpha x\| \leq n\}$ for $\alpha \in \mathcal{A}$, hence E_n is closed in E. By the pointwise boundedness hypothesis, for every $x \in E$ the quantity $\sup_\alpha \|A_\alpha x\|$ is finite, so $x \in E_N$ for some N. Therefore

$$E = \bigcup_{n=1}^{\infty} E_n. \tag{2.17}$$

Since E is a Banach space, it is a complete metric space. The Baire category theorem asserts that a complete metric space cannot be a countable union of nowhere dense closed sets. Hence there exists some $N \in \mathbb{N}$ such that E_N has nonempty interior. Thus there exist $x_0 \in E$ and $r > 0$ with the closed ball

$$\overline{B}(x_0, r) = \{x \in E : \|x - x_0\|_E \leq r\} \tag{2.18}$$

contained in E_N.

By definition of E_N, for every $x \in \overline{B}(x_0, r)$ and every $\alpha \in \mathcal{A}$ we have $\|A_\alpha x\|_F \leq N$. In particular, $\|A_\alpha x_0\|_F \leq N$ for all α, and therefore for any $y \in E$ with $\|y\|_E \leq r$ and any $\alpha \in \mathcal{A}$ we obtain

$$\|A_\alpha y\|_F = \|A_\alpha(y + x_0) - A_\alpha x_0\|_F \leq \|A_\alpha(y + x_0)\|_F + \|A_\alpha x_0\|_F \leq N + N = 2N, \tag{2.19}$$

since $y + x_0 \in \overline{B}(x_0, r)$. Thus

$$\sup_{\alpha \in \mathcal{A}} \|A_\alpha y\|_F \leq 2N \qquad \text{for all } y \in E \text{ with } \|y\|_E \leq r. \tag{2.20}$$

Now let $u \in E$ be arbitrary with $\|u\|_E = 1$. Take $y = ru$ (so $\|y\|_E = r$) and apply the estimate above to get

$$\sup_{\alpha \in \mathcal{A}} \|A_\alpha(ru)\|_F \leq 2N. \tag{2.21}$$

By homogeneity of each A_α,

$$\sup_{\alpha \in \mathcal{A}} \|A_\alpha u\|_F = \frac{1}{r} \sup_{\alpha \in \mathcal{A}} \|A_\alpha(ru)\|_F \leq \frac{2N}{r}. \tag{2.22}$$

Taking the supremum over all unit vectors $u \in E$ yields

$$\sup_{\alpha \in \mathcal{A}} \|A_\alpha\|_{L(E,F)} = \sup_\alpha \sup_{\|u\|_E = 1} \|A_\alpha u\|_F \leq \frac{2N}{r} < \infty. \tag{2.23}$$

This is precisely the desired uniform boundedness of the operator norms. □

This principle plays a subtle but important role in distinguishing Gâteaux from Fréchet differentiability.

2.3 Multilinear Maps and Their Norms

We generalize bounded linear maps to higher-order multilinear maps.

Definition 2.4. Let $E_1, \dots, E_k, F$ be Banach spaces. A k-linear map $A : E_1 \times \cdots \times E_k \to F$ is *bounded* if there exists $C \geq 0$ such that

$$\|A(x_1, \dots, x_k)\|_F \leq C \prod_{j=1}^{k} \|x_j\|_{E_j}, \qquad x_j \in E_j. \tag{2.24}$$

The least such C is the *operator norm*

$$\|A\| = \sup_{\|x_j\|_{E_j} \leq 1} \|A(x_1, \dots, x_k)\|_F. \tag{2.25}$$

We denote the space of all bounded k-linear maps by $L_k(E_1, \dots, E_k; F)$.

Proposition 2.5. *Let $E_1, \dots, E_k, F$ be Banach spaces and let $L_k(E_1, \dots, E_k; F)$ denote the space of all bounded k-linear maps $A : E_1 \times \cdots \times E_k \to F$, endowed with the operator norm*

$$\|A\| := \sup \Big\{ \|A(x_1, \dots, x_k)\|_F : \ \|x_j\|_{E_j} \leq 1, \ j = 1, \dots, k \Big\}. \tag{2.26}$$

Then $L_k(E_1, \dots, E_k; F)$ is a Banach space with respect to this norm.

Proof. We must show that $L_k(E_1, \dots, E_k; F)$ is complete under the norm $\|\cdot\|$. Let $\{A_n\}_{n \in \mathbb{N}}$ be a Cauchy sequence in $L_k(E_1, \dots, E_k; F)$. By definition, for every $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$\|A_n - A_m\| < \varepsilon \qquad \text{for all } n, m \geq N. \tag{2.27}$$

Step 1: Pointwise convergence. Fix arbitrary $(x_1, \dots, x_k) \in E_1 \times \cdots \times E_k$. For all $n, m \in \mathbb{N}$,

$$\begin{aligned} \|A_n(x_1, \dots, x_k) - A_m(x_1, \dots, x_k)\|_F &= \|(A_n - A_m)(x_1, \dots, x_k)\|_F \\ &\leq \|A_n - A_m\| \prod_{j=1}^{k} \|x_j\|_{E_j}, \end{aligned} \tag{2.28}$$

by the definition of the operator norm. Since $\{A_n\}$ is Cauchy in $L_k(E_1, \dots, E_k; F)$, the right-hand side of (2.28) tends to zero as $n, m \to \infty$. Hence, for each fixed $(x_1, \dots, x_k)$, the sequence $\{A_n(x_1, \dots, x_k)\}$ is Cauchy in F. Because F is complete, there exists a unique element $A(x_1, \dots, x_k) \in F$ such that

$$A(x_1, \dots, x_k) := \lim_{n \to \infty} A_n(x_1, \dots, x_k). \tag{2.29}$$

Step 2: k-linearity of the limit map. We show that A defined by (2.29) is k-linear. Fix $j \in \{1, \dots, k\}$ and let $x_j, y_j \in E_j$ and $\alpha, \beta \in \mathbb{K}$ (where $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$). For fixed $x_i \in E_i$, $i \neq j$, we have

$$
\begin{aligned}
A(x_1, \ldots, \alpha x_j + \beta y_j, \ldots, x_k) &= \lim_{n\to\infty} A_n(x_1, \ldots, \alpha x_j + \beta y_j, \ldots, x_k) \\
&= \lim_{n\to\infty} \left(\alpha A_n(x_1, \ldots, x_j, \ldots, x_k) + \beta A_n(x_1, \ldots, y_j, \ldots, x_k)\right) \\
&= \alpha A(x_1, \ldots, x_j, \ldots, x_k) + \beta A(x_1, \ldots, y_j, \ldots, x_k),
\end{aligned}
$$

where we used the k-linearity of each A_n and the continuity of limits in F. Since j was arbitrary, A is k-linear.

Step 3: Boundedness of the limit map. Let $(x_1, \ldots, x_k) \in E_1 \times \cdots \times E_k$ be arbitrary. For any $n \in \mathbb{N}$,

$$
\|A(x_1, \ldots, x_k)\|_F \le \|A(x_1, \ldots, x_k) - A_n(x_1, \ldots, x_k)\|_F + \|A_n(x_1, \ldots, x_k)\|_F. \tag{2.30}
$$

Taking the limit $n \to \infty$ and using (2.28) with $m \to \infty$, we obtain

$$
\|A(x_1, \ldots, x_k)\|_F \le \limsup_{n\to\infty} \|A_n\| \prod_{j=1}^{k} \|x_j\|_{E_j}. \tag{2.31}
$$

Since $\{A_n\}$ is Cauchy in $L_k(E_1, \ldots, E_k; F)$, it is bounded; hence there exists $M > 0$ such that $\|A_n\| \le M$ for all n. Therefore,

$$
\|A(x_1, \ldots, x_k)\|_F \le M \prod_{j=1}^{k} \|x_j\|_{E_j}, \tag{2.32}
$$

which shows that A is bounded and $A \in L_k(E_1, \ldots, E_k; F)$.

Step 4: Convergence in operator norm. Finally, we show that $A_n \to A$ in the norm $\|\cdot\|$. For any $(x_1, \ldots, x_k)$ with $\|x_j\|_{E_j} \le 1$,

$$
\|(A_n - A)(x_1, \ldots, x_k)\|_F = \lim_{m\to\infty} \|(A_n - A_m)(x_1, \ldots, x_k)\|_F \le \limsup_{m\to\infty} \|A_n - A_m\|. \tag{2.33}
$$

Taking the supremum over all such $(x_1, \ldots, x_k)$ yields

$$
\|A_n - A\| \le \limsup_{m\to\infty} \|A_n - A_m\|. \tag{2.34}
$$

Since $\{A_n\}$ is Cauchy in $L_k(E_1, \ldots, E_k; F)$, the right-hand side tends to zero as $n \to \infty$, hence

$$
\|A_n - A\| \to 0. \tag{2.35}
$$

Thus $\{A_n\}$ converges in $L_k(E_1, \ldots, E_k; F)$ to A, and therefore $L_k(E_1, \ldots, E_k; F)$ is complete. This proves that it is a Banach space. □

2.4 Gâteaux and Fréchet Derivatives

We now give precise definitions of the two fundamental notions of differentiability for maps between Banach spaces.

Definition 2.6 (Gâteaux derivative)**.** Let E, F be Banach spaces and $U \subset E$ an open set. A map $f : U \to F$ is *Gâteaux differentiable* at $x \in U$ if for every $h \in E$ the limit

$$D_G f(x)[h] = \lim_{t \to 0} \frac{f(x+th) - f(x)}{t} \tag{2.36}$$

exists in F. The map $h \mapsto D_G f(x)[h]$ is automatically homogeneous and additive if f is continuous, but is not necessarily bounded.

Definition 2.7 (Fréchet derivative)**.** Let $f : U \to F$ be a function between Banach spaces. It is *Fréchet differentiable* at $x \in U$ if there exists a bounded linear operator $A \in L(E, F)$ such that

$$\lim_{\|h\|_E \to 0} \frac{\|f(x+h) - f(x) - Ah\|_F}{\|h\|_E} = 0. \tag{2.37}$$

The operator A is unique and denoted $Df(x)$.

Remark 2.8. The Fréchet derivative implies the Gâteaux derivative and one has $Df(x)[h] = D_G f(x)[h]$. The converse is false in general and depends on geometric properties of the spaces involved and uniform boundedness of directional derivatives.

2.5 Distinguishing Gâteaux and Fréchet Differentiability

We prove that Fréchet differentiability is strictly stronger than Gâteaux differentiability.

Theorem 2.9 (Fréchet differentiability is strictly stronger than Gâteaux differentiability)**.** *Let E, F be Banach spaces, let $U \subset E$ be open, and let $f : U \to F$ be Gâteaux differentiable at a point $x \in U$. Assume the following:*

1. *The Gâteaux derivative*

   $$D_G f(x) : E \to F \tag{2.38}$$

   is linear and bounded.

2. *The directional difference quotients converge* uniformly on the unit ball*, i.e.*

   $$\sup_{\|h\| \le 1} \left\| \frac{f(x+th) - f(x)}{t} - D_G f(x)[h] \right\|_F \xrightarrow[t \to 0]{} 0. \tag{2.39}$$

Then f is Fréchet differentiable at x, and

$$Df(x) = D_G f(x). \tag{2.40}$$

Proof. We prove the statement directly from the definition of Fréchet differentiability.

Step 1: Recall of definitions.

Fréchet differentiability of f at x means that there exists a bounded linear operator $A \in L(E, F)$ such that

$$\lim_{\|h\| \to 0} \frac{\|f(x+h) - f(x) - Ah\|_F}{\|h\|_E} = 0. \tag{2.41}$$

Gâteaux differentiability at x means that for every $h \in E$, the limit

$$D_G f(x)[h] = \lim_{t \to 0} \frac{f(x+th) - f(x)}{t} \tag{2.42}$$

exists in F. Our goal is to show that (2.41) holds with $A = D_G f(x)$.

Step 2: Reduction to normalized increments.

Let $h \in E$, $h \neq 0$, be arbitrary and define

$$u := \frac{h}{\|h\|}, \qquad t := \|h\|. \tag{2.43}$$

Then $\|u\| = 1$ and $h = tu$. Note that $t \to 0$ if and only if $\|h\| \to 0$. We write

$$\frac{f(x+h) - f(x) - D_G f(x)[h]}{\|h\|} = \frac{f(x+tu) - f(x) - tD_G f(x)[u]}{t} \tag{2.44}$$

$$= \frac{f(x+tu) - f(x)}{t} - D_G f(x)[u]. \tag{2.45}$$

Thus, for all $h \neq 0$,

$$\frac{\|f(x+h) - f(x) - D_G f(x)[h]\|_F}{\|h\|_E} = \left\| \frac{f(x+tu) - f(x)}{t} - D_G f(x)[u] \right\|_F, \quad \|u\| = 1. \tag{2.46}$$

Step 3: Use of uniform convergence.

By the uniform convergence assumption (2.39), for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$|t| < \delta \quad \Longrightarrow \quad \sup_{\|h\| \le 1} \left\| \frac{f(x+th) - f(x)}{t} - D_G f(x)[h] \right\|_F < \varepsilon. \tag{2.47}$$

In particular, for every u with $\|u\| = 1$ and every $|t| < \delta$,

$$\left\| \frac{f(x+tu) - f(x)}{t} - D_G f(x)[u] \right\|_F < \varepsilon. \tag{2.48}$$

Step 4: Verification of the Fréchet limit.

Let $h \in E$ with $0 < \|h\| < \delta$. Then $t = \|h\| < \delta$, and by (2.46),

$$\frac{\|f(x+h) - f(x) - D_G f(x)[h]\|_F}{\|h\|_E} < \varepsilon. \tag{2.49}$$

Since $\varepsilon > 0$ is arbitrary, this proves that

$$\lim_{\|h\| \to 0} \frac{\|f(x+h) - f(x) - D_G f(x)[h]\|_F}{\|h\|_E} = 0. \tag{2.50}$$

Thus f is Fréchet differentiable at x with Fréchet derivative

$$Df(x) = D_G f(x). \tag{2.51}$$

Step 5: Strictness of the implication.

The theorem shows that Fréchet differentiability follows from Gâteaux differentiability *only* when the directional derivatives assemble into a bounded linear operator and the convergence is uniform in direction. In general, Gâteaux differentiability alone does *not* imply Fréchet differentiability, even when all directional derivatives exist.

□

2.6 Higher-Order Derivatives

Definition 2.10. A function $f : U \to F$ is *k-times Fréchet differentiable* at x if the $(k-1)$-st derivative $D^{k-1}f$ exists in a neighborhood of x and is Fréchet differentiable at x as a map into the Banach space $L_{k-1}(E, \dots, E; F)$. The k-linear map $D^k f(x) \in L_k(E, \dots, E; F)$ is the *k-th Fréchet derivative.*

Proposition 2.11. *Let E, F be Banach spaces, $U \subset E$ open, and let $f : U \to F$ be k-times Fréchet differentiable at $x \in U$. Then for arbitrary $h_1, \dots, h_k \in E$,*

$$D^k f(x)[h_1, \dots, h_k] = \frac{\partial^k}{\partial t_1 \cdots \partial t_k} f\big(x + t_1 h_1 + \cdots + t_k h_k\big)\Big|_{t_1 = \cdots = t_k = 0}. \tag{2.52}$$

Proof. We proceed by induction on k.

Base case $k = 1$. Let $h \in E$ and define $\Phi(t) = f(x + th)$. Since f is Fréchet differentiable at x, there exists a bounded linear operator $Df(x) \in L(E, F)$ such that

$$f(x + th) = f(x) + tDf(x)[h] + o(t) \quad \text{as } t \to 0. \tag{2.53}$$

Hence

$$\frac{\Phi(t) - \Phi(0)}{t} = Df(x)[h] + o(1), \tag{2.54}$$

which implies

$$\Phi'(0) = Df(x)[h]. \tag{2.55}$$

Thus the statement holds for $k = 1$.

Induction hypothesis: Assume the statement holds for some $k - 1 \geq 1$, that is, if f is $(k-1)$-times Fréchet differentiable at x, then for all $h_1, \dots, h_{k-1} \in E$,

$$D^{k-1} f(x)[h_1, \dots, h_{k-1}] = \frac{\partial^{k-1}}{\partial t_1 \cdots \partial t_{k-1}} f\big(x + t_1 h_1 + \cdots + t_{k-1} h_{k-1}\big)\Big|_{t_1 = \cdots = t_{k-1} = 0}. \tag{2.56}$$

Induction step. Assume now that f is k-times Fréchet differentiable at x and fix $h_1, \dots, h_k \in E$. Define

$$\Phi(t_1, \dots, t_k) = f\big(x + t_1 h_1 + \cdots + t_k h_k\big), \quad (t_1, \dots, t_k) \in \mathbb{R}^k. \tag{2.57}$$

By repeated application of the chain rule, Φ is k-times continuously differentiable in a neighborhood of $0 \in \mathbb{R}^k$. For each $(t_1, \ldots, t_k)$,

$$\frac{\partial \Phi}{\partial t_1}(t_1, \ldots, t_k) = Df(x + t_1 h_1 + \cdots + t_k h_k)[h_1]. \tag{2.58}$$

Define $G(y) = Df(y) \in L(E, F)$. Since f is k-times Fréchet differentiable at x, the map G is $(k-1)$-times Fréchet differentiable at x and satisfies $D^{k-1}G(x) = D^k f(x)$. Applying the induction hypothesis to the function

$$y \mapsto Df(y)[h_1], \tag{2.59}$$

we obtain

$$\frac{\partial^{k-1}}{\partial t_2 \cdots \partial t_k} Df(x + t_1 h_1 + \cdots + t_k h_k)[h_1]\Big|_{t_1 = \cdots = t_k = 0} = D^k f(x)[h_2, \ldots, h_k][h_1]. \tag{2.60}$$

Since $D^k f(x)$ is a symmetric k-linear map, the right-hand side equals $D^k f(x)[h_1, \ldots, h_k]$. Therefore,

$$\frac{\partial^k}{\partial t_1 \cdots \partial t_k} \Phi(0) = D^k f(x)[h_1, \ldots, h_k], \tag{2.61}$$

which is precisely the desired identity.

This completes the induction and hence the proof. □

2.7 Chain Rule

Theorem 2.12 (Chain Rule for Fréchet Derivatives). *Let E, F, G be Banach spaces. Let $U \subset E$ and $V \subset F$ be open sets, and let*

$$f : U \to F, \qquad g : V \to G \tag{2.62}$$

be Fréchet differentiable at a point $x \in U$ with $f(x) \in V$. Then the composition $g \circ f$ is Fréchet differentiable at x, and

$$D(g \circ f)(x) = Dg(f(x)) \circ Df(x). \tag{2.63}$$

Proof. We give a fully rigorous proof directly from the definition of Fréchet differentiability.

Step 1: Fréchet differentiability expansions.

Since f is Fréchet differentiable at x, there exists a bounded linear operator $Df(x) \in L(E, F)$ and a remainder mapping $r_f : E \to F$ such that

$$f(x+h) = f(x) + Df(x)h + r_f(h), \qquad \lim_{\|h\| \to 0} \frac{\|r_f(h)\|_F}{\|h\|_E} = 0. \tag{2.64}$$

Similarly, since g is Fréchet differentiable at $f(x)$, there exists $Dg(f(x)) \in L(F, G)$ and a remainder mapping $r_g : F \to G$ such that

$$g(f(x) + k) = g(f(x)) + Dg(f(x))k + r_g(k), \qquad \lim_{\|k\| \to 0} \frac{\|r_g(k)\|_G}{\|k\|_F} = 0. \tag{2.65}$$

Step 2: Substitute the increment of f into g.

Let $h \in E$ be sufficiently small so that $f(x+h) \in V$. Define

$$k := f(x+h) - f(x). \tag{2.66}$$

By (2.64),

$$k = Df(x)h + r_f(h). \tag{2.67}$$

Substituting into (2.65), we obtain

$$\begin{aligned} g(f(x+h)) &= g(f(x)) + Dg(f(x))(Df(x)h + r_f(h)) + r_g(Df(x)h + r_f(h)) && (2.68)\\ &= g(f(x)) + Dg(f(x))Df(x)h + Dg(f(x))r_f(h) + r_g(Df(x)h + r_f(h)). && (2.69)\end{aligned}$$

Step 3: Identify the candidate derivative.

Rewriting the above expression yields

$$(g \circ f)(x+h) = (g \circ f)(x) + (Dg(f(x)) \circ Df(x))h + R(h), \tag{2.70}$$

where the remainder term $R(h)$ is defined by

$$R(h) := Dg(f(x))r_f(h) + r_g(Df(x)h + r_f(h)). \tag{2.71}$$

Thus the natural candidate for the Fréchet derivative of $g \circ f$ at x is the bounded linear operator

$$Dg(f(x)) \circ Df(x) \in L(E, G). \tag{2.72}$$

Step 4: Estimate the remainder term.

We now show that $R(h) = o(\|h\|)$ in G. First, since $Dg(f(x))$ is bounded,

$$\frac{\|Dg(f(x))r_f(h)\|_G}{\|h\|_E} \le \|Dg(f(x))\|_{L(F,G)} \frac{\|r_f(h)\|_F}{\|h\|_E} \longrightarrow 0 \quad \text{as } \|h\| \to 0, \tag{2.73}$$

by (2.64). Next, note that

$$\|Df(x)h + r_f(h)\|_F \le \|Df(x)\| \, \|h\|_E + \|r_f(h)\|_F = O(\|h\|_E). \tag{2.74}$$

Hence

$$\lim_{\|h\|\to 0} \frac{\|Df(x)h + r_f(h)\|_F}{\|h\|_E} < \infty. \tag{2.75}$$

Using (2.65),

$$\frac{\|r_g(Df(x)h + r_f(h))\|_G}{\|h\|_E} = \frac{\|r_g(Df(x)h + r_f(h))\|_G}{\|Df(x)h + r_f(h)\|_F} \cdot \frac{\|Df(x)h + r_f(h)\|_F}{\|h\|_E}. \tag{2.76}$$

The first factor tends to zero as $\|h\| \to 0$, and the second factor remains bounded. Therefore,

$$\frac{\|r_g(Df(x)h + r_f(h))\|_G}{\|h\|_E} \longrightarrow 0. \tag{2.77}$$

Combining both estimates, we conclude that

$$\lim_{\|h\|\to 0} \frac{\|R(h)\|_G}{\|h\|_E} = 0. \tag{2.78}$$

Step 5: Conclusion.

By definition of Fréchet differentiability, equation (2.70) and the estimate above show that $g \circ f$ is Fréchet differentiable at x with derivative

$$D(g \circ f)(x) = Dg(f(x)) \circ Df(x). \tag{2.79}$$

□

2.8 Mean Value Theorem in Banach Spaces

Theorem 2.13 (Mean Value Inequality). *Let $f : U \subset E \to F$ be Fréchet differentiable on an open convex set. Then for all $x, y \in U$,*

$$\|f(y) - f(x)\|_F \leq \sup_{z\in[x,y]} \|Df(z)\|_{L(E,F)} \|y - x\|_E. \tag{2.80}$$

Proof. We proceed in several logically distinct steps.

Step 1: Reduction to a one-dimensional problem.

Since U is convex and $x, y \in U$, the line segment

$$[x, y] = \{x + t(y - x) : t \in [0, 1]\} \tag{2.81}$$

is contained in U. Define the curve

$$\gamma : [0, 1] \to U, \qquad \gamma(t) = x + t(y - x), \tag{2.82}$$

and define the Banach-valued function

$$\phi : [0, 1] \to F, \qquad \phi(t) = f(\gamma(t)) = f(x + t(y - x)). \tag{2.83}$$

Because f is Fréchet differentiable on U and γ is C^1, the chain rule for Fréchet derivatives implies that ϕ is continuously differentiable on $[0, 1]$, with derivative given by

$$\phi'(t) = Df(\gamma(t))[y - x] = Df(x + t(y - x))[y - x]. \tag{2.84}$$

Step 2: Fundamental theorem of calculus for Banach-valued functions.

Since $\phi \in C^1([0,1];F)$, the fundamental theorem of calculus for Banach-space-valued functions yields

$$\phi(1) - \phi(0) = \int_0^1 \phi'(t)\,dt, \tag{2.85}$$

where the integral is understood in the Bochner sense. Recalling the definition of ϕ, this identity reads

$$f(y) - f(x) = \int_0^1 Df(x + t(y-x))[y-x]\,dt. \tag{2.86}$$

Step 3: Norm estimate of the integral.

Taking norms in F and using the triangle inequality for Bochner integrals, we obtain

$$\|f(y) - f(x)\|_F \le \int_0^1 \|Df(x + t(y-x))[y-x]\|_F\,dt. \tag{2.87}$$

For each $t \in [0,1]$, the Fréchet derivative $Df(x + t(y-x)) \in L(E,F)$ is a bounded linear operator. Hence, by definition of the operator norm,

$$\|Df(x + t(y-x))[y-x]\|_F \le \|Df(x + t(y-x))\|_{L(E,F)}\,\|y-x\|_E. \tag{2.88}$$

Substituting this estimate into the integral yields

$$\|f(y) - f(x)\|_F \le \|y-x\|_E \int_0^1 \|Df(x + t(y-x))\|_{L(E,F)}\,dt. \tag{2.89}$$

Step 4: Uniform bound along the line segment.

Since the function

$$t \longmapsto \|Df(x + t(y-x))\|_{L(E,F)} \tag{2.90}$$

is continuous on the compact interval $[0,1]$, it attains its supremum. Therefore,

$$\|Df(x + t(y-x))\|_{L(E,F)} \le \sup_{z\in[x,y]} \|Df(z)\|_{L(E,F)} \quad \text{for all } t \in [0,1]. \tag{2.91}$$

Consequently,

$$\int_0^1 \|Df(x + t(y-x))\|_{L(E,F)}\,dt \le \sup_{z\in[x,y]} \|Df(z)\|_{L(E,F)}. \tag{2.92}$$

Step 5: Conclusion.

Combining the estimates from Steps 3 and 4, we conclude that

$$\|f(y) - f(x)\|_F \le \sup_{z\in[x,y]} \|Df(z)\|_{L(E,F)}\,\|y-x\|_E, \tag{2.93}$$

which is precisely the desired inequality.

□

2.9 Taylor's Theorem in Banach Spaces

Theorem 2.14 (Taylor's theorem)**.** *Let $f : U \subset E \to F$ be k-times continuously Fréchet differentiable on an open convex subset U and let $x, y \in U$. Then*

$$f(y) = f(x) + \sum_{j=1}^{k-1} \frac{1}{j!} D^j f(x)[(y-x)^j] + R_k(x,y), \tag{2.94}$$

where the remainder satisfies

$$\|R_k(x,y)\| \leq \frac{\|y-x\|^k}{(k-1)!} \sup_{z\in[x,y]} \|D^k f(z)\|. \tag{2.95}$$

Proof. Here $[x,y] = x + t(y-x) : t \in [0,1]$ and $D^j f(x) \in \mathcal{L}^j(E,F)$ is the j-linear Fréchet derivative. The first step is the reduction of this above theorem statement to a one-dimensional problem. Define the curve

$$\gamma(t) := x + t(y-x), \quad t \in [0,1] \tag{2.96}$$

which lies entirely in U by convexity. Define the function

$$\phi(t) := f(\gamma(t)) = f(x + t(y-x)), \quad t \in [0,1] \tag{2.97}$$

Since $f \in C^k(U;F)$, the chain rule for Fréchet derivatives implies:

$$\phi \in C^k([0,1];F) \tag{2.98}$$

The second step is to compute derivatives of ϕ. The First derivative by the chain rule is

$$\phi'(t) = Df(\gamma(t))[y-x]. \tag{2.99}$$

Similarly, let us derive the Higher derivatives by induction. Assume

$$\phi^{(j)}(t) = D^j f(\gamma(t))[(y-x)^j]. \tag{2.100}$$

Then differentiating once more:

$$\phi^{(j+1)}(t) = D^{j+1} f(\gamma(t))[(y-x)^{j+1}], \tag{2.101}$$

since differentiation acts on the argument $\gamma(t)$ and the direction vector is constant. Thus, for all $j \leq k$, we have

$$\boxed{\phi^{(j)}(t) = D^j f(x + t(y-x))[(y-x)^j]} \tag{2.102}$$

The third step is to apply one-dimensional Taylor's theorem (Banach-valued). Since $\phi \in C^k([0,1];F)$, the integral remainder form of Taylor's theorem in Banach spaces applies:

$$\phi(1) = \sum_{j=0}^{k-1} \frac{1}{j!} \phi^{(j)}(0) + \frac{1}{(k-1)!} \int_0^1 (1-t)^{k-1} \phi^{(k)}(t), dt. \tag{2.103}$$

Now we shall substitute the explicit formulas. The fourth step is to identify the Taylor polynomial. Note that we have

$$\phi^{(j)}(0) = D^j f(x)[(y-x)^j]. \tag{2.104}$$

Hence,

$$\sum_{j=0}^{k-1} \frac{1}{j!}\phi^{(j)}(0) = f(x) + \sum_{j=1}^{k-1} \frac{1}{j!} D^j f(x)[(y-x)^j]. \tag{2.105}$$

We now identify the remainder term. Define

$$R_k(x,y) := \frac{1}{(k-1)!}\int_0^1 (1-t)^{k-1} D^k f(x+t(y-x))[(y-x)^k], dt. \tag{2.106}$$

This is exactly the remainder produced by the 1D theorem applied to ϕ. Let's find the norm estimate of the remainder (core of the proof). We take norms in F:

$$|R_k(x,y)| \leq \frac{1}{(k-1)!}\int_0^1 (1-t)^{k-1} |D^k f(x+t(y-x))[(y-x)^k]|, dt. \tag{2.107}$$

Using the operator norm of multilinear maps,

$$|D^k f(z)[(y-x)^k]| \leq |D^k f(z)| \cdot |y-x|^k. \tag{2.108}$$

Thus,

$$|R_k(x,y)| \leq \frac{|y-x|^k}{(k-1)!}\int_0^1 (1-t)^{k-1} \sup_{z\in[x,y]} |D^k f(z)|, dt. \tag{2.109}$$

Since

$$\int_0^1 (1-t)^{k-1}\, dt = \frac{1}{k}, \tag{2.110}$$

we obtain

$$\boxed{|R_k(x,y)| \leq \frac{|y-x|^k}{(k-1)!} \sup_{z\in[x,y]} |D^k f(z)|} \tag{2.111}$$

The factor $1/k$ is absorbed because $(k-1)! \cdot k = k!$; the statement uses the standard sharp bound. Putting together Steps 4 and 5,

$$f(y) = f(x) + \sum_{j=1}^{k-1} \frac{1}{j!} D^j f(x)[(y-x)^j] + R_k(x,y), \tag{2.112}$$

with the stated norm bound on $R_k(x,y)$.

□

2.10 Inverse Function Theorem

Theorem 2.15 (Inverse function theorem)**.** *Let $f : U \subset E \to F$ be continuously Fréchet differentiable on an open set U. Suppose $Df(x_0)$ is invertible in $L(E,F)$ with bounded inverse. Then there exist open sets $V \ni x_0$ and $W \ni f(x_0)$ such that $f : V \to W$ is a C^1 diffeomorphism.*

Proof. The proof proceeds in several logically independent steps.

Step 1: Normalization.

Define the affine-linear map

$$T(x) := Df(x_0)^{-1}(f(x) - f(x_0)). \tag{2.113}$$

Then $T : U \to E$ is C^1, satisfies

$$T(x_0) = 0, \qquad DT(x_0) = I_E, \tag{2.114}$$

and f is locally invertible near x_0 if and only if T is locally invertible near x_0. Hence, without loss of generality, we may assume

$$x_0 = 0, \quad f(0) = 0, \quad Df(0) = I_E. \tag{2.115}$$

Step 2: Decomposition into identity plus perturbation.

Define

$$g(x) := f(x) - x. \tag{2.116}$$

Then $g \in C^1(U; E)$, $g(0) = 0$, and

$$Dg(0) = Df(0) - I_E = 0. \tag{2.117}$$

By continuity of Dg, there exists $r > 0$ such that the closed ball $\overline{B_r(0)} \subset U$ and

$$\sup_{x \in B_r(0)} \|Dg(x)\|_{L(E,E)} \leq \frac{1}{2}. \tag{2.118}$$

Step 3: Construction of a contraction.

Fix $y \in B_{r/2}(0)$ and define the map

$$\Phi_y(x) := y - g(x), \qquad x \in B_r(0). \tag{2.119}$$

We claim that Φ_y maps $B_r(0)$ into itself. Indeed, for $x \in B_r(0)$,

$$\|\Phi_y(x)\| \leq \|y\| + \|g(x)\|. \tag{2.120}$$

By the mean value inequality applied to g,

$$\|g(x)\| \leq \sup_{z \in [0,x]} \|Dg(z)\| \, \|x\| \leq \frac{1}{2} r. \tag{2.121}$$

Since $\|y\| \leq r/2$, it follows that

$$\|\Phi_y(x)\| \leq r. \tag{2.122}$$

Next, for $x_1, x_2 \in B_r(0)$,

$$\|\Phi_y(x_1) - \Phi_y(x_2)\| = \|g(x_2) - g(x_1)\|. \tag{2.123}$$

Again by the mean value inequality,

$$\|g(x_2) - g(x_1)\| \leq \sup_{z \in [x_1, x_2]} \|Dg(z)\| \, \|x_2 - x_1\| \leq \frac{1}{2}\|x_2 - x_1\|. \tag{2.124}$$

Thus Φ_y is a contraction on $B_r(0)$.

Step 4: Existence and uniqueness of local inverse.

By the Banach fixed-point theorem, for each $y \in B_{r/2}(0)$ there exists a unique $x \in B_r(0)$ such that

$$x = \Phi_y(x), \tag{2.125}$$

i.e.

$$f(x) = y. \tag{2.126}$$

Define $f^{-1}(y) := x$. This yields a well-defined inverse mapping

$$f^{-1} : B_{r/2}(0) \to B_r(0). \tag{2.127}$$

Step 5: Continuity of the inverse.

Let $y_1, y_2 \in B_{r/2}(0)$ with corresponding fixed points x_1, x_2. Then

$$x_1 - x_2 = y_1 - y_2 - (g(x_1) - g(x_2)). \tag{2.128}$$

Taking norms and using the contraction estimate,

$$\|x_1 - x_2\| \leq \|y_1 - y_2\| + \frac{1}{2}\|x_1 - x_2\|. \tag{2.129}$$

Hence

$$\|x_1 - x_2\| \leq 2\|y_1 - y_2\|, \tag{2.130}$$

which proves that f^{-1} is continuous (indeed Lipschitz).

Step 6: Differentiability of the inverse.

Let $y = f(x)$ with $x \in B_r(0)$. Since f is differentiable at x,

$$f(x+h) = f(x) + Df(x)h + o(\|h\|). \tag{2.131}$$

Set $k = f(x+h) - f(x)$. Then

$$h = Df(x)^{-1}k + o(\|k\|), \tag{2.132}$$

which shows that f^{-1} is Fréchet differentiable at y, with

$$Df^{-1}(y) = Df(x)^{-1}. \tag{2.133}$$

Step 7: Continuity of the derivative of the inverse.

Since $x \mapsto Df(x)$ is continuous and inversion is continuous on the open set of bounded linear isomorphisms, it follows that

$$y \mapsto Df^{-1}(y) \tag{2.134}$$

is continuous. Hence $f^{-1} \in C^1$.

Step 8: Conclusion.

Let

$$V := B_r(0), \qquad W := B_{r/2}(0). \tag{2.135}$$

Then $f : V \to W$ is bijective, both f and f^{-1} are C^1, and thus f is a C^1 diffeomorphism. □

2.11 Implicit Function Theorem

Theorem 2.16 (Implicit Function Theorem in Banach Spaces)**.** *Let E, F, G be Banach spaces, let $U \subset E \times F$ be open, and let*

$$\mathcal{F} : U \to G \tag{2.136}$$

be continuously Fréchet differentiable. Suppose that for some $(x_0, y_0) \in U$,

$$D_y\mathcal{F}(x_0, y_0) \in L(F, G) \tag{2.137}$$

is a bounded linear isomorphism with bounded inverse. Then there exist open neighborhoods $V \subset E$ of x_0 and $W \subset F$ of y_0 and a unique continuously Fréchet differentiable mapping

$$\phi : V \to W \tag{2.138}$$

such that

$$\mathcal{F}(x, \phi(x)) = 0 \quad \text{for all } x \in V, \qquad \phi(x_0) = y_0. \tag{2.139}$$

Moreover,

$$D\phi(x) = -\big(D_y\mathcal{F}(x, \phi(x))\big)^{-1} \circ D_x\mathcal{F}(x, \phi(x)). \tag{2.140}$$

Proof. The proof is divided into logically independent steps.

Step 1: Reduction to a normalized setting.

Define the map

$$T(x, y) := \big(x,\ D_y\mathcal{F}(x_0, y_0)^{-1}\mathcal{F}(x, y)\big), \tag{2.141}$$

which is C^1 on U. Define also

$$\widetilde{\mathcal{F}}(x, y) := D_y\mathcal{F}(x_0, y_0)^{-1}\mathcal{F}(x, y). \tag{2.142}$$

Then $\widetilde{\mathcal{F}}(x_0, y_0) = 0$ and

$$D_y\widetilde{\mathcal{F}}(x_0, y_0) = I_F. \tag{2.143}$$

Solving $\mathcal{F}(x, y) = 0$ is equivalent to solving $\widetilde{\mathcal{F}}(x, y) = 0$. Hence, without loss of generality, we may assume

$$\mathcal{F}(x_0, y_0) = 0, \qquad D_y\mathcal{F}(x_0, y_0) = I_F. \tag{2.144}$$

Step 2: Decomposition into linear and nonlinear parts.

Define

$$G(x, y) := \mathcal{F}(x, y) - y. \tag{2.145}$$

Then $G \in C^1(U; F)$ and

$$G(x_0, y_0) = 0, \qquad D_yG(x_0, y_0) = 0. \tag{2.146}$$

By continuity of D_yG, there exists a neighborhood $B_r(x_0) \times B_r(y_0) \subset U$ such that

$$\sup_{(x,y)\in B_r(x_0)\times B_r(y_0)} \|D_yG(x, y)\|_{L(F,F)} \leq \frac{1}{2}. \tag{2.147}$$

Step 3: Construction of a contraction.

Fix $x \in B_r(x_0)$ and define the map

$$\Phi_x(y) := -G(x, y), \qquad y \in B_r(y_0). \tag{2.148}$$

Observe that solving $\mathcal{F}(x, y) = 0$ is equivalent to solving

$$y = \Phi_x(y). \tag{2.149}$$

We first show that Φ_x maps $B_r(y_0)$ into itself. For $y \in B_r(y_0)$,

$$\|\Phi_x(y) - y_0\| \leq \|G(x, y) - G(x_0, y_0)\|. \tag{2.150}$$

By the Mean Value Inequality in Banach spaces,

$$\|G(x, y) - G(x_0, y_0)\| \leq \sup_{z\in[(x_0,y_0),(x,y)]} \|DG(z)\|(\|x - x_0\| + \|y - y_0\|). \tag{2.151}$$

For r sufficiently small, this quantity is bounded by r, hence $\Phi_x(B_r(y_0)) \subset B_r(y_0)$. Next, for $y_1, y_2 \in B_r(y_0)$,

$$\|\Phi_x(y_1) - \Phi_x(y_2)\| = \|G(x, y_2) - G(x, y_1)\|. \tag{2.152}$$

Applying the Mean Value Inequality in the y-variable and using (2.147),

$$\|\Phi_x(y_1) - \Phi_x(y_2)\| \leq \frac{1}{2}\|y_1 - y_2\|. \tag{2.153}$$

Thus Φ_x is a strict contraction on $B_r(y_0)$.

Step 4: Existence and uniqueness of the implicit function.

By the Banach Fixed Point Theorem, for each $x \in B_r(x_0)$ there exists a unique $y = \phi(x) \in B_r(y_0)$ such that

$$y = \Phi_x(y), \quad \text{i.e.} \quad \mathcal{F}(x, \phi(x)) = 0. \tag{2.154}$$

This defines a well-defined function

$$\phi : B_r(x_0) \to B_r(y_0), \qquad \phi(x_0) = y_0. \tag{2.155}$$

Step 5: Continuity of ϕ.

Let $x_1, x_2 \in B_r(x_0)$ and let $y_i = \phi(x_i)$. Then

$$y_1 - y_2 = -\big(G(x_1, y_1) - G(x_2, y_2)\big). \tag{2.156}$$

Using the Mean Value Inequality and the contraction estimate in y, one obtains

$$\|y_1 - y_2\| \le C\|x_1 - x_2\| + \frac{1}{2}\|y_1 - y_2\|, \tag{2.157}$$

for some constant $C > 0$. Hence

$$\|y_1 - y_2\| \le 2C\|x_1 - x_2\|, \tag{2.158}$$

which proves continuity (indeed Lipschitz continuity) of ϕ.

Step 6: Differentiability of ϕ.

Fix $x \in B_r(x_0)$ and write

$$\mathcal{F}(x, \phi(x)) = 0. \tag{2.159}$$

Let $h \in E$ be small. Then

$$0 = \mathcal{F}(x + h, \phi(x + h)) - \mathcal{F}(x, \phi(x)). \tag{2.160}$$

By Fréchet differentiability,

$$0 = D_x\mathcal{F}(x, \phi(x))h + D_y\mathcal{F}(x, \phi(x))(\phi(x + h) - \phi(x)) + o(\|h\|). \tag{2.161}$$

Since $D_y\mathcal{F}(x, \phi(x))$ is invertible for x near x_0, we solve for $\phi(x + h) - \phi(x)$ to obtain

$$\phi(x + h) - \phi(x) = -\big(D_y\mathcal{F}(x, \phi(x))\big)^{-1} D_x\mathcal{F}(x, \phi(x))h + o(\|h\|). \tag{2.162}$$

This proves that ϕ is Fréchet differentiable at x with derivative given by (2.140).

Step 7: Continuity of the derivative.

The mapping

$$(x, y) \longmapsto D_x\mathcal{F}(x, y), \qquad (x, y) \longmapsto D_y\mathcal{F}(x, y) \tag{2.163}$$

is continuous, and inversion is continuous on the group of bounded linear isomorphisms. Therefore,

$$x \longmapsto D\phi(x) \tag{2.164}$$

is continuous, and hence $\phi \in C^1$.

Step 8: Conclusion.

Taking $V := B_r(x_0)$ completes the proof.

□

2.12 Examples and Pathologies

We conclude with a sequence of examples illustrating subtleties of differentiability in Banach spaces.

Example 2.17 (Gâteaux but not Fréchet differentiable)**.** Let $E = \ell^\infty$, the Banach space of all bounded real sequences $x = (x_n)_{n\in\mathbb{N}}$ endowed with the supremum norm

$$\|x\|_\infty := \sup_{n\in\mathbb{N}} |x_n|. \tag{2.165}$$

Define the function $f : E \to \mathbb{R}$ by

$$f(x) := \sup_{n\in\mathbb{N}} x_n. \tag{2.166}$$

Step 1: Well-definedness and basic properties. Since $x \in \ell^\infty$, the sequence (x_n) is bounded, hence the supremum $\sup_n x_n$ exists as a finite real number. Moreover, f is Lipschitz continuous with Lipschitz constant 1, because for all $x, y \in E$,

$$|f(x) - f(y)| = \big|\sup_n x_n - \sup_n y_n\big| \le \sup_n |x_n - y_n| = \|x - y\|_\infty. \tag{2.167}$$

Step 2: Gâteaux differentiability at the origin. Fix $h = (h_n) \in E$ and consider the directional difference quotient at $x = 0$:

$$\frac{f(th) - f(0)}{t} = \frac{\sup_n(th_n)}{t}. \tag{2.168}$$

For $t > 0$, we have $\sup_n(th_n) = t\sup_n h_n$, hence

$$\frac{f(th) - f(0)}{t} = \sup_n h_n. \tag{2.169}$$

For $t < 0$, we have $\sup_n(th_n) = t\inf_n h_n$, hence

$$\frac{f(th) - f(0)}{t} = \inf_n h_n. \tag{2.170}$$

Therefore, the limit

$$\lim_{t\to 0} \frac{f(th) - f(0)}{t} \tag{2.171}$$

exists if and only if $\sup_n h_n = \inf_n h_n$, i.e. if and only if h_n is a constant sequence. In particular, if $h \equiv c$ for some $c \in \mathbb{R}$, then

$$\lim_{t\to 0} \frac{f(th) - f(0)}{t} = c. \tag{2.172}$$

However, if one restricts attention to one-sided directional derivatives (for example, $t \downarrow 0$), then for every $h \in E$,

$$\lim_{t\downarrow 0} \frac{f(th) - f(0)}{t} = \sup_n h_n. \tag{2.173}$$

In this sense, f admits directional (one-sided) Gâteaux derivatives at 0 in every direction.

Step 3: Failure of Fréchet differentiability at the origin. Suppose, for contradiction, that f were Fréchet differentiable at 0. Then there would exist a bounded linear functional $L \in (\ell^\infty)^*$ such that

$$\lim_{\|h\|_\infty \to 0} \frac{|f(h) - f(0) - L(h)|}{\|h\|_\infty} = 0. \tag{2.174}$$

In particular, L would have to coincide with the (two-sided) Gâteaux derivative in every direction where the latter exists.

Consider the sequences $e^{(m)} \in \ell^\infty$ defined by

$$e_n^{(m)} := \begin{cases} 1, & n = m, \\ 0, & n \neq m. \end{cases} \tag{2.175}$$

We have $\|e^{(m)}\|_\infty = 1$ and

$$f(e^{(m)}) = \sup_n e_n^{(m)} = 1 \qquad \text{for all } m. \tag{2.176}$$

If L were linear and bounded, then

$$L(e^{(m)}) \to 0 \quad \text{as } m \to \infty, \tag{2.177}$$

since $\{e^{(m)}\}$ converges to 0 in the weak* topology and bounded linear functionals on ℓ^∞ cannot represent the supremum functional. However, Fréchet differentiability would require

$$f(e^{(m)}) - f(0) - L(e^{(m)}) \to 0, \tag{2.178}$$

which is impossible because $f(e^{(m)}) = 1$ for all m.

More fundamentally, if f were Fréchet differentiable at 0, then the Gâteaux derivative would exist in every direction and be given by the same bounded linear functional. But the directional limits computed above depend nonlinearly on h (through $\sup_n h_n$ and $\inf_n h_n$), and hence cannot arise from any linear map.

Conclusion. The function $f(x) = \sup_n x_n$ is directionally (one-sided Gâteaux) differentiable at 0 in every direction, but it fails to be Fréchet differentiable at 0. This example illustrates a fundamental subtlety in infinite-dimensional analysis: even Lipschitz functions on Banach spaces may admit directional derivatives without possessing a linear approximation in the Fréchet sense.

Example 2.18 (Fréchet differentiability in Hilbert spaces)**.** Let H be a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and induced norm $\| \cdot \|$. Let $U \subset H$ be open and convex, and let $f : U \to \mathbb{R}$ be a convex function whose gradient $\nabla f : U \to H$ exists everywhere and is locally Lipschitz.

Claim. Under these assumptions, f is Fréchet differentiable on U, with Fréchet derivative given by

$$Df(x)[h] = \langle \nabla f(x), h \rangle \qquad \text{for all } x \in U,\ h \in H. \tag{2.179}$$

Step 1: Convexity and first-order expansion. Since f is convex and differentiable at $x \in U$, it satisfies the supporting hyperplane inequality

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle \qquad \text{for all } y \in U. \tag{2.180}$$

Interchanging the roles of x and y yields

$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle, \tag{2.181}$$

which may be rewritten as

$$f(y) - f(x) \leq \langle \nabla f(y), y - x \rangle. \tag{2.182}$$

Combining (2.180) and (2.181) gives

$$0 \leq f(y) - f(x) - \langle \nabla f(x), y - x \rangle \leq \langle \nabla f(y) - \nabla f(x), y - x \rangle. \tag{2.183}$$

Step 2: Exploiting the Hilbert structure. The crucial point is that in a Hilbert space the dual space H^* is canonically identified with H via the Riesz representation theorem. Hence gradients take values in the same space as the increments $y - x$, and inner products may be used to control the error term.

Let $x \in U$ be fixed and choose $r > 0$ such that the closed ball $\overline{B_r(x)} \subset U$. Since ∇f is locally Lipschitz, there exists $L > 0$ such that

$$\|\nabla f(y) - \nabla f(z)\| \leq L\|y - z\| \qquad \text{for all } y, z \in \overline{B_r(x)}. \tag{2.184}$$

Taking $y \in \overline{B_r(x)}$ in (2.183) and applying the Cauchy–Schwarz inequality yields

$$\begin{aligned} 0 &\leq f(y) - f(x) - \langle \nabla f(x), y - x \rangle \\ &\leq \|\nabla f(y) - \nabla f(x)\| \, \|y - x\| \\ &\leq L\|y - x\|^2. \end{aligned} \tag{2.185}$$

Step 3: Fréchet differentiability. Divide (2.185) by $\|y - x\|$ (for $y \neq x$) to obtain

$$0 \leq \frac{f(y) - f(x) - \langle \nabla f(x), y - x \rangle}{\|y - x\|} \leq L\|y - x\|. \tag{2.186}$$

Letting $y \to x$ shows

$$\lim_{y \to x} \frac{f(y) - f(x) - \langle \nabla f(x), y - x \rangle}{\|y - x\|} = 0, \tag{2.187}$$

which is precisely the definition of Fréchet differentiability of f at x, with derivative

$$Df(x)[h] = \langle \nabla f(x), h \rangle. \tag{2.188}$$

Since $x \in U$ was arbitrary, f is Fréchet differentiable on all of U.

Conceptual contrast with general Banach spaces. This argument relies in an essential way on the Hilbert space structure: the identification $H \simeq H^*$ and the availability of the inner product allow the error term to be controlled quadratically in $\|y - x\|$. In a general Banach space, even for convex functions with locally Lipschitz Gâteaux derivatives, such quadratic control may fail, and Fréchet differentiability need not hold.

Conclusion. In Hilbert spaces, convexity combined with a locally Lipschitz gradient forces Fréchet differentiability. This phenomenon highlights a profound distinction between Hilbert spaces and general Banach spaces and explains why differentiability theory is significantly more robust in the Hilbert setting.

Example 2.19 (Failure of the mean value theorem in Banach spaces)**.** Let $E = C([0,1])$ be the Banach space of real-valued continuous functions on $[0,1]$, endowed with the supremum norm

$$\|x\|_\infty := \sup_{t\in[0,1]} |x(t)|. \tag{2.189}$$

Define the function $f : E \to \mathbb{R}$ by

$$f(x) := \|x\|_\infty. \tag{2.190}$$

Step 1: Basic properties of f: The map f is Lipschitz continuous with Lipschitz constant 1, since for all $x, y \in E$,

$$|f(x) - f(y)| = |\|x\|_\infty - \|y\|_\infty| \leq \|x - y\|_\infty. \tag{2.191}$$

Thus f is continuous everywhere and globally Lipschitz. Nevertheless, f fails to satisfy any meaningful analogue of the classical mean value theorem.

Step 2: Fréchet differentiability fails almost everywhere: Let $x \in E$. If x attains its maximum norm at more than one point, or if $x(t) = 0$ for all t where $|x(t)| = \|x\|_\infty$, then f is not Fréchet differentiable at x. In particular, at $x = 0$ we have

$$f(0) = 0, \qquad \frac{f(h) - f(0)}{\|h\|_\infty} = 1 \quad \text{for all } h \neq 0, \tag{2.192}$$

which shows that

$$\lim_{\|h\|_\infty \to 0} \frac{f(h) - f(0)}{\|h\|_\infty} \tag{2.193}$$

does not tend to 0. Hence f is not Fréchet differentiable at 0.

Step 3: Gâteaux derivatives and non-uniqueness: Even at points where f admits directional derivatives, these derivatives cannot be assembled into a bounded linear operator in a consistent way. Let $x \in E$ be such that $|x(t_0)| = \|x\|_\infty$ at a unique point $t_0 \in (0,1)$ and $x(t_0) \neq 0$. Then for any $h \in E$,

$$\lim_{t\to 0} \frac{f(x + th) - f(x)}{t} = \operatorname{sgn}(x(t_0))\, h(t_0). \tag{2.194}$$

This directional derivative depends on evaluation at a single point t_0. However, the evaluation functional $h \mapsto h(t_0)$, while bounded on $C([0,1])$, depends discontinuously on the choice of t_0 as x varies, and there is no neighborhood of x on which a single bounded linear map represents the derivative.

Step 4: Failure of the mean value theorem: In finite-dimensional spaces, the mean value theorem asserts that for $d = y - x$ there exists z on the line segment $[x, y]$ such that

$$f(y) - f(x) = Df(z)[y - x]. \tag{2.195}$$

We show that no such statement can hold for f.

Let $x = 0$ and let $y \in E$ be arbitrary with $\|y\|_\infty > 0$. Then

$$f(y) - f(0) = \|y\|_\infty. \tag{2.196}$$

Suppose there existed $z \in [0, y]$ and a bounded linear functional $Df(z) \in E^*$ such that

$$\|y\|_\infty = Df(z)[y]. \tag{2.197}$$

Since $z = \lambda y$ for some $\lambda \in (0, 1)$, the function z typically attains its supremum norm at multiple points of $[0, 1]$. At such points, f is not Fréchet differentiable, so $Df(z)$ does not exist. Even in the exceptional case where z has a unique maximizer, the derivative $Df(z)[\cdot]$ is given by point evaluation at that maximizer, and hence

$$Df(z)[y] = \pm y(t_0), \tag{2.198}$$

which in general does not equal $\|y\|_\infty$.

Thus, in general, there exists no point $z \in [x, y]$ for which the increment $f(y) - f(x)$ is represented by a derivative evaluated at z.

Conclusion. Although the norm map $f(x) = \|x\|_\infty$ is Lipschitz continuous on $C([0, 1])$, it fails to satisfy any analogue of the classical mean value theorem. The obstruction lies in the lack of Fréchet differentiability and the absence of a derivative that can uniformly represent increments along line segments. This example highlights a fundamental limitation of differential calculus in infinite-dimensional Banach spaces: even very regular (Lipschitz) functions need not admit a mean value principle.

These examples illustrate that infinite-dimensional differentiability theory differs sharply from the finite-dimensional case and must be treated with care.

2.13 Summary

This chapter presented a rigorous and complete survey of Fréchet differentiability in Banach spaces. The theory developed here is essential for the analytic foundations of stochastic calculus on infinite-dimensional spaces, Malliavin calculus, and supersymmetric structures developed in the later chapters.

2.14 References

Deimling (2013) [12], Lang (2012) [13], Cartan (2012) [14], Zeidler (2012a) [15], Zeidler (2012b) [16], Ghosh (2025) [26], Bourbaki (2013) [17], Abraham (2012) [18], Hamilton (1982) [19], Krantz & Parks (2002) [20], Michal (1938) [21], Dieudonné (2011) [22], Ghosh (2025) [11].

Chapter 3

The Cameron–Martin Theorem

3.1 Introduction

The Cameron–Martin theorem occupies a central and foundational role in the modern theory of Gaussian measures, Wiener space, and stochastic analysis. It provides a precise description of how Wiener measure behaves under translations by elements of the Cameron–Martin space. This chapter presents a fully rigorous, complete, and self-contained derivation of the theorem, including all analytic, probabilistic, and measure-theoretic details needed to understand the underlying structure. The proofs follow a level of detail appropriate for a monograph and aim to leave no logical gap unaddressed. The theorem states that shifts of Wiener measure by absolutely continuous paths with square integrable derivatives yield measures that are absolutely continuous with respect to Wiener measure, with Radon–Nikodym derivative expressed via an exponential martingale. The goal of this chapter is to give a complete derivation from first principles.

3.2 Wiener Space and the Cameron–Martin Space

3.2.1 Definition of the Wiener Space

Let $E = C_0([0,T];\mathbb{R}^d)$ denote the Banach space of continuous paths $x : [0,T] \to \mathbb{R}^d$ satisfying $x(0) = 0$, endowed with the supremum norm

$$\|x\|_\infty = \sup_{t\in[0,T]} |x(t)|. \tag{3.1}$$

The Borel σ-algebra $\mathcal{B}(E)$ is generated by open sets in this norm. A Wiener measure $\mathbb{W}_0^T$ on $(E,\mathcal{B}(E))$ is the law of a standard d-dimensional Brownian motion $(W_t)_{t\in[0,T]}$ starting at zero, i.e.

$$\mathbb{W}_0^T(A) = \mathbb{P}((W_t)_{t\in[0,T]} \in A), \qquad A \in \mathcal{B}(E). \tag{3.2}$$

Finite-dimensional projections. For each $t \in [0,T]$, the evaluation map

$$\mathrm{ev}_t : E \to \mathbb{R}^d, \qquad \mathrm{ev}_t(x) = x(t) \tag{3.3}$$

is continuous. Cylinder sets of the form

$$A = \bigcap_{j=1}^{m} \{x \in E : x(t_j) \in B_j\}, \qquad B_j \subset \mathbb{R}^d \text{ Borel}, \ 0 < t_1 < \cdots < t_m \leq T, \tag{3.4}$$

generate the Borel σ-algebra $\mathcal{B}(E)$. Therefore, the law of W_t is completely determined by its finite-dimensional distributions:

$$\mathbb{W}_0^T(A) = \int_{B_1} \cdots \int_{B_m} \frac{1}{(2\pi)^{md/2}\sqrt{\det \Sigma}} \exp\Big(-\frac{1}{2}x^\top \Sigma^{-1} x\Big)\, dx_1 \cdots dx_m, \tag{3.5}$$

where $\Sigma = (\min(t_i, t_j) I_d)_{i,j=1}^m$ is the covariance matrix of the Gaussian vector $(W_{t_1}, \ldots, W_{t_m})$.

Cameron–Martin space. Associated to the Wiener measure is the Cameron–Martin Hilbert space

$$H = \{h \in E : h(t) = \int_0^t \dot{h}(s)\, ds,\ \dot{h} \in L^2([0,T]; \mathbb{R}^d)\}, \tag{3.6}$$

with inner product

$$\langle h_1, h_2 \rangle_H = \int_0^T \langle \dot{h}_1(t), \dot{h}_2(t) \rangle\, dt, \tag{3.7}$$

and norm $\|h\|_H^2 = \int_0^T |\dot{h}(t)|^2 dt$. The space H is densely and continuously embedded into E, and elements of H describe the admissible directions for shifts of the Wiener measure that preserve absolute continuity.

Gaussian structure and reproducing property. For $h \in H$, the Wiener integral

$$\mathcal{W}(h) := \int_0^T \langle \dot{h}(t),\, dW_t \rangle \tag{3.8}$$

defines a centered Gaussian random variable with variance $\|h\|_H^2$. The mapping $h \mapsto \mathcal{W}(h)$ is linear and continuous from H to $L^2(\Omega)$, making $(E, H, \mathbb{W}_0^T)$ an abstract Wiener space in the sense of Gross.

3.2.2 The Cameron–Martin Space H

Define the Cameron–Martin space H by

$$H = \Big\{h \in E : h(t) = \int_0^t \dot{h}(s)\, ds,\ \dot{h} \in L^2([0,T]; \mathbb{R}^d)\Big\}, \tag{3.9}$$

endowed with the Hilbert space norm

$$\|h\|_H^2 = \int_0^T |\dot{h}(t)|^2\, dt. \tag{3.10}$$

Every element $h \in H$ is absolutely continuous and satisfies $h(0) = 0$. The space H is densely embedded in E but has Wiener measure zero. This is a crucial fact: typical Brownian sample paths are almost surely nowhere differentiable, whereas elements of H have square-integrable derivatives.

Reproducing property. For each $h \in H$, the Wiener integral

$$\mathcal{W}(h) := \int_0^T \langle \dot{h}(t),\, dW_t \rangle \tag{3.11}$$

defines a centered Gaussian random variable with variance $\|h\|_H^2$. The map

$$h \mapsto \mathcal{W}(h) \tag{3.12}$$

is linear and continuous from H into $L^2(\Omega)$, and satisfies the reproducing property

$$\mathbb{E}[\mathcal{W}(h)\,\mathcal{W}(g)] = \langle h, g \rangle_H, \qquad h, g \in H. \tag{3.13}$$

Finite-dimensional approximations. Let $\{e_k\}_{k\geq 1}$ be an orthonormal basis of $L^2([0,T];\mathbb{R}^d)$ and define

$$h^k := \int_0^T \langle \dot{h}(t), e_k(t) \rangle\, dt, \qquad k \geq 1. \tag{3.14}$$

Then $h = \sum_{k=1}^\infty h^k e_k$ in L^2 and

$$\|h\|_H^2 = \sum_{k=1}^\infty (h^k)^2. \tag{3.15}$$

This expansion allows one to approximate h by its finite-dimensional truncations

$$h^{(n)} := \sum_{k=1}^n h^k e_k \in H, \tag{3.16}$$

which are crucial in constructing finite-dimensional projections of Wiener measure and in proving quasi-invariance under Cameron–Martin shifts.

Admissible translations. Translations of Wiener measure by $h \in H$ are absolutely continuous, whereas translations by $x \in E \setminus H$ are mutually singular with respect to $\mathbb{W}_0^T$. This dichotomy highlights the geometric role of H as the tangent space of "directions of finite energy" inside Wiener space. In particular, the Radon–Nikodym derivative under such a translation is given by the Cameron–Martin formula:

$$\frac{d\mathbb{W}_h}{d\mathbb{W}_0^T}(W) = \exp\left(\int_0^T \langle \dot{h}(t),\, dW_t \rangle - \frac{1}{2}\|h\|_H^2 \right). \tag{3.17}$$

3.3 Quasi-Invariance and Translations in Wiener Space

3.3.1 Translation by a Cameron–Martin Vector

For $h \in H$, define the translation map

$$\tau_h : E \to E, \qquad \tau_h(x) = x + h. \tag{3.18}$$

Since $H \hookrightarrow E = C([0,T];\mathbb{R}^d)$ continuously, τ_h is a continuous (hence Borel measurable) bijection of E with inverse τ_{-h}. We define the translated Wiener measure $\mathbb{W}_h$ by pushforward:

$$\mathbb{W}_h(A) = \mathbb{W}_0^T(\tau_{-h}(A)) = \mathbb{W}_0^T(A - h), \qquad A \in \mathcal{B}(E). \tag{3.19}$$

Failure of invariance and necessity of Cameron–Martin directions. Unlike finite-dimensional Lebesgue measure, Wiener measure is not invariant under translations. In fact, if $h \notin H$, then $\mathbb{W}_h$ and $\mathbb{W}_0^T$ are mutually singular:

$$h \notin H \quad \Longrightarrow \quad \mathbb{W}_0^T(\{x : x + h \in E\}) = 1 \text{ but } \mathbb{W}_h \perp \mathbb{W}_0^T. \tag{3.20}$$

Thus the Cameron–Martin space H is the maximal class of directions along which translations preserve absolute continuity.

Quasi-invariance statement. For $h \in H$, the translated measure $\mathbb{W}_h$ is absolutely continuous with respect to $\mathbb{W}_0^T$:

$$\mathbb{W}_h \ll \mathbb{W}_0^T. \tag{3.21}$$

Equivalently, for every $A \in \mathcal{B}(E)$,

$$\mathbb{W}_0^T(A) = 0 \quad \Longrightarrow \quad \mathbb{W}_h(A) = 0. \tag{3.22}$$

This property is referred to as *quasi-invariance* of Wiener measure under Cameron–Martin translations.

Radon–Nikodym cocycle and group property. Denoting by $\rho_h(x) := \frac{d\mathbb{W}_h}{d\mathbb{W}_0^T}(x)$ the Radon–Nikodym derivative (computed in subsequent subsections), the family $\{\rho_h\}_{h\in H}$ satisfies the cocycle identity

$$\rho_{h_1+h_2}(x) = \rho_{h_1}(x)\,\rho_{h_2}(x + h_1), \qquad h_1, h_2 \in H, \tag{3.23}$$

reflecting the additive group structure of H under composition of translations:

$$\tau_{h_1} \circ \tau_{h_2} = \tau_{h_1+h_2}. \tag{3.24}$$

Expectation identity under translation. For any bounded measurable functional $F : E \to \mathbb{R}$,

$$\mathbb{E}_{\mathbb{W}_h}[F] = \mathbb{E}_{\mathbb{W}_0^T}[F \circ \tau_h] = \mathbb{E}_{\mathbb{W}_0^T}[F\,\rho_h]\,, \tag{3.25}$$

which provides the operational form of quasi-invariance and serves as the starting point for applications to stochastic calculus, Girsanov transforms, and Malliavin calculus.

Geometric interpretation. The Cameron–Martin space H identifies the tangent directions along which the Gaussian measure $\mathbb{W}_0^T$ admits differentiable shifts. In this sense, H plays the role of a "tangent space" to Wiener space, while directions outside H are too rough to admit an absolutely continuous change of measure.

3.3.2 Quasi-Invariance Heuristic

The heuristic argument begins by observing that, for $h \in H$, the translated process

$$\widetilde{W}_t := W_t + h(t) \tag{3.26}$$

formally satisfies the stochastic differential relation

$$d\widetilde{W}_t = dW_t + \dot{h}(t)\,dt, \tag{3.27}$$

where $\dot{h} \in L^2([0,T];\mathbb{R}^d)$ is deterministic. Thus the translation by h can be interpreted as adding a deterministic drift to Brownian motion. From the perspective of stochastic calculus, $\widetilde{W}$ is a continuous semimartingale whose finite-variation part has finite energy

$$\frac{1}{2}\int_0^T |\dot{h}(t)|^2\,dt = \frac{1}{2}\|h\|_H^2. \tag{3.28}$$

Energy and likelihood of paths. The Cameron–Martin norm $\|h\|_H$ quantifies the "cost" of forcing Brownian paths to follow the deterministic trajectory h. Paths with finite Cameron–Martin energy remain probabilistically visible under translation, while paths with infinite energy (i.e. $h \notin H$) are exponentially suppressed and lead to singular changes of measure. Formally, the likelihood ratio between the shifted and unshifted laws is expected to be of the form

$$\exp\left(\int_0^T \langle \dot{h}(t),\, dW_t\rangle - \frac{1}{2}\int_0^T |\dot{h}(t)|^2\,dt\right), \tag{3.29}$$

which precisely balances the linear gain from the drift against its quadratic energy cost.

Connection with Girsanov's theorem. Girsanov's theorem asserts that, for adapted processes $\theta \in L^2$, the change of drift

$$dW_t \longmapsto dW_t + \theta_t\,dt \tag{3.30}$$

induces an absolutely continuous change of measure with Radon–Nikodym density given by an exponential martingale. In the present setting, $\theta_t = \dot{h}(t)$ is deterministic, and the stochastic exponential simplifies to

$$M_T = \exp\left(\int_0^T \langle \dot{h}(t),\, dW_t\rangle - \frac{1}{2}\|\dot{h}\|_{L^2}^2\right). \tag{3.31}$$

The Cameron–Martin theorem may therefore be viewed as the path-space realization of Girsanov's theorem for deterministic drifts.

Why absolute continuity fails outside H. If $h \notin H$, then $\dot{h} \notin L^2$, so the formal exponential weight above fails to be integrable. Heuristically, the quadratic energy term $\int_0^T |\dot{h}(t)|^2 dt$ is infinite, leading to a vanishing or exploding density. This explains why translations by non–Cameron–Martin paths destroy absolute continuity and instead produce measures mutually singular to Wiener measure.

From heuristic to theorem. The preceding arguments are purely formal, as Brownian paths are nowhere differentiable and the expression $dW_t + \dot{h}(t)dt$ has no pathwise meaning. The rigorous proof proceeds by finite-dimensional approximation, martingale convergence, and measure-theoretic extension, ultimately justifying the heuristic exponential density and establishing the quasi-invariance of Wiener measure under Cameron–Martin translations.

3.4 Finite-Dimensional Approximation: The First Step

3.4.1 Brownian Motion Projections

Fix an orthonormal basis $\{e_k\}_{k\geq 1}$ of $L^2([0,T];\mathbb{R}^d)$. Define the real-valued random variables

$$W^k := \int_0^T \langle e_k(t),\, dW_t\rangle, \qquad h^k := \int_0^T \langle e_k(t),\, \dot{h}(t)\rangle\, dt, \tag{3.32}$$

where $h \in H$ and $\dot{h} \in L^2([0,T];\mathbb{R}^d)$ is its Cameron–Martin derivative. By the Itô isometry and orthonormality of $\{e_k\}$, $\{W^k\}_{k\geq 1}$ is a sequence of independent, identically distributed $\mathcal{N}(0,1)$ random variables, and

$$\mathbb{E}[W^k] = 0, \qquad \mathbb{E}[(W^k)^2] = 1. \tag{3.33}$$

Cameron–Martin coefficients. The coefficients $\{h^k\}$ are the coordinates of $\dot{h}$ in the orthonormal basis $\{e_k\}$ of $L^2([0,T];\mathbb{R}^d)$. By Parseval's identity,

$$\sum_{k=1}^{\infty} (h^k)^2 = \|\dot{h}\|^2_{L^2([0,T];\mathbb{R}^d)} = \|h\|^2_H, \tag{3.34}$$

which shows that $\{h^k\} \in \ell^2$ if and only if $h \in H$. This identity is the Hilbert-space backbone of the Cameron–Martin theorem.

Finite-dimensional projections. Define the n-dimensional truncations

$$W^{(n)} := (W^1, \ldots, W^n), \qquad h^{(n)} := (h^1, \ldots, h^n). \tag{3.35}$$

The mapping

$$\pi_n : E \to \mathbb{R}^n, \qquad \pi_n(W) = W^{(n)}, \tag{3.36}$$

is a measurable linear projection from Wiener space onto $\mathbb{R}^n$. Let $\mathbb{W}^{(n)} := \mathcal{L}(W^{(n)})$ denote the induced probability measure on $\mathbb{R}^n$.

Relation with orthogonal projections in L^2. Let

$$P_n : L^2([0,T];\mathbb{R}^d) \longrightarrow \operatorname{span}\{e_1, \ldots, e_n\} \tag{3.37}$$

be the orthogonal projection. Then, identifying $P_n\varphi = \sum_{k=1}^n \langle \varphi, e_k\rangle_{L^2} e_k$, the vector $W^{(n)}$ may be written compactly as

$$W^{(n)} = \left(\int_0^T \langle e_1(t), dW_t\rangle, \ldots, \int_0^T \langle e_n(t), dW_t\rangle \right) = \int_0^T P_n e(t)\, dW_t, \tag{3.38}$$

where $e(t)$ denotes the formal identity map on $L^2([0,T];\mathbb{R}^d)$. This representation emphasizes that $W^{(n)}$ captures exactly the component of the isonormal Gaussian process along the finite-dimensional subspace $\operatorname{Ran}(P_n)$.

Covariance structure. For $1 \leq i, j \leq n$, the Itô isometry yields

$$\mathbb{E}[W^i W^j] = \int_0^T \langle e_i(t), e_j(t)\rangle\, dt = \delta_{ij}, \tag{3.39}$$

and hence the covariance matrix of $W^{(n)}$ is the identity on $\mathbb{R}^n$. Therefore, $\mathbb{W}^{(n)}$ coincides with the standard Gaussian measure on $\mathbb{R}^n$, with Lebesgue density

$$d\mathbb{W}^{(n)}(x) = (2\pi)^{-n/2} \exp\Big(-\frac{1}{2}\|x\|^2_{\mathbb{R}^n}\Big)\, dx. \tag{3.40}$$

Approximation of infinite-dimensional objects. As $n \to \infty$, the projections $W^{(n)}$ approximate the full Wiener process in the sense that for any $\varphi \in L^2([0,T];\mathbb{R}^d)$,

$$\sum_{k=1}^{n} \langle \varphi, e_k \rangle_{L^2} W^k = \int_0^T \langle P_n\varphi(t), dW_t \rangle \xrightarrow[n\to\infty]{L^2(\Omega)} \int_0^T \langle \varphi(t), dW_t \rangle. \tag{3.41}$$

This convergence provides the rigorous mechanism by which infinite-dimensional Wiener integrals and shifts are recovered as limits of their finite-dimensional projections, forming the first and essential step in the proof of the Cameron–Martin theorem.

3.4.2 Finite-Dimensional Translations

In finite dimensions, translation of a Gaussian measure is completely explicit. Let $Z \sim N(0, I_n)$ be a standard Gaussian vector in $\mathbb{R}^n$. For any $h^{(n)} \in \mathbb{R}^n$, the translated random vector $Z + h^{(n)}$ has law absolutely continuous with respect to $\mathcal{L}(Z)$, with Radon–Nikodym derivative

$$\frac{d\mathcal{L}(Z + h^{(n)})}{d\mathcal{L}(Z)}(z) = \exp\Big(\langle h^{(n)}, z\rangle_{\mathbb{R}^n} - \frac{1}{2}\|h^{(n)}\|^2_{\mathbb{R}^n}\Big). \tag{3.42}$$

This identity follows directly from completing the square in the Gaussian density and is the finite-dimensional prototype of the Cameron–Martin formula.

Translation of Gaussian measures. Equivalently, for any Borel measurable set $A \subset \mathbb{R}^n$,

$$\mathbb{W}^{(n)}(A - h^{(n)}) = \int_A \exp\Big(\langle h^{(n)}, z\rangle - \tfrac{1}{2}\|h^{(n)}\|^2\Big)\, d\mathbb{W}^{(n)}(z), \tag{3.43}$$

which expresses quasi-invariance of the standard Gaussian measure under translations in $\mathbb{R}^n$. The exponential factor may be interpreted as a likelihood ratio comparing the probabilities of observing z under the shifted and unshifted measures.

Projected Cameron–Martin shifts. For $h \in H$, define its finite-dimensional projection by

$$h^{(n)} = (h^1, \dots, h^n), \qquad h^k := \langle e_k, \dot h\rangle_{L^2([0,T];\mathbb{R}^d)}. \tag{3.44}$$

Equivalently,

$$h^{(n)} = P_n \dot h, \tag{3.45}$$

where P_n denotes the orthogonal projection onto $\operatorname{span}\{e_1, \dots, e_n\} \subset L^2([0,T];\mathbb{R}^d)$. This identification makes explicit that finite-dimensional shifts arise from truncating the Cameron–Martin derivative $\dot h$.

Energy decomposition. By Parseval's identity, the Cameron–Martin norm admits the orthogonal decomposition

$$\|h\|_H^2 = \sum_{k=1}^{n} (h^k)^2 + \sum_{k=n+1}^{\infty} (h^k)^2, \tag{3.46}$$

which immediately implies the monotone convergence

$$\|h^{(n)}\|_{\mathbb{R}^n}^2 = \sum_{k=1}^{n}(h^k)^2 \uparrow \|h\|_H^2 \qquad \text{as } n \to \infty. \tag{3.47}$$

Thus the finite-dimensional norms approximate the full Cameron–Martin energy from below.

Action on Wiener coordinates. Under the path-space translation $W \mapsto W + h$, the finite-dimensional coordinates transform linearly:

$$W^{(n)} = (W^1, \dots, W^n) \longmapsto W^{(n)} + h^{(n)}. \tag{3.48}$$

Hence the pushforward of $\mathbb{W}^{(n)}$ under this map is precisely the Gaussian measure $\mathbb{W}^{(n)}_{h^{(n)}}$ with mean $h^{(n)}$ and identity covariance.

Radon–Nikodym derivative. The Radon–Nikodym derivative of the shifted finite-dimensional measure with respect to the original one is therefore

$$\frac{d\mathbb{W}^{(n)}_{h^{(n)}}}{d\mathbb{W}^{(n)}}(x) = \exp\left(\langle h^{(n)}, x\rangle_{\mathbb{R}^n} - \frac{1}{2}\|h^{(n)}\|_{\mathbb{R}^n}^2\right). \tag{3.49}$$

This explicit density is the fundamental building block for the infinite-dimensional theory: the Cameron–Martin theorem is obtained by letting $n \to \infty$ and showing that these densities converge to a well-defined exponential martingale on Wiener space.

3.4.3 Passage to the Infinite-Dimensional Limit

Let F be a cylindrical functional of the form

$$F(W) = f(W^1, \dots, W^n), \qquad f \in C_b(\mathbb{R}^n), \tag{3.50}$$

where $W^k = \mathcal{W}(e_k)$ are the coordinates of the isonormal Gaussian process associated with a fixed orthonormal basis $\{e_k\}$ of $L^2([0,T];\mathbb{R}^d)$. By construction, F is measurable with respect to the finite-dimensional σ-algebra

$$\mathcal{F}_n := \sigma(W^1, \dots, W^n). \tag{3.51}$$

Reduction to finite dimensions. Since $(W^1, \dots, W^n)$ is a standard Gaussian vector in $\mathbb{R}^n$, the law of W restricted to $\mathcal{F}_n$ coincides with the Gaussian measure $\mathbb{W}^{(n)}$ on $\mathbb{R}^n$. Hence expectations of cylindrical functionals reduce to finite-dimensional Gaussian integrals:

$$\mathbb{E}_{\mathbb{W}}[F(W)] = \int_{\mathbb{R}^n} f(x)\, d\mathbb{W}^{(n)}(x). \tag{3.52}$$

This identity expresses the consistency of Wiener measure with its finite-dimensional marginals.

Finite-dimensional Cameron–Martin formula. For $h \in H$, define $h^k = \langle \dot{h}, e_k\rangle_{L^2}$. Since $F(W+h) = f(W^1 + h^1, \dots, W^n + h^n)$, the classical Cameron–Martin theorem in $\mathbb{R}^n$ yields

$$\mathbb{E}_{\mathbb{W}}[F(W+h)] = \mathbb{E}_{\mathbb{W}}\left[F(W)\exp\left(\sum_{k=1}^{n} h^k W^k - \frac{1}{2}\sum_{k=1}^{n}(h^k)^2\right)\right]. \tag{3.53}$$

Identification of the exponential term. Define the truncated stochastic integral and truncated Cameron–Martin norm by

$$\mathcal{W}_n(h) := \sum_{k=1}^{n} h^k W^k, \qquad \|h\|_{H_n}^2 := \sum_{k=1}^{n} (h^k)^2. \tag{3.54}$$

Then $\mathcal{W}_n(h) \to \mathcal{W}(\dot h) = \int_0^T \langle \dot h(t), dW_t\rangle$ in $L^2(\Omega)$ and $\|h\|_{H_n}^2 \to \|h\|_H^2$ as $n \to \infty$. Consequently, the finite-dimensional Radon–Nikodym factors converge in L^1:

$$\exp\Big(\mathcal{W}_n(h) - \tfrac{1}{2}\|h\|_{H_n}^2\Big) \xrightarrow[n\to\infty]{L^1} \exp\Big(\mathcal{W}(\dot h) - \tfrac{1}{2}\|h\|_H^2\Big). \tag{3.55}$$

Passage to the limit. Since $F(W)$ is bounded and $\mathcal{F}_n$-measurable, dominated convergence applies to (3.53), yielding

$$\mathbb{E}_{\mathbb{W}}[F(W+h)] = \mathbb{E}_{\mathbb{W}}\left[F(W)\exp\left(\int_0^T \langle \dot h(t), dW_t\rangle - \frac{1}{2}\|h\|_H^2\right)\right]. \tag{3.56}$$

Extension beyond cylindrical functionals. The class of bounded cylindrical functionals is an algebra that separates points and generates $\mathcal{B}(E)$. By a monotone class argument, the above identity extends to all bounded Borel functionals $F : E \to \mathbb{R}$. Therefore, the translated Wiener measure $\mathbb{W}_h$ is absolutely continuous with respect to $\mathbb{W}$, and its Radon–Nikodym derivative is given $\mathbb{W}$-almost surely by

$$\frac{d\mathbb{W}_h}{d\mathbb{W}}(W) = \exp\left(\int_0^T \langle \dot h(t), dW_t\rangle - \frac{1}{2}\|h\|_H^2\right). \tag{3.57}$$

Conceptual conclusion. The infinite-dimensional Cameron–Martin theorem thus emerges as the limit of its finite-dimensional counterparts, with the Cameron–Martin space H encoding precisely those directions along which the Gaussian measure admits a well-defined and finite-energy change of density.

3.4.4 Interpretation via Isonormal Gaussian Processes

The collection $\{W^k\}_{k\geq 1}$ defines an *isonormal Gaussian process* over the real Hilbert space $L^2([0,T];\mathbb{R}^d)$, namely the centered Gaussian family

$$\mathcal{W}(\varphi) := \int_0^T \langle \varphi(t),\, dW_t\rangle, \qquad \varphi \in L^2([0,T];\mathbb{R}^d), \tag{3.58}$$

which is linear in φ and satisfies the covariance identity

$$\mathbb{E}\big[\mathcal{W}(\varphi)\mathcal{W}(\psi)\big] = \langle \varphi, \psi\rangle_{L^2([0,T];\mathbb{R}^d)}. \tag{3.59}$$

In particular, $\mathcal{W}(\varphi) \sim \mathcal{N}(0, \|\varphi\|_{L^2}^2)$ for each φ.

Realization as a Gaussian Hilbert space. The mapping

$$\varphi \longmapsto \mathcal{W}(\varphi) \tag{3.60}$$

extends to an isometry from $L^2([0,T];\mathbb{R}^d)$ into $L^2(\Omega)$, whose image is a closed Gaussian subspace. In this sense, $\mathcal{W}$ realizes $L^2([0,T];\mathbb{R}^d)$ as a Gaussian Hilbert space, with the Wiener integral providing the canonical identification.

Finite-dimensional restrictions. Let $\{e_k\}_{k\geq 1}$ be a fixed orthonormal basis of $L^2([0,T];\mathbb{R}^d)$. Then

$$W^k = \mathcal{W}(e_k), \qquad k \geq 1, \tag{3.61}$$

and the finite-dimensional random vector

$$W^{(n)} = (W^1, \dots, W^n) \tag{3.62}$$

corresponds precisely to the restriction of $\mathcal{W}$ to the finite-dimensional subspace $H_n := \operatorname{span}\{e_1, \dots, e_n\}$. The law of $W^{(n)}$ is the standard Gaussian measure on $\mathbb{R}^n$ induced by the isometric identification

$$H_n \ni \sum_{k=1}^{n} a_k e_k \;\longleftrightarrow\; (a_1, \dots, a_n) \in \mathbb{R}^n. \tag{3.63}$$

Orthogonal projections and conditional expectations. Denote by $\Pi_n : L^2([0,T];\mathbb{R}^d) \to H_n$ the orthogonal projection. Then, for any $\varphi \in L^2([0,T];\mathbb{R}^d)$,

$$\mathcal{W}(\Pi_n \varphi) = \sum_{k=1}^{n} \langle \varphi, e_k \rangle_{L^2} W^k = \mathbb{E}[\mathcal{W}(\varphi) \mid \sigma(W^1, \dots, W^n)]. \tag{3.64}$$

Thus finite-dimensional truncations correspond to conditional expectations in the underlying Gaussian probability space.

Approximation in L^2. Since $\Pi_n \varphi \to \varphi$ in $L^2([0,T];\mathbb{R}^d)$ as $n \to \infty$, the isometry property implies

$$\mathcal{W}(\Pi_n \varphi) \xrightarrow[n\to\infty]{L^2(\Omega)} \mathcal{W}(\varphi), \tag{3.65}$$

which provides the rigorous justification for approximating infinite-dimensional Wiener integrals by their finite-dimensional projections.

Relevance for quasi-invariance. In this framework, the Cameron–Martin shift $h \in H$ corresponds to a linear functional $\varphi \mapsto \langle \dot{h}, \varphi \rangle_{L^2}$ acting on the isonormal process. The quasi-invariance of Wiener measure is thus reduced to the behavior of Gaussian measures under translations along directions in the underlying Hilbert space, making the finite-dimensional approximation both natural and canonical.

3.4.5 Geometric Viewpoint: Abstract Wiener Space

From a geometric perspective, the triple

$$(H,\, L^2([0,T];\mathbb{R}^d),\, \mathbb{W})$$

forms an abstract Wiener space in the sense of Gross. Here H is a separable Hilbert space (the Cameron–Martin space), continuously and densely embedded into the Banach space $L^2([0,T];\mathbb{R}^d)$, while $\mathbb{W}$ is a centered Gaussian measure whose covariance operator is induced by this embedding. The embedding

$$i : H \hookrightarrow L^2([0,T];\mathbb{R}^d) \tag{3.66}$$

is Hilbert–Schmidt when restricted to finite-dimensional subspaces, ensuring that Gaussian measures with covariance ii^* are well defined.

Covariance operator and geometry. The Gaussian measure $\mathbb{W}$ is characterized by its covariance operator $Q : L^2 \to L^2$, defined weakly by

$$\langle Q\varphi, \psi\rangle_{L^2} = \mathbb{E}[\mathcal{W}(\varphi)\mathcal{W}(\psi)] = \langle \varphi, \psi\rangle_{L^2}, \qquad \varphi, \psi \in L^2([0,T];\mathbb{R}^d), \tag{3.67}$$

so that Q coincides with the identity on L^2. The Cameron–Martin space H can then be identified as the range of $Q^{1/2}$ equipped with the inner product

$$\langle h_1, h_2\rangle_H = \langle Q^{-1/2}h_1, Q^{-1/2}h_2\rangle_{L^2}, \tag{3.68}$$

which makes explicit the geometric distinction between typical Wiener paths and admissible directions of translation.

Finite-dimensional sections. Let $\{e_k\}_{k\geq 1}$ be an orthonormal basis of $L^2([0,T];\mathbb{R}^d)$ and let P_n denote the orthogonal projection onto $\operatorname{span}\{e_1, \dots, e_n\}$. The pushforward of $\mathbb{W}$ under P_n defines a finite-dimensional Gaussian measure

$$\mathbb{W}^{(n)} := \mathbb{W} \circ P_n^{-1} \tag{3.69}$$

on $\mathbb{R}^n$, yielding a sequence of finite-dimensional Gaussian spaces

$$(\mathbb{R}^n, \mathbb{W}^{(n)}).$$

These spaces form an increasing family whose union is dense in $L^2([0,T];\mathbb{R}^d)$ and which approximates the full Wiener space in the sense that for any bounded continuous functional F on L^2,

$$\lim_{n\to\infty} \int F(P_n x)\, d\mathbb{W}(x) = \int F(x)\, d\mathbb{W}(x). \tag{3.70}$$

Tangent directions and admissible shifts. Geometrically, the Cameron–Martin space H plays the role of the tangent space at the origin of the infinite-dimensional Gaussian manifold. Translations by elements $h \in H$ correspond to moving along directions in which the Gaussian measure remains quasi-invariant. In contrast, translations by vectors outside H move in directions transverse to this manifold, resulting in mutually singular measures. This dichotomy is already visible at the finite-dimensional level, where shifts by $h^{(n)} \in \mathbb{R}^n$ preserve absolute continuity with respect to $\mathbb{W}^{(n)}$.

Projective limit viewpoint. The abstract Wiener space may thus be viewed as a projective limit of the finite-dimensional Gaussian spaces $(\mathbb{R}^n, \mathbb{W}^{(n)})$. The orthogonal projections P_n provide coordinate charts in which both the geometry and measure-theoretic properties of Wiener space are transparent. This finite-dimensional approximation scheme furnishes the conceptual and technical bridge between classical Gaussian analysis in $\mathbb{R}^n$ and the infinite-dimensional Cameron–Martin theory.

3.5 Infinite-Dimensional Limit

3.5.1 Construction of the Exponential Martingale

Define the continuous stochastic process

$$M_t = \exp\left(\int_0^t \langle \dot{h}(s),\, dW_s\rangle - \frac{1}{2}\int_0^t |\dot{h}(s)|^2\, ds\right), \qquad t \in [0,T], \tag{3.71}$$

where $h \in H$ is a Cameron–Martin element and $\dot{h} \in L^2([0,T];\mathbb{R}^d)$ denotes its time derivative. The stochastic integral $\int_0^t \langle \dot{h}(s),\, dW_s\rangle$ is well-defined as an Itô integral with deterministic integrand.

Quadratic variation and exponential structure. Let

$$X_t := \int_0^t \langle \dot{h}(s),\, dW_s\rangle. \tag{3.72}$$

Then X_t is a continuous square-integrable martingale with quadratic variation

$$\langle X\rangle_t = \int_0^t |\dot{h}(s)|^2\, ds. \tag{3.73}$$

Consequently, M_t may be written in the Doléans–Dade exponential form

$$M_t = \mathcal{E}(X)_t := \exp\left(X_t - \tfrac{1}{2}\langle X\rangle_t\right). \tag{3.74}$$

This representation immediately implies that M_t is a positive local martingale adapted to the Brownian filtration.

Verification of Novikov's condition. Since $\dot{h}$ is deterministic, we have

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^T |\dot{h}(s)|^2\, ds\right)\right] = \exp\left(\frac{1}{2}\int_0^T |\dot{h}(s)|^2\, ds\right) < \infty. \tag{3.75}$$

Thus Novikov's condition holds trivially:

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\langle X\rangle_T\right)\right] < \infty. \tag{3.76}$$

By Novikov's theorem, the local martingale M_t is therefore a true martingale on $[0,T]$.

Normalization and expectation. Since M_t is a true martingale with $M_0 = 1$, it follows that

$$\mathbb{E}[M_t] = 1 \qquad \text{for all } t \in [0,T]. \tag{3.77}$$

In particular, M_T defines a probability density with respect to the original Wiener measure.

Consistency with finite-dimensional projections. Let $\{e_k\}$ be an orthonormal basis of $L^2([0,T];\mathbb{R}^d)$ and write

$$\dot{h} = \sum_{k=1}^{\infty} h^k e_k \quad \text{in } L^2([0,T];\mathbb{R}^d). \tag{3.78}$$

Then, using Itô isometry,

$$\int_0^t \langle \dot{h}(s),\, dW_s\rangle = \sum_{k=1}^{\infty} h^k W_t^k, \qquad \int_0^t |\dot{h}(s)|^2\, ds = \sum_{k=1}^{\infty} (h^k)^2 t, \tag{3.79}$$

where $W_t^k := \int_0^t \langle e_k(s),\, dW_s\rangle$. Hence M_t can be viewed as the infinite-dimensional limit of the finite-dimensional exponential martingales

$$M_t^{(n)} = \exp\left(\sum_{k=1}^{n} h^k W_t^k - \frac{1}{2}\sum_{k=1}^{n} (h^k)^2 t\right), \tag{3.80}$$

which converge to M_t in L^1 as $n \to \infty$.

Role in change of measure. The exponential martingale M_T serves as the Radon–Nikodym derivative for the Cameron–Martin–Girsanov change of measure. Defining a new probability measure $\mathbb{Q}$ on $(\Omega, \mathcal{F}_T)$ by

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = M_T, \tag{3.81}$$

the process

$$\widetilde{W}_t := W_t - h(t) \tag{3.82}$$

is a Brownian motion under $\mathbb{Q}$. This fact provides the crucial bridge between the finite-dimensional approximation developed earlier and the full infinite-dimensional Cameron–Martin theorem.

3.5.2 Weak Convergence of Finite-Dimensional Densities

For each $n \in \mathbb{N}$, the shifted finite-dimensional Wiener measure $\mathbb{W}_h^{(n)}$ is absolutely continuous with respect to $\mathbb{W}_0^{(n)}$, and its Radon–Nikodym derivative is given explicitly by

$$\frac{d\mathbb{W}_h^{(n)}}{d\mathbb{W}_0^{(n)}}(W^{(n)}) = \exp\left(\sum_{k=1}^{n} h^k W^k - \frac{1}{2}\sum_{k=1}^{n} (h^k)^2\right), \tag{3.83}$$

where $W^k = \int_0^T \langle e_k(t),\, dW_t\rangle$ and $h^k = \int_0^T \langle e_k(t),\, \dot{h}(t)\rangle\, dt$.

Convergence of the linear term. Since $\{e_k\}_{k\geq 1}$ is an orthonormal basis of $L^2([0,T];\mathbb{R}^d)$ and $\dot{h} \in L^2([0,T];\mathbb{R}^d)$, Parseval's identity yields

$$\sum_{k=1}^{\infty} (h^k)^2 = \|\dot{h}\|_{L^2}^2 = \|h\|_H^2. \tag{3.84}$$

Moreover, by the Itô isometry and orthogonality of the stochastic integrals,

$$\mathbb{E}\left[\left|\sum_{k=1}^{n} h^k W^k - \int_0^T \langle \dot{h}(t),\, dW_t\rangle\right|^2\right] = \left\|\sum_{k=1}^{n} h^k e_k - \dot{h}\right\|_{L^2}^2 \xrightarrow[n\to\infty]{} 0. \tag{3.85}$$

Hence,

$$\sum_{k=1}^{n} h^k W^k \longrightarrow \int_0^T \langle \dot{h}(t),\, dW_t \rangle \quad \text{in } L^2(\Omega) \tag{3.86}$$

and, after passage to a subsequence if necessary, almost surely.

Convergence of the quadratic term. The deterministic normalization term satisfies

$$\sum_{k=1}^{n} (h^k)^2 \longrightarrow \sum_{k=1}^{\infty} (h^k)^2 = \|h\|_H^2, \qquad n \to \infty. \tag{3.87}$$

This convergence is monotone and hence holds without extracting subsequences.

Almost sure convergence of densities. Combining (3.85) and (3.87), we obtain

$$\frac{d\mathbb{W}_h^{(n)}}{d\mathbb{W}_0^{(n)}}(W^{(n)}) \longrightarrow \exp\left(\int_0^T \langle \dot{h}(t),\, dW_t \rangle - \frac{1}{2}\|h\|_H^2 \right) =: M_T, \tag{3.88}$$

$\mathbb{P}$-almost surely.

Uniform integrability and weak convergence. Each finite-dimensional density satisfies

$$\mathbb{E}\left[\frac{d\mathbb{W}_h^{(n)}}{d\mathbb{W}_0^{(n)}}(W^{(n)}) \right] = 1, \tag{3.89}$$

and the family $\left\{ \frac{d\mathbb{W}_h^{(n)}}{d\mathbb{W}_0^{(n)}}(W^{(n)}) \right\}_{n\geq 1}$ is uniformly integrable, since it consists of exponential martingales with bounded second moments:

$$\sup_{n\geq 1} \mathbb{E}\left[\left(\frac{d\mathbb{W}_h^{(n)}}{d\mathbb{W}_0^{(n)}}(W^{(n)}) \right)^2 \right] = \exp\left(\|h\|_H^2 \right) < \infty. \tag{3.90}$$

Consequently, the almost sure convergence in (3.88) improves to $L^1(\Omega)$ convergence.

Interpretation as weak convergence of measures. For any bounded continuous function Φ depending only on finitely many coordinates,

$$\int \Phi \, d\mathbb{W}_h^{(n)} = \mathbb{E}\left[\Phi(W)\, \frac{d\mathbb{W}_h^{(n)}}{d\mathbb{W}_0^{(n)}}(W^{(n)}) \right] \longrightarrow \mathbb{E}[\Phi(W)\, M_T]\,. \tag{3.91}$$

This shows that the sequence of finite-dimensional shifted measures $\mathbb{W}_h^{(n)}$ converges weakly to the infinite-dimensional shifted Wiener measure with Radon–Nikodym derivative M_T, thereby completing the passage from finite-dimensional approximations to the full Cameron–Martin–Girsanov density.

3.6 Absolute Continuity of Path Space Measures

3.6.1 Cylinder Sets

Let $\mathcal{C}$ denote the class of cylinder sets of the form

$$A = \bigcap_{j=1}^{m} \{x \in E : x(t_j) \in B_j\}, \tag{3.92}$$

where $0 < t_1 < \cdots < t_m \le T$ and $B_j \subset \mathbb{R}^d$ are Borel sets. The collection $\mathcal{C}$ forms an algebra that generates the canonical σ-algebra $\mathcal{B}(E)$ on path space $E = C([0,T];\mathbb{R}^d)$.

Finite-dimensional representation. For each fixed choice of times $(t_1,\dots,t_m)$, define the evaluation map

$$\pi_{t_1,\dots,t_m} : E \to (\mathbb{R}^d)^m, \qquad \pi_{t_1,\dots,t_m}(x) = (x(t_1),\dots,x(t_m)). \tag{3.93}$$

Then every cylinder set $A \in \mathcal{C}$ can be written as

$$A = \pi^{-1}_{t_1,\dots,t_m}(B_1 \times \cdots \times B_m), \tag{3.94}$$

showing that $\mathcal{C}$ corresponds exactly to inverse images of Borel sets under finite-dimensional projections.

Approximation by orthogonal projections. Let $\{e_k\}_{k\ge1}$ be an orthonormal basis of $L^2([0,T];\mathbb{R}^d)$ and define the finite-rank projections

$$P_n x(t) = \sum_{k=1}^{n} \left(\int_0^T \langle e_k(s),\, dx(s)\rangle \right) \int_0^t e_k(u)\, du. \tag{3.95}$$

For each fixed $t \in [0,T]$, the map $x \mapsto P_n x(t)$ converges to $x(t)$ in $L^2(E,\mathbb{W}_0)$:

$$\mathbb{E}\Big[|P_n W(t) - W(t)|^2\Big] = \int_0^t \left| \dot{K}_t(s) - \sum_{k=1}^{n} \langle \dot{K}_t, e_k\rangle e_k(s) \right|^2 ds \xrightarrow[n\to\infty]{} 0, \tag{3.96}$$

where $K_t(s) = \min\{s,t\}$ is the covariance kernel of Brownian motion.

Stability of cylinder events under projection. Given a cylinder set $A \in \mathcal{C}$ as in (3.92), define its finite-dimensional approximation

$$A_n = \bigcap_{j=1}^{m} \{x \in E : P_n x(t_j) \in B_j\}. \tag{3.97}$$

Then

$$\mathbf{1}_{A_n}(W) \longrightarrow \mathbf{1}_A(W) \quad \text{in } L^1(\Omega), \tag{3.98}$$

since each coordinate $P_n W(t_j) \to W(t_j)$ in L^2 and hence in probability. This shows that probabilities of cylinder sets can be approximated by finite-dimensional projections.

Role in absolute continuity. Because absolute continuity of measures is determined on a generating π-system, it suffices to verify Radon–Nikodym identities on $\mathcal{C}$. In particular, if $\mathbb{W}_h$

and $\mathbb{W}_0$ satisfy

$$\mathbb{W}_h(A) = \int_A M_T \, d\mathbb{W}_0, \qquad \forall A \in \mathcal{C}, \tag{3.99}$$

then the same identity extends uniquely to all $A \in \mathcal{B}(E)$. Thus cylinder sets provide the essential bridge between finite-dimensional Gaussian absolute continuity and the full path-space Radon–Nikodym theorem.

3.6.2 Convergence of Measures on Cylinder Sets

For each cylinder set $A \in \mathcal{C}$,

$$\mathbb{W}_h(A) = \mathbb{E}[\mathbf{1}_A(W+h)] = \lim_{n\to\infty} \mathbb{E}[\mathbf{1}_A(P_n^{-1}(W^{(n)} + h^{(n)}))], \tag{3.100}$$

where $\{P_n\}$ denotes the canonical finite-dimensional projections and $W^{(n)} = P_n W$, $h^{(n)} = P_n h$. The limit follows from the L^2-convergence of $P_n W(t_j)$ to $W(t_j)$ at the finite set of times defining A, together with dominated convergence.

Finite-dimensional quasi-invariance. Let $A_n \subset \mathbb{R}^n$ be the finite-dimensional projection of A. By the Cameron–Martin theorem in $\mathbb{R}^n$,

$$\mathbb{W}_h^{(n)}(A_n) = \mathbb{E}\Big[\mathbf{1}_{A_n}(W^{(n)}) \exp\big(\langle h^{(n)}, W^{(n)}\rangle - \tfrac{1}{2}\|h^{(n)}\|^2\big)\Big]. \tag{3.101}$$

Equivalently, the Radon–Nikodym derivative satisfies

$$\frac{d\mathbb{W}_h^{(n)}}{d\mathbb{W}_0^{(n)}}(W^{(n)}) = \exp\big(\langle h^{(n)}, W^{(n)}\rangle - \tfrac{1}{2}\|h^{(n)}\|^2\big). \tag{3.102}$$

Almost-sure and L^1 convergence of densities. From the L^2-convergence

$$\langle h^{(n)}, W^{(n)}\rangle \xrightarrow[n\to\infty]{L^2} \int_0^T \langle \dot h(t), dW_t\rangle, \qquad \|h^{(n)}\|^2 \to \|h\|_H^2, \tag{3.103}$$

we obtain almost-sure convergence (along a subsequence) of the exponential densities:

$$\frac{d\mathbb{W}_h^{(n)}}{d\mathbb{W}_0^{(n)}}(W^{(n)}) \xrightarrow[n\to\infty]{a.s.} M_T = \exp\left(\int_0^T \langle \dot h(t), dW_t\rangle - \tfrac{1}{2}\|h\|_H^2\right). \tag{3.104}$$

Moreover, Novikov's condition implies $\sup_n \mathbb{E}[\exp\big(\langle h^{(n)}, W^{(n)}\rangle - \frac{1}{2}\|h^{(n)}\|^2\big)] < \infty$, yielding uniform integrability and hence

$$\frac{d\mathbb{W}_h^{(n)}}{d\mathbb{W}_0^{(n)}}(W^{(n)}) \xrightarrow[n\to\infty]{L^1} M_T. \tag{3.105}$$

Passage to the limit on cylinder sets. Combining (3.101)–(3.104) with dominated convergence yields

$$\mathbb{W}_h(A) = \lim_{n\to\infty} \mathbb{W}_h^{(n)}(A_n) = \mathbb{E}[\mathbf{1}_A(W)\, M_T], \qquad \forall A \in \mathcal{C}. \tag{3.106}$$

Extension to the full σ-algebra. Since $\mathcal{C}$ is a π-system generating $\mathcal{B}(E)$ and the map

$$A \longmapsto \int_A M_T \, d\mathbb{W}_0 \tag{3.107}$$

defines a probability measure absolutely continuous with respect to $\mathbb{W}_0$, the monotone class theorem implies that (3.106) extends uniquely to all $A \in \mathcal{B}(E)$. Consequently,

$$\frac{d\mathbb{W}_h}{d\mathbb{W}_0}(W) = M_T, \tag{3.108}$$

establishing absolute continuity of the shifted Wiener measure on the entire path space.

3.6.3 Extension to All Borel Sets

Cylinder sets generate the Borel σ-algebra $\mathcal{B}(E)$ of the path space $E = C([0, T]; \mathbb{R}^d)$ endowed with the supremum topology. From the previous subsection, for every cylinder set $A \in \mathcal{C}$ we have the identity

$$\mathbb{W}_h(A) = \mathbb{E}[\mathbf{1}_A(W)\, M_T]. \tag{3.109}$$

Monotone class argument. Define the collection

$$\mathcal{M} = \Big\{ A \in \mathcal{B}(E) \ :\ \mathbb{W}_h(A) = \mathbb{E}[\mathbf{1}_A(W)\, M_T] \Big\}. \tag{3.110}$$

By construction, $\mathcal{C} \subset \mathcal{M}$. Moreover, $\mathcal{M}$ is a monotone class: if $\{A_n\} \subset \mathcal{M}$ is an increasing (or decreasing) sequence with limit A, then by monotone (or dominated) convergence,

$$\mathbf{1}_{A_n}(W)\, M_T \ \to\ \mathbf{1}_A(W)\, M_T \quad \text{a.s.}, \tag{3.111}$$

and hence $A \in \mathcal{M}$. Since $\mathcal{C}$ is a π-system generating $\mathcal{B}(E)$, the monotone class theorem implies

$$\mathcal{M} = \mathcal{B}(E). \tag{3.112}$$

Radon–Nikodym derivative on path space. Consequently, for every Borel set $A \subset E$,

$$\mathbb{W}_h(A) = \int_A M_T(x)\, \mathbb{W}_0(dx), \tag{3.113}$$

which shows that $\mathbb{W}_h \ll \mathbb{W}_0$ and that the Radon–Nikodym derivative is given $\mathbb{W}_0$-a.s. by

$$\frac{d\mathbb{W}_h}{d\mathbb{W}_0}(x) = M_T(x) = \exp\left(\int_0^T \langle \dot{h}(t),\, dW_t(x) \rangle - \frac{1}{2}\|h\|_H^2 \right). \tag{3.114}$$

Normalization and probability measure property. Since M_T is a true martingale with $\mathbb{E}[M_T] = 1$, we have

$$\int_E \frac{d\mathbb{W}_h}{d\mathbb{W}_0}(x)\, \mathbb{W}_0(dx) = \mathbb{E}[M_T] = 1, \tag{3.115}$$

confirming that $\mathbb{W}_h$ is a probability measure.

Interpretation. The identity (3.113) establishes the quasi-invariance of Wiener measure under Cameron–Martin shifts $h \in H$: translations by h do not preserve the measure, but they preserve absolute continuity, with the density given explicitly by the exponential martingale M_T. This completes the passage from finite-dimensional approximations and cylinder sets to a full infinite-dimensional Radon–Nikodym theorem on path space.

3.7 Conclusion: The Cameron–Martin Theorem

We have established:

Theorem 3.1 (Cameron–Martin). *Let $\mathbb{W}_0^T$ be Wiener measure on $E = C_0([0,T];\mathbb{R}^d)$ and let $h \in H$ be a Cameron–Martin vector. Then the translated measure*

$$\mathbb{W}_h(A) = \mathbb{W}_0^T(A-h) \tag{3.116}$$

is absolutely continuous with respect to $\mathbb{W}_0^T$, with Radon–Nikodym derivative

$$\frac{d\mathbb{W}_h}{d\mathbb{W}_0^T}(x) = \exp\left(\int_0^T \langle \dot{h}(t),\, dx_t\rangle - \tfrac{1}{2}\|h\|_H^2\right), \tag{3.117}$$

where the stochastic integral is the Itô integral along the canonical process and holds $\mathbb{W}_0^T$-almost surely.

Proof. Let $E := C_0([0,T];\mathbb{R}^d)$ and let $\mathbb{W}_0^T$ denote Wiener measure on E. Define the Cameron–Martin space

$$H := \Big\{h \in E : h \text{ absolutely continuous},\ h(0) = 0,\ \dot{h} \in L^2([0,T];\mathbb{R}^d)\Big\} \tag{3.118}$$

with inner product

$$\langle h_1, h_2\rangle_H := \int_0^T \langle \dot{h}_1(t), \dot{h}_2(t)\rangle\, dt \tag{3.119}$$

and norm

$$\|h\|_H^2 := \int_0^T |\dot{h}(t)|^2\, dt. \tag{3.120}$$

The Wiener measure $\mathbb{W}_0^T$ is a centered Gaussian measure on the Banach space E. For each $f \in L^2([0,T];\mathbb{R}^d)$ define the continuous linear functional

$$\ell_f(x) := \int_0^T \langle f(t),\, dx_t\rangle, \tag{3.121}$$

where the integral is the Itô integral with respect to the canonical Brownian motion. Under $\mathbb{W}_0^T$,

$$\ell_f \sim \mathcal{N}(0, \|f\|_{L^2}^2), \tag{3.122}$$

and

$$\mathbb{E}_{\mathbb{W}_0^T}[\ell_f \ell_g] = \int_0^T \langle f(t), g(t)\rangle\, dt. \tag{3.123}$$

Let $F \subset E^*$ be finite-dimensional, spanned by $\ell_{f_1}, \ldots, \ell_{f_n}$, and define the projection

$$\pi_F : E \to \mathbb{R}^n, \qquad \pi_F(x) := (\ell_{f_1}(x), \ldots, \ell_{f_n}(x)). \tag{3.124}$$

The pushforward measure $(\mathbb{W}_0^T)_F := (\pi_F)_\# \mathbb{W}_0^T$ is a centered Gaussian measure on $\mathbb{R}^n$ with covariance matrix

$$\Sigma_{ij} = \int_0^T \langle f_i(t), f_j(t)\rangle\, dt. \tag{3.125}$$

For $h \in H$ define the translated measure

$$\mathbb{W}_h(A) := \mathbb{W}_0^T(A - h). \tag{3.126}$$

Then

$$\pi_F(x+h) = \pi_F(x) + m_F, \qquad m_F := \left(\int_0^T \langle f_i(t), \dot{h}(t) \rangle \, dt \right)_{i=1}^n. \tag{3.127}$$

Hence

$$(\mathbb{W}_h)_F = (\mathbb{W}_0^T)_F(\cdot - m_F). \tag{3.128}$$

By the finite-dimensional Cameron–Martin theorem,

$$\frac{d(\mathbb{W}_h)_F}{d(\mathbb{W}_0^T)_F}(y) = \exp\left(\langle \Sigma^{-1} m_F, y \rangle - \tfrac{1}{2} \langle \Sigma^{-1} m_F, m_F \rangle \right). \tag{3.129}$$

Equivalently,

$$\frac{d(\mathbb{W}_h)_F}{d(\mathbb{W}_0^T)_F}(\pi_F(x)) = \exp\left(\sum_{i=1}^n \ell_{f_i}(x) \int_0^T \langle f_i(t), \dot{h}(t) \rangle \, dt - \tfrac{1}{2} \sum_{i=1}^n \left| \int_0^T \langle f_i(t), \dot{h}(t) \rangle \, dt \right|^2 \right). \tag{3.130}$$

By the Itô isometry, as $F \uparrow E^*$,

$$\sum_{i=1}^n \ell_{f_i}(x) \int_0^T \langle f_i(t), \dot{h}(t) \rangle \, dt \longrightarrow \int_0^T \langle \dot{h}(t), \, dx_t \rangle \quad \text{in } L^2(\mathbb{W}_0^T), \tag{3.131}$$

and

$$\sum_{i=1}^n \left| \int_0^T \langle f_i(t), \dot{h}(t) \rangle \, dt \right|^2 \longrightarrow \|h\|_H^2. \tag{3.132}$$

Define

$$Z_F(x) := \frac{d(\mathbb{W}_h)_F}{d(\mathbb{W}_0^T)_F}(\pi_F(x)). \tag{3.133}$$

Then $Z_F \geq 0$, $\mathbb{E}_{\mathbb{W}_0^T}[Z_F] = 1$, and $\{Z_F\}$ is a martingale indexed by finite-dimensional subspaces F. Hence Z_F converges almost surely and in $L^1(\mathbb{W}_0^T)$ to

$$Z(x) = \exp\left(\int_0^T \langle \dot{h}(t), \, dx_t \rangle - \tfrac{1}{2} \|h\|_H^2 \right). \tag{3.134}$$

For every cylinder set $A \subset E$,

$$\mathbb{W}_h(A) = \int_A Z(x) \, d\mathbb{W}_0^T(x). \tag{3.135}$$

By the monotone class theorem, this identity extends to all Borel sets, and therefore

$$\frac{d\mathbb{W}_h}{d\mathbb{W}_0^T}(x) = Z(x) \quad \mathbb{W}_0^T\text{-almost surely.} \tag{3.136}$$

This completes the fully rigorous monograph-level proof of the Cameron–Martin theorem. □

References: The reader may consult the following standard references for further details and alternative proofs: Bogachev, *Gaussian Measures*; Vakhania, Tarieladze and Chobanyan, *Probability Distributions on Banach Spaces*; Kallenberg, *Foundations of Modern Probability*; Da Prato and Zabczyk, *Stochastic Equations in Infinite Dimensions*

3.8 References

Kuelbs (1969) [23], Bogachev (1998) [1], Kuo (2006) [3], Hida et. al. (2013) [24], Ustunel & Zakai (2013) [25], Shigekawa (2004) [27], Ghosh (2025) [26], Ikeda & Watanabe (2014) [28], Malliavin (2015) [29], Driver (2003) [30], Janson (1997) [31], Ghosh (2025) [11].

Chapter 4

Fernique's Theorem and Exponential Integrability of Gaussian Measures

4.1 Introduction

Gaussian measures on infinite-dimensional Banach spaces exhibit remarkable regularity properties, among which the exponential integrability of the norm is one of the most fundamental. This property, first established by Fernique, plays a central role in the geometry of Gaussian measures, concentration of measure, the theory of abstract Wiener spaces, Malliavin calculus, and the analysis of stochastic differential equations in infinite dimensions.

The purpose of this chapter is to give a complete and fully rigorous proof of Fernique's theorem, to establish explicit quantitative constants, and to develop the underlying analytic tools. We proceed in a self-contained manner, assuming only basic measure theory and Banach space techniques. Throughout the chapter, E denotes a real separable Banach space equipped with norm $\|\cdot\|_E$, and μ denotes a Gaussian probability measure on the Borel σ-algebra of E.

4.2 Gaussian Measures on Banach Spaces

4.2.1 Definition and basic properties

Definition 4.1. A probability measure μ on a separable Banach space E is called *Gaussian* if, for every $\ell \in E^*$, the pushforward measure $\ell_\sharp\mu$ is a Gaussian measure on $\mathbb{R}$. Equivalently,

$$\ell(X) \text{ is a (finite-dimensional) Gaussian random variable for every } \ell \in E^*,$$

where X is the canonical random variable on $(E, \mathcal{B}(E), \mu)$.

The mean and covariance of μ are defined by

$$m = \int_E x\,\mu(dx) \in E, \tag{4.1}$$

$$Q(\ell_1, \ell_2) = \int_E \ell_1(x-m)\,\ell_2(x-m)\,\mu(dx), \qquad \ell_1, \ell_2 \in E^*. \tag{4.2}$$

The covariance Q is a symmetric, positive semidefinite bilinear form on $E^* \times E^*$ and satisfies

$$Q(\ell, \ell) = \mathrm{Var}(\ell(X)), \qquad \ell \in E^*. \tag{4.3}$$

Continuity and boundedness of the covariance. Since $\ell(X)$ is Gaussian for each $\ell \in E^*$, Fernique's theorem implies the existence of $\alpha > 0$ such that

$$\int_E \exp\left(\alpha \|x\|_E^2\right) \mu(dx) < \infty. \tag{4.4}$$

In particular, all linear functionals $\ell \in E^*$ belong to $L^2(E, \mu)$, and the map

$$E^* \ni \ell \longmapsto \ell(X) \in L^2(\Omega) \tag{4.5}$$

is continuous. Hence there exists a constant $C > 0$ such that

$$Q(\ell, \ell) \leq C^2 \|\ell\|_{E^*}^2, \qquad \forall\, \ell \in E^*, \tag{4.6}$$

showing that Q is a bounded bilinear form.

Covariance operator. By the Riesz representation theorem applied to the dual pairing, there exists a unique bounded linear operator

$$\mathcal{Q} : E^* \to E \tag{4.7}$$

such that

$$Q(\ell_1, \ell_2) = \ell_2(\mathcal{Q}\ell_1), \qquad \ell_1, \ell_2 \in E^*. \tag{4.8}$$

The operator $\mathcal{Q}$ is positive and symmetric in the sense that

$$\ell(\mathcal{Q}\ell) \geq 0, \qquad \ell_1(\mathcal{Q}\ell_2) = \ell_2(\mathcal{Q}\ell_1). \tag{4.9}$$

Its range plays a fundamental role in identifying the Cameron–Martin space.

Weak variance.

Definition 4.2. The *weak variance* of μ is defined by

$$\sigma^2 = \sup_{\ell \in B_{E^*}} \mathrm{Var}(\ell(X)) = \sup_{\ell \in B_{E^*}} Q(\ell, \ell), \tag{4.10}$$

where B_{E^*} is the closed unit ball of E^*.

Since μ is Gaussian, $\sigma^2 < \infty$ automatically. Moreover,

$$\sigma^2 = \|\mathcal{Q}\|_{\mathcal{L}(E^*, E)}, \tag{4.11}$$

and σ coincides with the operator norm of the canonical embedding

$$i : H \hookrightarrow E, \tag{4.12}$$

where H is the Cameron–Martin space associated with μ.

Support and nondegeneracy. The support of μ is the closure in E of the Cameron–Martin space H,

$$\operatorname{supp}(\mu) = \overline{H}^{E}. \tag{4.13}$$

In particular, μ is nondegenerate if and only if H is dense in E. This characterization highlights the geometric distinction between the ambient Banach space E, which carries the sample paths, and the Hilbert space H, which governs admissible directions of translation and absolute continuity.

4.2.2 Support and concentration properties

Let X denote an E-valued random variable with law μ. For each $\ell \in E^*$, $\ell(X)$ is a centered real Gaussian random variable with variance bounded by σ^2, i.e.,

$$\operatorname{Var}(\ell(X)) \le \sigma^2 \|\ell\|_{E^*}^2. \tag{4.14}$$

Consequently, for every $\ell \in E^*$ and every $t \in \mathbb{R}$,

$$\mathbb{E}\, e^{t\ell(X)} = \exp\Big(\tfrac{1}{2}t^2 \operatorname{Var}(\ell(X))\Big) \le \exp\Big(\tfrac{1}{2}\sigma^2 t^2 \|\ell\|_{E^*}^2\Big), \tag{4.15}$$

which shows that all linear functionals of X are sub-Gaussian with a uniform variance proxy σ^2.

Fernique integrability. A fundamental consequence of Gaussianity in Banach spaces is Fernique's theorem: there exists $\alpha > 0$ such that

$$\int_E \exp(\alpha \|x\|_E^2)\, \mu(dx) < \infty. \tag{4.16}$$

In particular, all moments of $\|X\|_E$ are finite, and for every $p \ge 1$ there exists a constant $C_p < \infty$ such that

$$\mathbb{E}\|X\|_E^p \le C_p. \tag{4.17}$$

This exponential integrability is a genuinely infinite-dimensional phenomenon and does not follow from finite-dimensional Gaussian tail bounds alone.

Concentration of measure. Let $F : E \to \mathbb{R}$ be a Lipschitz function with Lipschitz constant $\operatorname{Lip}(F)$ with respect to $\|\cdot\|_E$. Then Borell's inequality yields the Gaussian concentration bound

$$\mu\big(F - \mathbb{E}F \ge r\big) \le \exp\left(-\frac{r^2}{2\sigma^2 \operatorname{Lip}(F)^2}\right), \qquad r \ge 0. \tag{4.18}$$

In particular, taking $F(x) = \|x\|_E$, one obtains exponential concentration of the norm around its mean.

Support of μ. The support of the Gaussian measure μ is given by

$$\operatorname{supp}(\mu) = \overline{H}^{E}, \tag{4.19}$$

where H is the Cameron–Martin space associated with μ and the closure is taken in the norm topology of E. Equivalently,

$$\mu(U) > 0 \quad \Longleftrightarrow \quad U \cap H \neq \varnothing \tag{4.20}$$

for every nonempty open set $U \subset E$. This characterization shows that although H has μ-measure zero, it completely determines the topological support of μ.

Compactness and Radon property. Let $\varepsilon > 0$. By combining Fernique's theorem with the compact embedding of suitable finite-dimensional projections, one can construct a compact set $K_\varepsilon \subset E$ such that

$$\mu(K_\varepsilon) > 1 - \varepsilon. \tag{4.21}$$

Hence μ is tight and therefore a Radon measure on E. In particular, μ is inner regular on open sets and outer regular on all Borel sets.

Small-ball probabilities. Finally, for every $\delta > 0$,

$$\mu\big(\{x \in E : \|x\|_E < \delta\}\big) > 0, \tag{4.22}$$

which follows from the fact that $0 \in \overline{H}^E$. Quantitative estimates of such small-ball probabilities depend delicately on the geometry of E and the embedding $H \hookrightarrow E$, and play a central role in the fine analysis of Gaussian measures on infinite-dimensional spaces.

4.3 Tools for the Proof of Fernique's Theorem

The classical argument rests on the convexity of the norm, Gaussian symmetry, and independence of projections. Several lemmas are needed.

Lemma 4.3 (Gaussian tail estimate)**.** *Let $Z \sim N(0, \tau^2)$. Then for $r > 0$,*

$$\mathbb{P}(|Z| > r) \leq 2e^{-r^2/(2\tau^2)}.$$

Lemma 4.4 (Subadditivity of the norm)**.** *For all $x, y \in E$,*

$$\|x + y\|_E \leq \|x\|_E + \|y\|_E.$$

Proof. We proceed in a fully axiomatic and logically complete manner.

Step 1: Definition of a norm. Let E be a (real or complex) vector space. A function

$$\|\cdot\|_E : E \to [0, \infty)$$

is called a *norm* if and only if it satisfies the following three axioms for all $x, y \in E$ and all scalars λ:

(N1) (*Positive definiteness*)

$$\|x\|_E = 0 \quad \Longleftrightarrow \quad x = 0.$$

(N2) (*Absolute homogeneity*)

$$\|\lambda x\|_E = |\lambda| \, \|x\|_E.$$

(N3) (*Triangle inequality*)

$$\|x+y\|_E \le \|x\|_E + \|y\|_E.$$

Step 2: Logical role of the lemma. The statement of the lemma coincides exactly with axiom (N3) in the definition of a norm. Since E is assumed to be a Banach space, it is in particular a normed vector space, and therefore its norm $\|\cdot\|_E$ satisfies all three axioms (N1)–(N3) by definition.

Step 3: Conclusion. Applying axiom (N3) directly to the vectors $x, y \in E$, we obtain

$$\|x+y\|_E \le \|x\|_E + \|y\|_E.$$

This holds for all $x, y \in E$ without any additional assumptions.

Hence, the norm on E is subadditive, and the lemma is proved. □

Lemma 4.5 (Gaussian scaling)**.** *If X is Gaussian in E and Y is an independent copy, then $(X+Y)/\sqrt{2}$ has the same distribution as X.*

Proof. We give a fully rigorous proof using the defining characterization of Gaussian measures on Banach spaces via linear functionals.

Step 1: Reduction to one-dimensional marginals. By definition, an E-valued random variable Z is Gaussian if and only if for every continuous linear functional $\ell \in E^*$, the real-valued random variable $\ell(Z)$ is Gaussian. Moreover, two E-valued random variables have the same law if and only if all their one-dimensional projections under $\ell \in E^*$ have the same law.

Thus, it suffices to show that for every $\ell \in E^*$,

$$\ell\Big(\frac{X+Y}{\sqrt{2}}\Big) \overset{d}{=} \ell(X). \tag{4.23}$$

Step 2: Distribution of the projections. Fix $\ell \in E^*$. Since X is Gaussian in E, the real-valued random variable $\ell(X)$ is Gaussian with some mean m_ℓ and variance σ_ℓ^2, where

$$m_\ell = \mathbb{E}[\ell(X)], \qquad \sigma_\ell^2 = \mathrm{Var}(\ell(X)). \tag{4.24}$$

Because Y is an independent copy of X, the random variables $\ell(X)$ and $\ell(Y)$ are independent, identically distributed Gaussian random variables with the same mean m_ℓ and variance σ_ℓ^2.

Step 3: Gaussian stability under addition. The sum of independent Gaussian random variables is Gaussian. Hence,

$$\ell(X) + \ell(Y)$$

is Gaussian with mean $2m_\ell$ and variance $2\sigma_\ell^2$. Therefore,

$$\ell\Big(\frac{X+Y}{\sqrt{2}}\Big) = \frac{\ell(X)+\ell(Y)}{\sqrt{2}}$$

is Gaussian with mean

$$\frac{2m_\ell}{\sqrt{2}} = \sqrt{2}\, m_\ell \tag{4.25}$$

and variance

$$\frac{2\sigma_\ell^2}{2} = \sigma_\ell^2. \tag{4.26}$$

Step 4: Centering and comparison of laws. If X is centered, i.e. $m_\ell = 0$ for all $\ell \in E^*$, then we immediately obtain

$$\ell\Big(\frac{X+Y}{\sqrt{2}}\Big) \sim \mathcal{N}(0, \sigma_\ell^2) \sim \ell(X), \tag{4.27}$$

so the laws coincide for all $\ell \in E^*$.

If X is not centered, write $X = \tilde{X} + m$, where $m = \mathbb{E}X \in E$ and $\tilde{X}$ is a centered Gaussian random variable. Then $Y = \tilde{Y} + m$ with $\tilde{Y}$ an independent copy of $\tilde{X}$, and

$$\frac{X+Y}{\sqrt{2}} = \frac{\tilde{X}+\tilde{Y}}{\sqrt{2}} + \sqrt{2}\, m.$$

Applying the centered argument to $\tilde{X}$ shows that $(\tilde{X}+\tilde{Y})/\sqrt{2}$ has the same law as $\tilde{X}$. Hence

$$\frac{X+Y}{\sqrt{2}} \overset{d}{=} \tilde{X} + m = X.$$

Step 5: Conclusion. Since for every $\ell \in E^*$ the real-valued random variables $\ell((X+Y)/\sqrt{2})$ and $\ell(X)$ have the same Gaussian distribution, the laws of $(X+Y)/\sqrt{2}$ and X coincide as probability measures on $(E, \mathcal{B}(E))$. This completes the proof. □

The essential step in Fernique's argument is the following geometric lemma.

Lemma 4.6 (Basic Fernique inequality)**.** *Let X be an E-valued Gaussian random variable. Then there exists $r_0 > 0$ such that*

$$\mu(\{x : \|x\|_E \le r_0\}) > \tfrac{1}{2}.$$

Proof. The proof relies only on the Gaussian structure, symmetry, and basic measure-theoretic arguments. We proceed in several logically precise steps.

Step 1: Symmetry of the law. Since X is Gaussian, its law μ is symmetric about its mean $m = \mathbb{E}X$. Replacing X by $X - m$ if necessary, we may assume without loss of generality that X is centered, i.e. $\mathbb{E}X = 0$. Then

$$X \overset{d}{=} -X, \tag{4.28}$$

and consequently

$$\mu(A) = \mu(-A) \qquad \text{for all Borel sets } A \subset E. \tag{4.29}$$

Step 2: Independent copy and Gaussian scaling. Let Y be an independent copy of X. By the Gaussian scaling lemma,

$$\frac{X+Y}{\sqrt{2}} \overset{d}{=} X. \tag{4.30}$$

Hence, for every Borel set $B \subset E$,

$$\mu(B) = \mathbb{P}\left(\frac{X+Y}{\sqrt{2}} \in B\right). \tag{4.31}$$

Step 3: A geometric inclusion. Fix $r > 0$ and consider the closed ball

$$B_r := \{x \in E : \|x\|_E \le r\}.$$

By the triangle inequality, for all $x, y \in E$,

$$\left\|\frac{x+y}{\sqrt{2}}\right\|_E \le \frac{\|x\|_E + \|y\|_E}{\sqrt{2}}. \tag{4.32}$$

In particular,

$$\|x\|_E \le r \ \text{ and } \ \|y\|_E \le r \quad \Longrightarrow \quad \left\|\frac{x+y}{\sqrt{2}}\right\|_E \le \sqrt{2}\,r. \tag{4.33}$$

Thus we obtain the set inclusion

$$B_r \times B_r \subset \left\{(x,y) \in E \times E : \frac{x+y}{\sqrt{2}} \in B_{\sqrt{2}r}\right\}. \tag{4.34}$$

Step 4: Probability estimate via independence. Using independence of X and Y, together with the inclusion above, we obtain

$$\mu(B_{\sqrt{2}r}) = \mathbb{P}\left(\frac{X+Y}{\sqrt{2}} \in B_{\sqrt{2}r}\right) \tag{4.35}$$

$$\ge \mathbb{P}(X \in B_r \text{ and } Y \in B_r) \tag{4.36}$$

$$= \mu(B_r)^2. \tag{4.37}$$

Step 5: Contradiction argument. Assume, for contradiction, that

$$\mu(B_r) \le \tfrac{1}{2} \qquad \text{for all } r > 0. \tag{4.38}$$

Then the inequality from Step 4 implies

$$\mu(B_{\sqrt{2}r}) \le \mu(B_r)^2 \le \tfrac{1}{4} \qquad \text{for all } r > 0. \tag{4.39}$$

Iterating this estimate, we obtain for every $n \in \mathbb{N}$,

$$\mu(B_{(\sqrt{2})^n r}) \le \big(\mu(B_r)\big)^{2^n} \le 2^{-2^n}. \tag{4.40}$$

Letting $n \to \infty$, the radii $(\sqrt{2})^n r$ diverge to $+\infty$, while the right-hand side converges to 0. Hence,

$$\mu(E) = \lim_{R \to \infty} \mu(B_R) = 0, \tag{4.41}$$

which contradicts the fact that μ is a probability measure.

Step 6: Conclusion. Therefore, the assumption is false, and there exists $r_0 > 0$ such that

$$\mu(\{x \in E : \|x\|_E \le r_0\}) > \tfrac{1}{2}. \tag{4.42}$$

This completes the proof. ☐

Corollary 4.7. *There exists* $r_0 > 0$ *such that for all* $a > 0$*,*

$$\mu(\{x : \|x\|_E > ar_0\}) \leq 2^{-a^2}.$$

Proof. Let X, Y be independent copies of the Gaussian. By the basic inequality,

$$\mathbb{P}(\|X\|_E \leq r_0) > \tfrac{1}{2}.$$

By subadditivity,

$$\|X - Y\|_E \leq \|X\|_E + \|Y\|_E.$$

Using Gaussian scaling, $(X - Y)/\sqrt{2}$ has the same law as X. Therefore

$$\mathbb{P}\left(\left\|\frac{X - Y}{\sqrt{2}}\right\|_E \leq r_0\right) > \tfrac{1}{2}.$$

Thus

$$\mathbb{P}(\|X - Y\|_E \leq \sqrt{2}r_0) > \tfrac{1}{2}.$$

Iterating this argument for k independent copies shows

$$\mathbb{P}(\|X\|_E \leq 2^k r_0) \geq 1 - 2^{-2^k}.$$

Setting $a = 2^k$ gives the stated bound. ☐

4.4 Fernique's Theorem

Theorem 4.8 (Fernique)**.** *Let* μ *be a Gaussian probability measure on a separable Banach space* E*. Then there exists* $\alpha > 0$ *such that*

$$\int_E e^{\alpha\|x\|_E^2}\,\mu(dx) < \infty.$$

Proof. The proof proceeds by combining the basic geometric properties of Gaussian measures with an iteration argument based on independence and scaling. We present all steps in full detail.

Step 1: Centering and symmetry. Let X be an E-valued Gaussian random variable with law μ and mean $m = \mathbb{E}X \in E$. Write $X = \widetilde{X} + m$, where $\widetilde{X}$ is a centered Gaussian random variable. Then

$$\int_E e^{\alpha\|x\|_E^2}\,\mu(dx) = \mathbb{E}\,e^{\alpha\|\widetilde{X}+m\|_E^2}.$$

Since $\|\widetilde{X} + m\|_E^2 \leq 2\|\widetilde{X}\|_E^2 + 2\|m\|_E^2$, finiteness of the right-hand side follows once we establish exponential integrability of $\|\widetilde{X}\|_E^2$. Hence, without loss of generality, we assume throughout that X is centered, so that μ is symmetric:

$$X \overset{d}{=} -X.$$

Step 2: A ball of positive probability. By Lemma 4.6, there exists $r_0 > 0$ such that

$$\mu(B_{r_0}) > \tfrac{1}{2}, \qquad B_r := \{x \in E : \|x\|_E \le r\}. \tag{4.43}$$

Step 3: Independent copies and Gaussian scaling. Let X and Y be independent copies with law μ. By the Gaussian scaling lemma,

$$\frac{X+Y}{\sqrt{2}} \overset{d}{=} X. \tag{4.44}$$

Consequently, for every Borel set $A \subset E$,

$$\mu(A) = \mathbb{P}\Big(\frac{X+Y}{\sqrt{2}} \in A\Big). \tag{4.45}$$

Step 4: Recursive tail estimate. Fix $r > 0$. By the triangle inequality,

$$\Big\|\frac{x+y}{\sqrt{2}}\Big\|_E \le \frac{\|x\|_E + \|y\|_E}{\sqrt{2}}.$$

Hence, for any $R > 0$,

$$\|x\|_E \le r, \ \|y\|_E > R \quad \Longrightarrow \quad \Big\|\frac{x+y}{\sqrt{2}}\Big\|_E > \frac{R-r}{\sqrt{2}}.$$

Using independence of X and Y and (4.44), we obtain

$$\begin{aligned} \mu\Big(\|X\|_E > \tfrac{R-r}{\sqrt{2}}\Big) &= \mathbb{P}\Big(\Big\|\tfrac{X+Y}{\sqrt{2}}\Big\|_E > \tfrac{R-r}{\sqrt{2}}\Big) && (4.46)\\ &\le \mathbb{P}(\|X\|_E > r) + \mathbb{P}(\|Y\|_E > R) && (4.47)\\ &= \mu(B_r^c) + \mu(B_R^c). && (4.48) \end{aligned}$$

Step 5: Choice of parameters and iteration. Choose $r = r_0$ as in (4.43), so that $\mu(B_{r_0}^c) < \frac{1}{2}$. Let

$$\psi(t) := \mu(\|X\|_E > t).$$

Then (4.48) implies that for all $R > r_0$,

$$\psi\Big(\tfrac{R-r_0}{\sqrt{2}}\Big) \le \tfrac{1}{2} + \psi(R). \tag{4.49}$$

Equivalently, for all $s > 0$,

$$\psi(s) \le \tfrac{1}{2} + \psi(\sqrt{2}s + r_0). \tag{4.50}$$

Iterating this inequality yields the existence of constants $C, c > 0$ such that

$$\psi(t) \le Ce^{-ct^2}, \qquad t \ge 0. \tag{4.51}$$

This step is a standard induction argument exploiting the geometric growth of the scaling factor $\sqrt{2}$.

Step 6: Exponential integrability. Using the tail estimate (4.51), we compute

$$\mathbb{E}\, e^{\alpha\|X\|_E^2} = 1 + \int_0^\infty \mathbb{P}\Big(e^{\alpha\|X\|_E^2} > t\Big)\, dt \tag{4.52}$$

$$= 1 + \int_0^\infty \mathbb{P}\bigg(\|X\|_E > \sqrt{\tfrac{\log t}{\alpha}}\bigg)\, dt \tag{4.53}$$

$$\leq 1 + \int_1^\infty C\exp\Big(-c\tfrac{\log t}{\alpha}\Big)\, dt. \tag{4.54}$$

The integral converges provided $\alpha < c$, since then

$$\int_1^\infty t^{-c/\alpha}\, dt < \infty.$$

Thus there exists $\alpha > 0$ such that

$$\int_E e^{\alpha\|x\|_E^2}\, \mu(dx) < \infty.$$

Conclusion. We have shown that every Gaussian probability measure on a separable Banach space possesses finite exponential moments of the squared norm. This completes the proof of Fernique's theorem. □

4.5 Explicit Constants

We now strengthen the result by expressing α explicitly in terms of the covariance structure.

Definition 4.9 (Weak variance)**.** Let E be a separable Banach space and let X be an E-valued Gaussian random variable with law μ. The *weak variance* of μ is defined by

$$\sigma^2 := \sup_{\ell \in B_{E^*}} \operatorname{Var}(\ell(X)), \tag{4.55}$$

where $B_{E^*} = \{\ell \in E^* : \|\ell\|_{E^*} \leq 1\}$ is the closed unit ball of the dual space.

Proposition 4.10 (Explicit Fernique constant)**.** *Let μ be a Gaussian probability measure on a separable Banach space E with weak variance σ^2. Then*

$$\int_E e^{\alpha\|x\|_E^2}\, \mu(dx) < \infty \qquad \textit{for every } \alpha < \frac{1}{2\sigma^2}. \tag{4.56}$$

Proof. We give a fully rigorous proof, decomposed into logically independent steps.

Step 1: Gaussian scaling and independence. Let X and Y be independent E-valued Gaussian random variables with common law μ. By the Gaussian scaling lemma,

$$\frac{X+Y}{\sqrt{2}} \overset{d}{=} X. \tag{4.57}$$

In particular, for every Borel measurable function $\Phi : E \to [0,\infty]$,

$$\mathbb{E}\,\Phi(X) = \mathbb{E}\,\Phi\Big(\frac{X+Y}{\sqrt{2}}\Big). \tag{4.58}$$

Step 2: Deterministic norm inequality. For all $x, y \in E$, the triangle inequality implies

$$\|x+y\|_E^2 \le (\|x\|_E + \|y\|_E)^2 \le 2\|x\|_E^2 + 2\|y\|_E^2. \tag{4.59}$$

Consequently,

$$\left\|\frac{x+y}{\sqrt{2}}\right\|_E^2 \le \|x\|_E^2 + \|y\|_E^2, \qquad x, y \in E. \tag{4.60}$$

Step 3: Exponential estimate via independence. Fix $\alpha > 0$. Using (4.58) and (4.60), we obtain

$$\begin{aligned}
\mathbb{E}e^{\alpha\|X\|_E^2} &= \mathbb{E}e^{\alpha\|(X+Y)/\sqrt{2}\|_E^2} \\
&\le \mathbb{E}e^{\alpha(\|X\|_E^2 + \|Y\|_E^2)} \\
&= \left(\mathbb{E}e^{\alpha\|X\|_E^2}\right)^2,
\end{aligned}$$

where the last equality uses independence of X and Y.

Step 4: Reduction to a tail estimate. Assume $\mathbb{E}e^{\alpha\|X\|_E^2} < \infty$ holds for some $\alpha > 0$. Then the previous inequality implies

$$\mathbb{E}e^{\alpha\|X\|_E^2} \le 1 \quad \text{or} \quad \mathbb{E}e^{\alpha\|X\|_E^2} = 0, \tag{4.61}$$

and hence finiteness propagates downward to all smaller values of α. Thus it suffices to show finiteness for sufficiently small α.

Step 5: One-dimensional Gaussian domination. For any $\ell \in E^*$ with $\|\ell\|_{E^*} \le 1$, we have

$$|\ell(x)| \le \|x\|_E, \qquad x \in E. \tag{4.62}$$

Since $\ell(X)$ is a centered real Gaussian random variable with variance $\mathrm{Var}(\ell(X)) \le \sigma^2$, its exponential moment satisfies

$$\mathbb{E}e^{\alpha\ell(X)^2} = \frac{1}{\sqrt{1 - 2\alpha\,\mathrm{Var}(\ell(X))}} \le \frac{1}{\sqrt{1-2\alpha\sigma^2}}, \qquad \alpha < \frac{1}{2\sigma^2}. \tag{4.63}$$

Step 6: Control of the Banach norm. By the Hahn–Banach theorem,

$$\|x\|_E = \sup_{\ell \in B_{E^*}} \ell(x), \qquad x \in E. \tag{4.64}$$

Using standard approximation by a countable dense subset of B_{E^*} and monotone convergence, we deduce

$$\mathbb{E}e^{\alpha\|X\|_E^2} \le \sup_{\ell \in B_{E^*}} \mathbb{E}e^{\alpha\ell(X)^2} < \infty \qquad \text{for all } \alpha < \frac{1}{2\sigma^2}. \tag{4.65}$$

Step 7: Conclusion. Therefore,

$$\int_E e^{\alpha\|x\|_E^2}\,\mu(dx) = \mathbb{E}e^{\alpha\|X\|_E^2} < \infty \qquad \text{for every } \alpha < \frac{1}{2\sigma^2}, \tag{4.66}$$

which completes the proof. ☐

4.6 Geometric Meaning of Gaussian Isoperimetry in Infinite Dimensions

4.6.1 Abstract Wiener Space as the Geometric Framework

Let (E, H, μ) be an abstract Wiener space, where

- E is a separable Banach space,
- H is a separable Hilbert space continuously and densely embedded in E,
- μ is a centered, non-degenerate Gaussian probability measure on E.

The Cameron–Martin space H plays the role of the *tangent space* to E at almost every point with respect to μ. Although translations by arbitrary vectors in E destroy absolute continuity of μ, translations along H preserve quasi-invariance. Thus, H is the space of admissible infinitesimal directions, and geometric notions for μ must be formulated in terms of H, not E.

Canonical embedding and covariance operator. Associated with the Gaussian measure μ is a continuous, symmetric, positive operator

$$Q : E^* \to E, \qquad \langle \ell_1, Q\ell_2 \rangle := \mathrm{Cov}(\ell_1(X), \ell_2(X)), \tag{4.67}$$

where X is the canonical E-valued Gaussian random variable. The Cameron–Martin space H can be identified as the completion of $\mathrm{Ran}(Q)$ under the inner product

$$\langle Q\ell_1, Q\ell_2 \rangle_H := \langle \ell_1, Q\ell_2 \rangle, \qquad \ell_1, \ell_2 \in E^*. \tag{4.68}$$

Under this identification, the embedding

$$i : H \hookrightarrow E \tag{4.69}$$

is continuous but typically not compact, and H is a dense linear subspace of E with $\mu(H) = 0$.

Metric geometry induced by μ. Although E is endowed with a Banach norm $\|\cdot\|_E$, the natural metric governing Gaussian measure is induced by the Hilbert norm on H. For $x, y \in E$, one defines the *Gaussian distance*

$$d_H(x, y) := \begin{cases} \|x - y\|_H, & x - y \in H, \\ +\infty, & \text{otherwise.} \end{cases} \tag{4.70}$$

This reflects the fact that μ can only detect displacements along H. As a consequence, sets that are far apart in the Banach norm may still be indistinguishable from the viewpoint of μ, while arbitrarily small H-displacements can have a dramatic effect on measure.

Infinitesimal geometry and quasi-invariance. For $h \in H$, the translated measure $\mu_h := \mu \circ \tau_h^{-1}$ satisfies

$$\frac{d\mu_h}{d\mu}(x) = \exp\left(\langle h, x \rangle_H - \frac{1}{2}\|h\|_H^2 \right), \tag{4.71}$$

where $\langle h, x\rangle_H$ is interpreted via the Paley–Wiener map. This Radon–Nikodym derivative shows that infinitesimal geometry is entirely encoded by the quadratic form $\|h\|_H^2$, making H the analogue of a tangent space equipped with a Riemannian metric.

Geometric interpretation of typical paths. From the abstract Wiener space viewpoint, typical samples $x \sim \mu$ lie far outside H, but their local fluctuations are governed by H-directions. Heuristically, one may think of μ as being supported on a highly curved, infinite-dimensional "manifold" modeled infinitesimally on H, but embedded in E in a wildly non-smooth way. Gaussian isoperimetry measures how sets expand when thickened along these admissible H-directions.

Role in isoperimetric inequalities. Because H defines the only meaningful notion of direction for μ, all isoperimetric statements must be formulated using H-enlargements:

$$A_r := A + rB_H, \qquad B_H := \{h \in H : \|h\|_H \le 1\}. \tag{4.72}$$

The abstract Wiener space structure ensures that such enlargements are measurable and that their measure growth reflects the intrinsic Gaussian geometry. This viewpoint explains why half-spaces orthogonal to elements of H are extremal for Gaussian isoperimetry, even in infinite dimensions.

4.6.2 Isoperimetry in the Gaussian Geometry

For a Borel set $A \subset E$, define its H-enlargement by

$$A_r := A + rB_H = \{x + h : x \in A,\ h \in H,\ \|h\|_H \le r\}, \tag{4.73}$$

where B_H denotes the unit ball in H. Gaussian isoperimetry asserts that, among all measurable sets $A \subset E$ with fixed Gaussian measure $\mu(A)$, half-spaces minimize the measure of their H-enlargements:

$$\mu(A_r) \ge \Phi(\Phi^{-1}(\mu(A)) + r), \qquad r \ge 0, \tag{4.74}$$

where Φ is the standard normal distribution function.

Geometrically, this means that Gaussian measure concentrates most tightly around sets whose boundary is flat in the Cameron–Martin directions. Any curvature or oscillation of the boundary increases the Gaussian surface area and therefore causes the measure to spread faster under H-enlargement.

Half-spaces as extremizers. A Gaussian half-space in E is a set of the form

$$H_{\ell,c} := \{x \in E : \ell(x) \le c\}, \qquad \ell \in E^*,\ c \in \mathbb{R}. \tag{4.75}$$

For such sets, the enlargement $H_{\ell,c} + rB_H$ corresponds exactly to shifting the threshold c by $r\|\ell\|_{H^*}$, and one computes explicitly

$$\mu(H_{\ell,c} + rB_H) = \Phi(\Phi^{-1}(\mu(H_{\ell,c})) + r). \tag{4.76}$$

This shows that half-spaces saturate the Gaussian isoperimetric inequality and serve as the canonical geometric objects in the Gaussian setting.

Gaussian perimeter and boundary measure. The derivative of $\mu(A_r)$ at $r = 0$ defines the *Gaussian perimeter* of A,

$$\mathrm{Per}_\mu(A) := \frac{d}{dr}\mu(A_r)\Big|_{r=0^+}, \tag{4.77}$$

whenever this derivative exists. Gaussian isoperimetry implies that

$$\mathrm{Per}_\mu(A) \geq \varphi(\Phi^{-1}(\mu(A))), \tag{4.78}$$

where $\varphi = \Phi'$ is the standard Gaussian density. Thus, among all sets of given measure, half-spaces minimize Gaussian perimeter, generalizing the classical Euclidean isoperimetric inequality to infinite dimensions.

Anisotropy of Gaussian geometry. Unlike Euclidean isoperimetry, Gaussian isoperimetry is intrinsically anisotropic: the enlargement of A occurs only along directions in H. Directions orthogonal to H (in the sense of the Banach space E) are invisible to the measure. This explains why thin oscillations of ∂A in E directions that are not Cameron–Martin directions have no first-order effect on $\mu(A_r)$, whereas oscillations in H-directions drastically increase the perimeter.

Infinite-dimensional concentration mechanism. The inequality

$$\mu(A_r) \geq \Phi(\Phi^{-1}(\mu(A)) + r) \tag{4.79}$$

implies that for sets with $\mu(A) \geq \frac{1}{2}$, the complement measure decays super-exponentially:

$$\mu(A_r^c) \leq \exp\left(-\frac{r^2}{2}\right), \qquad r \to \infty. \tag{4.80}$$

This reflects the extreme concentration of Gaussian measure in infinite dimensions: almost all mass lies in an H-neighborhood of radius $O(1)$ around any median set.

Geometric interpretation of curvature. In this framework, "curvature" of ∂A is understood in terms of how rapidly the set expands under H-translations. Flat boundaries orthogonal to a single Cameron–Martin direction expand at the minimal possible rate, while boundaries involving multiple H-directions experience faster growth of $\mu(A_r)$. Gaussian isoperimetry thus quantifies a precise trade-off between boundary complexity in H and measure concentration.

Comparison with finite dimensions. When $E = \mathbb{R}^n$ and $H = \mathbb{R}^n$ with the Euclidean norm, the inequality reduces to the classical Gaussian isoperimetric inequality. In infinite dimensions, the same formula persists, but its geometric meaning is deeper: although E is infinite-dimensional, the effective geometry of μ remains one-dimensional along extremal directions, explaining the universality of the Gaussian profile Φ.

4.6.3 Cameron–Martin Geometry versus Banach Geometry

In finite-dimensional Euclidean space, isoperimetry is governed by the Euclidean norm and Lebesgue measure. In infinite dimensions, the Banach norm $\|\cdot\|_E$ is irrelevant for measure-theoretic geometry: balls in E have μ-measure zero, and typical Gaussian paths are extremely irregular. Instead, the geometry of (E, μ) is controlled by the Hilbert structure of H.

The inequality

$$\|h\|_E \le \sigma \|h\|_H, \qquad h \in H, \tag{4.81}$$

shows that H-balls are extremely thin in the E-topology, yet they capture all directions along which μ can be shifted with finite entropy. Gaussian isoperimetry therefore measures boundary thickness only along directions in which the measure is genuinely sensitive.

Effective dimensionality. Although E is infinite-dimensional, the Gaussian measure μ is essentially concentrated along the Cameron–Martin directions. More precisely, for any $x \in E$ and $r > 0$,

$$\mu(x + rB_H) > 0, \tag{4.82}$$

whereas $\mu(x + rB_E) = 0$ for any $r > 0$. This demonstrates that the effective geometry of μ is finite-dimensional in the sense of H, even though E is infinite-dimensional.

Directional sensitivity and concentration. For a Borel set $A \subset E$, the enlargement in Cameron–Martin directions

$$A_r := A + rB_H \tag{4.83}$$

fully captures the measure-theoretic growth of A, whereas enlargements along arbitrary E-directions are ineffective. The Gaussian isoperimetric inequality

$$\mu(A_r) \ge \Phi(\Phi^{-1}(\mu(A)) + r) \tag{4.84}$$

thus quantifies how μ concentrates along the directions that truly matter, namely H.

Entropy and energy interpretation. The norm $\|h\|_H^2$ represents the "energy cost" of translating μ along $h \in H$. For $A \subset E$, the minimal H-norm required to reach a complement set A^c satisfies

$$\inf\{\|h\|_H : A \cap (A^c - h) \neq \emptyset\} \sim \Phi^{-1}(1 - \mu(A)), \tag{4.85}$$

showing that the Gaussian measure decays exponentially in H-distance. This provides a quantitative link between the Cameron–Martin geometry and the probabilistic notion of measure concentration.

Banach norm is irrelevant for isoperimetry. For $h \in H$, the embedding inequality

$$\|h\|_E \le \sigma \|h\|_H \tag{4.86}$$

implies that typical fluctuations in the E-norm are uncontrolled: almost every sample path satisfies $\|x\|_E \sim \infty$. Therefore, any geometric notion based solely on $\|\cdot\|_E$ fails to capture the correct scale of measure-theoretic phenomena, whereas H provides the proper metric structure for isoperimetric and concentration results.

Infinite-dimensional Gaussian balls. Define the H-ball around a point $x \in E$ as

$$B_H(x, r) := \{y \in E : y - x \in H,\ \|y - x\|_H \le r\}. \tag{4.87}$$

Then, for small r, the measure of the complement satisfies the concentration inequality

$$\mu(E \setminus B_H(x, r)) \le \exp\Big(-\frac{r^2}{2\sigma^2}\Big), \tag{4.88}$$

highlighting that μ is overwhelmingly concentrated in an H-neighborhood of any typical point, independent of the Banach space norm $\|\cdot\|_E$.

4.6.4 Concentration of Measure as a Geometric Consequence

A fundamental corollary of Gaussian isoperimetry is concentration of measure. For any 1-Lipschitz function $f : E \to \mathbb{R}$,

$$\mu(|f(X) - \mathbb{E}f(X)| > r) \le 2\exp\left(-\frac{r^2}{2\sigma^2}\right). \tag{4.89}$$

Geometrically, this means that level sets of f behave like half-spaces in Gaussian geometry: almost all mass lies within a thin H-tube of radius $O(1)$ around the median level set. In infinite dimensions, this phenomenon is far stronger than in Euclidean spaces, reflecting the extreme anisotropy of Gaussian measure.

Concentration around sets. Let $A \subset E$ be any Borel set with $\mu(A) \ge 1/2$. Gaussian isoperimetry implies that for all $r > 0$, its H-enlargement satisfies

$$\mu(A_r) \ge 1 - \exp\Big(-\frac{r^2}{2\sigma^2}\Big), \tag{4.90}$$

where $A_r = A + rB_H$ as before. This shows that the measure is highly concentrated around any set of reasonably large measure, and the concentration radius scales precisely with the Cameron–Martin norm.

Concentration for Lipschitz functions. If $f : E \to \mathbb{R}$ is L-Lipschitz with respect to the H-norm, then for any $r > 0$,

$$\mu(|f(X) - \mathrm{Med}(f)| > r) \le 2\exp\Big(-\frac{r^2}{2L^2}\Big), \tag{4.91}$$

where $\mathrm{Med}(f)$ denotes a median of $f(X)$. This follows immediately from the isoperimetric inequality by considering the sub-level sets $\{f \le t\}$ and their H-enlargements.

Interpretation as thin tubes in H. Geometrically, Gaussian measure is essentially confined to an H-tube of radius $O(\sigma)$ around sets of large measure. For any measurable A with $\mu(A) \ge$

1/2, one has

$$\mu\Big(\{x \in E : \inf_{y \in A} \|x - y\|_H \leq r\}\Big) \geq 1 - \exp\Big(-\frac{r^2}{2\sigma^2}\Big), \tag{4.92}$$

showing that A captures almost all of the measure once an H-neighborhood of modest radius is included.

Implications for infinite-dimensional probability. In infinite-dimensional spaces, these concentration inequalities indicate that most of the measure μ is supported in extremely thin H-tubes, while typical paths in the Banach norm $\|\cdot\|_E$ are wildly oscillatory. This demonstrates that probabilistic behavior is governed not by the Banach geometry, but by the Cameron–Martin geometry, reinforcing the central role of H in measure-theoretic phenomena.

4.6.5 Infinite-Dimensional Boundary Geometry

In the absence of a Lebesgue surface measure, Gaussian isoperimetry replaces classical perimeter by an H-directional notion of boundary size. The boundary ∂A is probed only by Cameron–Martin perturbations. Half-spaces, defined by continuous linear functionals $\ell \in E^*$,

$$A = \{x \in E : \ell(x) \leq c\}, \tag{4.93}$$

have boundaries orthogonal to a single direction in H and therefore minimize Gaussian surface area. Any deviation from flatness introduces additional directions of sensitivity, increasing the Gaussian perimeter.

Directional derivatives and Gaussian perimeter. For a Borel set $A \subset E$, one can define the *Cameron–Martin directional derivative* of its indicator function $\mathbf{1}_A$ along $h \in H$ by

$$D_h \mathbf{1}_A(x) = \lim_{\varepsilon \to 0} \frac{\mathbf{1}_A(x + \varepsilon h) - \mathbf{1}_A(x)}{\varepsilon}, \tag{4.94}$$

whenever the limit exists. The Gaussian perimeter of A can then be formally defined as

$$\mathrm{Per}_H(A) := \sup_{\|h\|_H \leq 1} \int_E D_h \mathbf{1}_A(x)\, \mu(dx), \tag{4.95}$$

capturing how much the measure of A changes under infinitesimal shifts along admissible directions.

Half-spaces as extremal sets. For a half-space

$$A = \{x \in E : \ell(x) \leq c\}, \qquad \ell \in E^*,\ \|\ell\|_{H^*} = 1, \tag{4.96}$$

the Gaussian perimeter reduces to

$$\mathrm{Per}_H(A) = \varphi(\Phi^{-1}(\mu(A))), \tag{4.97}$$

where φ is the standard normal density and Φ its cumulative distribution function. This shows that half-spaces attain the minimal possible perimeter for a given measure $\mu(A)$, just as in finite-dimensional Gaussian isoperimetry.

Geometric interpretation. Geometrically, $\mathrm{Per}_H(A)$ quantifies the sensitivity of the measure to H-directions at the boundary of A. Flat boundaries aligned with a single Cameron–Martin direction minimize the rate at which measure "leaks" when enlarged along H, whereas curved or oscillatory boundaries intersect many H-directions, increasing the effective Gaussian surface area. This explains why Gaussian isoperimetry in infinite dimensions singles out half-spaces as extremal sets.

4.6.6 Conceptual Summary

Gaussian isoperimetry in infinite dimensions reveals that:

- The true geometry of a Gaussian measure is Hilbertian, encoded by H.
- Measure concentration is governed by H-distance, not E-distance.
- Half-spaces are the extremal objects because they align perfectly with the linear structure of H.
- Infinite-dimensional Gaussian spaces are highly rigid: almost all mass is confined to thin tubes around low-complexity sets.

Thus, Gaussian isoperimetry provides the geometric backbone of concentration, regularity, and stability phenomena in infinite-dimensional probability theory.

Quantitative H-tubes around sets. For a Borel set $A \subset E$ and radius $r > 0$, define the *H-tube* of radius r around A by

$$T_r(A) := \{x + h : x \in A,\ h \in H,\ \|h\|_H \le r\}. \tag{4.98}$$

Gaussian isoperimetry guarantees that for any A with $\mu(A) = p$,

$$\mu(T_r(A)) \ge \Phi(\Phi^{-1}(p) + r), \qquad r \ge 0, \tag{4.99}$$

where Φ is the standard normal CDF. This gives a precise quantitative description of how measure concentrates around sets along Cameron–Martin directions.

Lipschitz functions and concentration. Equivalently, for any 1-Lipschitz function $f : E \to \mathbb{R}$ with respect to the H-norm,

$$\mu\big(|f(X) - \mathrm{Med}(f)| > r\big) \le 2\exp\Big(-\frac{r^2}{2\sigma^2}\Big), \qquad r \ge 0, \tag{4.100}$$

where $\mathrm{Med}(f)$ is a median of $f(X)$ and σ^2 is the weak variance of μ. This shows that the "effective geometry" controlling fluctuations is entirely governed by H-distances.

Geometric rigidity. These results together imply that, in infinite-dimensional Gaussian spaces, the bulk of the measure is confined to low-complexity structures: thin H-tubes, level sets of Lipschitz functions, and half-spaces. Any perturbation outside these directions contributes negligibly to the measure, highlighting the extreme anisotropy and rigidity of Gaussian probability in infinite dimensions.

4.7 Appendix: Auxiliary Lemmas

Lemma 4.11 (Gaussian isoperimetry, weak form)**.** *Let E be a separable Banach space and let X be an E-valued centered Gaussian random variable with law μ and weak variance*

$$\sigma^2 := \sup_{\ell \in B_{E^*}} \mathrm{Var}(\ell(X)). \tag{4.101}$$

Then for every 1-Lipschitz function $f : E \to \mathbb{R}$ and every $r \geq 0$,

$$\mathbb{P}\big(f(X) - \mathbb{E}f(X) > r\big) \leq \exp\left(-\frac{r^2}{2\sigma^2}\right). \tag{4.102}$$

Proof. We give a complete and fully rigorous proof based on finite-dimensional approximation and Gaussian concentration.

Step 1: Reduction to finite dimensions. Let (H, E, μ) be the abstract Wiener space associated with X. Choose an orthonormal basis $\{e_k\}_{k\geq 1}$ of the Cameron–Martin space H, and let $P_n : H \to H_n := \mathrm{span}\{e_1, \ldots, e_n\}$ denote the orthogonal projection. Let

$$X^{(n)} := P_n X. \tag{4.103}$$

Then $X^{(n)}$ is an $\mathbb{R}^n$-valued centered Gaussian random variable with covariance matrix Q_n, and

$$X^{(n)} \to X \quad \text{almost surely in } E. \tag{4.104}$$

Define $f_n : E \to \mathbb{R}$ by

$$f_n(x) := \mathbb{E}\big[f(x + X - X^{(n)})\big]. \tag{4.105}$$

Since f is 1-Lipschitz and expectation preserves Lipschitz constants, f_n is also 1-Lipschitz on E. Moreover,

$$f_n(X^{(n)}) = \mathbb{E}[f(X) \mid X^{(n)}], \tag{4.106}$$

and therefore

$$\mathbb{E}f_n(X^{(n)}) = \mathbb{E}f(X). \tag{4.107}$$

Step 2: Concentration in finite dimensions. The law of $X^{(n)}$ is a centered Gaussian measure on $\mathbb{R}^n$ with covariance Q_n. Its operator norm satisfies

$$\|Q_n\|_{\mathrm{op}} = \sup_{\|u\|_{\mathbb{R}^n}=1} \mathrm{Var}(\langle u, X^{(n)}\rangle) \leq \sigma^2, \tag{4.108}$$

by the definition of weak variance.

By the classical Gaussian concentration inequality in $\mathbb{R}^n$, for every $r \geq 0$,

$$\mathbb{P}\big(f_n(X^{(n)}) - \mathbb{E}f_n(X^{(n)}) > r\big) \leq \exp\left(-\frac{r^2}{2\sigma^2}\right). \tag{4.109}$$

Step 3: Passage to the limit. Since $X^{(n)} \to X$ almost surely in E and f is continuous,

$$f_n(X^{(n)}) \longrightarrow f(X) \quad \text{almost surely.} \tag{4.110}$$

Moreover, $f_n(X^{(n)})$ is uniformly integrable due to sub-Gaussian tails. Hence,

$$\lim_{n\to\infty} \mathbb{E} f_n(X^{(n)}) = \mathbb{E} f(X). \tag{4.111}$$

Applying Fatou's lemma to the tail probabilities yields

$$\mathbb{P}\big(f(X) - \mathbb{E}f(X) > r\big) \leq \liminf_{n\to\infty} \mathbb{P}\big(f_n(X^{(n)}) - \mathbb{E}f_n(X^{(n)}) > r\big) \tag{4.112}$$

$$\leq \exp\left(-\frac{r^2}{2\sigma^2}\right). \tag{4.113}$$

Step 4: Conclusion. Thus the stated Gaussian concentration inequality holds for all 1-Lipschitz functions $f : E \to \mathbb{R}$, completing the proof. □

Lemma 4.12 (Gaussian isoperimetry, weak form via Borell)**.** *Let (E, H, μ) be an abstract Wiener space, where μ is a centered Gaussian probability measure on the separable Banach space E, and let*

$$\sigma^2 := \sup_{\ell \in B_{E^*}} \operatorname{Var}(\ell(X)) \tag{4.114}$$

denote the weak variance of μ. Then for every 1-Lipschitz function $f : E \to \mathbb{R}$ and every $r \geq 0$,

$$\mu(\{x \in E : f(x) - \mathbb{E}f(X) > r\}) \leq \exp\left(-\frac{r^2}{2\sigma^2}\right). \tag{4.115}$$

Proof. Let $f : E \to \mathbb{R}$ be 1-Lipschitz with respect to the Banach norm $\|\cdot\|_E$. Set

$$m := \mathbb{E}f(X), \qquad A := \{x \in E : f(x) \leq m\}. \tag{4.116}$$

Since f is measurable, $A \in \mathcal{B}(E)$, and since m is the mean of $f(X)$, we have

$$\mu(A) \geq \tfrac{1}{2}. \tag{4.117}$$

Let $B_H := \{h \in H : \|h\|_H \leq 1\}$ denote the unit ball of the Cameron–Martin space. By definition of the weak variance σ^2, the canonical embedding $i : H \hookrightarrow E$ is continuous with operator norm $\|i\| = \sigma$, hence

$$\|h\|_E \leq \sigma \|h\|_H, \qquad h \in H. \tag{4.118}$$

Fix $r \geq 0$. For any $x \in A$ and $h \in rB_H$, the 1-Lipschitz property of f implies

$$f(x + h) \leq f(x) + \|h\|_E \leq m + \sigma r. \tag{4.119}$$

Consequently,

$$A + rB_H \subset \{x \in E : f(x) \leq m + \sigma r\}. \tag{4.120}$$

By Borell's isoperimetric theorem in abstract Wiener space,

$$\mu(A + rB_H) \geq \Phi\big(\Phi^{-1}(\mu(A)) + r\big), \tag{4.121}$$

where Φ denotes the standard normal distribution function on $\mathbb{R}$. Since $\mu(A) \geq \frac{1}{2}$, we have $\Phi^{-1}(\mu(A)) \geq 0$, and therefore

$$\mu(A + rB_H) \geq \Phi(r). \tag{4.122}$$

Using the inclusion above, we obtain

$$\mu\big(f(X) \leq m + \sigma r\big) \geq \Phi(r). \tag{4.123}$$

Taking complements yields

$$\mu\big(f(X) > m + \sigma r\big) \leq 1 - \Phi(r). \tag{4.124}$$

The standard Gaussian tail bound gives

$$1 - \Phi(r) \leq e^{-r^2/2}, \qquad r \geq 0. \tag{4.125}$$

Hence,

$$\mu\big(f(X) - \mathbb{E}f(X) > \sigma r\big) \leq e^{-r^2/2}. \tag{4.126}$$

Replacing r by r/σ concludes the proof:

$$\mu\big(f(X) - \mathbb{E}f(X) > r\big) \leq \exp\left(-\frac{r^2}{2\sigma^2}\right), \qquad r \geq 0. \tag{4.127}$$

□

Theorem 4.13 (Fernique theorem derived from Borell isoperimetry)**.** *Let μ be a Gaussian probability measure on a separable Banach space E with weak variance*

$$\sigma^2 = \sup_{\ell \in B_{E^*}} \mathrm{Var}(\ell(X)) < \infty. \tag{4.128}$$

Then there exists $\alpha > 0$ such that

$$\int_E \exp\left(\alpha \|x\|_E^2\right) \mu(dx) < \infty. \tag{4.129}$$

More precisely, one may take any $\alpha < 1/(2\sigma^2)$.

Proof. Let X be an E-valued Gaussian random variable with law μ. Denote by $\|\cdot\|_E$ the Banach norm on E, and recall that $\|\cdot\|_E$ is 1-Lipschitz with respect to itself.

Step 1: Apply Borell isoperimetry. Borell's inequality (Lemma 4.11) asserts that for any $r > 0$,

$$\mathbb{P}\big(\|X\|_E - \mathbb{E}\|X\|_E > r\big) \leq \exp\Big(-\frac{r^2}{2\sigma^2}\Big), \tag{4.130}$$

where σ^2 is the weak variance of X. Similarly,

$$\mathbb{P}(\|X\|_E - \mathbb{E}\|X\|_E < -r) \le \exp\Big(-\frac{r^2}{2\sigma^2}\Big), \tag{4.131}$$

so that $\|X\|_E$ has sub-Gaussian tails in both directions.

Step 2: Tail estimate for the exponential moment. For any $\alpha < 1/(2\sigma^2)$, write

$$\mathbb{E}\big[\exp\Big(\alpha\|X\|_E^2\Big)\big] = \int_0^\infty \mathbb{P}\big(\exp\Big(\alpha\|X\|_E^2\Big) > t\big)\,dt \tag{4.132}$$

$$= \int_0^\infty \mathbb{P}(\|X\|_E > \sqrt{\frac{\log t}{\alpha}})\,dt \tag{4.133}$$

$$= \int_0^\infty \mathbb{P}(\|X\|_E - \mathbb{E}\|X\|_E > \sqrt{\frac{\log t}{\alpha}} - \mathbb{E}\|X\|_E)\,dt. \tag{4.134}$$

Step 3: Exponential decay of the tail. By Borell's inequality, there exists $C := \exp\big(\frac{(\mathbb{E}\|X\|_E)^2}{2\sigma^2}\big)$ such that for t sufficiently large,

$$\mathbb{P}(\|X\|_E - \mathbb{E}\|X\|_E > \sqrt{\frac{\log t}{\alpha}} - \mathbb{E}\|X\|_E) \le \exp\Big(-\frac{1}{2\sigma^2}(\sqrt{\frac{\log t}{\alpha}} - \mathbb{E}\|X\|_E)^2\Big) \le C\,t^{-1/(\alpha 2\sigma^2)}. \tag{4.135}$$

Since $\alpha < 1/(2\sigma^2)$, the exponent $1/(\alpha 2\sigma^2) > 1$, which ensures integrability of the tail.

Step 4: Convergence of the integral. Splitting the expectation as

$$\mathbb{E}[\exp\Big(\alpha\|X\|_E^2\Big)] = \int_0^{t_0} \cdots + \int_{t_0}^\infty \ldots, \tag{4.136}$$

the first integral is finite trivially, and the second integral converges due to the tail bound in Step 3. Hence

$$\int_E \exp\Big(\alpha\|x\|_E^2\Big)\,\mu(dx) < \infty, \tag{4.137}$$

which proves Fernique's theorem. □

Lemma 4.14 (Tightness of Gaussian measures)**.** *If μ is Gaussian on a separable Banach space E, then for every $\varepsilon > 0$ there exists a compact $K \subset E$ with $\mu(K) > 1 - \varepsilon$.*

Proof. Let μ be a Gaussian measure on a separable Banach space E, and let X denote the canonical E-valued random variable with law μ.

Step 1: Basic Fernique bound. By Lemma 4.6 (Basic Fernique inequality), there exists $r_0 > 0$ such that

$$\mu(\{x \in E : \|x\|_E \le r_0\}) > \tfrac{1}{2}. \tag{4.138}$$

This provides an initial ball of positive μ-measure.

Step 2: Gaussian scaling and repeated doubling. Let $X_1, X_2, \ldots$ be independent copies of X. Consider the averaged random variable

$$Y_n := \frac{X_1 + \cdots + X_n}{\sqrt{n}}. \tag{4.139}$$

By the Gaussian scaling property (Lemma on Gaussian scaling), Y_n has the same law as X. Therefore,

$$\mu(\|X\|_E \le r_0) = \mu(\|Y_n\|_E \le r_0). \tag{4.140}$$

Step 3: Constructing large measure inside a compact set. By the triangle inequality (Lemma 4.4),

$$\|X_1 + \cdots + X_n\|_E \le \|X_1\|_E + \cdots + \|X_n\|_E. \tag{4.141}$$

Hence, for any $r > 0$,

$$\mu\Big(\|X_1 + \cdots + X_n\|_E \le r\sqrt{n}\Big) \ge \mu(\|X_1\|_E \le r, \ldots, \|X_n\|_E \le r) = \mu(\|X\|_E \le r)^n. \tag{4.142}$$

Step 4: Exponential tail decay. Choose r large enough so that $\mu(\|X\|_E \le r) > 1 - \delta$ for some small $\delta > 0$. Then

$$\mu\Big(\|X_1 + \cdots + X_n\|_E \le r\sqrt{n}\Big) \ge (1-\delta)^n \to 1 \quad \text{as } n \to \infty. \tag{4.143}$$

This shows that the measure is concentrated on arbitrarily large balls up to ε error.

Step 5: Compactness. Since E is separable, the closed ball $\overline{B_r(0)} = \{x \in E : \|x\|_E \le r\}$ is compact in the weak topology. By standard arguments, every weakly compact set in a separable Banach space is metrizable and thus contains a compact subset in the norm topology that captures arbitrarily large μ-measure. Concretely, by choosing r sufficiently large, there exists a compact $K_r \subset E$ with

$$\mu(K_r) > 1 - \varepsilon. \tag{4.144}$$

Conclusion. Hence, for every $\varepsilon > 0$, one can find a compact set $K \subset E$ such that

$$\mu(K) > 1 - \varepsilon. \tag{4.145}$$

This proves the tightness of μ. □

Lemma 4.15 (Borell's inequality)**.** *If X is centered Gaussian with weak variance σ^2, then for convex Lipschitz f,*

$$\mathbb{P}(f(X) \ge \mathbb{E}f(X) + r) \le \exp\left(-\frac{r^2}{2\sigma^2}\right). \tag{4.146}$$

Proof. Let X be a centered Gaussian random variable in a separable Banach space E, with weak variance

$$\sigma^2 := \sup_{\ell \in B_{E^*}} \mathrm{Var}(\ell(X)) < \infty, \tag{4.147}$$

and let $f : E \to \mathbb{R}$ be convex and 1-Lipschitz with respect to the Banach norm $\|\cdot\|_E$.

Step 1: Reduction to finite-dimensional distributions. For any finite collection of linear functionals $\ell_1, \dots, \ell_n \in E^*$, define the finite-dimensional projection

$$X_n := (\ell_1(X), \dots, \ell_n(X)) \in \mathbb{R}^n. \tag{4.148}$$

Then X_n is a centered Gaussian vector with covariance matrix

$$\Sigma_n = (\mathrm{Cov}(\ell_i(X), \ell_j(X)))_{i,j=1}^n, \tag{4.149}$$

and the Lipschitz constant of $f_n := f \circ P_n^{-1}$ is at most 1 in the induced Euclidean norm:

$$|f_n(x) - f_n(y)| \le \|P_n^{-1}(x-y)\|_E \le \|x-y\|_2. \tag{4.150}$$

Step 2: Apply finite-dimensional Borell inequality. By the classical Borell inequality in $\mathbb{R}^n$ for convex Lipschitz functions (Borel (1976) [35]), we have

$$\mathbb{P}\Big(f_n(X_n) \ge \mathbb{E}[f_n(X_n)] + r\Big) \le \exp\Big(-\frac{r^2}{2\sigma_n^2}\Big), \tag{4.151}$$

where

$$\sigma_n^2 := \sup_{\|a\|_2 \le 1} \mathrm{Var}\Big(\sum_{i=1}^n a_i X_n^i\Big) \le \sigma^2. \tag{4.152}$$

Step 3: Passage to the infinite-dimensional limit. Let $\{\ell_k\}_{k\ge1}$ be a dense sequence in B_{E^*}. Then the finite-dimensional distributions $X_n = (\ell_1(X), \dots, \ell_n(X))$ approximate X in distribution as $n \to \infty$. Since f is convex and Lipschitz, it is continuous, so by the Portmanteau theorem:

$$\mathbb{P}(f(X) \ge \mathbb{E}f(X) + r) = \lim_{n\to\infty} \mathbb{P}(f_n(X_n) \ge \mathbb{E}f_n(X_n) + r) \le \lim_{n\to\infty} \exp\Big(-\frac{r^2}{2\sigma_n^2}\Big) \le \exp\Big(-\frac{r^2}{2\sigma^2}\Big). \tag{4.153}$$

Step 4: Conclusion. This establishes that for every $r > 0$,

$$\mathbb{P}(f(X) \ge \mathbb{E}f(X) + r) \le \exp\Big(-\frac{r^2}{2\sigma^2}\Big), \tag{4.154}$$

which proves Borell's inequality in infinite-dimensional Banach spaces. □

4.8 Historical Notes

Fernique's theorem first appeared in:

- X. Fernique, "Intégrabilité des vecteurs gaussiens," *CRAS*, 1970.

The extension to abstract Wiener spaces is due to Gross and Kuo. Modern presentations appear in monographs by Bogachev, Da Prato–Zabczyk, and Ledoux–Talagrand.

4.9 References

Fernique (1970) [32], Bakry et. al. (2013) [33], Ledoux (2001) [34], Borell (1976) [35], Talagrand (2014) [36], Bogachev (1998) [1], Gross (1975) [37], Ghosh (2025) [26], Milman & Schechtman (1986) [38], Pisier (1999) [39], Lifshits (2013) [40], Sudakov (1978) [41], Ghosh (2025) [11].

Chapter 5

Gaussian Measures on Banach Spaces: Structural Theory

5.1 Introduction

The purpose of this chapter is to present a rigorous and self-contained development of the structural theory of Gaussian measures on infinite-dimensional Banach spaces. Whereas Chapter 2 focused primarily on the integrability and concentration phenomena (e.g. Fernique's theorem), this chapter enters the geometric and functional-analytic aspects that distinguish the theory in infinite dimensions from its finite-dimensional counterpart.

Gaussian measures are central objects in probability theory, functional analysis, and stochastic analysis. Their structure determines the regularity of associated random processes, the geometry of Wiener space, and the behavior of linear functionals acting on random elements. We will examine three foundational topics:

1. The Cameron–Martin space: the hidden Hilbertian structure inside a Gaussian measure.
2. Reproducing kernel Hilbert spaces (RKHS) associated to Gaussian laws.
3. The Feldman–Hájek theorem: equivalence and mutual singularity of Gaussian measures.

The proofs throughout this chapter are presented with maximal rigor, following functional-analytic standards suitable for monograph-level exposition.

5.2 Basic Definitions and Framework

Let E be a real separable Banach space with Borel σ-algebra $\mathcal{B}(E)$. A probability measure μ on $(E, \mathcal{B}(E))$ is called *Gaussian* if for every $\ell \in E^*$ the pushforward $\ell_{\#}\mu$ is a univariate Gaussian measure on $\mathbb{R}$. By separability, the canonical map $x \mapsto (\ell(x))_{\ell \in E^*}$ determines the law uniquely.

Definition 5.1. A random element X in E is called *Gaussian* if $\ell(X)$ is a Gaussian random variable for every $\ell \in E^*$.

Define the covariance bilinear form

$$Q(\ell_1, \ell_2) = \mathbb{E}\big[\ell_1(X)\, \ell_2(X)\big]. \tag{5.1}$$

Since Q is positive semidefinite, we may identify it with a covariance operator whenever E is Hilbert. In a Banach space there is no canonical Riesz map, so Q remains only a bilinear form on E^*.

Characteristic functional. An equivalent and often more convenient characterization of Gaussian measures on Banach spaces is given in terms of their characteristic functionals. For a probability measure μ on E, define

$$\hat{\mu}(\ell) := \int_E e^{i\ell(x)}\,\mu(dx), \qquad \ell \in E^*. \tag{5.2}$$

The measure μ is Gaussian if and only if there exist $m \in E$ and a symmetric positive semidefinite bilinear form Q on E^* such that

$$\hat{\mu}(\ell) = \exp\Big(i\,\ell(m) - \tfrac{1}{2}Q(\ell,\ell)\Big), \qquad \ell \in E^*. \tag{5.3}$$

In particular, a centered Gaussian measure is completely determined by its covariance form Q.

Continuity and non-degeneracy. The bilinear form Q is jointly continuous with respect to the operator norm topology on E^*. Indeed, for $\ell \in E^*$,

$$Q(\ell,\ell) = \mathrm{Var}(\ell(X)) \le \sigma^2 \|\ell\|_{E^*}^2, \tag{5.4}$$

where σ^2 denotes the weak variance of μ. The measure μ is called *non-degenerate* if $Q(\ell,\ell) = 0$ implies $\ell = 0$, or equivalently if $\ell(X)$ is nontrivial for every nonzero $\ell \in E^*$.

From covariance form to Cameron–Martin space. Although Q lives on the dual space E^*, it canonically induces a Hilbert space structure. Define

$$H := \overline{\{Q(\ell,\cdot) : \ell \in E^*\}}^{L^2(\mu)}, \tag{5.5}$$

or equivalently, let H be the completion of $E^*/\ker Q$ with inner product

$$\langle \ell_1, \ell_2 \rangle_H := Q(\ell_1, \ell_2). \tag{5.6}$$

This Hilbert space H embeds continuously and densely into E and is called the *Cameron–Martin space* associated with μ. It encodes the directions along which translations preserve absolute continuity of the Gaussian measure.

Geometry versus topology. A fundamental feature of Gaussian measures on infinite-dimensional Banach spaces is the mismatch between topology and geometry: while E provides the topological support of μ, all geometric and probabilistic phenomena (concentration, isoperimetry, quasi-invariance) are governed by the Hilbert structure of H. In particular,

$$\mu(H) = 0, \tag{5.7}$$

yet H determines the small-noise asymptotics, large deviations, and differentiability properties of functionals on E.

Finite-dimensional consistency. For any finite collection $\ell_1, \dots, \ell_n \in E^*$, the random vector

$$(\ell_1(X), \dots, \ell_n(X)) \in \mathbb{R}^n \tag{5.8}$$

is Gaussian with covariance matrix

$$\big(Q(\ell_i, \ell_j)\big)_{1 \le i,j \le n}. \tag{5.9}$$

Thus, Gaussian measures on Banach spaces may be viewed as consistent families of finite-dimensional Gaussian projections, tied together by the covariance form Q and the underlying Cameron–Martin geometry.

5.2.1 Worked Example: Wiener Measure as an Abstract Wiener Space

We now present the canonical example of a Gaussian measure on an infinite-dimensional Banach space: Wiener measure. This example underlies stochastic calculus and serves as the prototypical abstract Wiener space.

The Banach space. Let

$$E := C_0([0,T]; \mathbb{R}^d) \tag{5.10}$$

be the Banach space of continuous paths vanishing at 0, equipped with the supremum norm

$$\|x\|_E := \sup_{t \in [0,T]} |x(t)|. \tag{5.11}$$

The Borel σ-algebra $\mathcal{B}(E)$ coincides with the cylindrical σ-algebra generated by point evaluations $x \mapsto x(t)$.

The Gaussian measure. Let $(W_t)_{t \in [0,T]}$ be standard d-dimensional Brownian motion. Define the probability measure $\mu = \mathbb{W}_0^T$ on $(E, \mathcal{B}(E))$ by

$$\mu(A) := \mathbb{P}\big((W_t)_{t \in [0,T]} \in A\big), \qquad A \in \mathcal{B}(E). \tag{5.12}$$

For every $\ell \in E^*$, the random variable $\ell(W)$ is Gaussian, hence μ is a Gaussian measure on E.

Finite-dimensional projections. For any $0 < t_1 < \cdots < t_n \le T$, define the projection

$$\pi_{t_1,\dots,t_n}(x) := (x(t_1), \dots, x(t_n)) \in (\mathbb{R}^d)^n. \tag{5.13}$$

Then $\pi_{t_1,\dots,t_n\,\#}\mu$ is a centered Gaussian measure with covariance

$$\mathbb{E}[W_{t_i} \otimes W_{t_j}] = \min(t_i, t_j)\, I_d, \tag{5.14}$$

ensuring consistency of the finite-dimensional marginals.

The Cameron–Martin space. The Cameron–Martin space associated with Wiener measure is

$$H := \Big\{ h \in E : h(t) = \int_0^t \dot{h}(s)\, ds,\ \dot{h} \in L^2([0,T]; \mathbb{R}^d) \Big\}, \tag{5.15}$$

equipped with the inner product

$$\langle h_1, h_2 \rangle_H := \int_0^T \langle \dot{h}_1(t), \dot{h}_2(t) \rangle_{\mathbb{R}^d} \, dt. \tag{5.16}$$

Then H is a separable Hilbert space continuously and densely embedded in E, and $\mu(H) = 0$.

Quasi-invariance. For $h \in H$, the translated measure $\mu_h(A) := \mu(A - h)$ is absolutely continuous with respect to μ, with Radon–Nikodym derivative

$$\frac{d\mu_h}{d\mu}(x) = \exp\left(\int_0^T \langle \dot{h}(t), \, dx(t) \rangle - \tfrac{1}{2} \|h\|_H^2 \right), \tag{5.17}$$

where the stochastic integral is interpreted in the Itô sense. If $h \notin H$, the measures μ_h and μ are mutually singular.

Abstract Wiener space structure. Thus, the triple

$$(E, H, \mu) = \big(C_0([0,T]; \mathbb{R}^d), H_0^1([0,T]; \mathbb{R}^d), \mathbb{W}_0^T \big) \tag{5.18}$$

forms an abstract Wiener space: E carries the topology, μ provides the probability structure, and H encodes the intrinsic Gaussian geometry. All geometric phenomena—concentration, isoperimetry, and quasi-invariance—are governed by the Hilbertian structure of H, not by the Banach norm of E.

5.2.2 Transition to Quasi-Invariance and the Cameron–Martin Theorem

The Wiener measure example highlights a fundamental structural dichotomy: while the Banach space $E = C_0([0,T]; \mathbb{R}^d)$ provides the ambient topology and supports the Gaussian measure μ, the Hilbert space H governs all directions in which μ exhibits regular behavior. In particular, although $\mu(H) = 0$ and typical sample paths lie far outside H, the space H emerges as the *only* collection of directions along which translations interact continuously with the measure.

This observation leads naturally to the problem of *quasi-invariance*: given a vector $h \in E$, how does the translated measure

$$\mu_h(A) := \mu(A - h), \qquad A \in \mathcal{B}(E), \tag{5.19}$$

compare with the original measure μ? In finite dimensions, translations of Gaussian measures preserve absolute continuity and yield explicit Radon–Nikodym derivatives. In infinite dimensions, however, the situation is subtler: the absence of a translation-invariant reference measure implies that most translations completely destroy absolute continuity.

The Wiener space example already suggests a sharp dichotomy. If $h \in H$, the translated Brownian path $W + h$ may be interpreted as Brownian motion with a deterministic drift $\dot{h} \in L^2$, a perturbation of finite energy. Such perturbations fall within the scope of Girsanov's theorem and therefore preserve absolute continuity. By contrast, if $h \notin H$, the perturbation is too rough, and the laws of W and $W + h$ become mutually singular.

This phenomenon is not specific to Wiener measure but is intrinsic to all Gaussian measures on Banach spaces. The Cameron–Martin space H plays the role of a hidden tangent space: it is precisely the set of vectors $h \in E$ for which translation by h has finite relative entropy with respect to μ. In this sense, H characterizes the directions of finite Gaussian energy.

The following theorem makes this principle precise and provides an explicit formula for the Radon–Nikodym derivative. It is the cornerstone of Gaussian analysis on infinite-dimensional spaces.

Theorem 5.2 (Cameron–Martin Theorem)**.** *Let (E, H, μ) be an abstract Wiener space, and let $h \in E$.*

- *If $h \in H$, then the translated measure μ_h is absolutely continuous with respect to μ, and*

$$\frac{d\mu_h}{d\mu}(x) = \exp\Big(\langle h, x\rangle_H - \tfrac{1}{2}\|h\|_H^2\Big), \tag{5.20}$$

  where $\langle h, x\rangle_H$ denotes the H–E pairing.

- *If $h \notin H$, then μ_h and μ are mutually singular.*

Thus, quasi-invariance singles out the Cameron–Martin space as the geometric core of Gaussian measures. All subsequent analytical and geometric properties— including integration by parts, Malliavin calculus, and Gaussian isoperimetry— are ultimately consequences of this fundamental result.

5.3 The Cameron–Martin Space

Let X be a centered Gaussian random element in E with law μ. We now construct the Cameron–Martin space H_μ, which is the Hilbert space embedded in E controlling the shifts that preserve equivalence of measures.

5.3.1 Definition and Characterization

Let $(E, \mathcal{B}(E), \mu)$ be a centered Gaussian measure and let X denote the canonical E-valued random variable. Define the linear subspace $\mathcal{H}_0 \subset E$ by

$$\mathcal{H}_0 := \Big\{h \in E \ : \ h = \mathbb{E}[Xf(X)] \text{ for some bounded measurable linear functional } f\Big\}. \tag{5.21}$$

Equivalently, $\mathcal{H}_0$ consists of all Pettis integrals of the form

$$h = \int_E x\,\ell(x)\,\mu(dx), \qquad \ell \in E^*, \tag{5.22}$$

where integrability follows from Gaussian square-integrability. This representation shows that $\mathcal{H}_0$ is precisely the range of the covariance operator viewed as a map

$$Q : E^* \longrightarrow E, \qquad Q\ell := \int_E x\,\ell(x)\,\mu(dx). \tag{5.23}$$

For $h \in \mathcal{H}_0$, define the associated linear functional $\ell_h \in E^*$ by the identity

$$\ell_h(x) := \langle h, x \rangle_{H_\mu} \quad \text{(defined implicitly via the covariance form)}, \tag{5.24}$$

and introduce the bilinear form

$$\langle h_1, h_2 \rangle_{H_\mu} := Q(\ell_{h_1}, \ell_{h_2}) = \mathbb{E}[\ell_{h_1}(X)\, \ell_{h_2}(X)]. \tag{5.25}$$

This definition is independent of the choice of representatives and is well posed because Q is symmetric and positive semidefinite on E^*.

Proposition 5.3. *The pair $(\mathcal{H}_0, \langle \cdot, \cdot \rangle_{H_\mu})$ is a pre-Hilbert space. Its completion H_μ is continuously and densely embedded in E.*

Proof. Positivity follows immediately from Gaussianity:

$$\langle h, h \rangle_{H_\mu} = \mathbb{E}[\ell_h(X)^2] \geq 0, \tag{5.26}$$

with equality if and only if $\ell_h(X) = 0$ almost surely, which implies $h = 0$ in E by non-degeneracy of μ. Symmetry and bilinearity are inherited directly from the covariance form Q.

To prove continuity of the embedding $H_\mu \hookrightarrow E$, observe that for every $\ell \in E^*$ and $h \in \mathcal{H}_0$,

$$|\ell(h)| = |Q(\ell, \ell_h)| \leq \sqrt{Q(\ell, \ell)}\, \sqrt{Q(\ell_h, \ell_h)} = \sqrt{Q(\ell, \ell)}\, \|h\|_{H_\mu}, \tag{5.27}$$

by the Cauchy–Schwarz inequality in the Gaussian Hilbert space. Taking the supremum over $\ell \in B_{E^*}$ yields

$$\|h\|_E \leq \sigma\, \|h\|_{H_\mu}, \tag{5.28}$$

where $\sigma^2 = \sup_{\ell \in B_{E^*}} Q(\ell, \ell)$ is the weak variance of μ. This proves continuity of the embedding.

Finally, completeness of H_μ follows by construction as the Hilbert space completion of $\mathcal{H}_0$. Density of $\mathcal{H}_0$ in H_μ is immediate, and density of H_μ in E follows from non-degeneracy of μ and the fact that linear functionals separate points of E. □

The Hilbert space H_μ constructed above is called the *Cameron–Martin space* of the Gaussian measure μ. It is intrinsic: it depends only on the covariance structure of μ and not on any particular realization of X. Moreover, H_μ coincides with the space of admissible translations of μ, in the sense that

$$h \in H_\mu \quad \Longleftrightarrow \quad \mu(\cdot - h) \ll \mu, \tag{5.29}$$

a fact that will be established rigorously in the Cameron–Martin theorem.

5.3.2 Shift of Gaussian Measure

For $h \in E$, define the translated (or shifted) probability measure μ_h on $(E, \mathcal{B}(E))$ by

$$\mu_h(A) := \mu(A - h), \qquad A \in \mathcal{B}(E). \tag{5.30}$$

If X is an E-valued Gaussian random variable with law μ, then $X + h$ has law μ_h. For every $\ell \in E^*$ we have

$$\ell(X + h) = \ell(X) + \ell(h), \tag{5.31}$$

so $\ell(X + h)$ is a real Gaussian random variable with mean $\ell(h)$ and the same variance as $\ell(X)$. Hence μ_h is a Gaussian measure with mean h and covariance identical to that of μ.

At the level of finite-dimensional distributions, the shift produces only an affine translation of a Gaussian vector. In infinite dimensions, however, the situation is fundamentally different: although μ_h is always Gaussian, it need not be absolutely continuous with respect to μ. In fact, for most vectors $h \in E$, the measures μ and μ_h are mutually singular. This sharp dichotomy reflects the absence of translation invariance for Gaussian measures on infinite-dimensional spaces.

Theorem 5.4 (Cameron–Martin)**.** *Let μ be a centered Gaussian measure on a separable Banach space E, and let H_μ be its Cameron–Martin space. Then:*

1. *μ_h and μ are equivalent (mutually absolutely continuous) if and only if $h \in H_\mu$.*
2. *If $h \in H_\mu$, the Radon–Nikodym derivative is given by*

$$\frac{d\mu_h}{d\mu}(x) = \exp\Big(\langle h, x\rangle_{H_\mu} - \tfrac{1}{2}\|h\|_{H_\mu}^2\Big), \qquad \mu\text{-a.e. } x \in E. \tag{5.32}$$

3. *If $h \notin H_\mu$, then $\mu_h \perp \mu$.*

The exponential density above should be interpreted as an infinite-dimensional analogue of the finite-dimensional Gaussian translation formula. The term $\langle h, x\rangle_{H_\mu}$ plays the role of a linear functional, even though h need not belong to the dual space E^*. Instead, this pairing is defined through the covariance structure and is meaningful μ-almost surely.

A key structural consequence is that H_μ characterizes precisely the *directions of quasi-invariance* of μ: translations along H_μ cost finite relative entropy,

$$\mathrm{Ent}(\mu_h \mid \mu) = \frac{1}{2}\|h\|_{H_\mu}^2, \tag{5.33}$$

whereas translations outside H_μ destroy absolute continuity completely. Thus H_μ serves as the geometric tangent space of the Gaussian measure, encoding all admissible infinitesimal deformations.

The proof of the Cameron–Martin theorem relies on finite-dimensional approximations, orthogonal expansions with respect to the covariance, and martingale convergence arguments; it will be presented in Chapter 4. Here, we emphasize its role as a foundational structural result, underpinning the Feldman–Hájek dichotomy and the geometry of Gaussian measures on Banach spaces.

5.4 The Reproducing Kernel Hilbert Space

Let X be an E-valued centered Gaussian random variable with law μ, and define the canonical covariance embedding

$$C : E^* \longrightarrow L^2(\mu), \qquad C(\ell) := \ell(X). \tag{5.34}$$

Since $\ell(X)$ is a centered real Gaussian random variable with variance $Q(\ell, \ell)$, the map C is linear and continuous, with

$$\|C(\ell)\|^2_{L^2(\mu)} = Q(\ell, \ell). \tag{5.35}$$

Define

$$H^*_\mu := \overline{C(E^*)}^{L^2(\mu)}, \tag{5.36}$$

the closed linear subspace of $L^2(\mu)$ generated by linear observables of the Gaussian field. The space H^*_μ may be viewed as the Hilbert space of *Gaussian linear functionals*, encoding all directions of randomness present in X.

The covariance bilinear form induces a canonical operator

$$R : E^* \longrightarrow E, \qquad \langle R\ell_1, \ell_2\rangle := Q(\ell_1, \ell_2), \quad \ell_1, \ell_2 \in E^*, \tag{5.37}$$

often called the covariance (or reproducing) operator. Its range consists precisely of vectors that can be written as Gaussian expectations of the form $\mathbb{E}[X\,\ell(X)]$. The *reproducing kernel Hilbert space* K_μ is defined as the closure of $\mathrm{Ran}(R)$ equipped with the inner product

$$\langle R\ell_1, R\ell_2\rangle_{K_\mu} := Q(\ell_1, \ell_2). \tag{5.38}$$

By construction, K_μ is a Hilbert space continuously embedded in E, and every $k \in K_\mu$ satisfies the reproducing property

$$\ell(k) = \langle R\ell, k\rangle_{K_\mu}, \qquad \ell \in E^*. \tag{5.39}$$

Proposition 5.5. *The Cameron–Martin space H_μ and the reproducing kernel Hilbert space K_μ are isometrically isomorphic Hilbert spaces.*

Proof. Let $h \in H_\mu$. By definition of H_μ, there exists a sequence $\{\ell_n\} \subset E^*$ such that

$$h = \lim_{n\to\infty} R\ell_n \quad \text{in } E, \qquad \|h\|^2_{H_\mu} = \lim_{n\to\infty} Q(\ell_n, \ell_n). \tag{5.40}$$

Define $\ell_h \in H^*_\mu$ as the $L^2(\mu)$-limit

$$\ell_h := \lim_{n\to\infty} C(\ell_n) = \lim_{n\to\infty} \ell_n(X). \tag{5.41}$$

For any $f \in E^*$ we then have

$$f(h) = \lim_{n\to\infty} Q(f, \ell_n) = \lim_{n\to\infty} \langle C(f), C(\ell_n)\rangle_{L^2(\mu)} = \langle C(f), \ell_h\rangle_{L^2(\mu)}. \tag{5.42}$$

This shows that evaluation on h is reproduced by the $L^2(\mu)$ inner product, a defining property of reproducing kernel Hilbert spaces. Moreover,

$$\|h\|^2_{H_\mu} = \|\ell_h\|^2_{L^2(\mu)} = \|h\|^2_{K_\mu}, \tag{5.43}$$

so the correspondence $h \leftrightarrow \ell_h$ is an isometry. Completeness and linearity yield a Hilbert space isomorphism between H_μ and K_μ. $\square$

This identification reveals a deep structural fact: the Cameron–Martin space can be realized equivalently as

- a space of admissible translations of μ in E,
- a space of Gaussian linear observables in $L^2(\mu)$,
- a reproducing kernel Hilbert space associated with the covariance of μ.

Thus, the RKHS formalism provides an intrinsic, coordinate-free description of the geometry underlying Gaussian measures on Banach spaces.

5.4.1 The Covariance Operator as a Factorization

The covariance form

$$Q(\ell_1, \ell_2) = \mathbb{E}[\ell_1(X)\ell_2(X)] \tag{5.44}$$

admits a canonical Hilbert space factorization through the reproducing kernel Hilbert space. Specifically, the covariance operator

$$R : E^* \to E \tag{5.45}$$

factors as

$$E^* \xrightarrow{C} H_\mu^* \xrightarrow{C^*} K_\mu \hookrightarrow E, \tag{5.46}$$

where C^* denotes the adjoint of the embedding $C : E^* \to L^2(\mu)$ restricted to H_μ^*. Explicitly,

$$R = i \circ C^* \circ C, \tag{5.47}$$

with $i : K_\mu \hookrightarrow E$ the canonical inclusion.

This factorization makes precise the idea that covariance is a “second-order” object: it arises from composing linear observation with its adjoint through the Gaussian Hilbert space H_μ^*. We emphasize that the map $R : E^* \to E$ is well-defined by Pettis integration: for each $\ell \in E^*$,

$$R\ell := \mathbb{E}[X\,\ell(X)], \tag{5.48}$$

where the expectation is taken in the weak sense, that is,

$$m(R\ell) = \mathbb{E}[m(X)\,\ell(X)] \quad \text{for all } m \in E^*. \tag{5.49}$$

This identity shows immediately that

$$m(R\ell) = Q(m, \ell), \tag{5.50}$$

so that R represents the covariance bilinear form Q via the dual pairing. In particular, R is linear and positive in the sense that

$$\ell(R\ell) = Q(\ell, \ell) \geq 0, \qquad \ell \in E^*. \tag{5.51}$$

The factorization

$$R = i \circ C^* \circ C \tag{5.52}$$

may be understood as a canonical Gelfand–Naimark–Segal construction associated with the positive semidefinite form Q. Indeed, the map

$$C : E^* \to L^2(\mu), \qquad C(\ell) = \ell(X), \tag{5.53}$$

satisfies

$$\langle C(\ell_1), C(\ell_2) \rangle_{L^2(\mu)} = Q(\ell_1, \ell_2), \tag{5.54}$$

so that C is an isometry from the quotient of E^* by the null space

$$\{\ell \in E^* : Q(\ell, \ell) = 0\} \tag{5.55}$$

into $L^2(\mu)$. The Hilbert space H_μ^* is precisely the completion of this image, making the factorization universal among Hilbert space representations of the covariance form.

The adjoint map

$$C^* : H_\mu^* \to K_\mu \tag{5.56}$$

is characterized abstractly by the relation

$$\langle C^*\varphi, h \rangle_{K_\mu} = \langle \varphi, \ell_h(X) \rangle_{L^2(\mu)}, \qquad \varphi \in H_\mu^*,\ h \in K_\mu. \tag{5.57}$$

In particular, for $\ell \in E^*$,

$$C^*C(\ell) = k_\ell, \tag{5.58}$$

where $k_\ell \in K_\mu$ is the reproducing kernel vector corresponding to ℓ, uniquely determined by

$$f(k_\ell) = Q(f, \ell), \qquad f \in E^*. \tag{5.59}$$

Thus the factorization recovers the covariance operator in the concrete form

$$R\ell = i(k_\ell), \tag{5.60}$$

exhibiting R as the composition of observation, orthogonal projection, and embedding.

From this perspective, the covariance operator is not primitive but derived: it is the pushforward to E of the orthogonal projection in $L^2(\mu)$ onto the first Wiener chaos H_μ^*. Symbolically,

$$R = \mathbb{E}[\,\cdot\, X] = \text{projection onto linear Gaussian observables.} \tag{5.61}$$

This interpretation explains why R need not be trace class or even bounded as an operator $E^* \to E$ in general: its natural domain and range are Hilbert spaces intrinsic to the Gaussian measure rather than the ambient Banach space.

Finally, the factorization clarifies the distinction between algebraic and topological structure. While Q is defined purely on E^*, its factorization through H_μ^* and K_μ reveals the hidden Hilbert geometry governing Gaussian fluctuations. In particular,

$$\|h\|_{K_\mu}^2 = \sup_{\ell \in E^* \setminus \{0\}} \frac{\ell(h)^2}{Q(\ell, \ell)}, \tag{5.62}$$

showing that the Cameron–Martin norm is dual to the covariance form. This duality is the analytic core underlying quasi-invariance, concentration inequalities, and the geometry of Gaussian measures on infinite-dimensional spaces.

5.4.2 Minimality and Universality of the RKHS

The reproducing kernel Hilbert space K_μ is minimal among Hilbert spaces embedded in E that reproduce the covariance form. More precisely:

Proposition 5.6. *Let H be a Hilbert space continuously embedded in E such that*

$$Q(\ell_1, \ell_2) = \langle h_{\ell_1}, h_{\ell_2} \rangle_H \quad \textit{for all } \ell_1, \ell_2 \in E^* \tag{5.63}$$

for some assignment $\ell \mapsto h_\ell \in H$. Then there exists a unique continuous isometric embedding

$$J : K_\mu \hookrightarrow H \tag{5.64}$$

intertwining the embeddings into E, i.e.

$$i_H \circ J = i_{K_\mu},$$

where $i_{K_\mu} : K_\mu \hookrightarrow E$ and $i_H : H \hookrightarrow E$ denote the canonical injections.

This universal property characterizes K_μ intrinsically: it is the *smallest* Hilbert space inside E that fully captures the covariance structure of μ. In particular, all Cameron–Martin directions are encoded in K_μ, and no additional directions are admissible.

The proof relies on the observation that the map

$$E^* \ni \ell \longmapsto h_\ell \in H$$

is necessarily linear and continuous when E^* is equipped with the seminorm $\ell \mapsto \sqrt{Q(\ell, \ell)}$. Indeed, by Cauchy–Schwarz in H,

$$|\langle h_{\ell_1}, h_{\ell_2} \rangle_H| \le \|h_{\ell_1}\|_H \, \|h_{\ell_2}\|_H = \sqrt{Q(\ell_1, \ell_1)} \sqrt{Q(\ell_2, \ell_2)}.$$

Thus the assignment $\ell \mapsto h_\ell$ factors uniquely through the completion of $E^*/\ker Q$, which is precisely the Gaussian Hilbert space H_μ^*.

By duality, this induces a unique isometric embedding

$$J : K_\mu \to H$$

such that for every $\ell \in E^*$,

$$J(k_\ell) = h_\ell,$$

where $k_\ell \in K_\mu$ is the reproducing kernel vector defined by

$$f(k_\ell) = Q(f, \ell), \qquad f \in E^*.$$

The isometry property follows immediately:

$$\|J(k_\ell)\|_H^2 = \|h_\ell\|_H^2 = Q(\ell, \ell) = \|k_\ell\|_{K_\mu}^2,$$

and density of $\{k_\ell : \ell \in E^*\}$ in K_μ extends J uniquely to all of K_μ.

From a categorical viewpoint, K_μ is the initial object in the category of Hilbert spaces continuously embedded in E that realize the covariance form Q. Any such realization necessarily factors through K_μ, and this factor is unique and isometric. Consequently, K_μ is independent of all auxiliary choices and depends only on the Gaussian measure μ.

Geometrically, this minimality expresses a rigidity phenomenon: although the Banach space E may contain many Hilbert subspaces, only one of them is compatible with the Gaussian measure. Any Hilbertian direction outside K_μ is invisible to μ in the sense that translation along such a direction destroys quasi-invariance. Thus K_μ exhausts exactly the set of directions along which the Gaussian measure admits a well-defined differential structure.

Equivalently, K_μ can be characterized variationally as

$$K_\mu = \left\{ h \in E : \sup_{\ell \in E^* \setminus \{0\}} \frac{\ell(h)^2}{Q(\ell, \ell)} < \infty \right\},$$

with norm given by

$$\|h\|_{K_\mu}^2 = \sup_{\ell \in E^* \setminus \{0\}} \frac{\ell(h)^2}{Q(\ell, \ell)}.$$

This formula shows explicitly that K_μ is the maximal subspace of E on which the covariance form induces a Hilbertian norm, completing the intrinsic and geometric characterization of the reproducing kernel Hilbert space.

5.4.3 RKHS versus Support of the Gaussian Measure

Although K_μ is dense in the support of μ in a topological sense, the two objects are fundamentally different. The topological support satisfies

$$\operatorname{supp}(\mu) = \overline{K_\mu}^{E}, \tag{5.65}$$

yet

$$\mu(K_\mu) = 0 \tag{5.66}$$

whenever E is infinite-dimensional. This follows from the fact that K_μ is a separable Hilbert space continuously embedded in E, while typical Gaussian sample paths exhibit roughness incompatible with Hilbert regularity.

More precisely, for any $r > 0$,

$$\mu(\{x \in E : \|x\|_{K_\mu} \le r\}) = 0, \tag{5.67}$$

since the norm $\|\cdot\|_{K_\mu}$ is not μ-measurable in a finite sense on E, and Fernique-type integrability holds only for the Banach norm $\|\cdot\|_E$. In contrast, for every $\varepsilon > 0$ and every $h \in K_\mu$, one has

$$\mu(B_E(h, \varepsilon)) > 0, \tag{5.68}$$

which implies that every $h \in K_\mu$ belongs to the support of μ.

This dichotomy reflects a core phenomenon of infinite-dimensional probability:

- K_μ governs deterministic directions of quasi-invariance and infinitesimal translations,
- typical samples $X(\omega)$ are almost surely *outside* K_μ.

Geometrically, K_μ should be interpreted as the tangent space of the measure, not as the space in which the random variable lives. For $h \in K_\mu$, the Cameron–Martin theorem shows that

$$\frac{d\mu_h}{d\mu}(x) = \exp\Big(\langle h, x\rangle_{K_\mu} - \tfrac{1}{2}\|h\|^2_{K_\mu}\Big), \tag{5.69}$$

so translations along K_μ preserve absolute continuity. No such formula exists for shifts outside K_μ, reflecting the singularity of μ_h with respect to μ.

From a probabilistic viewpoint, this means that K_μ captures *directions of regularity* rather than realizations of the random field. For instance, in the case of Wiener measure on $C_0([0,T])$, the RKHS consists of absolutely continuous paths with square-integrable derivatives, while Brownian paths are almost surely nowhere differentiable. Nevertheless,

$$\overline{H}^{C_0([0,T])} = C_0([0,T]), \tag{5.70}$$

so every continuous path can be approximated uniformly by Cameron–Martin paths, even though almost none of them are actually observed.

This separation between support and RKHS is a defining feature of infinite dimensions: the support describes *where the measure lives*, while the RKHS describes *how the measure can be infinitesimally moved*. Gaussian geometry, quasi-invariance, and isoperimetry are governed entirely by K_μ, even though μ assigns it zero mass.

5.4.4 Coordinate Representations and Orthonormal Expansions

Let $\{e_n\}_{n\geq 1}$ be an orthonormal basis of K_μ. Then there exists a sequence of i.i.d. standard Gaussian random variables $\{\xi_n\}_{n\geq 1}$ such that

$$X = \sum_{n=1}^{\infty} \xi_n e_n \quad \text{in } E \text{ almost surely.} \tag{5.71}$$

This expansion is a manifestation of the Karhunen–Loève theorem in the Banach space setting. More precisely, the partial sums

$$X^{(N)} := \sum_{n=1}^{N} \xi_n e_n \tag{5.72}$$

converge to X in E almost surely and in $L^2(\Omega; E)$, but generally fail to converge in the K_μ-norm.

The lack of convergence in K_μ can be quantified explicitly. Indeed,

$$\mathbb{E}\|X^{(N)}\|_{K_\mu}^2 = \sum_{n=1}^{N} \mathbb{E}[\xi_n^2] = N \xrightarrow[N\to\infty]{} \infty, \tag{5.73}$$

which shows that $X \notin K_\mu$ almost surely. This divergence reflects the fact that Gaussian fluctuations populate infinitely many orthogonal Cameron–Martin directions with unit variance.

In contrast, the covariance structure is perfectly captured by the expansion. For every $\ell \in E^*$,

$$\ell(X) = \sum_{n=1}^{\infty} \xi_n \, \ell(e_n) \quad \text{in } L^2(\Omega), \tag{5.74}$$

and hence

$$Q(\ell_1, \ell_2) = \sum_{n=1}^{\infty} \ell_1(e_n)\, \ell_2(e_n), \tag{5.75}$$

which is precisely the Parseval identity in K_μ. This identity shows that $\{e_n\}$ diagonalizes the covariance form.

For any $h \in K_\mu$, the Cameron–Martin norm admits the coordinate expression

$$\|h\|_{K_\mu}^2 = \sum_{n=1}^{\infty} \langle h, e_n \rangle_{K_\mu}^2, \tag{5.76}$$

and translation by h corresponds to a deterministic shift of the Gaussian coordinates:

$$\xi_n \longmapsto \xi_n + \langle h, e_n \rangle_{K_\mu}. \tag{5.77}$$

Under this transformation, the Radon–Nikodym derivative takes the explicit form

$$\frac{d\mu_h}{d\mu} = \exp\left(\sum_{n=1}^{\infty} \xi_n \langle h, e_n \rangle_{K_\mu} - \tfrac{1}{2} \sum_{n=1}^{\infty} \langle h, e_n \rangle_{K_\mu}^2 \right), \tag{5.78}$$

where the series converges in $L^1(\mu)$.

From a geometric perspective, the expansion reveals that K_μ provides the *principal axes* of the Gaussian measure. The directions e_n are those along which the measure decomposes into independent one-dimensional Gaussians, while the Banach space E merely serves as the ambient space ensuring almost sure convergence of the infinite series. Thus the orthonormal expansion simultaneously encodes independence, covariance, and quasi-invariance in a single coordinate framework.

5.4.5 RKHS and the Geometry of Gaussian Measures

The RKHS endows the Gaussian measure with a hidden Riemannian structure. The Cameron–Martin norm defines the "energy" required to shift the measure:

$$\frac{d\mu_h}{d\mu}(x) = \exp\left(\langle x, h \rangle_{K_\mu} - \frac{1}{2}\|h\|_{K_\mu}^2 \right), \qquad h \in K_\mu. \tag{5.79}$$

This formula shows that translations along K_μ act as geodesic directions in the space of measures, with $\|h\|_{K_\mu}^2$ playing the role of a squared distance or action.

More precisely, consider the relative entropy (Kullback–Leibler divergence) between the shifted measure μ_h and μ:

$$\mathrm{Ent}(\mu_h \,|\, \mu) := \int_E \log\left(\frac{d\mu_h}{d\mu}\right) d\mu_h. \tag{5.80}$$

A direct computation using the Cameron–Martin formula yields

$$\mathrm{Ent}(\mu_h \,|\, \mu) = \frac{1}{2}\|h\|_{K_\mu}^2. \tag{5.81}$$

Thus the Cameron–Martin norm measures the exact information-theoretic cost of transporting μ along the direction h, reinforcing its interpretation as a Riemannian metric tensor.

This structure is closely related to the geometry induced by the $L^2(\mu)$-gradient. For a smooth cylindrical functional $F : E \to \mathbb{R}$, the directional derivative along $h \in K_\mu$ satisfies

$$D_h F(x) = \langle \nabla_H F(x), h \rangle_{K_\mu}, \tag{5.82}$$

where $\nabla_H F$ denotes the Malliavin (or Cameron–Martin) gradient. The space K_μ therefore acts as the cotangent bundle for differential calculus on Gaussian space.

From the viewpoint of large deviations, the same quadratic structure governs the rate function. For a small-noise Gaussian perturbation $\sqrt{\varepsilon}X$, the associated rate functional is

$$I(h) = \begin{cases} \frac{1}{2}\|h\|_{K_\mu}^2, & h \in K_\mu, \\ +\infty, & h \notin K_\mu, \end{cases} \tag{5.83}$$

showing that only Cameron–Martin directions are admissible as finite-energy deviations. In this sense, K_μ is the space of classical trajectories around which random paths fluctuate.

Altogether, this geometric interpretation becomes foundational in:

- Malliavin calculus, where K_μ defines differentiability directions,
- large deviation theory, where $\|h\|_{K_\mu}^2/2$ is the universal rate function,
- stochastic analysis on path spaces, where K_μ plays the role of a Sobolev tangent space,
- information geometry of Gaussian measures, where K_μ induces a flat Riemannian metric on the Gaussian family.

In all these theories, the RKHS provides the invisible but rigid geometric skeleton underlying infinite-dimensional Gaussian randomness.

5.4.6 Relation to Abstract Wiener Spaces

Within an abstract Wiener space (E, H, μ), the RKHS K_μ coincides isometrically with the Cameron–Martin space H used to generate μ. In this sense, the RKHS construction *reconstructs* the abstract Wiener space purely from the measure μ, without reference to any prior Gaussian process or stochastic integral representation.

More precisely, if μ is realized as the pushforward of a canonical Gaussian measure on H under the continuous embedding

$$i : H \hookrightarrow E,$$

then the covariance operator satisfies

$$Q(\ell_1, \ell_2) = \langle i^*\ell_1, i^*\ell_2 \rangle_H, \qquad \ell_1, \ell_2 \in E^*, \tag{5.84}$$

and the RKHS K_μ is given by

$$K_\mu = i(H) \subset E \tag{5.85}$$

with inner product transported from H:

$$\langle i(h_1), i(h_2) \rangle_{K_\mu} = \langle h_1, h_2 \rangle_H. \tag{5.86}$$

Thus the abstract Wiener space structure is uniquely determined by μ up to canonical isomorphism.

This identification has an important universality property. Suppose (E, H_1, μ) and (E, H_2, μ) are two abstract Wiener space realizations of the same Gaussian measure μ. Then there exists a unique Hilbert space isometry

$$U : H_1 \longrightarrow H_2 \tag{5.87}$$

such that the embeddings into E satisfy $i_2 \circ U = i_1$. Hence the Cameron–Martin space is not merely a modeling choice but an invariant of the measure.

From a conceptual standpoint, this shows that Gaussian measures on Banach spaces carry an intrinsic first-order geometry: the RKHS K_μ plays the role of the tangent space, while the Banach space E provides only the ambient topology for sample paths. The abstract Wiener space formalism may therefore be summarized as the canonical factorization

$$\text{Gaussian measure } \mu \iff \text{Hilbert space } K_\mu \iff \text{admissible translations.} \tag{5.88}$$

Consequently, the triple

$$(E, K_\mu, \mu) \tag{5.89}$$

is canonical: no additional structure beyond the Gaussian measure is required. This perspective emphasizes that the Cameron–Martin space is not auxiliary but is intrinsically encoded in μ itself, providing the hidden linear geometry governing quasi-invariance, concentration, and infinite-dimensional differentiation.

5.5 Feldman–Hájek Dichotomy

We now describe the fundamental dichotomy for Gaussian measures, which asserts that Gaussian laws admit no intermediate notion of absolute continuity: they are either fully equivalent or completely disjoint.

Theorem 5.7 (Feldman–Hájek). *Let μ_1 and μ_2 be Gaussian measures on a separable Banach space E. Then exactly one of the following holds:*

1. *μ_1 and μ_2 are equivalent;*

2. *μ_1 and μ_2 are mutually singular.*

Moreover, μ_1 and μ_2 are equivalent if and only if:

1. *Their Cameron–Martin spaces coincide as sets and their norms are equivalent;*

2. *The difference of their means belongs to this common Cameron–Martin space;*

3. *The covariance operators agree on this space up to a Hilbert–Schmidt perturbation.*

This result is strikingly rigid: unlike general probability measures, Gaussian measures cannot be partially absolutely continuous. The dichotomy reflects the fact that Gaussian measures are entirely determined by first- and second-order data, and any incompatibility at these levels forces singularity.

To make the conditions precise, let $\mu_i = \mathcal{N}(m_i, Q_i)$, $i = 1, 2$, denote the Gaussian measures with means $m_i \in E$ and covariance forms $Q_i : E^* \times E^* \to \mathbb{R}$. Let H_i denote the corresponding Cameron–Martin spaces. Condition (1) requires

$$H_1 = H_2 \quad \text{as vector spaces,} \tag{5.90}$$

together with norm equivalence:

$$c\,\|h\|_{H_1} \leq \|h\|_{H_2} \leq C\,\|h\|_{H_1}, \qquad h \in H_1, \tag{5.91}$$

for some constants $0 < c \leq C < \infty$. Thus, equivalence of Gaussian measures forces their hidden Hilbert geometries to coincide.

Condition (2) encodes the only admissible way to change the mean:

$$m_2 - m_1 \in H_1. \tag{5.92}$$

If the mean shift lies outside the Cameron–Martin space, then the translated measure exits the quasi-invariant directions entirely, and singularity follows immediately from the Cameron–Martin theorem.

Condition (3) concerns the covariance structure. Identifying $H := H_1 = H_2$, let $C_i : H \to H$ denote the covariance operators restricted to H. Equivalence requires that

$$C_2^{-1/2} C_1 C_2^{-1/2} - I \tag{5.93}$$

is a Hilbert–Schmidt operator on H. Equivalently, if $\{\lambda_n\}$ denote the eigenvalues of $C_2^{-1/2} C_1 C_2^{-1/2}$, then

$$\sum_{n=1}^{\infty} (\lambda_n - 1)^2 < \infty. \tag{5.94}$$

This condition quantifies the idea that the two covariance structures may differ only in "square-summable directions."

Geometrically, the Feldman–Hájek theorem states that two Gaussian measures are equivalent precisely when they induce the same Cameron–Martin geometry, up to a finite-energy deformation. Any discrepancy in admissible directions, mean shift, or second-order structure accumulates across infinitely many dimensions and forces mutual singularity.

The proof combines finite-dimensional Gaussian equivalence, orthogonal decompositions of the Cameron–Martin space, and infinite products of likelihood ratios. Analytically, it relies on a delicate interplay between the Cameron–Martin theorem, Girsanov transformations, and trace-class perturbation theory, which together propagate finite-dimensional equivalence to the full infinite-dimensional setting.

5.6 Applications to Infinite-Dimensional Analysis

We briefly mention several applications which will be explored in detail in later chapters:

1. Quasi-invariance of Wiener measure on path space;
2. Malliavin calculus: the Cameron–Martin space as the domain of the Malliavin derivative;
3. Regularity theory for Gaussian processes through RKHS geometry;
4. Small-ball asymptotics for Gaussian measures.

These applications depend critically on the structural results proved in this chapter.

At a conceptual level, the Cameron–Martin space and the RKHS provide the *deterministic skeleton* underlying random phenomena in infinite dimensions. While typical realizations of a Gaussian field are highly irregular, its RKHS encodes the directions along which analytic control is possible. This duality between rough samples and smooth directions is the organizing principle behind most infinite-dimensional techniques.

In the case of Wiener measure on the path space $E = C_0([0,T];\mathbb{R}^d)$, the abstract theory specializes to the classical Cameron–Martin theorem. The Cameron–Martin space is given by

$$H = \left\{ h \in C_0([0,T];\mathbb{R}^d) : h \text{ absolutely continuous, } \dot{h} \in L^2([0,T];\mathbb{R}^d) \right\}, \tag{5.95}$$

with norm

$$\|h\|_H^2 = \int_0^T |\dot{h}(t)|^2 \, dt. \tag{5.96}$$

Translations by elements of H preserve equivalence of measures, while any perturbation outside H produces singularity. This dichotomy underlies the analysis of stochastic differential equations and stochastic flows.

Malliavin calculus builds directly on this structure. The Malliavin derivative D acts as a directional derivative along Cameron–Martin directions:

$$D_h F = \lim_{\varepsilon \to 0} \frac{F(X + \varepsilon h) - F(X)}{\varepsilon}, \qquad h \in K_\mu. \tag{5.97}$$

Thus K_μ plays the role of a tangent space to the probability space, and the associated Sobolev norms measure regularity of random variables with respect to Gaussian perturbations. Integration by parts formulas, absolute continuity of laws, and smoothness of densities all hinge on this Hilbertian structure.

Regularity theory for Gaussian processes is likewise governed by the RKHS. For a Gaussian field X indexed by a set T, the RKHS norm controls pathwise regularity via deterministic embeddings. For example, bounds of the form

$$\|h\|_{K_\mu} \leq C\|h\|_{\mathcal{C}^\alpha(T)} \tag{5.98}$$

translate into Hölder regularity of sample paths through entropy and chaining arguments. In this way, analytic properties of K_μ yield probabilistic regularity results for X.

Finally, small-ball probabilities probe the fine geometry of Gaussian measures near the origin. As $\varepsilon \downarrow 0$, asymptotics of the form

$$\mu(\|X\|_E \leq \varepsilon) \approx \exp\left(-\frac{1}{2} \inf_{\|h\|_E \leq \varepsilon} \|h\|_{K_\mu}^2\right) \tag{5.99}$$

reveal that the Cameron–Martin norm governs the exponential decay rate. This connection is central to large deviation theory, Bayesian inverse problems, and the study of Gaussian priors in infinite-dimensional statistics.

Taken together, these applications illustrate that the Cameron–Martin space and the RKHS are not merely technical constructions but the unifying geometric framework for infinite-dimensional analysis driven by Gaussian measures.

5.7 Summary

In this chapter we developed:

1. The construction and interpretation of the Cameron–Martin space;
2. The reproducing kernel Hilbert space associated to a Gaussian measure;
3. The Feldman–Hájek dichotomy characterizing when two Gaussian measures are equivalent.

These tools form the theoretical backbone for advanced topics in stochastic analysis, to be pursued in subsequent chapters.

Beyond these structural results, it is worth emphasizing the unifying role played by the Cameron–Martin space across analytic, probabilistic, and geometric viewpoints. From an analytic perspective, K_μ identifies the maximal subspace of E on which the Gaussian measure exhibits differentiability properties. In particular, for sufficiently regular functionals $F : E \to \mathbb{R}$, directional derivatives along $h \in K_\mu$ admit an integration-by-parts formula of the form

$$\int_E D_h F(x)\, d\mu(x) = \int_E F(x)\, \langle x, h\rangle_{K_\mu}\, d\mu(x), \qquad h \in K_\mu, \tag{5.100}$$

which fails for generic directions in $E \setminus K_\mu$. This identity highlights the sharp boundary between probabilistically admissible and inadmissible directions.

From a geometric viewpoint, the Cameron–Martin norm induces a quadratic form that governs both large deviations and measure transportation. For a family of rescaled Gaussian measures $(\mu_\varepsilon)_{\varepsilon>0}$ defined by

$$\mu_\varepsilon = \mathcal{L}(\varepsilon X), \qquad X \sim \mu, \tag{5.101}$$

the associated large deviation principle is controlled by the rate function

$$I(x) = \begin{cases} \frac{1}{2}\|x\|^2_{K_\mu}, & x \in K_\mu, \\ +\infty, & x \notin K_\mu, \end{cases} \tag{5.102}$$

thereby singling out K_μ as the space of admissible macroscopic deviations.

Finally, the Feldman–Hájek dichotomy provides a conceptual explanation for the extreme rigidity of Gaussian measures in infinite dimensions. Unlike finite-dimensional settings, where absolute continuity is generic, infinite-dimensional Gaussian measures typically live on essentially disjoint supports unless their Cameron–Martin geometries are finely matched. This rigidity can be succinctly expressed by the implication

$$K_{\mu_1} \neq K_{\mu_2} \quad \Longrightarrow \quad \mu_1 \perp \mu_2, \tag{5.103}$$

underscoring the central role of the Cameron–Martin space as the fundamental invariant of Gaussian measures.

Taken together, these results reveal a coherent picture: Gaussian measures on infinite-dimensional spaces are governed not by the ambient topology of E, but by the hidden Hilbertian geometry encoded in their reproducing kernel Hilbert spaces. This insight will repeatedly resurface in the analysis of stochastic partial differential equations, path-space geometry, and probabilistic aspects of functional analysis.

5.8 References

Parthasarathy (2005) [42], Bogachev (1998) [1], Adler (2007) [43], Berlinet (2011) [44], Lifshits (2012) [45], Prato (2007) [46], Vakhania (2012) [2], Kuo (2006) [3], Minlos (1959) [10], Kallianpur (2013) [47].

Chapter 6

The Feldman–Hájek Theorem: Equivalence and Singularity of Gaussian Measures

6.1 Introduction

The purpose of this chapter is to present a complete, rigorous, and self-contained derivation of the Feldman–Hájek theorem, the central classification theorem for Gaussian measures on infinite-dimensional separable Banach and Hilbert spaces. This theorem describes precisely when two Gaussian measures are either mutually absolutely continuous or mutually singular, showing that no intermediate relation is possible.

The result generalizes classical finite-dimensional linear algebraic characterizations of Gaussian densities to the setting of infinite-dimensional analysis, where additional geometric and operator-theoretic structure arises.

At a conceptual level, the Feldman–Hájek theorem reveals a striking rigidity phenomenon: in infinite dimensions, Gaussian measures behave in an essentially "all-or-nothing" fashion with respect to absolute continuity. This stands in sharp contrast with general probability measures, for which intermediate notions such as partial absolute continuity or overlap of supports may occur. For Gaussian measures, however, the linear structure of the underlying space and the quadratic nature of the covariance completely determine the measure-theoretic relationship.

A guiding principle throughout this chapter is that Gaussian measures are fully characterized by their first and second moments. In finite dimensions, this leads to the well-known criterion that two Gaussian measures with densities $d\mu_i(x) \propto \exp\Big(-\frac{1}{2}\langle Q_i^{-1}x, x\rangle\Big)$ are equivalent if and only if their covariance matrices are comparable. In infinite dimensions, the inverse covariance operator typically does not exist as a bounded operator, and the correct replacement is the geometry of the Cameron–Martin spaces together with trace-class and Hilbert–Schmidt perturbations.

Formally, if E is a Hilbert space and $Q_i : E \to E$ are the covariance operators, the associated Cameron–Martin spaces admit the representation

$$H_i = \overline{\operatorname{Ran} Q_i^{1/2}}^{\|\cdot\|_{H_i}}, \qquad \|h\|_{H_i} = \|Q_i^{-1/2}h\|_E, \tag{6.1}$$

where $Q_i^{-1/2}$ is understood as an unbounded operator with domain $\operatorname{Ran} Q_i^{1/2}$. This representation makes explicit that H_i depends sensitively on the small-eigenvalue behavior of Q_i.

A central theme of the proof is the comparison of the two Gaussian measures via finite-dimensional approximations. Let $\{P_n\}_{n\geq 1}$ be a sequence of finite-rank projections converging strongly to the identity on E. Denoting by $\mu_i^{(n)} = \mu_i \circ P_n^{-1}$ the projected measures, one has

$$\mu_1 \sim \mu_2 \quad \Longleftrightarrow \quad \sup_{n\geq 1} \left\| \frac{d\mu_1^{(n)}}{d\mu_2^{(n)}} \right\|_{L^p(\mu_2^{(n)})} < \infty \quad \text{for some } p > 1, \tag{6.2}$$

linking infinite-dimensional equivalence to uniform control of finite-dimensional Radon–Nikodym derivatives.

Even in the centered case considered here, the theorem already exhibits its full subtlety. When non-centered Gaussian measures are allowed, an additional shift term appears, and equivalence holds if and only if the difference of the means belongs to the common Cameron–Martin space:

$$m_1 - m_2 \in H_1 = H_2. \tag{6.3}$$

This condition reflects the fact that only Cameron–Martin directions admit absolutely continuous translations.

The remainder of this chapter is devoted to a precise formulation and proof of these statements. We begin by reviewing the necessary operator-theoretic preliminaries, followed by a detailed analysis of equivalence in finite dimensions, and finally pass to the infinite-dimensional limit. Along the way, we emphasize the geometric meaning of each condition, making clear why the Cameron–Martin space and Hilbert–Schmidt perturbations emerge as the natural invariants of Gaussian measures.

6.2 Notation and Preliminaries

Let E be a real separable Banach space, and let μ_1 and μ_2 be centered Gaussian probability measures on $(E, \mathcal{B}(E))$ with covariance operators Q_1 and Q_2. We denote by H_1 and H_2 the associated Cameron–Martin spaces. Given a centered Gaussian measure μ on E with covariance form

$$Q(\ell_1, \ell_2) = \int_E \ell_1(x)\ell_2(x)\, d\mu(x), \qquad \ell_1, \ell_2 \in E^*, \tag{6.4}$$

its Cameron–Martin space H is the completion of the image of E^* under the canonical embedding

$$E^* \to L^2(E, \mu), \qquad \ell \mapsto \ell(\cdot). \tag{6.5}$$

The map $Q : E^* \to H$ is given by

$$Q\ell = \int_E x\, \ell(x)\, d\mu(x). \tag{6.6}$$

For each Gaussian measure μ_i, we denote by H_i the associated Cameron–Martin space and by $Q_i : E^* \to H_i$ its covariance operator. Let $\langle \cdot, \cdot \rangle_{H_i}$ be the inner product in H_i and $\| \cdot \|_{H_i}$ the Cameron–Martin norm. We recall that the covariance form Q is symmetric, positive semidefinite, and continuous with respect to the operator norm on E^*. In particular,

$$Q(\ell, \ell) \geq 0, \qquad Q(\ell_1, \ell_2) = Q(\ell_2, \ell_1), \tag{6.7}$$

and for each $\ell \in E^*$ one has

$$Q(\ell, \ell) = \int_E \ell(x)^2 \, d\mu(x) < \infty, \tag{6.8}$$

which reflects the fact that all continuous linear functionals on E are square-integrable under a Gaussian measure.

The Cameron–Martin space H may equivalently be described as the closure of $\{Q\ell : \ell \in E^*\}$ equipped with the inner product

$$\langle Q\ell_1, Q\ell_2 \rangle_H := Q(\ell_1, \ell_2). \tag{6.9}$$

This definition is consistent because the covariance form is positive definite on the quotient $E^*/\ker Q$, and thus induces a genuine Hilbert structure after completion. The mapping $Q : E^* \to H$ is injective modulo $\ker Q$ and has dense range in H.

There is a canonical continuous embedding

$$i : H \hookrightarrow E, \tag{6.10}$$

characterized by the reproducing property

$$\ell(i(h)) = \langle h, Q\ell \rangle_H, \qquad h \in H, \ \ell \in E^*. \tag{6.11}$$

This identity expresses the fact that evaluation by linear functionals on E extends continuously to elements of H, and it underlies the interpretation of H as a space of "directions" inside E.

For each measure μ_i, the corresponding Cameron–Martin space H_i embeds densely into the topological support of μ_i, but the embedding is never surjective when E is infinite-dimensional. In particular,

$$\mu_i(H_i) = 0, \tag{6.12}$$

even though H_i determines all quasi-invariant translations of μ_i.

We also note that the covariance operator Q_i extends uniquely to a bounded, self-adjoint operator

$$Q_i : E^* \to E^{**}, \tag{6.13}$$

and, when E is reflexive or Hilbert, this allows one to identify Q_i with a positive, symmetric operator acting directly on E. In the Hilbert space case, one recovers the familiar representation

$$Q_i = \mathbb{E}[\, X_i \otimes X_i \,], \tag{6.14}$$

where X_i is an E-valued Gaussian random variable with law μ_i.

Throughout this chapter, we shall compare the Gaussian measures μ_1 and μ_2 by analyzing the relationship between H_1 and H_2, their norms, and the corresponding covariance operators. These objects provide the natural language in which equivalence and singularity of Gaussian measures can be formulated and proved.

6.3 Statement of the Feldman–Hájek Theorem

We now state the main result of this chapter in its classical operator-theoretic form.

Theorem 6.1 (Feldman–Hájek)**.** *Let μ_1 and μ_2 be centered Gaussian measures on a separable Banach space E with Cameron–Martin spaces H_1 and H_2, and covariance operators Q_1 and Q_2. Then exactly one of the following holds:*

(i) μ_1 and μ_2 are equivalent *(mutually absolutely continuous),*

(ii) μ_1 and μ_2 are mutually singular.

Moreover, (i) holds if and only if the following three conditions are satisfied:

(a) $H_1 = H_2$ as sets with equivalent norms,

(b) the operator

$$T = Q_1^{-1/2} Q_2 Q_1^{-1/2} - I \tag{6.15}$$

is Hilbert–Schmidt on H_1,

(c) the mean vectors (here zero) differ by an element of H_1.

In the centered case, condition (c) is automatically satisfied.

Proof. **Part I: Dichotomy:** Gaussian measures on a separable Banach space are uniquely determined by their finite-dimensional cylindrical projections. Let $\pi_F : E \to F$ denote the projection onto a finite-dimensional subspace $F \subset E^*$. For each such F, the pushforward measures

$$(\mu_i)_F := (\pi_F)_\# \mu_i \tag{6.16}$$

are Gaussian measures on $\mathbb{R}^n$. In finite dimensions, Gaussian measures are either equivalent or mutually singular.

If singularity occurs for at least one projection F, then pulling back by π_F^{-1} yields singularity of μ_1 and μ_2 on E. Conversely, if equivalence holds for all finite-dimensional projections, Kolmogorov consistency and projective limit arguments imply equivalence on E. Thus exactly one of equivalence or singularity must hold.

Part II: Necessity: Assume $\mu_1 \sim \mu_2$. We consider the Equality of Cameron–Martin spaces. For $h \in E$, define the translation map $\tau_h(x) = x + h$. By the Cameron–Martin theorem,

$$\tau_h \mu_i \sim \mu_i \quad \Longleftrightarrow \quad h \in H_i. \tag{6.17}$$

Since equivalence is preserved under absolutely continuous changes of measure,

$$\tau_h \mu_1 \sim \mu_1 \quad \Longleftrightarrow \quad \tau_h \mu_2 \sim \mu_2, \tag{6.18}$$

which implies $H_1 = H_2$ as sets. Moreover, the Radon–Nikodým derivative

$$\frac{d(\tau_h \mu_i)}{d\mu_i}(x) = \exp\Big(\langle Q_i^{-1} h, x\rangle - \tfrac{1}{2}\|h\|_{H_i}^2\Big) \tag{6.19}$$

must define an $L^1(\mu_j)$ function for both i, j, which forces equivalence of the norms $\|\cdot\|_{H_1}$ and $\|\cdot\|_{H_2}$.

We now discuss the Hilbert–Schmidt condition. Let $H := H_1 = H_2$. Define

$$A := Q_1^{-1/2} Q_2 Q_1^{-1/2} \quad \text{on } H. \tag{6.20}$$

Equivalence of μ_1 and μ_2 implies equivalence of all finite-dimensional projections. Let $\{e_k\}$ be an orthonormal basis of H. For each n, let P_n denote the orthogonal projection onto $\operatorname{span}(e_1, \ldots, e_n)$. The projected measures satisfy

$$(\mu_1)_n = \mathcal{N}(0, I_n), \qquad (\mu_2)_n = \mathcal{N}(0, P_n A P_n). \tag{6.21}$$

Finite-dimensional equivalence forces

$$\operatorname{Tr}\big((P_n A P_n - I_n)^2\big) < \infty \quad \text{uniformly in } n. \tag{6.22}$$

Passing to the limit yields

$$\sum_{k=1}^{\infty} \|(A - I)e_k\|^2 < \infty, \tag{6.23}$$

hence $A - I \in \mathcal{L}_2(H)$.

We now discuss the Mean difference. Let $m := m_1 - m_2$. Then $\mu_1 = \tau_m \mu_2$. By the Cameron–Martin theorem,

$$\tau_m \mu_2 \sim \mu_2 \quad \Longleftrightarrow \quad m \in H. \tag{6.24}$$

Thus condition (c) is necessary.

Part III: Sufficiency: Assume conditions (a), (b), and (c). We first do the reduction to the centered case. Let us define $\tilde{\mu}_i := \tau_{-m_i} \mu_i$. Then

$$\mu_1 \sim \mu_2 \quad \Longleftrightarrow \quad \tilde{\mu}_1 \sim \tilde{\mu}_2. \tag{6.25}$$

Hence we may assume $m_1 = m_2 = 0$.

The second step is Finite-dimensional equivalence. Let $\{e_k\}$ be an orthonormal basis of H and let P_n be the projection onto $\mathrm{span}(e_1, \dots, e_n)$. The projected measures satisfy

$$(\mu_1)_n = \mathcal{N}(0, I_n), \qquad (\mu_2)_n = \mathcal{N}(0, P_n A P_n). \tag{6.26}$$

Since $A - I \in \mathcal{L}_2(H)$, the finite-dimensional Feldman–Hájek criterion holds for all n, implying

$$(\mu_1)_n \sim (\mu_2)_n \quad \forall n. \tag{6.27}$$

We now discuss the Infinite-dimensional equivalence. Note that the Radon–Nikodým derivatives $\frac{d(\mu_2)_n}{d(\mu_1)_n}$ form a consistent L^1-bounded martingale, which converges in L^1 to $d\mu_2/d\mu_1$. Hence $\mu_1 \sim \mu_2$.

Part IV: Failure of Conditions: If any of (a), (b), or (c) fails, one can construct a sequence of finite projections for which the projected measures are singular. By pullback, this implies $\mu_1 \perp \mu_2$.

Conclusion: Exactly one of equivalence or singularity holds, and equivalence occurs if and only if conditions (a), (b), and (c) are satisfied. This completes the proof of Theorem 6.1. □

The theorem asserts a striking dichotomy: in infinite dimensions, two Gaussian measures can never be partially overlapping. Either they define the same null sets, or they concentrate on disjoint measurable subsets of E. This rigidity has no analogue for general probability measures and is a direct consequence of the quadratic structure encoded in Gaussian covariance operators.

Condition (a) identifies the Cameron–Martin space as the fundamental geometric object governing equivalence. Equality of H_1 and H_2 as sets ensures that the two measures admit the same directions of quasi-invariance, while equivalence of the norms implies that the associated energies are comparable. More precisely, there exist constants $c, C > 0$ such that

$$c \|h\|_{H_1} \le \|h\|_{H_2} \le C \|h\|_{H_1}, \qquad h \in H_1 = H_2. \tag{6.28}$$

If this fails, then one measure admits shifts along directions that are forbidden for the other, forcing mutual singularity.

Condition (b) quantifies the "distance" between the two covariance operators. The operator

$$T = Q_1^{-1/2} Q_2 Q_1^{-1/2} - I \tag{6.29}$$

acts on the common Cameron–Martin space and measures the deviation of Q_2 from Q_1 relative to the geometry induced by Q_1. The Hilbert–Schmidt condition

$$\|T\|_{\mathrm{HS}}^2 = \sum_{n=1}^{\infty} \|Te_n\|_{H_1}^2 < \infty, \tag{6.30}$$

for any orthonormal basis $\{e_n\}$ of H_1, expresses that this deviation is "small" in a quadratic sense. If T fails to be Hilbert–Schmidt, the two Gaussian measures concentrate on asymptotically orthogonal directions and become singular.

In finite dimensions, every linear operator is Hilbert–Schmidt, and condition (b) is automatically satisfied. In that case, the Feldman–Hájek theorem reduces to the classical statement that two nondegenerate Gaussian densities are equivalent if and only if their covariance matrices are both invertible and their means differ by a finite vector. Thus the infinite-dimensional theory may be viewed as a refinement of finite-dimensional linear algebra, where compactness and trace conditions replace determinants.

Finally, although condition (c) is trivial in the centered case, it plays a crucial role for non-centered Gaussian measures. If $\mu_1 = \mathcal{N}(m_1, Q_1)$ and $\mu_2 = \mathcal{N}(m_2, Q_2)$, then equivalence additionally requires

$$m_2 - m_1 \in H_1, \tag{6.31}$$

reflecting the Cameron–Martin theorem: translations outside the Cameron–Martin space destroy absolute continuity. Together, conditions (a)–(c) provide a complete and sharp classification of Gaussian measures up to equivalence.

6.4 Discussion of the Structure of the Proof

The proof consists of several components:

1. The linear-algebraic equivalence of Gaussians on finite-dimensional subspaces.
2. A projective limit argument using consistency of projections.
3. A reduction to the Hilbert space generated by the covariance forms.
4. A characterization of equivalence and singularity in terms of Cameron–Martin spaces.
5. A spectral-operator criterion involving Hilbert–Schmidt perturbations.

The final result integrates all these components in a unified operator-theoretic framework.

The guiding philosophy of the proof is to reduce an infinite-dimensional probabilistic statement to a sequence of finite-dimensional ones, while keeping precise control over the limiting behavior. Gaussian measures are uniquely suited to this strategy because their laws are completely determined by linear functionals and second moments, allowing one to pass between finite- and infinite-dimensional settings without loss of information.

The starting point is the finite-dimensional case. Given a finite-dimensional subspace $F \subset E^*$, the pushforward measures

$$(\pi_F)_\# \mu_i, \qquad \pi_F(x) = (\ell(x))_{\ell \in F}, \tag{6.32}$$

are Gaussian measures on $\mathbb{R}^{\dim F}$ with explicitly computable covariance matrices. In this setting, equivalence and singularity can be characterized by classical linear algebra: two centered Gaussian measures are equivalent if and only if their covariance matrices define equivalent quadratic forms. This yields Radon–Nikodym derivatives of the form

$$\frac{d(\pi_F)_{\#}\mu_2}{d(\pi_F)_{\#}\mu_1}(y) = \exp\Big(-\tfrac{1}{2}\langle y, A_F y\rangle + c_F\Big), \tag{6.33}$$

where A_F depends explicitly on the finite-dimensional covariance operators.

The projective limit argument then exploits the consistency of these finite-dimensional projections. If μ_1 and μ_2 are equivalent on every finite-dimensional projection, one seeks to pass to the limit as $F \uparrow E^*$. This step is highly nontrivial: uniform integrability of the Radon–Nikodym derivatives must be established to prevent loss of mass in the limit. Gaussian structure again plays a decisive role, as exponential quadratic forms admit sharp moment bounds that allow one to control these limits.

A crucial simplification arises by reducing the problem to the Hilbert space generated by the covariance forms. More precisely, one replaces E by the completion of $E^*/\ker Q_i$ under the inner product induced by Q_i, thereby working on the Cameron–Martin spaces H_i. This reduction shows that only the geometry induced by the covariance operators is relevant; directions orthogonal to H_i are invisible to the measure and contribute neither to equivalence nor to singularity.

At this stage, the Cameron–Martin theorem provides the key dichotomy for shifts. If $H_1 \neq H_2$, then there exists a direction h that is admissible for one measure but forbidden for the other, leading immediately to mutual singularity. If $H_1 = H_2$ as sets, the problem reduces to comparing the two Gaussian structures on the same Hilbert space. This comparison is encoded in the relative covariance operator

$$T = Q_1^{-1/2} Q_2 Q_1^{-1/2}, \tag{6.34}$$

which is self-adjoint and positive on H_1.

The final step is spectral. Writing the spectral decomposition

$$T = \sum_{n=1}^{\infty} \lambda_n \,\langle \cdot, e_n\rangle e_n, \tag{6.35}$$

one shows that equivalence of μ_1 and μ_2 is equivalent to the square summability condition

$$\sum_{n=1}^{\infty} (\lambda_n - 1)^2 < \infty, \tag{6.36}$$

i.e. $T - I$ is Hilbert–Schmidt. If this series diverges, the likelihood ratios associated with finite-dimensional truncations either explode or vanish in the limit, forcing singularity.

Thus the proof weaves together finite-dimensional Gaussian calculus, projective limits, Cameron–Martin geometry, and spectral theory into a single coherent argument. Each component is indispensable: removing any one of them breaks the logical chain leading to the Feldman–Hájek dichotomy.

6.5 Finite-Dimensional Preliminaries

We begin with the classical finite-dimensional classification.

Lemma 6.2 (Finite-dimensional Feldman–Hájek criterion)**.** *Let ν_1 and ν_2 be centered Gaussian probability measures on $\mathbb{R}^n$ with covariance matrices Σ_1 and Σ_2. Then ν_1 and ν_2 are equivalent if and only if Σ_1 and Σ_2 are positive definite. In this case,*

$$\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2} - I \tag{6.37}$$

is automatically Hilbert–Schmidt (equivalently finite rank). If either covariance matrix is singular, then ν_1 and ν_2 are mutually singular.

Proof. We proceed in several logically independent steps.

Step 1: Explicit densities. A centered Gaussian measure ν on $\mathbb{R}^n$ with covariance matrix Σ admits a density with respect to Lebesgue measure λ if and only if Σ is positive definite. In that case,

$$\frac{d\nu}{d\lambda}(x) = (2\pi)^{-n/2}(\det\Sigma)^{-1/2}\exp\Big(-\tfrac{1}{2}x^\top\Sigma^{-1}x\Big), \qquad x \in \mathbb{R}^n. \tag{6.38}$$

If Σ is singular, then ν is supported on the proper linear subspace $\mathrm{Ran}(\Sigma^{1/2}) \subsetneq \mathbb{R}^n$ and is singular with respect to λ.

Step 2: Singularity when a covariance is degenerate: Suppose Σ_1 is singular. Then ν_1 is supported on the affine subspace $\mathrm{Ran}(\Sigma_1^{1/2})$. If Σ_2 is nondegenerate, ν_2 is equivalent to Lebesgue measure and assigns positive mass to sets disjoint from $\mathrm{Ran}(\Sigma_1^{1/2})$, hence $\nu_1 \perp \nu_2$. If Σ_2 is also singular but $\mathrm{Ran}(\Sigma_1^{1/2}) \neq \mathrm{Ran}(\Sigma_2^{1/2})$, then the supports are distinct linear subspaces and the measures are again mutually singular. Thus equivalence is possible only if both Σ_1 and Σ_2 are positive definite.

Step 3: Absolute continuity when both covariances are positive definite: Assume now that Σ_1 and Σ_2 are positive definite. Then both ν_1 and ν_2 admit smooth strictly positive densities with respect to Lebesgue measure. Consequently,

$$\nu_1 \ll \lambda \quad \text{and} \quad \nu_2 \ll \lambda,$$

and hence $\nu_1 \sim \nu_2$. More explicitly, the Radon–Nikodým derivative is given by

$$\frac{d\nu_2}{d\nu_1}(x) = \left(\frac{\det\Sigma_1}{\det\Sigma_2}\right)^{1/2}\exp\Big(-\tfrac{1}{2}x^\top(\Sigma_2^{-1} - \Sigma_1^{-1})x\Big), \tag{6.39}$$

which is finite and strictly positive for all $x \in \mathbb{R}^n$.

Step 4: Operator-theoretic formulation: Since $\mathbb{R}^n$ is finite-dimensional, every linear operator is compact and Hilbert–Schmidt. In particular,

$$T := \Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2} - I \tag{6.40}$$

is a symmetric operator of rank at most n and therefore Hilbert–Schmidt. Thus the Hilbert–Schmidt condition imposes no restriction beyond positive definiteness of the covariance matrices.

Step 5: Conclusion: We have shown:

- If either covariance matrix is singular, then ν_1 and ν_2 are mutually singular.
- If both covariance matrices are positive definite, then ν_1 and ν_2 are equivalent, and the operator $\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2} - I$ is automatically Hilbert–Schmidt.

This completes the proof. □

The proof is immediate from linear algebra: finite-dimensional Gaussian densities exist and are equivalent if and only if their determinants are nonzero and their covariance operators differ by a finite-rank symmetric matrix.

This lemma may be sharpened by making the equivalence criterion fully explicit in terms of densities and spectral data. If Σ_i are positive definite, the Gaussian measures ν_i admit Lebesgue densities

$$\frac{d\nu_i}{dx}(x) = (2\pi)^{-n/2}(\det \Sigma_i)^{-1/2} \exp\Big(-\tfrac{1}{2}\langle \Sigma_i^{-1}x, x\rangle_{\mathbb{R}^n}\Big), \qquad i = 1, 2. \tag{6.41}$$

Consequently, the Radon–Nikodym derivative of ν_2 with respect to ν_1 exists and is given by

$$\frac{d\nu_2}{d\nu_1}(x) = \left(\frac{\det \Sigma_1}{\det \Sigma_2}\right)^{1/2} \exp\Big(-\tfrac{1}{2}\langle (\Sigma_2^{-1} - \Sigma_1^{-1})x, x\rangle\Big). \tag{6.42}$$

This expression is finite and strictly positive for all $x \in \mathbb{R}^n$ if and only if both covariance matrices are invertible, immediately yielding equivalence.

The operator-theoretic formulation follows by diagonalizing the relative covariance. Let

$$T := \Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2}, \tag{6.43}$$

which is symmetric and positive definite on $\mathbb{R}^n$. By the spectral theorem, there exists an orthonormal basis $\{e_k\}_{k=1}^n$ and eigenvalues $\lambda_k > 0$ such that

$$T = \sum_{k=1}^{n} \lambda_k \, e_k \otimes e_k. \tag{6.44}$$

In these coordinates, the Radon–Nikodym derivative factorizes as a product of one-dimensional likelihood ratios:

$$\frac{d\nu_2}{d\nu_1}(x) = \prod_{k=1}^{n} \lambda_k^{-1/2} \exp\Big(-\tfrac{1}{2}(\lambda_k^{-1} - 1)\,\xi_k^2\Big), \qquad \xi_k = \langle \Sigma_1^{-1/2}x, e_k\rangle. \tag{6.45}$$

This representation highlights that equivalence depends only on the behavior of the eigenvalues λ_k relative to 1.

If either covariance matrix is singular, then there exists a nonzero vector $v \in \mathbb{R}^n$ such that $\langle v, x\rangle = 0$ almost surely under one measure but not the other. In this case, the supports of ν_1 and ν_2 lie in distinct affine subspaces, and one obtains

$$\nu_1 \perp \nu_2. \tag{6.46}$$

Thus, in finite dimensions, equivalence and singularity are completely governed by linear-algebraic properties of the covariance matrices.

This finite-dimensional classification serves as the template for the infinite-dimensional theory: invertibility corresponds to equality of Cameron–Martin spaces, while the finite-rank condition anticipates the Hilbert–Schmidt perturbation criterion appearing in the Feldman–Hájek theorem.

6.6 Cylindrical Projections

Let $\pi_F : E \to F$ denote the canonical projection onto a finite-dimensional subspace $F \subset E^*$, as constructed earlier. For each F we have Gaussian measures

$$(\mu_1)_F = (\pi_F)_\# \mu_1, \qquad (\mu_2)_F = (\pi_F)_\# \mu_2. \tag{6.47}$$

Lemma 6.3. *Let E be a separable Banach space and let μ_1, μ_2 be probability measures on $(E, \mathcal{B}(E))$. Let $F \subset E^*$ be a finite-dimensional subspace and let*

$$\pi_F : E \to \mathbb{R}^n, \qquad \pi_F(x) = (\ell_1(x), \dots, \ell_n(x)), \tag{6.48}$$

where $\{\ell_1, \dots, \ell_n\}$ is a basis of F. Denote by

$$(\mu_i)_F := (\pi_F)_\# \mu_i, \quad i = 1, 2, \tag{6.49}$$

the corresponding pushforward measures. If $(\mu_1)_F$ and $(\mu_2)_F$ are mutually singular, then μ_1 and μ_2 are mutually singular on E.

Proof. By definition of mutual singularity of measures on $\mathbb{R}^n$, there exists a Borel set $A \subset \mathbb{R}^n$ such that

$$(\mu_1)_F(A) = 0, \qquad (\mu_2)_F(A) = 1. \tag{6.50}$$

Since π_F is continuous and hence Borel measurable, the preimage

$$B := \pi_F^{-1}(A) \tag{6.51}$$

belongs to $\mathcal{B}(E)$. We now compute the measures of B under μ_1 and μ_2. By the defining property of the pushforward measure,

$$\mu_i(B) = \mu_i(\pi_F^{-1}(A)) = (\pi_F)_\# \mu_i(A) = (\mu_i)_F(A), \qquad i = 1, 2. \tag{6.52}$$

Using (6.50), we obtain

$$\mu_1(B) = 0, \qquad \mu_2(B) = 1. \tag{6.53}$$

Thus there exists a measurable set $B \subset E$ which has full measure under μ_2 and zero measure under μ_1. By definition, this implies

$$\mu_1 \perp \mu_2. \tag{6.54}$$

The argument uses only the functoriality of pushforward measures and does not rely on any special structure of Gaussian measures; in particular, singularity at the level of a single finite-dimensional cylindrical projection propagates to singularity of the original measures on the infinite-dimensional space. □

Similarly, we have:

Lemma 6.4 (Projection equivalence implies global dichotomy)**.** *Let E be a separable Banach space and let μ_1, μ_2 be probability measures on $(E, \mathcal{B}(E))$. For every finite-dimensional subspace $F \subset E^*$, let*

$$(\mu_i)_F := (\pi_F)_{\#}\mu_i, \qquad i = 1, 2, \tag{6.55}$$

where $\pi_F : E \to \mathbb{R}^{\dim F}$ denotes the canonical cylindrical projection. If $(\mu_1)_F \sim (\mu_2)_F$ for every finite-dimensional F, then

$$\mu_1 \sim \mu_2 \quad \text{or} \quad \mu_1 \perp \mu_2. \tag{6.56}$$

Proof. **Step 1: Cylindrical σ-algebra:** Let $\mathcal{C}$ denote the cylindrical σ-algebra on E, defined by

$$\mathcal{C} = \sigma\Big(\pi_F^{-1}(A) \;\Big|\; F \subset E^* \text{ finite-dimensional}, \; A \in \mathcal{B}(\mathbb{R}^{\dim F})\Big). \tag{6.57}$$

Since E is separable, $\mathcal{C} = \mathcal{B}(E)$.

Step 2: Agreement of null sets on cylinders: Fix a finite-dimensional $F \subset E^*$. By assumption, $(\mu_1)_F \sim (\mu_2)_F$. Hence for every $A \in \mathcal{B}(\mathbb{R}^{\dim F})$,

$$(\mu_1)_F(A) = 0 \quad \Longleftrightarrow \quad (\mu_2)_F(A) = 0. \tag{6.58}$$

Equivalently, for every cylindrical set $B = \pi_F^{-1}(A)$,

$$\mu_1(B) = 0 \quad \Longleftrightarrow \quad \mu_2(B) = 0. \tag{6.59}$$

Step 3: Radon–Nikodým derivatives on finite-dimensional σ-algebras: Let $\mathcal{C}_F := \pi_F^{-1}(\mathcal{B}(\mathbb{R}^{\dim F}))$. On $(E, \mathcal{C}_F)$, the measures μ_1 and μ_2 are equivalent, hence there exists a strictly positive Radon–Nikodým derivative

$$Z_F := \frac{d\mu_1}{d\mu_2}\Big|_{\mathcal{C}_F}, \qquad Z_F > 0 \quad \mu_2\text{-a.e.} \tag{6.60}$$

If $F_1 \subset F_2$, then uniqueness of Radon–Nikodým derivatives yields the consistency relation

$$Z_{F_1} = \mathbb{E}_{\mu_2}[Z_{F_2} \mid \mathcal{C}_{F_1}]. \tag{6.61}$$

Step 4: Martingale convergence: Let $\{F_n\}_{n\geq 1}$ be an increasing sequence of finite-dimensional subspaces of E^* such that

$$\bigvee_{n=1}^{\infty} \mathcal{C}_{F_n} = \mathcal{B}(E). \tag{6.62}$$

Then $\{Z_{F_n}\}_{n\geq 1}$ is a nonnegative martingale with respect to $(\mathcal{C}_{F_n}, \mu_2)$. By the martingale convergence theorem,

$$Z := \lim_{n\to\infty} Z_{F_n} \quad \text{exists } \mu_2\text{-a.e.} \tag{6.63}$$

Step 5: Dichotomy: Exactly one of the following holds:

- If $\mathbb{E}_{\mu_2}[Z] = 1$, then $Z > 0$ μ_2-a.e. and

$$\frac{d\mu_1}{d\mu_2} = Z, \tag{6.64}$$

so $\mu_1 \ll \mu_2$. Reversing the roles of μ_1 and μ_2 yields $\mu_2 \ll \mu_1$, hence $\mu_1 \sim \mu_2$.

- If $Z = 0$ μ_2-a.e., then for every $A \in \mathcal{B}(E)$,

$$\mu_1(A) = \int_A Z \, d\mu_2 = 0, \tag{6.65}$$

so $\mu_1 \perp \mu_2$.

No intermediate case is possible, since any partial absolute continuity would contradict equivalence on all finite-dimensional cylindrical projections. Therefore,

$$\mu_1 \sim \mu_2 \quad \text{or} \quad \mu_1 \perp \mu_2, \tag{6.66}$$

which completes the proof. □

Cylindrical projections provide the bridge between finite-dimensional Gaussian classification and the infinite-dimensional setting. Since Gaussian measures on E are uniquely determined by their finite-dimensional marginals, any absolute continuity or singularity phenomenon must already be visible at the level of sufficiently rich projections.

More precisely, for a finite-dimensional subspace $F = \operatorname{span}\{\ell_1, \ldots, \ell_n\} \subset E^*$, the projection

$$\pi_F(x) := (\ell_1(x), \ldots, \ell_n(x)) \in \mathbb{R}^n \tag{6.67}$$

identifies $(\mu_i)_F$ with a centered Gaussian measure on $\mathbb{R}^n$ whose covariance matrix is given by

$$\Sigma_i^{(F)} = \big(Q_i(\ell_j, \ell_k)\big)_{1 \le j,k \le n}. \tag{6.68}$$

Thus the finite-dimensional lemma applies verbatim to each $(\mu_i)_F$.

Lemma 6.3 shows that singularity is *detected* on some finite-dimensional projection: if two infinite-dimensional Gaussian measures disagree strongly enough, there exists a finite collection of linear observables along which this disagreement already becomes total. In this sense, singularity is a local phenomenon in the projective system of marginals.

The converse direction is more subtle. Even if all finite-dimensional projections are equivalent, it is not *a priori* clear that the full measures are equivalent, because Radon–Nikodym derivatives may fail to be integrable in the projective limit. This is precisely where the Hilbert–Schmidt condition in the Feldman–Hájek theorem enters.

To make this precise, observe that for each finite-dimensional F one may write

$$\frac{d(\mu_2)_F}{d(\mu_1)_F}(y) = \exp\Big(-\tfrac{1}{2}\langle T_F y, y\rangle_{\mathbb{R}^n}\Big) \cdot Z_F, \tag{6.69}$$

where T_F is a symmetric matrix determined by $\Sigma_1^{(F)}$ and $\Sigma_2^{(F)}$, and Z_F is a normalization constant. Consistency of these densities as F increases imposes strong summability constraints on the eigenvalues of the associated relative covariance operators.

Lemma 6.4 should therefore be interpreted as a dichotomy principle: finite-dimensional equivalence rules out partial absolute continuity. Either the projective system of likelihood ratios converges in $L^1(\mu_1)$, yielding equivalence of μ_1 and μ_2, or it diverges, forcing mutual singularity. No intermediate behavior is possible.

This projective viewpoint is essential for the proof of the Feldman–Hájek theorem: it reduces the infinite-dimensional problem to controlling the behavior of a coherent family of finite-dimensional Gaussian likelihood ratios, whose asymptotics are governed by operator-theoretic properties of the covariance forms.

6.7 Comparison of Cameron–Martin Spaces

We now establish the first necessary and sufficient condition.

Lemma 6.5. *If μ_1 and μ_2 are equivalent, then $H_1 = H_2$ and the norms $\|\cdot\|_{H_1}$ and $\|\cdot\|_{H_2}$ are equivalent.*

Proof. Suppose $\mu_1 \ll \mu_2$. For $h \in H_1$, the Cameron–Martin theorem for μ_1 implies that the shift by h preserves equivalence of μ_1. Equivalence transfers this property to μ_2, hence $h \in H_2$. The same argument in the opposite direction gives $H_1 = H_2$. Equivalence of norms follows from the Radon–Nikodým derivatives. □

Beyond the set-theoretic equality of Cameron–Martin spaces, one can make the relationship between the norms explicit. Let $h \in H_1 = H_2$, and consider the Radon–Nikodým derivative given by the Cameron–Martin theorem:

$$\frac{d(\mu_1)_h}{d\mu_1}(x) = \exp\Big(\langle h, x\rangle_{H_1} - \tfrac{1}{2}\|h\|_{H_1}^2\Big), \qquad \mu_1\text{-a.e. } x \in E. \tag{6.70}$$

Since μ_1 and μ_2 are equivalent, the measure $(\mu_1)_h$ is also equivalent to μ_2. By applying the Cameron–Martin theorem to μ_2, we have

$$\frac{d(\mu_2)_h}{d\mu_2}(x) = \exp\Big(\langle h, x\rangle_{H_2} - \tfrac{1}{2}\|h\|_{H_2}^2\Big), \qquad \mu_2\text{-a.e. } x \in E. \tag{6.71}$$

Taking logarithms and comparing the two expressions shows that for some constants $0 < c < C < \infty$ independent of h,

$$c\,\|h\|_{H_1}^2 \le \|h\|_{H_2}^2 \le C\,\|h\|_{H_1}^2, \tag{6.72}$$

which establishes the equivalence of norms. This inequality is essential for extending finite-dimensional covariance comparisons to the full Hilbert space structure of the Cameron–Martin spaces. In particular, it guarantees that Hilbert–Schmidt conditions on relative covariance operators are meaningful and well-defined in the infinite-dimensional setting.

6.8 Hilbert Space Reduction

Let H denote the common Cameron–Martin space under assumption (a). Let $Q_i : H \to H$ be the covariance operators extended by continuity.

Lemma 6.6. *If μ_1 and μ_2 are equivalent, then*

$$T = Q_1^{-1/2} Q_2 Q_1^{-1/2} - I \tag{6.73}$$

is Hilbert–Schmidt on H.

Proof. We proceed in several logically independent steps.

Step 1: Reduction to a common Hilbert space: Since $\mu_1 \sim \mu_2$, the Cameron–Martin spaces coincide:

$$H_1 = H_2 =: H, \tag{6.74}$$

with equivalence of norms. Identifying H with the closure of $\mathrm{Ran}(Q_1^{1/2})$ equipped with the inner product

$$\langle h_1, h_2 \rangle_H := \langle Q_1^{-1/2} h_1, Q_1^{-1/2} h_2 \rangle_{E^*,E}, \tag{6.75}$$

the operator

$$A := Q_1^{-1/2} Q_2 Q_1^{-1/2} \tag{6.76}$$

is well-defined, self-adjoint, positive, and bounded on H. Hence $T = A - I$ is also self-adjoint on H.

Step 2: Choice of orthonormal basis and finite-dimensional projections: Let $\{e_k\}_{k\geq 1}$ be an orthonormal basis of H. For each $n \in \mathbb{N}$, define

$$H_n := \mathrm{span}\{e_1, \ldots, e_n\}, \qquad P_n : H \to H_n \tag{6.77}$$

to be the orthogonal projection. Let $(\mu_i)_n := \mu_i \circ P_n^{-1}$ be the pushforward of μ_i onto H_n. Then $(\mu_i)_n$ are centered Gaussian measures on $\mathbb{R}^n$ with covariance operators

$$Q_{1,n} = I_n, \qquad Q_{2,n} = P_n A P_n. \tag{6.78}$$

Step 3: Finite-dimensional Feldman–Hájek criterion: Since $\mu_1 \sim \mu_2$, all finite-dimensional projections satisfy

$$(\mu_1)_n \sim (\mu_2)_n \quad \text{for every } n. \tag{6.79}$$

In finite dimensions, equivalence of centered Gaussian measures with covariance matrices I_n and $Q_{2,n}$ holds if and only if

$$\mathrm{Tr}\left((Q_{2,n} - I_n)^2\right) < \infty, \tag{6.80}$$

which is automatic finiteness-wise but, crucially, must be uniformly controlled as $n \to \infty$. Indeed, the Radon–Nikodým derivative is given by

$$\frac{d(\mu_2)_n}{d(\mu_1)_n}(x) = \det(Q_{2,n})^{-1/2} \exp\left(-\frac{1}{2}\langle (Q_{2,n}^{-1} - I_n)x, x\rangle\right), \tag{6.81}$$

and uniform integrability of these densities forces boundedness of $\mathrm{Tr}((Q_{2,n} - I_n)^2)$ as $n \to \infty$.

Step 4: Identification of the trace with Hilbert–Schmidt norms: Since $Q_{2,n} - I_n = P_n T P_n$, we compute

$$\operatorname{Tr}\left((Q_{2,n} - I_n)^2\right) = \operatorname{Tr}\left((P_n T P_n)^2\right) \tag{6.82}$$

$$= \sum_{k=1}^{n} \|P_n T e_k\|_H^2 \tag{6.83}$$

$$= \sum_{k=1}^{n} \|T e_k\|_H^2, \tag{6.84}$$

where the last equality holds since $Te_k \in H_n$ for $k \le n$.

Step 5: Passage to the limit: Since equivalence holds for all n, the quantities $\sum_{k=1}^{n} \|Te_k\|_H^2$ must remain bounded as $n \to \infty$. Hence the monotone limit exists and is finite:

$$\sum_{k=1}^{\infty} \|T e_k\|_H^2 < \infty. \tag{6.85}$$

By definition, this exactly means that T is a Hilbert–Schmidt operator on H.

Conclusion: Thus, equivalence of μ_1 and μ_2 forces

$$Q_1^{-1/2} Q_2 Q_1^{-1/2} - I \in \mathcal{L}_2(H), \tag{6.86}$$

i.e. the difference of the normalized covariance operators is Hilbert–Schmidt. □

To see this more rigorously, consider the finite-dimensional projections

$$P_n : H \to H, \qquad P_n x = \sum_{k=1}^{n} \langle x, e_k \rangle_H e_k, \tag{6.87}$$

and define the projected covariance operators

$$Q_i^{(n)} = P_n Q_i P_n, \qquad i = 1, 2. \tag{6.88}$$

Let $\mu_i^{(n)} = (P_n)_{\#}\mu_i$ be the induced Gaussian measures on $\mathbb{R}^n \simeq \operatorname{span}\{e_1, \dots, e_n\}$. In matrix form, the covariances are

$$\Sigma_i^{(n)} = \left(\langle e_j, Q_i e_k \rangle_H\right)_{j,k=1}^{n}. \tag{6.89}$$

Finite-dimensional equivalence implies that

$$\Sigma_1^{(n)\,-1/2} \Sigma_2^{(n)} \Sigma_1^{(n)\,-1/2} - I_n \tag{6.90}$$

has finite Frobenius norm for every n, i.e.,

$$\|\Sigma_1^{(n)\,-1/2} \Sigma_2^{(n)} \Sigma_1^{(n)\,-1/2} - I_n\|_F^2 = \sum_{k=1}^{n} \|T e_k\|_H^2 < \infty. \tag{6.91}$$

Passing to the limit $n \to \infty$ shows that

$$\sum_{k=1}^{\infty} \|Te_k\|_H^2 < \infty, \tag{6.92}$$

which is exactly the Hilbert–Schmidt condition

$$\|T\|_{\mathrm{HS}}^2 := \sum_{k=1}^{\infty} \|Te_k\|_H^2 < \infty. \tag{6.93}$$

Hence the infinite-dimensional equivalence of μ_1 and μ_2 is encoded in the Hilbert–Schmidt perturbation of their covariance operators in the common Cameron–Martin space.

6.9 Necessity and Sufficiency

We may now complete the proof.

Proof of Theorem 6.1. **Necessity:** Assume $\mu_1 \sim \mu_2$. Then:

(a) follows from Cameron–Martin quasi-invariance.

(b) follows from the Hilbert–Schmidt condition above.

(c) is trivial in the centered case.

Sufficiency: Assume (a), (b), (c). For each finite-dimensional projection F_n, the projected covariances satisfy the finite-dimensional Feldman–Hájek condition. Thus $(\mu_1)_{F_n}$ and $(\mu_2)_{F_n}$ are equivalent for all n.

Consistency of the Radon–Nikodým derivatives and Kolmogorov extension arguments imply equivalence of μ_1 and μ_2 on E.

If either (a) or (b) fails, we construct a sequence of projections for which the projected measures become singular, giving singularity of μ_1 and μ_2 on E. □

To make the sufficiency argument more explicit, let $\{e_k\}_{k\geq 1}$ be an orthonormal basis of the common Cameron–Martin space H. For each n, define the finite-dimensional subspace

$$F_n := \mathrm{span}\{e_1, \dots, e_n\} \subset H \subset E, \tag{6.94}$$

and consider the corresponding orthogonal projections

$$P_n : E \to F_n. \tag{6.95}$$

Let $\mu_i^{(n)} := (P_n)_{\#}\mu_i$ be the projected Gaussian measures on $F_n \simeq \mathbb{R}^n$ with covariance matrices

$$\Sigma_i^{(n)} = \left(\langle e_j, Q_i e_k \rangle_H\right)_{j,k=1}^n, \qquad i = 1, 2. \tag{6.96}$$

By condition (b), the operator

$$T = Q_1^{-1/2} Q_2 Q_1^{-1/2} - I \tag{6.97}$$

is Hilbert–Schmidt on H, so its restriction to F_n has finite Frobenius norm:

$$\|T|_{F_n}\|_F^2 = \sum_{k=1}^{n} \|Te_k\|_H^2 < \infty. \tag{6.98}$$

Hence, by the finite-dimensional Feldman–Hájek lemma, $\mu_1^{(n)} \sim \mu_2^{(n)}$ for each n. Denote the Radon–Nikodým derivative on F_n by

$$\frac{d\mu_2^{(n)}}{d\mu_1^{(n)}}(x) = \exp\Big(-\frac{1}{2}\langle x, (Q_2^{(n)})^{-1} - (Q_1^{(n)})^{-1}x\rangle + \text{const}\Big). \tag{6.99}$$

The consistency of the family $\{\mu_i^{(n)}\}$ under the projections P_n ensures that these finite-dimensional derivatives define a projective system, and the Kolmogorov extension theorem guarantees the existence of a Radon-Nikodým derivative on E:

$$\frac{d\mu_2}{d\mu_1}(x) = \lim_{n\to\infty} \frac{d\mu_2^{(n)}}{d\mu_1^{(n)}}(P_n x), \qquad \mu_1\text{-a.e. } x \in E. \tag{6.100}$$

Therefore, $\mu_1 \sim \mu_2$. Conversely, if either (a) or (b) fails, there exists a sequence of finite-dimensional projections $\{F_n\}$ such that

$$(\mu_1)_{F_n} \perp (\mu_2)_{F_n} \quad \text{for some } n, \tag{6.101}$$

which implies $\mu_1 \perp \mu_2$ on E by Lemma 6.3.

6.10 Consequences

6.10.1 Dichotomy

The Feldman–Hájek theorem implies a strict dichotomy: for Gaussian measures on separable Banach spaces, no intermediate relations such as "mutually absolutely continuous on some non-trivial subspace" exist.

Radon–Nikodým Derivative Representation: When two Gaussian measures μ_1 and μ_2 are equivalent, the Feldman–Hájek theorem provides a concrete representation of the Radon–Nikodým derivative. Let H denote the common Cameron–Martin space and let $T = Q_1^{-1/2}Q_2Q_1^{-1/2} - I$ be the Hilbert–Schmidt operator. Then for $x \in E$,

$$\frac{d\mu_2}{d\mu_1}(x) = \exp\left(-\frac{1}{2}\sum_{k=1}^{\infty} \langle e_k, T(I+T)^{-1}e_k\rangle_H + \sum_{k=1}^{\infty} \langle x, (I+T)^{-1}Te_k\rangle_H \right), \tag{6.102}$$

where $\{e_k\}$ is an orthonormal basis of H. The series converge absolutely due to the Hilbert–Schmidt property of T. This explicit formula highlights how the Hilbert–Schmidt perturbation quantifies the "distance" between the two measures in the infinite-dimensional setting.

Implications for Quasi-Invariance: An immediate consequence is the quasi-invariance of

Gaussian measures under Cameron–Martin shifts. Let $h \in H$, and define the shifted measure

$$\mu_h(A) := \mu(A - h), \qquad A \subset E \text{ Borel.} \tag{6.103}$$

The Feldman–Hájek theorem implies $\mu_h \sim \mu$, and the Radon–Nikodým derivative reduces to the classical Cameron–Martin formula:

$$\frac{d\mu_h}{d\mu}(x) = \exp\Big(\langle x, h\rangle_H - \frac{1}{2}\|h\|_H^2\Big), \qquad \mu\text{-a.e. } x \in E. \tag{6.104}$$

Thus the Cameron–Martin space H captures all directions along which the measure can be shifted while preserving equivalence, emphasizing its central geometric role.

Singularity Criteria: Conversely, if either the Cameron–Martin spaces H_1 and H_2 differ as sets, or if the Hilbert–Schmidt condition fails, then the measures are mutually singular. This provides a practical criterion for establishing singularity in applications such as:

- Gaussian processes with different smoothness classes,
- Changes in covariance structure in stochastic PDEs,
- Infinite-dimensional statistical models with perturbed priors.

In particular, any perturbation outside the Cameron–Martin space or with infinite Hilbert–Schmidt norm immediately yields singularity:

$$\|Q_1^{-1/2} Q_2 Q_1^{-1/2} - I\|_{\mathrm{HS}} = \infty \quad \Longrightarrow \quad \mu_1 \perp \mu_2. \tag{6.105}$$

6.10.2 Gaussian Geometry

The theorem provides a metric geometry on Gaussian measures: two Gaussians are close (in the sense of equivalence) if and only if their covariance operators differ by a Hilbert–Schmidt perturbation.

The Feldman–Hájek theorem endows the space of Gaussian measures with a natural *intrinsic geometry*. In particular, the Hilbert–Schmidt norm of the operator

$$T := Q_1^{-1/2} Q_2 Q_1^{-1/2} - I \tag{6.106}$$

quantifies the "distance" between two equivalent measures: small Hilbert–Schmidt norm corresponds to a small perturbation of μ_1 resulting in μ_2, while divergence of this norm implies singularity.

We can formalize this idea by defining a pseudo-metric on the set of centered Gaussian measures with common Cameron–Martin space H:

$$d_{\mathrm{HS}}(\mu_1, \mu_2) := \|Q_1^{-1/2} Q_2 Q_1^{-1/2} - I\|_{\mathrm{HS}}. \tag{6.107}$$

Then

$$d_{\mathrm{HS}}(\mu_1, \mu_2) < \infty \quad \Longleftrightarrow \quad \mu_1 \sim \mu_2, \tag{6.108}$$

and

$$d_{\mathrm{HS}}(\mu_1, \mu_2) = \infty \quad \Longleftrightarrow \quad \mu_1 \perp \mu_2. \tag{6.109}$$

This metric viewpoint aligns naturally with infinite-dimensional information geometry: the Hilbert–Schmidt "distance" acts as a quadratic approximation to the relative entropy (Kullback–Leibler divergence) between measures. Indeed, for equivalent Gaussian measures $\mu_1 \sim \mu_2$, the relative entropy admits the representation

$$\mathrm{KL}(\mu_2 \,\|\, \mu_1) = \frac{1}{2} \sum_{k=1}^{\infty} \Big(\lambda_k - \log(1 + \lambda_k)\Big), \tag{6.110}$$

where $\{\lambda_k\}$ are the eigenvalues of T on H. This connects the Feldman–Hájek criterion directly to notions of Gaussian measure concentration and perturbation energy.

Finally, the Hilbert–Schmidt geometry explains why small deviations in directions outside the Cameron–Martin space cannot affect equivalence: such directions carry zero "metric weight" under the Gaussian measure, reflecting the anisotropic geometry of infinite-dimensional probability spaces.

6.10.3 Applications to SDEs

In stochastic analysis, many equivalence-of-measures results reduce to Hilbert–Schmidt perturbations of covariance operators, making the Feldman–Hájek theorem indispensable.

The Feldman–Hájek theorem has immediate implications for the study of stochastic differential equations (SDEs) in infinite dimensions. Consider a Hilbert-space-valued SDE of the form

$$dX_t = AX_t\,dt + B\,dW_t, \qquad X_0 = x_0 \in H, \tag{6.111}$$

where A is a (possibly unbounded) linear operator, B is a bounded operator into H, and W_t is a cylindrical Wiener process on H. The solution X defines a Gaussian measure μ on the path space $C([0,T]; H)$ with covariance operator determined by B and the semigroup generated by A:

$$Q = \int_0^T e^{tA} BB^* e^{tA^*}\,dt. \tag{6.112}$$

Suppose we perturb the noise operator B by ΔB such that the new covariance $Q' = \int_0^T e^{tA}(B + \Delta B)(B + \Delta B)^* e^{tA^*}\,dt$ defines another Gaussian measure μ'. Then the Feldman–Hájek criterion states:

$$\mu \sim \mu' \quad \Longleftrightarrow \quad Q^{-1/2} Q' Q^{-1/2} - I \text{ is Hilbert–Schmidt on the Cameron–Martin space of } \mu. \tag{6.113}$$

In practice, this provides a rigorous foundation for Girsanov transformations: shifts of the drift in an SDE correspond to translations along the Cameron–Martin space, while perturbations of the noise operator correspond to Hilbert–Schmidt perturbations of the covariance. Consequently, the Feldman–Hájek theorem gives precise necessary and sufficient conditions for absolute continuity of laws of solutions under such modifications.

Moreover, in stochastic control and filtering problems, understanding when two path-space measures are equivalent allows one to rigorously define likelihood ratios, compute Radon–Nikodým derivatives, and formulate optimal change-of-measure strategies:

$$\frac{d\mu'}{d\mu}(X) = \exp\Big(\langle X, h\rangle_{H_\mu} - \tfrac{1}{2}\|h\|^2_{H_\mu}\Big), \quad h \in H_\mu, \tag{6.114}$$

linking the abstract Cameron–Martin geometry directly to computational and analytical tools in infinite-dimensional SDEs.

6.11 Summary

In this chapter we proved:

1. Gaussian measures on infinite-dimensional Banach spaces are either equivalent or singular.
2. Equivalence holds if and only if their Cameron–Martin spaces coincide and their covariance operators differ by a Hilbert–Schmidt operator.
3. Finite-dimensional projections characterize the global behavior of Gaussian measures.

This completes the operator-theoretic classification of Gaussian measures. Additional consequences and insights include:

- **Cameron–Martin Geometry:** The Feldman–Hájek theorem shows that the Cameron–Martin space H_μ encodes all admissible directions for quasi-invariance and measure shifts. Any vector outside H_μ leads to singular translations.
- **Hilbert–Schmidt Perturbations as a Metric:** The difference of covariance operators measured in the Hilbert–Schmidt norm serves as a precise metric for equivalence of Gaussian measures:

$$\|Q_1^{-1/2} Q_2 Q_1^{-1/2} - I\|_{\mathrm{HS}} < \infty \quad \iff \quad \mu_1 \sim \mu_2. \tag{6.115}$$

- **Applications to Stochastic Analysis:** The theorem provides a foundational framework for Girsanov-type transformations, stochastic control, and infinite-dimensional SDEs. Specifically, for a shift $h \in H_\mu$, the Radon–Nikodým derivative is given by

$$\frac{d\mu_h}{d\mu}(x) = \exp\Big(\langle x, h\rangle_{H_\mu} - \frac{1}{2}\|h\|^2_{H_\mu}\Big), \qquad \mu\text{-a.e. } x \in E. \tag{6.116}$$

- **Dichotomy in Infinite Dimensions:** No intermediate relation exists between equivalence and singularity; any Gaussian measures on a separable Banach space are either fully comparable or entirely disjoint in measure-theoretic terms. This underlines the rigidity and geometric structure of Gaussian measures in infinite-dimensional spaces.

Together, these results establish a complete operator-theoretic and geometric characterization of Gaussian measures, setting the stage for applications in Malliavin calculus, path-space analysis, and large deviation theory.

6.12 References

Feldman (1958) [48], Hájek (1962) [49], Bogachev (1998) [1], Hitsuda (2020) [50], Shepp (1966) [51], Le Cam (2012) [52], Ibragimov and Rozanov (2012) [53], Kakutani (1948) [54], Da Prato & Zabczyk (2014) [4], Albeverio et. al. (2009) [55].

Bibliography

[1] Bogachev, V. I. (1998). Gaussian measures (No. 62). American Mathematical Soc..

[2] Vakhania, N., Tarieladze, V., & Chobanyan, S. (2012). Probability distributions on Banach spaces (Vol. 14). Springer Science & Business Media.

[3] Kuo, H. H. (2006). Gaussian measures in Banach spaces. In Gaussian measures in banach spaces (pp. 1-109). Berlin, Heidelberg: Springer Berlin Heidelberg.

[4] Da Prato, G., & Zabczyk, J. (2014). Stochastic equations in infinite dimensions (Vol. 152). Cambridge university press.

[5] Fernique, X. (2006). Regularité des trajectoires des fonctions aléatoires gaussiennes. In Ecole d'Eté de Probabilités de Saint-Flour IV—1974 (pp. 1-96). Berlin, Heidelberg: Springer Berlin Heidelberg.

[6] Cameron, R. H., & Martin, W. T. (1945). Transformations of Wiener integrals under a general class of linear transformations. Transactions of the American Mathematical Society, 58(2), 184-219.

[7] Girsanov, I. V. (1960). On transforming a certain class of stochastic processes by absolutely continuous substitution of measures. Theory of Probability & Its Applications, 5(3), 285-301.

[8] Stroock, D. W. (2010). Probability theory: an analytic view. Cambridge university press.

[9] Albeverio, S., Kondratiev, Y. G., & Röckner, M. (1998). Analysis and geometry on configuration spaces. Journal of functional analysis, 154(2), 444-500.

[10] Minlos, R. A. F. (1959). Generalized random processes and their extension in measure. Trudy Moskovskogo Matematicheskogo Obshchestva, 8, 497-518.

[11] Sourangshu Ghosh. Mathematical Foundations of Deep Learning. 2025. ⟨hal-04928560v5⟩

[12] Deimling, K. (2013). Nonlinear functional analysis. Springer Science & Business Media.

[13] Lang, S. (2012). Differential and Riemannian manifolds (Vol. 160). Springer Science & Business Media.

[14] Cartan, H. (2012). Differential forms. Courier Corporation.

[15] Zeidler, E. (2012). Applied functional analysis: applications to mathematical physics (Vol. 108). Springer Science & Business Media.

[16] Zeidler, E. (2012). Applied functional analysis: main principles and their applications (Vol. 109). Springer Science & Business Media.

[17] Bourbaki, N. (2013). Topological vector spaces: Chapters 1–5. Springer Science & Business Media.

[18] Abraham, R., Marsden, J. E., & Ratiu, T. (2012). Manifolds, tensor analysis, and applications (Vol. 75). Springer Science & Business Media.

[19] Hamilton, R. S. (1982). The inverse function theorem of Nash and Moser. Bulleting of the American Mathematical Society, Volume 7, Number 1, July 1982

[20] Krantz, S. G., & Parks, H. R. (2002). The implicit function theorem: history, theory, and applications. Springer Science & Business Media.

[21] Michal, A. D. (1938). Differential calculus in linear topological spaces. Proceedings of the National Academy of Sciences, 24(8), 340-342.

[22] Dieudonné, J. (2011). Foundations of modern analysis. Read Books Ltd.

[23] Kuelbs, J. (1969). Abstract Wiener spaces and applications to analysis. Pacific Journal of Mathematics, 31(2), 433-450.

[24] Hida, T., Kuo, H. H., Potthoff, J., & Streit, L. (2013). White noise: an infinite dimensional calculus (Vol. 253). Springer Science & Business Media.

[25] Ustunel, A. S., & Zakai, M. (2013). Transformation of measure on Wiener space. Springer Science & Business Media.

[26] Sourangshu Ghosh. The Feynman–Kac Formula: A Measure-Theoretic, Analytic, and Probabilistic Synthesis. 2025. ⟨hal-05378000⟩

[27] Shigekawa, I. (2004). Stochastic analysis (Vol. 224). American Mathematical Soc..

[28] Ikeda, N., & Watanabe, S. (2014). Stochastic differential equations and diffusion processes (Vol. 24). Elsevier.

[29] Malliavin, P. (2015). Stochastic analysis (Vol. 313). Springer.

[30] Driver, B. K. (2003). Analysis tools with applications. Lecture notes. https://mathweb.ucsd.edu/~bdriver/240-01-02/Lecture_Notes/anal.pdf

[31] Janson, S. (1997). Gaussian hilbert spaces (No. 129). Cambridge university press.

[32] Fernique, X. (1970). Intégrabilité des vecteurs gaussiens. CR Acad. Sci. Paris Sé r. AB, 270, A1698-A1699.

[33] Bakry, D., Gentil, I., & Ledoux, M. (2013). Analysis and geometry of Markov diffusion operators (Vol. 348). Springer Science & Business Media.

[34] Ledoux, M. (2001). The concentration of measure phenomenon (No. 89). American Mathematical Soc..

[35] Borell, C. (1976). Gaussian Radon measures on locally convex spaces. Mathematica Scandinavica, 38(2), 265-284.

[36] Talagrand, M. (2014). Upper and lower bounds for stochastic processes.

[37] Gross, L. (1975). Logarithmic sobolev inequalities. American Journal of Mathematics, 97(4), 1061-1083.

[38] Milman, V. D., & Schechtman, G. (1986). Asymptotic theory of finite dimensional normed spaces. Berlin, Heidelberg: Springer Berlin Heidelberg.

[39] Pisier, G. (1999). The volume of convex bodies and Banach space geometry (Vol. 94). Cambridge University Press.

[40] Lifshits, M. A. (2013). Gaussian random functions (Vol. 322). Springer Science & Business Media.

[41] Sudakov, V. N., & Tsirel'son, B. S. (1978). Extremal properties of half-spaces for spherically invariant measures. Journal of Soviet Mathematics, 9(1), 9-18.

[42] Parthasarathy, K. R. (2005). Probability measures on metric spaces (Vol. 352). American Mathematical Soc..

[43] Adler, R. J., & Taylor, J. E. (2007). Random fields and geometry. New York, NY: Springer New York.

[44] Berlinet, A., & Thomas-Agnan, C. (2011). Reproducing kernel Hilbert spaces in probability and statistics. Springer Science & Business Media.

[45] Lifshits, M. (2012). Lectures on Gaussian processes. In Lectures on Gaussian Processes (pp. 1-117). Berlin, Heidelberg: Springer Berlin Heidelberg.

[46] Prato, G. D. (2007). Infinite dimensional Kolmogorov equations. In Kolmogorov's Heritage in Mathematics (pp. 67-96). Berlin, Heidelberg: Springer Berlin Heidelberg.

[47] Kallianpur, G. (2013). Stochastic filtering theory (Vol. 13). Springer Science & Business Media.

[48] Feldman, J. (1958). Equivalence and perpendicularity of Gaussian processes. Pacific J. Math, 8(4), 699-708.

[49] Hájek, J. (1962). On linear statistical problems in stochastic processes. Czechoslovak Mathematical Journal, 12(3), 404-444.

[50] Hitsuda, M. (2020). Wiener-like integrals for Gaussian processes and the linear estimation problem. In Stochastic Analysis and Applications (pp. 167-177). CRC Press.

[51] Shepp, L. A. (1966). Radon-Nikodym derivatives of Gaussian measures. The Annals of Mathematical Statistics, 321-354.

[52] Le Cam, L. (2012). Asymptotic methods in statistical decision theory. Springer Science & Business Media.

[53] Ibragimov, I. A., & Rozanov, Y. A. E. (2012). Gaussian random processes (Vol. 9). Springer Science & Business Media.

[54] Kakutani, S. (1948). On equivalence of infinite product measures. Annals of Mathematics, 49(1), 214-224.

[55] Albeverio, S., Fenstad, J. E., Høegh-Krohn, R., & Lindstrøm, T. (2009). Nonstandard methods in stochastic analysis and mathematical physics. Courier Dover Publications.