

GAUSSIAN MEASURES IN TRADITIONAL AND NOT SO TRADITIONAL SETTINGS

DANIEL W. STROOCK

ABSTRACT. This article is intended to provide non-specialists with an introduction to integration theory on pathspace.

§0: INTRODUCTION

Let \mathbb{Q}_1 denote the set of rational numbers $q \in [0, 1]$, and, for $a \in \mathbb{Q}_1$, define the translation map τ_a on \mathbb{Q}_1 by addition modulo 1:

$$\tau_a b = \begin{cases} a + b & \text{if } a + b \leq 1 \\ a + b - 1 & \text{if } a + b > 1 \end{cases} \quad \text{for } b \in \mathbb{Q}_1.$$

Is there a probability measure $\lambda_{\mathbb{Q}_1}$ on \mathbb{Q}_1 which is translation invariant in the sense that

$$\lambda_{\mathbb{Q}_1}(\Gamma) = \lambda_{\mathbb{Q}_1}(\tau_a^{-1}\Gamma)$$

for every $a \in \mathbb{Q}_1$?

To answer this question, it is necessary to be precise about what it is that one is seeking. Indeed, the answer will depend on the level of ambition at which the question is asked. For example, if \mathcal{A} denotes the *algebra* of subsets $\Gamma \subseteq \mathbb{Q}_1$ generated by intervals

$$[a, b]_{\mathbb{Q}_1} \equiv \{q \in \mathbb{Q}_1 : a \leq q \leq b\}$$

for $a \leq b$ from \mathbb{Q}_1 , then it is easy to check that there is a unique, *finitely additive* way to define $\lambda_{\mathbb{Q}_1}$ on \mathcal{A} . In fact,

$$\lambda_{\mathbb{Q}_1}([a, b]_{\mathbb{Q}_1}) = b - a, \quad \text{for } a \leq b \text{ from } \mathbb{Q}_1.$$

On the other hand, if one is more ambitious and asks for a countably additive $\lambda_{\mathbb{Q}_1}$ on the σ -algebra generated by \mathcal{A} (equivalently, all subsets of \mathbb{Q}_1), then one is asking for too much. In fact, since $\lambda_{\mathbb{Q}_1}(\{q\}) = 0$ for every $q \in \mathbb{Q}_1$, countable additivity forces

$$\lambda_{\mathbb{Q}_1}(\mathbb{Q}_1) = \sum_{q \in \mathbb{Q}_1} \lambda_{\mathbb{Q}_1}(\{q\}) = 0,$$

which is obviously irreconcilable with $\lambda_{\mathbb{Q}_1}(\mathbb{Q}_1) = 1$. Thus, in order to achieve countable additivity, one has to leave \mathbb{Q}_1 behind. In particular, one can complete \mathbb{Q}_1 and pass to $[0, 1]$, where Lebesgue and his measure come to the rescue. Of course, one pays a heavy price for this transition: \mathbb{Q}_1 is invisible to Lebesgue measure!

1991 *Mathematics Subject Classification*. Primary 60J65; Secondary 35K05, 53C99.

The preceding is a somewhat primitive example of an issue which arises quite often in the construction of measures. Namely, one has a space on which it appears reasonable to construct a measure which reflects, in a natural way, one or more of the defining properties of that space. Moreover, it is clear that a finitely additive representative exists. However, one discovers an insurmountable obstruction to extending this finitely additive representative as a countably additive measure. Of course, in general, there will be no *deus ex machina* which comes to the rescue. On the other hand, as N. Wiener was the first to discover, there is an interesting class of examples for which the sort of subterfuge involved in passing from \mathbb{Q}_1 to $[0, 1]$ makes it possible (again at the same sort of heavy price paid above) to overcome what looks, at first, like a fatal obstruction.

The problem to which Wiener addressed himself was that of putting a *natural* probability measure on the space of \mathbb{R} -valued paths on $[0, 1]$, and, guided by physics, *natural* meant that the paths should be distributed in the way that Einstein and Smoluchowski predicted that trajectories of Brownian particles are.¹ Thanks to R. Feynman, I.E. Segal, and others, we now understand that Wiener's construction is just one example of a whole class of constructions growing out of the attempt to put Gaussian measures on an infinite dimensional Hilbert space; and thanks to K. Itô, we also know how to extend the construction to cover "Gaussian measures" on certain non-linear structures, like the space of paths with values in a Riemannian manifold.

The goal of this survey is to explain to the uninitiated what Wiener and his successors have done.

§1: THE CLASSICAL SETTING

If \mathbf{H} is a finite dimensional, real Hilbert space with norm $\|\cdot\|_{\mathbf{H}}$, then the *standard Gauss* measure $\mu_{\mathbf{H}}$ on \mathbf{H} is the Borel, probability measure given by

$$(1) \quad \mu_{\mathbf{H}}(d\mathbf{h}) = (\sqrt{2\pi})^{-\dim(\mathbf{H})} \exp\left(-\frac{\|\mathbf{h}\|_{\mathbf{H}}^2}{2}\right) \lambda_{\mathbf{H}}(d\mathbf{h}),$$

where $\lambda_{\mathbf{H}}$ is the Lebesgue measure on \mathbf{H} , in the sense that $\lambda_{\mathbf{H}}$ is translation invariant and, for any choice of orthonormal basis $(\mathbf{e}_1, \dots, \mathbf{e}_{\dim(\mathbf{H})})$ in \mathbf{H} , the corresponding unit cube has measure one:

$$(2) \quad \lambda_{\mathbf{H}}\left(\{\mathbf{h} : 0 \leq (\mathbf{h}, \mathbf{e}_m)_{\mathbf{H}} \leq 1 \text{ for } 1 \leq m \leq \dim(\mathbf{H})\}\right) = 1.$$

Thus, in the jargon of probability theory, when $\mathbf{H} = \mathbb{R}$ and \mathbb{R} is given the standard Euclidean norm, $\mu_{\mathbb{R}}$ is the distribution of a *standard, normal* random variable.

In spite of the obvious objections (a normalizing constant which vanishes and a translation invariant measure which does not exist) to doing so, one might try

¹Wiener was neither the only nor even the first mathematician to grapple with this problem. Indeed, it is clear that the French school of analysis had a head start on Wiener and, in many ways, they *understood* the abstract problem better than he did. In particular, they seem to have realized that the measure would have to place most of its mass in a finite dimensional part of the pathspace (cf. the discussion of Ulam's lemma below). To what extent Wiener was aware of such considerations is unclear. Certainly, he never addressed them directly, and perhaps his ignorance of them played an essential role in his ultimate success.

to adopt the same definition even when \mathbf{H} is infinite dimensional.² Indeed, one might think that it is possible to circumvent the objections by replacing (1) with a statement in which only finite dimensional objects appear. That is, ask that $\mu_{\mathbf{H}}$ be the Borel measure on \mathbf{H} with the property that

$$(3) \quad \mu_{\mathbf{H}} \circ \Pi_{\mathbf{F}}^{-1} = \mu_{\mathbf{F}} \quad \text{for every finite dimensional subspace } \mathbf{F} \subseteq \mathbf{H},$$

where $\Pi_{\mathbf{F}}$ denotes orthogonal projection onto \mathbf{F} and the notation $\mu \circ f^{-1}$ denotes the *pushforward* or *image* of the measure μ under the map f :

$$\mu \circ f^{-1}(\Gamma) \equiv \mu(f^{-1}\Gamma).$$

Certainly (3) holds when \mathbf{H} is finite dimensional. Moreover, even in the infinite dimensional case, (3) still determines a unique *finitely additive* measure on the algebra of subsets of the form $\Pi_{\mathbf{F}}^{-1}(\Gamma)$, where \mathbf{F} is a finite dimensional subspace and Γ is a Borel subset of \mathbf{F} . However, when \mathbf{H} is infinite dimensional, *there is no countably additive Borel measure $\mu_{\mathbf{H}}$ which both satisfies (3) and lives on \mathbf{H} .*

There are several ways in which to see what goes wrong in infinite dimensions, but they all turn on Ulam's Lemma, which says that a Borel, probability measure on a Polish (i.e., complete, separable) metric space must be *nearly supported* on a compact subset.³ To see that (3) is incompatible with Ulam's Lemma, observe that, were it to exist, $\mu_{\mathbf{H}}$ would have to be *rotation invariant*, and there are no non-trivial, rotation invariant, compact subsets of an infinite dimensional Hilbert space.

At this point, one has to make a decision: either live with finite additivity or abandon \mathbf{H} . We will opt for the latter. To explain what we have in mind,⁴ let an infinite dimensional \mathbf{H} be given, ignore the warning contained in the preceding paragraph, and pretend that $\mu_{\mathbf{H}}$ on \mathbf{H} exists. Next, choose an orthonormal basis $\{\mathbf{h}_n\}_0^\infty$ for \mathbf{H} , and observe that, under $\mu_{\mathbf{H}}$,

$$\mathbf{h} \in \mathbf{H} \longmapsto \left((\mathbf{h}, \mathbf{h}_0)_{\mathbf{H}}, \dots, (\mathbf{h}, \mathbf{h}_n)_{\mathbf{H}}, \dots \right) \in \mathbb{R}^{\mathbb{N}}$$

is a sequence of mutually independent, standard normal random variables:

$$\mu_{\mathbf{H}}\left(\{\mathbf{h} : (\mathbf{h}, \mathbf{h}_0)_{\mathbf{H}} \in \Gamma_0, \dots, (\mathbf{h}, \mathbf{h}_n)_{\mathbf{H}} \in \Gamma_n\}\right) = \prod_0^n \mu_{\mathbb{R}}(\Gamma_m)$$

for all $n \in \mathbb{N}$ and all Borel subsets $\{\Gamma_m\}_0^n$ of \mathbb{R} . Thus, to realize $\mu_{\mathbf{H}}$, it is reasonable to run this observation backwards. That is,

(i) Define γ to be the infinite product measure $\mu_{\mathbb{R}}^{\mathbb{N}}$ on $\mathbb{R}^{\mathbb{N}}$. Under γ , the coordinates of $\boldsymbol{\xi} \in \mathbb{R}^{\mathbb{N}}$ form a sequence of mutually independent, standard normal random variables. That is,

$$\gamma(\Gamma_0 \times \dots \times \Gamma_n \times \mathbb{R} \times \dots) = \prod_{m=0}^n \mu_{\mathbb{R}}(\Gamma_m)$$

for any $n \in \mathbb{N}$ and Borel subsets $\{\Gamma_m\}_0^n$ of \mathbb{R} .

²Throughout, all Hilbert spaces will be real and separable.

³More precisely, for each $\epsilon \in (0, 1)$, there must exist a compact K_ϵ for which $\mu(K_\epsilon) \geq 1 - \epsilon$.

⁴Details of this construction can be found in §4.2 of [S].

(ii) Let $\ell^2(\mathbb{N})$ denote the subset of square summable elements of $\mathbb{R}^{\mathbb{N}}$, and define

$$(4) \quad \boldsymbol{\eta} \in \ell^2(\mathbb{N}) \mapsto H(\mathbf{h}) \equiv \sum_{n=0}^{\infty} \eta_n \mathbf{h}_n \in \mathbf{H}.$$

(iii) Show that $\gamma(\ell^2(\mathbb{N})) = 1$, and take $\mu_{\mathbf{H}} = \gamma \circ H^{-1}$.

Obviously, it is at step (iii) that the scheme comes apart. Namely, not only does $\ell^2(\mathbb{N})$ fail to have full measure under γ , it has γ -measure 0. To see this, simply note that

$$\int_{\mathbb{R}^{\mathbb{N}}} \exp\left(-\sum_0^{\infty} \xi_m^2\right) \gamma(d\boldsymbol{\xi}) = \prod_0^{\infty} \int_{\mathbb{R}} e^{-\xi_m^2} \mu_{\mathbb{R}}(d\xi_m) = 0.$$

Thus, on the one hand, if we are going to get anywhere, we must learn how to extend the map H in (4) to a set which does have full γ -measure. On the other hand, it is only for $\boldsymbol{\eta} \in \ell^2(\mathbb{N})$ that $H(\boldsymbol{\eta})$ can be in \mathbf{H} . However, there is no *a priori* reason to doubt that an extension exists if we take a more liberal attitude toward what *existence* means, that is, if we look for an extension whose values lie in a space which is strictly larger than \mathbf{H} itself.

With the preceding ideas in mind, we will specialize to the particularly important special case when $\mathbf{H} = \mathbf{H}_1(\mathbb{R}^d)$, the space of absolutely continuous $\mathbf{h} : [0, 1] \rightarrow \mathbb{R}^d$ with $\mathbf{h}(0) = \mathbf{0}$ and derivative $\dot{\mathbf{h}} \in L^2([0, 1]; \mathbb{R}^d)$. Clearly, $\mathbf{H}_1(\mathbb{R}^d)$ becomes a separable Hilbert space when we take

$$\|\mathbf{h}\|_{\mathbf{H}_1(\mathbb{R}^d)} = \|\dot{\mathbf{h}}\|_{L^2([0, 1]; \mathbb{R}^d)}.$$

Thus, if it made sense, $\mu_{\mathbf{H}_1(\mathbb{R}^d)}$ would be given by

$$(5) \quad \mu_{\mathbf{H}_1(\mathbb{R}^d)}(d\mathbf{h}) = (\sqrt{2\pi})^{-\dim(\mathbf{H}_1(\mathbb{R}^d))} \exp\left(-\frac{1}{2} \int_0^1 |\dot{\mathbf{h}}(t)|^2 dt\right) \lambda_{\mathbf{H}_1(\mathbb{R}^d)}(d\mathbf{h}).$$

In order to implement the sort of program suggested at the end of the preceding paragraph, choose an orthonormal basis $\{\mathbf{h}_n\}_0^{\infty}$ for $\mathbf{H}_1(\mathbb{R}^d)$ and define $H : \ell^2(\mathbb{N}) \rightarrow \mathbf{H}_1(\mathbb{R}^d)$ as in (4). Next, for $N \in \mathbb{N}$ and $\boldsymbol{\xi} \in \mathbb{R}^{\mathbb{N}}$, let $\boldsymbol{\xi}^{(N)}$ be the element of $\ell^2(\mathbb{N})$ which is obtained from $\boldsymbol{\xi}$ by replacing ξ_n with 0 for all $n > N$. What we want to do is find a space $\mathfrak{W}(\mathbb{R}^d) \supset \mathbf{H}$ with the property that

$$(6) \quad H(\boldsymbol{\xi}) \equiv \lim_{N \rightarrow \infty} H(\boldsymbol{\xi}^{(N)}) \text{ exists in } \mathfrak{W}(\mathbb{R}^d) \text{ for } \gamma\text{-almost every } \boldsymbol{\xi} \in \mathbb{R}^{\mathbb{N}}.$$

The celebrated fact revealed by N. Wiener in [W] is that one can take $\mathfrak{W}(\mathbb{R}^d)$ to be the Banach (a term to which Wiener himself might have taken exception) space of continuous $\mathbf{w} : [0, 1] \rightarrow \mathbb{R}^d$ with $\mathbf{w}(0) = \mathbf{0}$ and norm $\|\mathbf{w}\|_{\mathfrak{W}(\mathbb{R}^d)} = \sup_{t \in [0, 1]} |\mathbf{w}(t)|$.

Actually, Wiener dealt with a particular choice of orthonormal basis and with the lacunary subsequence $H(\boldsymbol{\xi}^{(2^n)})$. Thus, it is important to know that the measure at which the above procedure arrives does not depend on the basis $\{\mathbf{h}_n\}_0^{\infty}$, the subsequence extracted, or any other peculiarity. In addition, one should be curious about the extent to which the resulting measure is a representation of the idea expressed in (5). In other words, what is a direct characterization of $\mu_{\mathbf{H}_1(\mathbb{R}^d)}$ on

$\mathfrak{W}(\mathbb{R}^d)$ which captures the spirit of (5)? To answer this, note that, if it means anything at all, (5) is trying to say that, for each $\mathbf{g} \in \mathbf{H}_1(\mathbb{R}^d)$,

$$\mathbf{h} \in \mathbf{H}_1(\mathbb{R}^d) \longmapsto (\mathbf{h}, \mathbf{g})_{\mathbf{H}_1(\mathbb{R}^d)} \in \mathbb{R} \quad \text{under } \mu_{\mathbf{H}_1(\mathbb{R}^d)}$$

is a Gaussian random variable with mean 0 and variance $\|\mathbf{g}\|_{\mathbf{H}_1(\mathbb{R}^d)}^2$.⁵ Thus, one should hope to have the characterization of $\mu_{\mathbf{H}_1(\mathbb{R}^d)}$ say something like: for each $\mathbf{g} \in \mathbf{H}_1(\mathbb{R}^d)$,

$$\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d) \longmapsto (\mathbf{w}, \mathbf{g})_{\mathbf{H}_1(\mathbb{R}^d)} \in \mathbb{R}$$

is Gaussian with mean 0 and variance $\|\mathbf{g}\|_{\mathbf{H}_1(\mathbb{R}^d)}^2$. However, there is a problem here: $\mu_{\mathbf{H}_1(\mathbb{R}^d)}$ -almost no \mathbf{w} is an element of $\mathbf{H}_1(\mathbb{R}^d)$, and, so, what is the meaning of $(\mathbf{w}, \mathbf{g})_{\mathbf{H}_1(\mathbb{R}^d)}$? To get around this technicality, we restrict our attention to \mathbf{g} 's for which one can use integration by parts to interpret $(\mathbf{w}, \mathbf{g})_{\mathbf{H}_1(\mathbb{R}^d)}$. Namely, by the Riesz representation theorem, we can identify the dual space $\mathfrak{W}(\mathbb{R}^d)^*$ with the space of totally finite, \mathbb{R}^d -valued, Borel measures $\boldsymbol{\lambda}$ on $(0, 1]$. Next, for $\boldsymbol{\lambda} \in \mathfrak{W}(\mathbb{R}^d)^*$, let $\mathbf{g}_\boldsymbol{\lambda}$ be the element of $\mathbf{H}_1(\mathbb{R}^d)$ determined by

$$\mathbf{g}_\boldsymbol{\lambda}(t) = \int_0^t \boldsymbol{\lambda}((\tau, 1]) \, d\tau, \quad t \in [0, 1].$$

Since, for $\mathbf{h} \in \mathbf{H}_1(\mathbb{R}^d)$,

$$(\mathbf{h}, \mathbf{g}_\boldsymbol{\lambda})_{\mathbf{H}_1(\mathbb{R}^d)} = \langle \mathbf{h}, \boldsymbol{\lambda} \rangle \equiv \int_{(0,1]} (\mathbf{h}(t), \boldsymbol{\lambda}(dt))_{\mathbb{R}^d},$$

it is clear that

$$\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d) \longmapsto \langle \mathbf{w}, \boldsymbol{\lambda} \rangle \in \mathbb{R}$$

provides the unique continuous extension of

$$\mathbf{h} \in \mathbf{H}_1(\mathbb{R}^d) \longmapsto (\mathbf{h}, \mathbf{g}_\boldsymbol{\lambda})_{\mathbf{H}_1(\mathbb{R}^d)} \in \mathbb{R}$$

to $\mathfrak{W}(\mathbb{R}^d)$. Hence, the interpretation of (5) for $\mu_{\mathbf{H}_1(\mathbb{R}^d)}$ as a measure on $\mathfrak{W}(\mathbb{R}^d)$ ought to be that, under $\mu_{\mathbf{H}_1(\mathbb{R}^d)}$,

$$(7) \quad \begin{aligned} &\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d) \longmapsto \langle \mathbf{w}, \boldsymbol{\lambda} \rangle \in \mathbb{R} \quad \text{is Gaussian with mean 0} \\ &\text{and variance } \|\mathbf{g}_\boldsymbol{\lambda}\|_{\mathbf{H}_1(\mathbb{R}^d)}^2 = \iint_{(0,1]^2} (s \wedge t) (\boldsymbol{\lambda}(ds), \boldsymbol{\lambda}(dt))_{\mathbb{R}^d} \\ &\hspace{15em} \text{for every } \boldsymbol{\lambda} \in \mathfrak{W}(\mathbb{R}^d)^*. \end{aligned}$$

It is an easy matter to check that there is at most one measure on $\mathfrak{W}(\mathbb{R}^d)$ for which (7) holds. Moreover, suppose that, for some choice of orthonormal basis

⁵That is, a random variable which one gets by multiplying a standard, normal random variable by $\|\mathbf{g}\|_{\mathbf{H}}$.

$\{\mathbf{h}_n\}_0^\infty$ and subsequence $\{N_m\}_{m=0}^\infty$, $H : \mathbb{R}^\mathbb{N} \rightarrow \mathfrak{W}(\mathbb{R}^d)$ is a measurable extension of the map in (4) with the property that

$$H(\boldsymbol{\xi}) = \lim_{m \rightarrow \infty} H(\boldsymbol{\xi}^{(N_m)}) \quad \text{for } \gamma\text{-almost every } \boldsymbol{\xi} \in \mathbb{R}^\mathbb{N}.$$

Then, for each $\boldsymbol{\lambda} \in \mathfrak{W}(\mathbb{R}^d)^*$ and γ -almost every $\boldsymbol{\xi} \in \mathbb{R}^\mathbb{N}$,

$$\left(H(\boldsymbol{\xi}^{(N_m)}), \mathbf{g}\boldsymbol{\lambda} \right)_{\mathbf{H}_1(\mathbb{R}^d)} = \langle H(\boldsymbol{\xi}^{(N_m)}), \boldsymbol{\lambda} \rangle \rightarrow \langle H(\boldsymbol{\xi}), \boldsymbol{\lambda} \rangle.$$

At the same time, if \mathbf{F}_m denotes the span of $\{\mathbf{h}_0, \dots, \mathbf{h}_{N_m}\}$, then, under γ ,

$$\boldsymbol{\xi} \in \mathbb{R}^\mathbb{N} \mapsto \left(H(\boldsymbol{\xi}^{(N_m)}), \mathbf{g}\boldsymbol{\lambda} \right)_{\mathbf{H}_1(\mathbb{R}^d)} \in \mathbb{R}$$

is Gaussian with mean 0 and variance (cf. (3)) $\|\Pi_{\mathbf{F}_m} \mathbf{g}\boldsymbol{\lambda}\|_{\mathbf{H}_1(\mathbb{R}^d)}^2$. Hence, under γ ,

$$\boldsymbol{\xi} \in \mathbb{R}^\mathbb{N} \mapsto \langle H(\boldsymbol{\xi}), \boldsymbol{\lambda} \rangle \in \mathbb{R}$$

is Gaussian with mean 0 and variance $\|\mathbf{g}\boldsymbol{\lambda}\|_{\mathbf{H}_1(\mathbb{R}^d)}^2$. In other words, $\gamma \circ H^{-1}$ satisfies (7).

However one arrives at it, the measure $\mu_{\mathbf{H}_1(\mathbb{R}^d)}$ on $\mathfrak{W}(\mathbb{R}^d)$ satisfying (7) is called *Wiener measure*, and, because Wiener was motivated by the Einstein–Smoluchowski model of Brownian motion, “ $\mu_{\mathbf{H}_1(\mathbb{R}^d)}$ -typical” elements of $\mathfrak{W}(\mathbb{R}^d)$ are sometimes called *Brownian paths*.

Before moving on, it may be helpful to mention one of the basic properties of Brownian paths. We already know that almost none of them is in $\mathbf{H}_1(\mathbb{R}^d)$, but matters are even worse. To wit, $\mu_{\mathbf{H}_1(\mathbb{R}^d)}$ -almost no $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d)$ is differentiable anywhere, a fact which, as we will see below, haunts every attempt to deal with Brownian paths. The reason why this unfortunate property is not fatal is that Brownian paths compensate for their lack of differentiability by having remarkable cancellation properties, of which we have already seen an example. Namely, although we side-stepped the issue of defining $(\mathbf{w}, \mathbf{h})_{\mathbf{H}_1(\mathbb{R}^d)}$ for arbitrary $\mathbf{h} \in \mathbf{H}_1(\mathbb{R}^d)$, we now have a way to deal with it. Indeed, a dividend of the preceding paragraph is the observation that

$$\mathbf{g}\boldsymbol{\lambda} \in \mathbf{H}_1(\mathbb{R}^d) \mapsto “(\cdot, \mathbf{g}\boldsymbol{\lambda})_{\mathbf{H}_1(\mathbb{R}^d)}” \equiv \langle \cdot, \boldsymbol{\lambda} \rangle \in L^2(\mu_{\mathbf{H}_1(\mathbb{R}^d)})$$

is an isometry. Hence, since $\{\mathbf{g}\boldsymbol{\lambda} : \boldsymbol{\lambda} \in \mathfrak{W}(\mathbb{R}^d)^*\}$ is dense in $\mathbf{H}_1(\mathbb{R}^d)$, there is a unique way to extend this map as a continuous mapping from $\mathbf{H}_1(\mathbb{R}^d)$ to $L^2(\mu_{\mathbf{H}_1(\mathbb{R}^d)})$. That is, even though $\mu_{\mathbf{H}_1(\mathbb{R}^d)}(\mathbf{H}_1(\mathbb{R}^d)) = 0$, the inner product $(\mathbf{w}, \mathbf{h})_{\mathbf{H}_1(\mathbb{R}^d)}$ makes (some) sense for each $\mathbf{h} \in \mathbf{H}_1(\mathbb{R}^d)$ and $\mu_{\mathbf{H}_1(\mathbb{R}^d)}$ -almost every $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d)$. As it turns out, this extension, which is due to Paley and Wiener, is essential as soon as one attempts doing calculus involving $\mu_{\mathbf{H}_1(\mathbb{R}^d)}$. Namely, elementary manipulation of Gaussian integrals quickly leads to the conclusion (cf. Lemma 4.2.15 in [S]) that $\mu_{\mathbf{H}_1(\mathbb{R}^d)}$ is *quasi-translation invariant in the directions of* $\mathbf{H}_1(\mathbb{R}^d)$. More precisely, if $\mathbf{h} \in \mathbf{H}_1(\mathbb{R}^d)$ and

$$\mathbf{R}_{\mathbf{h}}(\mathbf{w}) \equiv \exp\left((\mathbf{w}, \mathbf{h})_{\mathbf{H}_1(\mathbb{R}^d)} - \frac{1}{2}\|\mathbf{h}\|_{\mathbf{H}_1(\mathbb{R}^d)}^2\right),$$

then (just as (1) would seem to be predicting)

$$\mu_{\mathbf{H}_1(\mathbb{R}^d)}(\{\mathbf{w} : \mathbf{h} + \mathbf{w} \in \Gamma\}) = \int_{\Gamma} \mathbf{R}_{\mathbf{h}}(\mathbf{w}) \mu_{\mathbf{H}_1(\mathbb{R}^d)}(\mathbf{w})$$

for all Borel subsets Γ of $\mathfrak{M}(\mathbb{R}^d)$. This important observation was made originally by R.H. Cameron and T. Martin and is the reason why $\mathbf{H}_1(\mathbb{R}^d)$ is sometimes called the *Cameron–Martin space* of $\mu_{\mathbf{H}_1(\mathbb{R}^d)}$. The complementary assertion that $\mu_{\mathbf{H}_1(\mathbb{R}^d)}$ becomes singular to itself when translated in directions outside of $\mathbf{H}_1(\mathbb{R}^d)$ was first proved by I.E. Segal (cf. Exercise 5.2.38 in [S]).

§2: BROWNIAN MOTION ON A RIEMANNIAN MANIFOLD

Up to this point, all the ideas with which we have been dealing stem from Wiener, although our treatment benefits immeasurably from the interpretation given them by I.E. Segal and his school (especially, L. Gross). Due, in part, to the efforts of Segal et al., Wiener’s ideas had a renaissance during the heyday of constructive quantum field theory. Most notably, in the early ’60s, K. Symanzik initiated the study of *Euclidean quantum field theory*,⁶ and his program was taken up by a host of mathematicians, especially E. Nelson, A. Jaffe, and J. Glimm. However, we will take a quite different tack. Namely, following in K. Itô’s footsteps, we will explain how to transfer Wiener’s ideas about Brownian paths in a linear space, like \mathbb{R}^d , to “Brownian” paths on a non-linear space. Thus, this section is devoted to the construction of Brownian motion on a d -dimensional, Riemannian manifold M .⁷

To get started, we adopt the same heuristic approach as we did in \mathbb{R}^d . Thus, choose an initial reference point $x_0 \in M$ (alias $\mathbf{0} \in \mathbb{R}^d$), and take $\mathbf{H}_1(M)$ to be the space of continuous paths $h : [0, 1] \rightarrow M$ with

$$(8) \quad h(0) = x_0 \quad \text{and} \quad \|h\|_{\mathbf{H}_1(M)} \equiv \left(\int_{[0,1]} |\dot{h}(t)|_{T_{h(t)}(M)}^2 dt \right)^{\frac{1}{2}} < \infty.$$

Our goal is to make sense out of specifying a measure $\mu_{\mathbf{H}_1(M)}$ by the “Gaussian” prescription

$$(9) \quad \mu_{\mathbf{H}_1(M)}(dh) = (\sqrt{2\pi})^{-\dim(\mathbf{H}_1(M))} \exp\left(-\frac{1}{2} \int_0^1 |\dot{h}(t)|_{T_{h(t)}(M)}^2 dt\right) \lambda_{\mathbf{H}_1(M)}(dh).$$

⁶Symanzik’s idea (cf. [Sym]) was to start with a Gaussian measure on an appropriate function space over $\mathbb{R} \times \mathbb{R}^N$, which represents (*imaginary time*) \times *space*. If one uses analytic continuation to define the moment generating function for such a *Gaussian field* on $\sqrt{-1}\mathbb{R} \times \mathbb{R}^N$, (*real time*) \times *space*, then one gets what physicists would call a *Schwinger function*. More precisely, they would call it a *trivial* Schwinger function. To use this procedure to arrive at *non-trivial* Schwinger functions, Symanzik knew that the original Gaussian field must be replaced by a non-Gaussian one, and his idea was to “perturb” the Gaussian field. However, except when $N = 0$ and one is basically dealing with Brownian motion, the “perturbation” is highly singular and the computations are formidable. For introductions to this fascinating but technically horrendous topic, the books [GJ], by Glimm and Jaffe, and [Sim], by B. Simon, are hard to beat.

⁷To avoid technical difficulties, the reader should assume that M is connected and compact. As will be apparent, connectedness is not so important, since nothing is lost by restricting attention to a connected component. On the other hand, compactness simplifies matters considerably. In particular, it eliminates concerns about possible *explosion*.

If (5) seemed problematic, then (9) has to be considered downright ridiculous! In the first place, $\mathbf{H}_1(M)$ has no linear structure, and so, even if we were to suspend our disbelief, what is (9) supposed to be saying? For example, in view of the absence of any linear structure on $\mathbf{H}_1(M)$, how is one to give even a fanciful interpretation to $\lambda_{\mathbf{H}_1(M)}$? (Presumably $\dim(\mathbf{H}_1(M)) = \infty$, for whatever good that does.) Thus, no matter what one's political persuasion, (9) calls for quite a lot of explanation.

Our explanation will rely on the typical mathematical subterfuge of mapping the problem at hand to the problem which we have already solved. We begin with the remark that $\mathbf{H}_1(M)$ can be easily⁸ made into a Polish space in which convergence of $\{h_n\}_1^\infty$ to h is equivalent to

$$\|\varphi \circ h_n - \varphi \circ h\|_{\mathbf{H}_1(\mathbb{R})} \quad \text{for every } \varphi \in C^1(M; \mathbb{R}).$$

Thus, at least it is a reasonable place in which to discuss measures. More important, although $\mathbf{H}_1(M)$ appears to possess no linear structure of its own, the Levi–Civita connection provides a canonical relation between $\mathbf{H}_1(M)$ and $\mathbf{H}_1(\mathbb{R}^d)$. Namely, first choose and fix an orthonormal basis $\mathbf{e}_0 = (E_1(0), \dots, E_d(0))$ in $T_{x_0}(M)$. Next, for each $h \in \mathbf{H}_1(M)$, let

$$t \in [0, 1] \longmapsto \mathbf{e}(t) = (E_1(t), \dots, E_d(t)) \in (T_{h(t)}(M))^d$$

be defined so that, for each $1 \leq k \leq d$ and $t \in [0, 1]$, $E_k(t)$ is obtained from $E_k(0)$ by parallel (relative to the Levi–Civita connection) transport along $h \upharpoonright [0, t]$.⁹ Finally, define $\mathbf{h} \in \mathbf{H}_1(\mathbb{R}^d)$ so that¹⁰

$$(10) \quad \dot{h}(t) = \mathbf{e}(t)\dot{\mathbf{h}}(t), \quad t \in [0, 1].$$

The passage from $h \in \mathbf{H}_1(M)$ to $\mathbf{h} \in \mathbf{H}_1(\mathbb{R}^d)$ is a familiar one in geometry. However, geometers usually prefer to stay in $T_{x_0}(M)$, instead of moving all the way to \mathbb{R}^d . Thus, they would consider the path

$$t \in [0, 1] \longmapsto \mathbf{e}_0\mathbf{h}(t) \in T_{x_0}(M)$$

and would call this path the *development* of h in $T_{x_0}(M)$.¹¹ Moreover, the development map is homeomorphic, its inverse being the map which *rolls* a curve in

⁸For instance, one might embed M in \mathbb{R}^N (for some sufficiently large N) so that $x_0 = \mathbf{0}$, and then give $\mathbf{H}_1(M)$ the metric which it inherits as a closed subset of $\mathbf{H}_1(\mathbb{R}^N)$.

⁹For the reader who is wondering whether this parallel transport makes good sense along a curve h from $\mathbf{H}_1(M)$, as opposed to $C^1([0, 1]; M)$, be assured that an easy approximation procedure shows that it does. Also, be forewarned that things are only going to get worse.

¹⁰Here and elsewhere, we think of $\mathfrak{f} \in \mathcal{O}(M)$ as an isometry from \mathbb{R}^d onto $T_{\pi(\mathfrak{f})}(M)$. Thus, if $\mathfrak{f} = (x, \mathbf{e})$ and $\mathbf{v} \in \mathbb{R}^d$, then $\mathfrak{f}\mathbf{v}$ is the element of $T_x(M)$ for which \mathbf{v} are the coordinates relative to the basis \mathbf{e} .

¹¹An intuitive description of the development can be given as follows. Take a marker and lay down an excess of ink along the path h . Next, thinking of M as a submanifold in \mathbb{R}^N , take the tangent space $T_{x_0}(M)$ to M at x_0 , and *press* it down along the path h in such a way that it remains tangent at its point of contact with M . The development of h is the curve which gets inked onto $T_{x_0}(M)$.

$T_{x_0}(M)$ onto M . That is, given $\mathbf{h} \in \mathbf{H}_1(\mathbb{R}^d)$, transfer it to $T_{x_0}(M)$ via $\boldsymbol{\epsilon}_0$, and roll the resulting curve onto M by solving the system of equations

$$(11) \quad \begin{aligned} \dot{h}(t) &= \boldsymbol{\epsilon}(t)\dot{\mathbf{h}}(t) && \text{with } h(0) = x_0 \\ \nabla_{\dot{h}(t)}\boldsymbol{\epsilon}(t) &= \mathbf{0} && \text{with } \boldsymbol{\epsilon}(0) = \boldsymbol{\epsilon}_0, \end{aligned}$$

where $\nabla_X Y$ denotes the covariant derivative of Y with respect to X .

With the preceding discussion in mind, we now give $\mathbf{H}_1(M)$ the linear structure which it inherits from $\mathbf{H}_1(\mathbb{R}^d)$ via the homeomorphism just described. In this way, we can interpret $\lambda_{\mathbf{H}_1(M)}$ heuristically as *the image of $\lambda_{\mathbf{H}_1(\mathbb{R}^d)}$ under the rolling map*; and, because $\|h\|_{\mathbf{H}_1(M)} = \|\mathbf{h}\|_{\mathbf{H}_1(\mathbb{R}^d)}$ if $\mathbf{h} \in \mathbf{H}_1(\mathbb{R}^d)$ gets rolled onto $h \in \mathbf{H}_1(M)$, it is now only consistent to interpret (9) as saying that $\mu_{\mathbf{H}_1(M)}$ on $\mathbf{H}_1(M)$ is *the image of $\mu_{\mathbf{H}_1(\mathbb{R}^d)}$ on $\mathbf{H}_1(\mathbb{R}^d)$ under the rolling map*. Equivalently, under $\mu_{\mathbf{H}_1(M)}$, *the distribution of developed paths is the Gaussian measure $\mu_{\mathbf{H}_1(\mathbb{R}^d)}$* . Of course, since neither $\lambda_{\mathbf{H}_1(\mathbb{R}^d)}$ nor even $\mu_{\mathbf{H}_1(\mathbb{R}^d)}$ actually exists on $\mathbf{H}_1(\mathbb{R}^d)$, this interpretation does not really solve our problem. On the other hand, it represents progress and tells us what still has to be done. Specifically, we must show that *the rolling map admits an extension to a subset of $\mathfrak{W} \in \mathfrak{W}(\mathbb{R}^d)$ having full $\mu_{\mathbf{H}_1(\mathbb{R}^d)}$ -measure*. Obviously, we cannot expect the image of this extension to be $\mathbf{H}_1(M)$. Instead, we should hope that it will be

$$\mathfrak{W}(M) \equiv \{w \in C([0, 1]; M) : w(0) = x_0\}.$$

In other words, we are now at the same place in the present program as we were earlier at the end of step (ii) in the construction of $\mu_{\mathbf{H}_1(\mathbb{R}^d)}$ from γ , and our strategy for taking the next step will resemble the one which we adopted there.

The procedure with which we will extend the rolling map derives from a general scheme developed by K. Itô.¹² In more or less the same way as Paley and Wiener carried out their extension (cf. the last paragraph of §1), what Itô's scheme provides is a machine for extending maps which are originally defined on $\mathbf{H}_1(\mathbb{R}^d)$ to ones which are defined, up to a $\mu_{\mathbf{H}_1(\mathbb{R}^d)}$ -null set, on $\mathfrak{W}(\mathbb{R}^d)$. However, in order to apply Itô's machine, it is important to have the original map arise as the solution to a first order, ordinary differential equation into which $\dot{\mathbf{h}}$ enters linearly. For this reason, it is important for us to develop such a representation of the rolling map, and, to facilitate matters, it is best¹³ to move all our considerations to the bundle $\mathcal{O}(M)$ of orthonormal frames.¹⁴ That is, as an abstract set, $\mathcal{O}(M)$ is the collection of all pairs $(x, \boldsymbol{\epsilon})$ as x varies through M and, for a given $x \in M$, $\boldsymbol{\epsilon}$ runs over all the orthonormal bases in $T_x(M)$. Clearly, the fiber map $\pi : \mathcal{O}(M) \rightarrow M$, which takes $(x, \boldsymbol{\epsilon})$ to x , is onto, and, for each $x \in M$, the fiber $\pi^{-1}(x)$ is an isomorphic image of the orthogonal group $O(d)$. In fact, given a coordinate chart (U, ψ) for M at x ,

¹²The cognoscenti will realize that Stratonovich introduced the flavor of Itô's methodology which we are going to use.

¹³The origins of this important observation are a little murky. However, the credit should be shared by J. Eells, E.D. Elworthy, and P. Malliavin (the significance of the order being only alphabetical). All three mentioned the idea in lectures given in the early '70s. Papers followed: [EE] and [M]. In a less elegant form, their program was anticipated by R. Gangolli [G]. Of course, the first words on the subject are Itô's in [I1] and [I2].

¹⁴The author's own introduction to the orthonormal frame bundle was provided by [BC], and the notation which follows is derived from that source.

use the Gram–Schmidt orthogonalization procedure to produce a smoothly varying section

$$y \in U \longmapsto \mathbf{e}(y) \in (T_y(M))^d$$

so that $\mathbf{e}(y)$ is an orthonormal basis in $T_y(M)$ for each $y \in U$. Then the mapping

$$(y, \mathbf{e}) \in \pi^{-1}(U) \longmapsto (y, \mathbf{e}(y)^{-1}\mathbf{e}) \in U \times O(d)$$

is an isomorphism, and so, by declaring that such maps be diffeomorphisms, we get a differentiable structure on $\mathcal{O}(M)$. Next, let $\mathfrak{f} = (x, \mathbf{e}) \in \mathcal{O}(M)$, and observe that the tangent space $T_{\mathfrak{f}}(\mathcal{O}(M))$ contains a subspace $V_{\mathfrak{f}}(\mathcal{O}(M))$ of vertical directions: the tangents at \mathfrak{f} to curves in $\mathcal{O}(M)$ along which π is constant. Clearly, $V_{\mathfrak{f}}(\mathcal{O}(M))$ is Lie algebraically isomorphic to the Lie algebra $\mathfrak{o}(d)$ of $O(d)$, and every complementary subspace of $V_{\mathfrak{f}}(\mathcal{O}(M))$ will be linearly isomorphic to $T_x(M)$. Moreover, the Levi–Civita connection provides us with a canonical choice of complementary subspace. Namely, given $X_x \in T_x(M)$, choose $\gamma : \mathbb{R} \rightarrow M$ to be a smooth curve satisfying $\gamma(0) = x$ and $\dot{\gamma}(0) = X_x$, and define

$$s \in \mathbb{R} \longmapsto \mathbf{e}(s) \in (T_{\gamma(s)}(M))^d$$

so that $\mathbf{e}(s)$ is obtained from \mathbf{e} via parallel translation along γ . Then $s \in \mathbb{R} \mapsto \Gamma(s) = (\gamma(s), \mathbf{e}(s)) \in \mathcal{O}(M)$ is a smooth curve in $\mathcal{O}(M)$ which passes through \mathfrak{f} at time 0, and we will call $\hat{X}_{\mathfrak{f}} \equiv \dot{\Gamma}(0) \in T_{\mathfrak{f}}(\mathcal{O}(M))$ the *horizontal lift* of X_x to \mathfrak{f} .¹⁵ Finally, the *horizontal subspace* $H_{\mathfrak{f}}(\mathcal{O}(M))$ of $T_{\mathfrak{f}}(\mathcal{O}(M))$ consists of the vectors $\hat{X}_{\mathfrak{f}}$ obtained by horizontally lifting every $X_x \in T_x(M)$. Note that the distribution $\mathfrak{f} \in \mathcal{O}(M) \mapsto H_{\mathfrak{f}}(\mathcal{O}(M))$ admits a globally defined family of sections which form a basis at each point. To see this, for each $\mathbf{v} \in \mathbb{R}^d$, define $\mathfrak{f} \in \mathcal{O}(M) \mapsto \mathfrak{E}(\mathbf{v})_{\mathfrak{f}} \in H_{\mathfrak{f}}(\mathcal{O}(M))$ so that $\mathfrak{E}(\mathbf{v})_{\mathfrak{f}}$ is the horizontal lift of $\mathfrak{f}\mathbf{v}$. Clearly, if $\{\mathbf{e}_1, \dots, \mathbf{e}_d\}$ is an orthonormal basis in \mathbb{R}^d and

$$(12) \quad \mathfrak{E}_k(\mathfrak{f}) \equiv \mathfrak{E}(\mathbf{e}_k)_{\mathfrak{f}}, \quad 1 \leq k \leq d \text{ and } \mathfrak{f} \in \mathcal{O}(M),$$

then $\{\mathfrak{E}_1(\mathfrak{f}), \dots, \mathfrak{E}_d(\mathfrak{f})\}$ is a basis in $H_{\mathfrak{f}}(\mathcal{O}(M))$ for each $\mathfrak{f} \in \mathcal{O}(M)$.

The virtue of these considerations is revealed in their application to the discussion of quantities like the rolling map. To see their relevance, let $\mathbf{h} \in \mathbf{H}_1(\mathbb{R}^d)$ be given, and determine the path $\mathfrak{F}_{\mathfrak{f}_0}(\cdot, \mathbf{h})$ on $\mathcal{O}(M)$ from

$$(13) \quad \dot{\mathfrak{F}}_{\mathfrak{f}_0}(t, \mathbf{h}) = \mathfrak{E}(\dot{\mathbf{h}}(t))_{\mathfrak{F}_{\mathfrak{f}_0}(t, \mathbf{h})} \quad \text{with } \mathfrak{F}_{\mathfrak{f}_0}(0, \mathbf{h}) = \mathfrak{f}_0 \equiv (x_0, \mathbf{e}_0).$$

After tracing through all the definitions, one discovers that

$$(14) \quad \mathcal{R}_{\mathfrak{f}_0}(\mathbf{h}) \equiv \pi \circ \mathfrak{F}_{\mathfrak{f}_0}(\cdot, \mathbf{h})$$

is precisely the path $h \in \mathbf{H}_1(M)$ described in (11). That is, $\mathcal{R}_{\mathfrak{f}_0}(\mathbf{h})$ is the path in M which results from rolling $t \in [0, 1] \mapsto \mathbf{e}_0\mathbf{h}(t) \in T_{x_0}(M)$ onto M , and so we

¹⁵It is important to note that this definition does not depend on the particular choice of γ but only on \mathfrak{f} and X_x . In addition, one should check that $X_x \in T_x(M) \mapsto \hat{X}_{\mathfrak{f}} \in \hat{T}_{\mathfrak{f}}(\mathcal{O}(M))$ is linear and that, if $x \in U \mapsto X_x \in T_x(M)$ is a smooth vector field on the open set U , then the vector field $\mathfrak{f} \in \pi^{-1}(U) \mapsto \hat{X}_{\mathfrak{f}} \in T_{\mathfrak{f}}(\mathcal{O}(M))$ is also smooth.

have achieved our goal of representing the rolling map in terms of a solution to a first order, ordinary differential into which \mathbf{h} enters linearly.

Although the remaining step is far from trivial, it is completely standard ever since Itô taught probabilists how to make it. In general terms, the procedure runs as follows. For $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d)$ and $n \in \mathbb{N}$, use \mathbf{w}_n to denote the polygonal path which is obtained from \mathbf{w} by linear interpolation between points of the form $m2^{-n}$, $0 \leq m \leq 2^n$. What Itô's theory facilitates is the proof that

$$(15) \quad \mathfrak{F}_{f_0}(\cdot, \mathbf{w}) \equiv \lim_{n \rightarrow \infty} \mathfrak{F}_{f_0}(\cdot, \mathbf{w}_n) \quad \text{exists in } \mathfrak{W}(M)$$

for $\mu_{\mathbf{H}_1(\mathbb{R}^d)}$ -almost every $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d)$,

where the limit is taken in the topology of uniform convergence.¹⁶ Thus, in view of (14), this means that we can find a measurable $\mathcal{R}_{f_0} : \mathfrak{W}(\mathbb{R}^d) \rightarrow \mathfrak{W}(M)$ so that

$$(16) \quad \mathcal{R}_{f_0}(\mathbf{w}) = \lim_{n \rightarrow \infty} \mathcal{R}_{f_0}(\mathbf{w}_n) \quad \text{for } \mu_{\mathbf{H}_1(\mathbb{R}^d)}\text{-almost every } \mathbf{w} \in \mathfrak{W}(\mathbb{R}^d).$$

It remains only to take

$$(17) \quad \mu_{\mathbf{H}_1(M)} = \mu_{\mathbf{H}_1(\mathbb{R}^d)} \circ (\mathcal{R}_{f_0})^{-1}.$$

To complete the analogy with the Euclidean case, a $\mu_{\mathbf{H}_1(M)}$ -typical $w \in \mathfrak{W}(M)$ is said to be a *Brownian path* on M .¹⁷

§3: ANOTHER POINT OF VIEW

Thus far, the slant of this article has been devoutly Gaussian. The Gaussian slant has the advantage that it encourages one to write down intuitively appealing, if not entirely meaningful, expressions like that in (9). The disadvantage is that, except in the Euclidean setting, the Gaussian approach leaves one high and dry when it comes to computations. In fact, at the end of its construction, we know so little about the measure $\mu_{\mathbf{H}_1(M)}$ that we cannot even characterize it as “the only measure on $\mathfrak{W}(M)$ with the property that...” All we know is that, after development, the paths under $\mu_{\mathbf{H}_1(M)}$ are distributed according to the Gaussian measure $\mu_{\mathbf{H}_1(\mathbb{R}^d)}$. However, this is cold comfort when one considers just how hard it must be to actually develop a nowhere differentiable path.

In order to carry out computations with $\mu_{\mathbf{H}_1(M)}$, it is helpful to broaden one's perspective and consider the *Markovian* aspects of $\mu_{\mathbf{H}_1(M)}$ along with the Gaussian. Loosely speaking, a Markov process is a generalized dynamical system in which the underlying dynamical mechanism may be random, and, in the case of Brownian motion, the analogy with classical dynamical systems can be made quite compelling. In fact, the goal of this section is to persuade the reader that the assertion

$$(*) \quad \begin{aligned} &\mu_{\mathbf{H}_1(M)} \text{ is the unique probability measure on } \mathfrak{W}(M) \text{ under which typical} \\ &\text{paths } w \text{ are integral curves of the Laplace–Beltrami operator } \frac{1}{2}\Delta_M \\ &\text{issuing from } x_0 \end{aligned}$$

holds water.

¹⁶See [ST] for a treatment of Itô's theory from the point of view adopted here.

¹⁷Brownian motion has been constructed on other, and more exotic, non-linear spaces. In particular, for reasons which elude this author, physicists want to put Brownian motion on fractals. Whether or not they make sense physically, beautiful constructions of Brownian motion on the Sierpinski gasket and related fractals have been carried out by S. Kusuoka [K] and improved by M. Barlow and E. Perkins [BP].

The preceding assertion is based on the following version of the renowned¹⁸ *Itô's formula*. To be precise, let $\mathcal{B}_s(\mathbb{R}^d)$ denote the smallest σ -algebra over $\mathfrak{W}(\mathbb{R}^d)$ with respect to which $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d) \mapsto \mathbf{w}(\sigma) \in \mathbb{R}^d$ is measurable for every $\sigma \in [0, s]$. (In words, $\mathcal{B}_s(\mathbb{R}^d)$ contains the history of the paths during $[0, s]$.) Then, for all $0 \leq s < t \leq 1$, $A \in \mathcal{B}_s(\mathbb{R}^d)$, and $\Psi \in C^\infty(\mathcal{O}(M); \mathbb{R})$:

$$(18) \quad \begin{aligned} & \int_A \left(\Psi(\mathfrak{F}_{f_0}(t, \mathbf{w})) - \Psi(\mathfrak{F}_{f_0}(s, \mathbf{w})) \right) \mu_{\mathbf{H}_1(\mathbb{R}^d)}(d\mathbf{w}) \\ &= \int_A \left(\int_s^t \frac{1}{2} \Delta_{\mathcal{O}(M)} \Psi(\mathfrak{F}_{f_0}(\tau, \mathbf{w})) d\tau \right) \mu_{\mathbf{H}_1(\mathbb{R}^d)}(d\mathbf{w}), \end{aligned}$$

where (cf. (12)) $\Delta_{\mathcal{O}(M)} \equiv \sum_1^d \mathfrak{E}_k^2$ is Bochner's Laplacian on the orthonormal frame bundle. While proving (18), we may and will assume that $2^N s$ and $2^N t$ are both integers for some $N \in \mathbb{N}$, in which case¹⁹

$$\Psi \circ \mathfrak{F}_{f_0}(t, \mathbf{w}) - \Psi \circ \mathfrak{F}_{f_0}(s, \mathbf{w}) = \sum_{m=[2^N s]}^{[2^N t]-1} D_{n,m}(\mathbf{w})$$

$$\text{where } D_{n,m}(\mathbf{w}) \equiv \Psi \circ \mathfrak{F}_{f_0}(2^{-n}(m+1), \mathbf{w}) - \Psi \circ \mathfrak{F}_{f_0}(2^{-n}m, \mathbf{w})$$

for every $n > N$. Next, set $\mathbf{w}_{n,m} = \mathbf{w}(2^{-n}(m+1)) - \mathbf{w}(2^{-n}m)$ and apply Taylor's Theorem to write

$$\begin{aligned} D_{n,m}(\mathbf{w}) &= \mathfrak{E}(\mathbf{w}_{n,m})_{\mathfrak{F}_{f_0}(2^{-n}m, \mathbf{w})} \Psi + \frac{1}{2} \mathfrak{E}(\mathbf{w}_{n,m})_{\mathfrak{F}_{f_0}(2^{-n}m, \mathbf{w})}^2 \Psi + E_{n,m}(\mathbf{w}) \\ &\text{where } |E_{n,m}(\mathbf{w})| \leq C |\mathbf{w}_{n,m}|^3. \end{aligned}$$

The key observation is that, because $\mathbf{w}_{n,m}$ is independent of $\mathcal{B}_{2^{-n}m}(\mathbb{R}^d)$ and because $2^{\frac{n}{2}} \mathbf{w}_{n,m}$ has distribution $\mu_{\mathbb{R}^d}$,

$$\begin{aligned} & \left| \int_A D_{n,m}(\mathbf{w}) \mu_{\mathbf{H}_1(\mathbb{R}^d)}(d\mathbf{w}) - 2^{-n} \int_A \frac{1}{2} \Delta_{\mathcal{O}(M)} \Psi(\mathfrak{F}_{f_0}(2^{-n}m, \mathbf{w})) \mu_{\mathbf{H}_1(\mathbb{R}^d)}(d\mathbf{w}) \right| \\ & \leq C 2^{-\frac{3n}{2}} \end{aligned}$$

and therefore, after summation and a passage to the limit as $n \rightarrow \infty$, we get (18).

¹⁸Itô's formula is now the bread and butter of the "quant" department of several major financial institutions. Actually, the application of what we now call *Brownian motion* to finance antedates its application to Brownian motion and goes back to the thesis, written at the turn of the century (five years before Einstein's famous paper about the kinetic theory of gases) by L. Bachelier. Bachelier was trying to model the fluctuations of prices on the Bourse. From a purely mathematical standpoint, his insights are far more penetrating than anyone else's prior to Wiener. In addition, their practical impact on all our lives is also far more penetrating. In fact, models, like that of Black & Scholes, which are the form in which Bachelier's ideas have been incarnated, constitute the basis on which modern business makes decisions about how everything from stocks and bonds to pork belly futures should be priced. The role that Itô's formula plays in all this is very much the same as the one which it plays in our considerations. Namely, Itô's formula provides the link between various stochastic quantities and differential equations of which those quantities are the solution.

¹⁹We use $[s]$ to denote the largest integer dominated by $s \in \mathbb{R}$.

To explain what (18) has to do with our assertion (*), we show that (18) specializes to a statement about $\mu_{\mathbf{H}_1(M)}$ when we apply (18) to Ψ 's which come from a function on M , in the sense that $\Psi = \psi \circ \pi$. For this purpose, first note that $\Delta_{\mathcal{O}(M)}(\psi \circ \pi)$ is not only well-defined as a function on M (i.e., is constant on each fiber), it is equal to $\Delta_M \psi$, the action of the Laplace–Beltrami operator Δ_M on ψ . Second, observe that, for each $n \in \mathbb{N}$ and $0 \leq m \leq 2^n$, $\mathfrak{F}_{f_0}(2^{-n}m, \mathbf{w}_n)$ is $\mathcal{B}_{2^{-n}m}(\mathbb{R}^d)$ -measurable. Hence, if the σ -algebra $\mathcal{B}_s(M)$ is defined over $\mathfrak{W}(M)$ in the same way as $\mathcal{B}_s(\mathbb{R}^d)$ was over $\mathfrak{W}(\mathbb{R}^d)$, then it is not hard to see that, for each $s \in [0, 1]$ and $A \in \mathcal{B}_s(M)$, $\{\mathbf{w} : \mathfrak{F}_{f_0}(\cdot, \mathbf{w}) \in A\} \in \mathcal{B}_s(\mathbb{R}^d)$. Therefore, after elementary manipulation, we conclude that (18) specializes to the statement that

$$(19) \quad \int_A \left(\psi(w(t)) - \psi(w(s)) \right) \mu_{\mathbf{H}_1(M)}(dw) = \int_A \left(\int_s^t \frac{1}{2} \Delta_M \psi(w(\tau)) d\tau \right) \mu_{\mathbf{H}_1(M)}(dw),$$

for all $0 \leq s < t \leq 1$, $A \in \mathcal{B}_s(M)$, and $\psi \in C^\infty(M; \mathbb{R})$.

To appreciate why (19) supports the claim in (*), recall that w is an integral curve of the vector field X on M means that, for each $0 \leq s < t \leq 1$ and $\psi \in C^\infty(M; \mathbb{R})$,

$$\psi(w(t)) - \psi(w(s)) - \int_s^t X\psi(w(\tau)) d\tau = 0.$$

By comparison, (19) says that, at least on the average, the analogous relationship holds between $\frac{1}{2}\Delta_M$ and paths distributed according to $\mu_{\mathbf{H}_1(M)}$. This analogy becomes even more compelling after one introduces the terminology of prediction theory. Namely, because (19) holds for every $A \in \mathcal{B}_s(M)$, it says that the *conditional expectation value* of

$$\psi(w(t)) - \psi(w(s)) - \int_s^t \frac{1}{2} \Delta_M \psi(w(\tau)) d\tau$$

under $\mu_{\mathbf{H}_1(M)}$ given $\mathcal{B}_s(M)$ is 0. In other words, based on information about $w \upharpoonright [0, s]$, the *best prediction* that one can make about the value of

$$\psi(w(t)) - \psi(w(s)) - \int_s^t \frac{1}{2} \Delta_M \psi(w(\tau)) d\tau$$

is that it will be 0. Alternatively, one can say that, under $\mu_{\mathbf{H}_1(M)}$, the quantity

$$\psi(w(t)) - \int_0^t \frac{1}{2} \Delta_M \psi(w(\tau)) d\tau$$

is *conditionally constant* in the sense that, for each $0 \leq s < t \leq 1$ its conditional expectation value given $\mathcal{B}_s(M)$ is

$$\psi(w(s)) - \int_0^s \frac{1}{2} \Delta_M \psi(w(\tau)) d\tau.$$

Because they arise in so many contexts, families of random variables which are conditionally constant (in the sense just described) have been the subject of intensive study by probabilists, who call such a family a *martingale*. Adopting this terminology, we can summarize the statements in (18) and (19) as saying that

$$(20) \quad \left(\Psi(\mathfrak{F}_{f_0}(t, \mathbf{w})) - \frac{1}{2} \int_0^t \Delta_{\mathcal{O}(M)} \Psi(\mathfrak{F}_{f_0}(\tau, \mathbf{w})) d\tau, \mathcal{B}_t(\mathbb{R}^d), \mu_{\mathbf{H}_1(\mathbb{R}^d)} \right)$$

is a martingale for every $\Psi \in C^\infty(\mathcal{O}(M); \mathbb{R})$

and

$$(21) \quad \left(\psi(w(t)) - \int_0^t \frac{1}{2} \Delta_M \psi(w(\tau)) d\tau, \mathcal{B}_t(M), \mu_{\mathbf{H}_1(M)} \right)$$

is a martingale for every $\psi \in C^\infty(M; \mathbb{R})$.

For future reference, it will be convenient to have remarked that (20) and (21) self-improve to cover functions which depend on time as well. In fact, they lead, respectively, to

$$(20') \quad \left(\Psi(t, \mathfrak{F}_{f_0}(t, \mathbf{w})) - \frac{1}{2} \int_0^t \sum_1^d \left(\frac{\partial}{\partial \tau} + \frac{1}{2} \Delta_{\mathcal{O}(M)} \right) \Psi(\tau, \mathfrak{F}_{f_0}(\tau, \mathbf{w})) d\tau, \mathcal{B}_t(\mathbb{R}^d), \mu_{\mathbf{H}_1(\mathbb{R}^d)} \right)$$

is a martingale for every $\Psi \in C^{1,2}([0, 1] \times \mathcal{O}(M); \mathbb{R})$

and

$$(21') \quad \left(\psi(t, w(t)) - \int_0^t \left(\frac{\partial}{\partial \tau} + \frac{1}{2} \Delta_M \right) \psi(\tau, w(\tau)) d\tau, \mathcal{B}_t(M), \mu_{\mathbf{H}_1(M)} \right)$$

is a martingale for every $\psi \in C^{1,2}([0, 1] \times M; \mathbb{R})$.

The immediate importance of these considerations for us is that (21) characterizes $\mu_{\mathbf{H}_1(M)}$. That is, $\mu_{\mathbf{H}_1(M)}$ is the only probability measure on $\mathfrak{W}(M)$ with the property described in (21). To see this, we introduce the *heat flow semigroup* $\{P_t : t \in [0, 1]\}$ generated by $\frac{1}{2} \Delta_M$: if $\psi_0 \in C^2(M; \mathbb{R})$ and $u(t, \cdot) = P_t \psi_0$, then u solves the heat equation

$$(22) \quad \begin{aligned} \frac{\partial u}{\partial t}(t, x) &= \frac{1}{2} [\Delta_M u(t, \cdot)](x) & \text{for } (t, x) \in (0, \infty) \times M \\ u(0, x) &= \psi_0(x) & \text{for } x \in M \end{aligned}$$

with initial data ψ_0 . Now suppose that μ is any probability measure with the property described in (21), and therefore (21') as well. If, for some $T \in (0, 1]$, we apply (21') with $\psi(t, \cdot) = P_{T-t} \psi_0$, then we arrive at

$$(23) \quad \int_A \psi_0(w(T)) d\mu = \int_A [P_{T-t} \psi_0](w(t)) d\mu \quad \text{for all } t \in [0, T) \text{ and } A \in \mathcal{B}_t(M).$$

Proceeding inductively and using (23) at each stage, one finds that, for all $n \geq 1$, $\psi_1, \dots, \psi_n \in C^2(M; \mathbb{R})$, and $0 = t_0 < t_1 < \dots < t_n \leq 1$:

$$\int_{\mathfrak{W}} \psi_1(w(t_1)) \cdots \psi_n(w(t_n)) \mu(dw) = \left[P_{\tau_n} \psi_n \left(P_{\tau_{n-1}} \psi_{n-1} \left(\cdots \left(P_{\tau_1} \psi_1 \right) \cdots \right) \right) \right](x_0),$$

where $\tau_m = t_m - t_{m-1}$. Thus (21), or its equivalent (21'), determines

$$\mu \left(\{w : w(t_1) \in \Gamma_1, \dots, w(t_n) \in \Gamma_n\} \right)$$

for all $n \geq 1$, $0 < t_1 < \dots < t_n \leq 1$, and Borel subsets $\Gamma_1, \dots, \Gamma_n$ of M . Since, at least in theory, the probability of an arbitrary Borel subset of $\mathfrak{W}(M)$ can be

determined from such information, we now know that there is at most one Borel probability measure μ for which (21) holds.

More generally, (21) and (21') act as a bridge between $\mu_{\mathbf{H}_1(M)}$ and equations which, like the heat equation, involve Δ_M . Unfortunately for probabilists, most of the traffic on this bridge flows from the analytic to the probabilistic side. Indeed, the existence and uniqueness of a μ satisfying (21) constitute little more than a fancy, albeit often convenient, way in which to catalogue information which is already implicit in properties of solutions to the heat equation (23). In fact, if all that one wants is to produce such a μ , there is a far more direct route than the circuitous one which we have taken. For example, as soon as one shows that a heat flow semigroup exists and possesses a few elementary properties, one can follow the route mapped out by A.N. Kolmogorov (cf. [D] or [SV]) and one will arrive quickly at both the existence and uniqueness statements. Thus, it is not until one takes advantage of particular, probabilistically inspired representations of the measure μ that one can hope that probability theory will have something really new to contribute. Not surprisingly, the equation

$$(24) \quad \mu_{\mathbf{H}_1(M)} = \mu_{\mathbf{H}_1(\mathbb{R}^d)} \circ (\pi \circ \mathfrak{F}_\epsilon)^{-1}$$

is a wonderful example of just such a representation: it is built by taking an intrinsic, geometric construction for smooth paths and using probabilistic machinery to extend that construction to paths which are not smooth. Some examples of its virtues are the subject of the next section.

§4: SOME APPLICATIONS

The basic connection between probability theory and partial differential equations resides in formulae like

$$(25) \quad [P_T \psi](x_0) = \int_{\mathfrak{W}(M)} \psi \circ \pi(\mathfrak{F}_{f_0}(T, \mathbf{w})) \mu_{\mathbf{H}_1(\mathbb{R}^d)}(d\mathbf{w}).$$

However, in spite of the intuitively appealing picture it provides of heat flowing along random paths, there remains the question whether useful information about solutions to partial differential equations can be gleaned from (25). The answer is that (25) does contain interesting information, even if it divulges it with some reluctance.

As our first example, we will show how (25) leads to a particularly attractive formulation of *Bochner's identity*. For this purpose, we define $\nabla_{\mathcal{O}(M)} \Psi : \mathcal{O}(M) \rightarrow \mathbb{R}^d$ for $\Psi \in C^1(\mathcal{O}(M); \mathbb{R})$ so that (cf. (12) and try to avoid any confusion arising from our earlier use of ∇ for covariant differentiation)

$$[\nabla_{\mathcal{O}(M)} \Psi](f) = \sum_1^d (\mathbf{e}_k \Psi)(f) \mathbf{e}_k,$$

and set

$$\nabla_M \psi = \nabla(\psi \circ \pi) \quad \text{when } \psi \in C^1(M; \mathbb{R}).$$

(Notice that, although $\Delta_{\mathcal{O}(M)}(\psi \circ \pi)$ is well-defined on M when $\psi \in C^2(M; \mathbb{R})$, $\nabla_M \psi$ is not.) Finally, define the Ricci curvature matrix $\text{Ric} : \mathcal{O}(M) \rightarrow \mathbb{R}^d \otimes \mathbb{R}^d$ so that

$$(\mathbf{v}, \text{Ric}(f)\mathbf{v}')_{\mathbb{R}^d} = \sum_{k=1}^d \left(\text{Riem}(f\mathbf{e}_k, f\mathbf{v})\mathbf{e}_k, \mathbf{v}' \right)_{T_{\pi(f)}(M)}, \quad \mathbf{v}, \mathbf{v}' \in \mathbb{R}^d,$$

where $\text{Riem} : T_x(M)^2 \rightarrow \text{Hom}(T_x(M); T_x(M))$ is the Riemann curvature (cf. [doC]).

In the preceding notation, Bochner's identity becomes the commutation relation:

$$(26) \quad [\nabla_{\mathcal{O}(M)}, \Delta_{\mathcal{O}(M)}](\psi \circ \pi) = -\text{Ric} \nabla_M \psi \quad \text{for } \psi \in C^3(M; \mathbb{R}).$$

After combining (22) with (26), we get

$$(27) \quad \frac{d}{dt} \nabla P_t \psi = \frac{1}{2} \nabla_M \Delta_M P_t \psi = \frac{1}{2} \Delta_M \nabla P_t \psi - \frac{1}{2} \text{Ric} \nabla_M P_t \psi.$$

In other words, $\nabla P_t \psi$ is a solution to a d -dimensional system of heat equations in which *coupling* occurs only in the zeroth order terms, and this is precisely the sort of equation to which the following mild extension²⁰ of the famous *Feynman–Kac formula* applies.

Lemma F–K. *Suppose that $\mathbf{M} : [0, 1] \times \mathfrak{W}(\mathbb{R}^d) \rightarrow \mathbb{R}^d$ and $\mathbf{A} : [0, 1] \times \mathfrak{W}(\mathbb{R}^d) \rightarrow \text{Hom}(\mathbb{R}^d; \mathbb{R}^d)$ are bounded, measurable functions which are adapted.²¹ Further, assume that*

$$\mathbf{M}(\cdot, \mathbf{w}) \in C([0, 1]; \mathbb{R}^d) \quad \text{and} \quad \mathbf{A}(\cdot, \mathbf{w}) \in C^1([0, 1]; \text{Hom}(\mathbb{R}^d; \mathbb{R}^d))$$

for $\mu_{\mathbf{H}_1(\mathbb{R}^d)}$ -almost every $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d)$ and that the uniform norm of $\dot{\mathbf{A}}(\cdot, \mathbf{w})\|_{\text{op}}$ is in $L^\infty(\mu_{\mathbf{H}_1(\mathbb{R}^d)})$. If $(\mathbf{M}(t, \mathbf{w}), \mathcal{B}_t(\mathbb{R}^d), \mu_{\mathbf{H}_1(\mathbb{R}^d)})$ is a martingale, then so is

$$\left(\mathbf{A}(t, \mathbf{w})\mathbf{M}(t, \mathbf{w}) - \int_0^t \dot{\mathbf{A}}(\tau, \mathbf{w})\mathbf{M}(\tau, \mathbf{w}) d\tau, \mathcal{B}_t(\mathbb{R}^d), \mu_{\mathbf{H}_1(\mathbb{R}^d)} \right).$$

To apply Lemma F–K here, set $\mathbf{U}(t, f) = [\nabla P_{1-t}\psi](f)$, note that, by (27),

$$\frac{\partial \mathbf{U}}{\partial \tau} + \frac{1}{2} \Delta_{\mathcal{O}(M)} \mathbf{U} = \frac{1}{2} \text{Ric} \mathbf{U},$$

and use (20') to conclude that

$$\left(\mathbf{M}(t, \mathbf{w}) \equiv \mathbf{U}(t, \mathfrak{F}_{f_0}(t, \mathbf{w})) - \frac{1}{2} \int_0^t \text{Ric}(\mathfrak{F}_{f_0}(\tau, \mathbf{w})) \mathbf{U}(\mathfrak{F}_{f_0}(\tau, \mathbf{w})) d\tau, \mathcal{B}_t(\mathbb{R}^d), \mu_{\mathbf{H}_1(\mathbb{R}^d)} \right)$$

²⁰Actually, one can think of it equally well as a special case of Itô's formula. For an elementary proof, see Theorem 7.1.19 in [S].

²¹A function F on $[0, 1] \times \mathfrak{W}(\mathbb{R}^d)$ is said to be adapted if, for each $t \in [0, 1]$, $F(t, \cdot)$ is $\mathcal{B}_t(\mathbb{R}^d)$ -measurable.

is a martingale. Next, define the *Ricci flow* $\mathbf{A} : [0, 1] \times \mathfrak{W}(\mathbb{R}^d) \longrightarrow \text{Hom}(\mathbb{R}^d; \mathbb{R}^d)$ so that

$$(28) \quad \dot{\mathbf{A}}(t, \mathbf{w}) + \frac{1}{2} \mathbf{A}(t, \mathbf{w}) \text{Ric}(\mathfrak{F}_{f_0}(t, \mathbf{w})) = \mathbf{0} \quad \text{and} \quad \mathbf{A}(0, \mathbf{w}) = \mathbf{I}.$$

Then, because

$$\begin{aligned} & \mathbf{A}(t, \mathbf{w}) \mathbf{U}(t, \mathfrak{F}_{f_0}(t, \mathbf{w})) \\ &= \mathbf{A}(t, \mathbf{w}) \mathbf{M}(t, \mathbf{w}) + \mathbf{A}(t, \mathbf{w}) \int_0^t \frac{1}{2} \text{Ric}(\mathfrak{F}_{f_0}(\tau, \mathbf{w})) \mathbf{U}(\tau, \mathfrak{F}_{f_0}(\tau, \mathbf{w})) d\tau \\ &= \mathbf{A}(t, \mathbf{w}) \mathbf{M}(t, \mathbf{w}) - \int_0^t \dot{\mathbf{A}}(\tau, \mathbf{w}) \mathbf{M}(\tau, \mathbf{w}) d\tau, \end{aligned}$$

we know that

$$(29) \quad \left(\mathbf{A}(t, \mathbf{w}) [\nabla_M P_{1-t} \psi](\mathfrak{F}_{f_0}(t, \mathbf{w})), \mathcal{B}_t(\mathbb{R}^d), \mu_{\mathbf{H}_1(\mathbb{R}^d)} \right)$$

is a martingale for each $\psi \in C^3(M; \mathbb{R})$.

In particular,

$$t \in [0, 1] \longmapsto \int_{\mathfrak{W}(\mathbb{R}^d)} \mathbf{A}(t, \mathbf{w}) [\nabla_M P_{1-t} \psi](\mathfrak{F}_{f_0}(t, \mathbf{w})) \mu_{\mathbf{H}_1(\mathbb{R}^d)}(d\mathbf{w}) \in \mathbb{R}$$

is constant, and so

$$(30) \quad [\nabla_M P_1 \psi](f_0) = \int_{\mathfrak{W}(\mathbb{R}^d)} \mathbf{A}(1, \mathbf{w}) [\nabla_M \psi](\mathfrak{F}_{f_0}(1, \mathbf{w})) \mu_{\mathbf{H}_1(\mathbb{R}^d)}(d\mathbf{w}).$$

Clearly, (30) can be used to gain estimates on the gradient of the heat flow semigroup in terms of the gradient of the initial data. To wit, from its definition in (28), it is an easy matter to obtain

$$(31) \quad \|\mathbf{A}(1, \mathbf{w})\|_{\text{op}} \leq e^{-\kappa} \quad \text{where} \quad \kappa \equiv \inf_{\substack{f \in \mathcal{O}(M) \\ \xi \in \mathbb{S}^{d-1}}} (\text{Ric}(f)\xi, \xi)_{\mathbb{R}^d},$$

which, together with (30), leads immediately to the estimate

$$(32) \quad \begin{aligned} |[\nabla_M P_1 \psi](f_0)| &\leq e^{-\kappa} \int_{\mathfrak{W}(\mathbb{R}^d)} |[\nabla_M \psi](\mathfrak{F}_{f_0}(1, \mathbf{w}))| \mu_{\mathbf{H}_1(\mathbb{R}^d)}(d\mathbf{w}) \\ &= e^{-\kappa} [P_1(|\nabla_M \psi|)](f_0), \end{aligned}$$

which, at least when $\kappa > 0$, has interesting implications about the ergodic properties of $\{P_t : t > 0\}$ (cf. [BE]). On the other hand, (32) is a poor advertisement for (30): it can be proved far more easily, via the weak maximum principle, directly from (27) (cf. [LY]). Thus, before one can gain a full appreciation of (30), one has to go further. The *further* here entails the removal of the derivatives from the ψ in the

integrand on the right-hand side of (30). Obviously, if this can be done at all, one would expect to do it by some sort of *integration by parts*.

Actually, when one considers the situation at hand, integration by parts is a quite reasonable suggestion. Namely, the measure $\mu_{\mathbf{H}_1(\mathbb{R}^d)}$ is a *smooth* (i.e., behaves well under translation in the directions of $\mathbf{H}_1(\mathbb{R}^d)$ in the sense explained at the end of §1) on $\mathfrak{W}(\mathbb{R}^d)$ and the map $\pi \circ \mathfrak{F}_{f_0}(1, \cdot)$ is *non-degenerate* in the sense that a (formal) computation indicates that its Jacobian $d\pi D\mathfrak{F}_{f_0}(1, \mathbf{w})$ maps \mathbf{H} onto $T_{\mathfrak{F}_{f_0}(1, \mathbf{w})}(M)$. Hence, one ought to be able to use $\mathfrak{F}_{f_0}(1, \cdot)$ to first *pull back* the vector fields on M to vector fields on $\mathfrak{W}(\mathbb{R}^d)$, then integrate by parts using $\mu_{\mathbf{H}_1(\mathbb{R}^d)}$, and finally *push forward* the resulting expression. These are the principles which underlie a collection of techniques known, somewhat loosely, as *Malliavin's calculus*,²² but, unfortunately, rigorous application of them entails quite a lot of rather tedious work. Thus, we are lucky that, for the particular case under consideration, there is an alternate way to proceed. Namely, let $\mathbf{B} : [0, 1] \times \mathfrak{W}(\mathbb{R}^d) \rightarrow \text{Hom}(\mathbb{R}^d, \mathbb{R}^d)$ be any bounded, adapted, measurable map with the properties that $\mathbf{B}(\cdot, \mathbf{w})$ is twice continuously differentiable for $\mu_{\mathbf{H}_1(\mathbb{R}^d)}$ -almost every $\mathbf{w} \in \mathfrak{W}(\mathbb{R}^d)$, $\sup_{t \in [0, 1]} \|\dot{\mathbf{B}}(t, \cdot)\|_{\text{op}} \in L^\infty(\mu_{\mathbf{H}_1(\mathbb{R}^d)})$, $\mathbf{B}(0, \mathbf{w}) = \mathbf{0}$, and $\mathbf{B}(1, \mathbf{w}) = \mathbf{I}$. Next apply Lemma F–K to see (cf. (29)) that

$$\begin{aligned} & \mathbf{B}(t, \mathbf{w})\mathbf{A}(t, \mathbf{w})[\nabla_M P_{1-t}\psi](\mathfrak{F}_{f_0}(t, \mathbf{w})) \\ & \quad - \int_0^t \dot{\mathbf{B}}(\tau, \mathbf{w})\mathbf{A}(\tau, \mathbf{w})[\nabla_M P_{1-\tau}\psi](\mathfrak{F}_{f_0}(\tau, \mathbf{w})) d\tau \end{aligned}$$

is a $\mu_{\mathbf{H}_1(\mathbb{R}^d)}$ -martingale. In particular, by comparing expectation values taken at $t = 0$ and $t = 1$ and using (30), we get the equation

$$\begin{aligned} (33) \quad & [\nabla_M P_1\psi](f_0) \\ & = \int_{\mathfrak{W}(\mathbb{R}^d)} \left(\int_0^1 \dot{\mathbf{B}}(\tau, \mathbf{w})\mathbf{A}(\tau, \mathbf{w})[\nabla_M P_{1-\tau}\psi](\mathfrak{F}_{f_0}(\tau, \mathbf{w})) d\tau \right) \mu_{\mathbf{H}_1(\mathbb{R}^d)}(d\mathbf{w}). \end{aligned}$$

Of course, (33) still has derivatives on the right-hand side and therefore may not appear to represent much of an improvement over (30). On the other hand, after staring at (33) for a second, one realizes that the derivatives on the right-hand side of (33) are getting integrated with respect to time. Thus, there is a chance that one will be able to remove the derivatives by integrating by parts in time (as opposed to $\mathfrak{W}(\mathbb{R}^d)$). In fact, that is precisely what can be done, but in order to carry out the details (see, [SZ]) one must know a little stochastic integration theory as well as the full statement of Itô's formula. For this reason, we will simply write down the answer:

$$\begin{aligned} (34) \quad & [\nabla_M P_1\psi](f_0) \\ & = \int_{\mathfrak{W}(\mathbb{R}^d)} \left(\int_0^1 \Xi_{\mathbf{B}}(t, \mathbf{w})(\mathbf{w}(1) - \mathbf{w}(t)) dt \right) \psi \circ \pi(\mathfrak{F}_{f_0}(1, \mathbf{w})) \mu_{\mathbf{H}_1(\mathbb{R}^d)}(d\mathbf{w}) \\ & \quad \text{where } \Xi_{\mathbf{B}}(t, \mathbf{w}) \equiv \dot{\mathbf{B}}(t, \mathbf{w})\text{Ric}(\mathfrak{F}_{f_0}(t, \mathbf{w}))\mathbf{A}(t, \mathbf{w}) + \ddot{\mathbf{B}}(t, \mathbf{w})\mathbf{A}(t, \mathbf{w}). \end{aligned}$$

²²The idea of integrating by parts in Wiener space antedates P. Malliavin. Most notably, first R.H. Cameron and then his student M.D. Donsker developed integration by parts formulae for Wiener measure, and their formulae were re-discovered and sharpened by constructive quantum field theorists. However, it was not until Malliavin got into the game that anyone had the courage to apply these formulae to functions as complicated as $\mathfrak{F}_{f_0}(1, \cdot)$.

In particular, by taking $\mathbf{B}(t, \mathbf{w}) = t\mathbf{I}$, one recovers the integrated form of a beautiful formula due to J.-M. Bismut (see [B]).

Equation (34) is the starting place for a variety of different calculations. To give a sense of what these are, let $p_T(x, y)$ denote the fundamental solution to the heat equation (22). That is,

$$[P_T\psi](x) = \int_M \psi(y)p_T(x, y) \lambda_M(dy),$$

where λ_M is the Riemann measure on M . As is well-known (in fact derivable from the sort of considerations discussed here), $p_T(x, y)$ is a positive, smooth function of $(T, x, y) \in (0, 1] \times M \times M$, and, for each $T \in (0, 1]$, it is symmetric with respect to $(x, y) \in M^2$. What (34) does is provide, for each choice of \mathbf{B} , a representation of the logarithmic gradient $[\nabla_M \log p_1(\cdot, y)](f_0)$ as the conditional expectation value under $\mu_{\mathbf{H}_1(\mathbb{R}^d)}$ of

$$\int_0^1 \Xi_{\mathbf{B}}(t, \mathbf{w})(\mathbf{w}(1) - \mathbf{w}(t)) dt \quad \text{given } \pi \circ (\mathfrak{F}_{f_0}(1, \mathbf{w})) = y.$$

Indeed, it is not until (34) (with $\mathbf{B}(t, \mathbf{w}) = t\mathbf{I}$) is interpreted this way that one has *Bismut's formula* in all its glory. More generally, after rescaling time, one gets analogous expressions for $[\nabla_M \log p_T(\cdot, y)](f_0)$ with general $T \in (0, \infty)$, and from these one can derive effective estimates on the gradient of $p_T(\cdot, y)$ in terms of $p_T(\cdot, y)$ itself (cf. [SZ]). Perhaps more interesting are the asymptotic results which one can obtain by combining Bismut's formula with the theory of large deviations. For instance, one can show (cf. [MS]) that if x_0 is outside the cut locus (cf. [doC]) of y , then

$$\lim_{T \searrow 0} 2T [\nabla_M \log p_T(\cdot, y)](f_0) = -\theta(f_0, y)$$

where $f_0\theta(f_0, y)$ is the initial velocity of the minimal geodesic on $[0, 1]$ running from x_0 to y . In particular,

$$x_0 \text{ outside the cut-locus of } y \implies \lim_{T \searrow 0} 2T \left| [\nabla_M \log p_T(\cdot, y)](f_0) \right| = \text{dist}(x_0, y).$$

By contrast, if x_0 lies in the cut locus of y , then, under additional technical conditions, one can show that

$$\lim_{T \searrow 0} 2T \left| [\nabla_M \log p_T(\cdot, y)](f_0) \right| < \text{dist}(x_0, y).$$

Other, and more intriguing, results can be proved from higher derivative analogs of Bismut's formula. (However, to develop such formulae, it seems that one has to invoke the sort of reasoning alluded to earlier as Malliavin's calculus.) Moreover, the applications of these ideas is not restricted to the generation of estimates on $p_T(\cdot, y)$. For instance, throughout the last decade, Malliavin [FG] and his school have been applying them to the analysis of diffusions on loop spaces, and, more recently, B. Driver ([Dr1] and [Dr2]) has successfully constructed the *number operator* on loop space. (See also [H], [ES1], and [ES2].)

REFERENCES

- [BE] Bakry, D. & Emery, M., *Hypercontractivité de semi-groupes des diffusions*, C.R. Acad. Sci. Paris t. 299 Serie 1 (1984), 775–777. MR **86f**:60097
- [BP] Barlow, M. & Perkins, E., *Brownian motion on the Sierpinski gasket*, Prob. Theory & Related Fields # 79 (1988), 543–623. MR **89g**:60241
- [BC] Bishop, R. & Crittenden, R., *Geometry of Manifolds*, Pure & Appl. Math. Series # 15, Academic Press, N.Y., 1964. MR **29**:6401
- [B] Bismut, J.M., *Large Deviations and the Malliavin Calculus*, Progress in Math. #45, Birkhäuser, Cambridge USA, 1984. MR **86f**:58150
- [doC] do Carmo, M., *Riemannian Geometry*, Birkhäuser, 1992. MR **92i**:53001
- [Dr1] Driver, B., *A Cameron-Martin type quasi-invariance theorem for Brownian motion on compact Riemannian manifolds*, J. Funct. Anal. **110** (1992), 272–376. MR **94a**:58214
- [Dr2] ———, *A Cameron-Martin type quasi-invariance theorem for pinned Brownian motion on compact Riemannian manifolds*, Trans. Amer. Math. Soc. **342** (1994), 375–395. MR **94e**:60072
- [D] Dynkin, E., *Markov Processes, I & II*, Grundlehren # 121 & 122, Springer-Verlag, 1965. MR **33**:1887
- [EE] Eells, J. & Elworthy, D., *Stochastic dynamical systems*, Control Theory and Topics in Functional Analysis, Vol. III, Intern. Atomic Energy Agency, Vienna, 1979, pp. 179–185. MR **58**:24390
- [ES1] Enchev, O. & Stroock, D., *Towards a Riemannian geometry on the path space over a Riemannian manifold* (to appear in J. Funct. Anal.).
- [ES2] ———, *Pinned Brownian and its perturbations* (to appear in Adv. in Math.).
- [FG] Fang, S. & Malliavin, P., *Stochastic analysis on the path space of a Riemannian manifold*, J. Funct. Anal. **118** #1 (1993), 249–274. MR **94i**:58209
- [G] Gangolli, R., *On the construction of certain diffusions on a differentiable manifold*, Zeit. Wahr. **2** (1964), 209–219. MR **29**:2870
- [GJ] Glimm, J. & Jaffe, A., *Quantum Physics: a Functional Integration Point of View*, 2nd Edition, Springer-Verlag, 1987. MR **89k**:81001
- [H] Hsu, E., *Logarithmic Sobolev inequalities on path space*, C. R. Acad. Sci. Paris Sér. I Math. **320** (1995), 1009–1012. MR **1**:328 728
- [It] Itô, K., *On stochastic differential equations on a differentiable manifold I*, Nagoya Math. J. **1** (1950), 35–47. MR **12**:425g
- [I2] ———, *The Brownian motion and tensor fields on Riemannian manifold*, Proc. Intern. Congr. Math., Stockholm (1963), 536–539. MR **31**:772
- [K] Kusuoka, S., *A diffusion process on a fractal*, Probability Methods in Mathematical Physics (K. Itô & N. Ikeda, eds.), Proceedings of 1985 Taniguchi Conf., North Holland, 1987, pp. 251–274. MR **89e**:60149
- [LY] Li, P. & Yau, S.T., *On the parabolic kernel of the Schrödinger operator*, Acta Math. **156** (1986), 153–201. MR **87f**:58156
- [M] Malliavin, P., *Formule de la moyenne, calcul des perturbations, et théorème d’annulation pour les formes harmoniques*, J. Funct. Anal. **17** (1974), 274–291. MR **52**:6791
- [MS] Malliavin, P. and Stroock, D., *Short time behavior of the heat kernel and its logarithmic derivatives*, submitted to J. Differential Geom.
- [Sim] Simon, B., *Functional Integration and Quantum Physics*, Pure & Appl. Math. Series # 86, Academic Press, 1979. MR **84m**:81066
- [S] Stroock, D., *Probability Theory, An Analytic View*, Cambridge Univ. Press, Cambridge U.K. and NYC, USA, 1994. MR **95f**:60003
- [ST] Stroock, D. & Taniguchi, S., *Diffusions as integral curves, or Stratonovich without Itô*, The Dynkin Festschrift (M. Freidlin, ed.), Progress in Probability # 34, Birkhäuser, 1994, pp. 333–369. MR **1**:311 729
- [SV] Stroock, D. & Varadhan, S., *Multidimensional Diffusion Processes*, Grundlehren # 233, Springer-Verlag, 1979. MR **81f**:60108

- [SZ] Stroock, D. & Zeitouni, O., *Variations on a theme by Bismut* (to appear).
- [Sym] Symansik, K., *Euclidean quantum field theory, I. equations for a scalar model*, J. Math. Phys. **7** (1966), 510–525.
- [W] Wiener, N., *Differential Space*, J. Math. Physics **2** (1923), 131–174.

MIT 2-272, 77 MASSACHUSETTS AVE., CAMBRIDGE, MASSACHUSETTS 02139
E-mail address: `dws@math.mit.edu`