COLLOQUIUM MATHEMATICUM

VOL. 118 2010 NO. 1

SOME THOUGHTS ABOUT SEGAL'S ERGODIC THEOREM

BY

DANIEL W. STROOCK (Cambridge, MA)

Abstract. Over fifty years ago, Irving Segal proved a theorem which leads to a characterization of those orthogonal transformations on a Hilbert space which induce ergodic transformations. Because Segal did not present his result in a way which made it readily accessible to specialists in ergodic theory, it was difficult for them to appreciate what he had done. The purpose of this note is to state and prove Segal's result in a way which, I hope, will win it the recognition which it deserves.

1. Background. Unless a separable, real Hilbert space H is finite-dimensional, there is no *standard Gauss measure* for H. That is, there is no Borel probability measure μ on H whose Fourier transform $\hat{\mu}$ is given by

$$\hat{\mu}(h) \equiv \int_H e^{\sqrt{-1}(h,g)_H}\,\mu(dg) = e^{-\|h\|_H^2/2}, \qquad h \in H.$$

Indeed, if H is infinite-dimensional and such a μ were to exist, then, for any orthonormal basis $\{e_k : k \geq 0\}$ in H, the random variables $g \in H \mapsto X_m(g) = (e_k, g)_H$ would be independent, Gaussian random variables with mean 0 and variance 1. In particular, the strong law of large numbers would imply that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{m=0}^{n-1} X_m^2 = 1 \qquad \mu\text{-almost surely},$$

and we would have $\|g\|_H^2 = \sum_{m=0}^{\infty} X_m(g)^2 = \infty$ for μ-almost every $g \in H$. In other words, H is just too small to accommodate μ.

In order to get around the problem described above, L. Gross [1] introduced the notion of an abstract Wiener space. To describe Gross's idea, suppose that Θ is a separable, real Banach space in which H is continuously embedded as a dense subspace. Then, for each $\lambda \in \Theta^*$, there is a unique $h_\lambda \in H$ such that (1) $(g, h_\lambda)_H = \langle g, \lambda\rangle$ for all $g \in H$. Indeed, $\lambda \rightsquigarrow h_\lambda$ is continuous from the weak* topology on Θ^* into the weak topology on H, and $\{h_\lambda : \lambda \in \Theta^*\}$ is dense in H. What Gross showed is that for each H there

2010 *Mathematics Subject Classification*: 60G15, 60G17.
Key words and phrases: abstract Wiener space, ergodic theory.

(1) We use $\langle \theta, \lambda\rangle$ to denote the action of $\lambda \in \Theta^*$ on $\theta \in \Theta$.

exists such a Banach space Θ on which there is a unique Borel probability measure $\mathcal{W}_H$ with the property that

$$\widehat{\mathcal{W}_H}(\lambda) \equiv \int_{\Theta} e^{\sqrt{-1}\,\langle \theta,\lambda\rangle}\,\mathcal{W}_H(d\theta) = e^{-\|h_\lambda\|_H^2/2}, \qquad \lambda \in \Theta^*. \tag{1}$$

Further, in recognition of the fact that N. Wiener was the first to carry out this sort of construction, he called the triple $(H, \Theta, \mathcal{W}_H)$ an *abstract Wiener space.*

There are various ways to prove Gross's theorem (cf. [4]). For our purposes here, the best way is to base the proof on the following theorem, a derivation of which can be found in [4]. In its statement, γ is the standard Gaussian measure on $\mathbb{R}$, that is, $\gamma(dx) = (2\pi)^{-1/2}e^{-x^2/2}dx$.

THEOREM 1. *Suppose that H is continuously embedded as a dense subspace of Θ. Then there exists a $\mathcal{W}_H$ on Θ for which $(H, \Theta, \mathcal{W}_H)$ is an abstract Wiener space if and only if, for each orthonormal basis $\{e_k : k \geq 0\}$, the series*

$$\sum_{k=0}^{\infty} x_k e_k \quad \textit{converges in } \Theta$$

for $\gamma^{\mathbb{N}}$-almost every $\mathbf{x} \in \mathbb{R}^{\mathbb{N}}$, in which case $\mathcal{W}_H$ is the $\gamma^{\mathbb{N}}$-distribution of the series.

Knowing Theorem 1, it is an easy matter to give a crude proof of Gross's theorem. Namely, given an orthonormal basis $\{e_k : k \geq 0\}$ in H, take Θ to be the completion of H with respect to the Hilbert norm

$$\|g\|_\Theta = \sqrt{\sum_{m=0}^{\infty} (m+1)^{-2}(g, e_m)_H^2},$$

and note that

$$\sum_{m=0}^{\infty} \frac{x_m^2}{(m+1)^2} < \infty \qquad \gamma^{\mathbb{N}}\text{-almost surely.}$$

Of course, as Wiener showed, in particular cases there are far better choices of Θ. Indeed, if Wiener had made this choice, he would have ended up with Brownian paths which are no better than locally square integrable. To recover Wiener's famous result that Brownian paths are continuous, one needs to use more sophisticated reasoning (cf. [4]).

Given an abstract Wiener space $(H, \Theta, \mathcal{W}_H)$, there is an isometry, known as the *Paley–Wiener map*, $h \in H \mapsto \mathcal{I}(h) \in L^2(\mathcal{W}_H; \mathbb{R})$ such that $\mathcal{I}(h_\lambda)(\theta) = \langle \theta, \lambda\rangle$ for all $\lambda \in \Theta^*$. Indeed, for each $\lambda \in \Theta^*$, $\theta \rightsquigarrow \langle\theta, \lambda\rangle$ under $\mathcal{W}_H$ is a Gaussian random variable with mean 0 and variance $\|h_\lambda\|_H^2$. Thus, since $\{h_\lambda : \lambda \in \Theta^*\}$ is dense in H, the existence of $\mathcal{I}$ is obvious. Furthermore, since limits of Gaussian random variables are again Gaussian, for all $h \in H$, $\mathcal{I}(h)$ under $\mathcal{W}_H$ is Gaussian with mean 0 and variance $\|h\|_H^2$.

From our point of view, the importance of the Paley–Wiener map is that $\mathcal{I}(h)$ provides an extension of the inner product $(\cdot, h)_H$ to Θ. With this in mind, we can understand in what sense an abstract Wiener space represents a resolution of the problem posed at the outset. Namely,

$$\mathbb{E}^{\mathcal{W}_H}\big[e^{\sqrt{-1}\,\mathcal{I}(g)}\big] = e^{-\|g\|_H^2/2}, \qquad g \in H,$$

which, if $\mathcal{I}(g)$ were really $(\cdot, g)_H$, is exactly what we wanted.

In the same sense, the Paley–Wiener map explains the origin of the following corollary to Theorem 1. Again, a proof is given in [4].

Corollary 2. *Let $(H, \Theta, \mathcal{W}_H)$ be an abstract Wiener space. Then for each orthogonal transformation $\mathcal{O}$ in H, there is a $\mathcal{W}_H$-almost surely unique, Borel measurable map $T_{\mathcal{O}} : \Theta \to \Theta$ such that $\mathcal{I}(h) \circ T_{\mathcal{O}} = \mathcal{I}(\mathcal{O}^\top h)$ $\mathcal{W}_H$-almost surely for each $h \in H$. Moreover, $T_{\mathcal{O}}$ is $\mathcal{W}_H$-measure preserving.*

Remark. Say that $D \subseteq \Theta^*$ is *determining* if $\theta = \theta'$ whenever $\langle \theta, \lambda \rangle = \langle \theta', \lambda \rangle$ for all $\lambda \in D$. Next, suppose that $\mathcal{O}$ is an orthogonal transformation on H and that $F : \Theta \to \Theta$ has the properties that $F \restriction H = \mathcal{O}$ and that $\theta \rightsquigarrow \langle F(\theta), \lambda \rangle$ is continuous for all λ's from a determining set D. Then $T_{\mathcal{O}}\theta = F(\theta)$ for $\mathcal{W}_H$-almost every $\theta \in \Theta$. In particular, if $\mathcal{O}$ admits a continuous extension to Θ, then $T_{\mathcal{O}}$ can be taken to be that extension.

We can now formulate the problem to which the rest of this article is dedicated. Given an abstract Wiener space $(H, \Theta, \mathcal{W}_H)$ and an orthogonal transformation $\mathcal{O}$ on H, it is natural to ask when, if ever, the associated, $\mathcal{W}_H$-measure preserving map $T_{\mathcal{O}}$ is ergodic. Clearly, if H is finite-dimensional, in which case $H = \Theta$, the answer is *never*, since every function f on H which depends only on $\|h\|_H$ is $T_{\mathcal{O}}$-invariant. More generally, if $\mathcal{O}$ admits a non-trivial, finite-dimensional subspace L and if $\{e_k : 1 \le k \le l\}$ is an orthonormal basis for L, then the same argument applied to a function of $\sum_{k=1}^{l} \mathcal{I}(e_k)^2$ shows that $T_{\mathcal{O}}$ cannot be ergodic. On the other hand, reasoning as we did when we showed that there is no standard Gauss measure on an infinite-dimensional H, one might suspect that $T_{\mathcal{O}}$ is ergodic if $\mathcal{O}$ admits no non-trivial, finite-dimensional subspace. In fact, I. E. Segal [3] proved such a result in a general, abstract setting. However, because Segal strove for maximal generality, both the statement of his result as well as his reasoning look a little obscure to probabilists. As a consequence, his result is not well known. The central goal of this note is to translate Segal's ideas into the language of abstract Wiener space, where, I hope, they will be more palatable to a wider audience. If nothing else, I hope to convince my readers that Segal's Theorem deserves recognition as a profound contribution to Gaussian analysis.

2. Segal's Theorem. Given an orthogonal transformation $\mathcal{O}$ on a separable, real Hilbert space H, say that $\mathcal{O}$ is *atomic* if there is an $\mathcal{O}$-invariant subspace L with $\dim(L) \in \mathbb{Z}^+$. Equivalently, if H_c is the complexification of H and $\mathcal{O}_c$ is the unitary operator on H_c determined by $\mathcal{O}$, then $\mathcal{O}$ is atomic if and only if $\mathcal{O}_c$ admits a non-trivial eigenvector. Thus, in order for $\mathcal{O}$ to be atomic, it must admit an invariant subspace L of dimension either 1 or 2. Orthogonal transformations which are not atomic will be said to be *non-atomic*.

THEOREM 3 (Segal). *Let $(H, \Theta, \mathcal{W}_H)$ be an abstract Wiener space. If $\mathcal{O}$ is an orthogonal transformation and $T_{\mathcal{O}}$ is the associated measure preserving transformation on $(\Theta, \mathcal{B}_\Theta, \mathcal{W}_H)$, then $T_{\mathcal{O}}$ is ergodic if and only if $\mathcal{O}$ is non-atomic.*

Since we have already seen that $T_{\mathcal{O}}$ cannot be ergodic if $\mathcal{O}$ is atomic, all that we have to show is that $\mathcal{O}$ must be atomic if $T_{\mathcal{O}}$ is not ergodic. That is, we need to show that if $F \in L^2(\mathcal{W}_H; \mathbb{R})$ is non-constant and $F = F \circ T_{\mathcal{O}}$ $\mathcal{W}_H$-almost surely, then $\mathcal{O}$ admits an invariant subspace L with $\dim(L) \in \mathbb{Z}^+$. For this purpose, we will make use of the following simple lemma.

LEMMA 4. *Let $\mathcal{O}$ be an orthogonal transformation on the real Hilbert space H. Then $\mathcal{O}$ is atomic if and only if there is a real Hilbert space $\tilde{H}$, an orthogonal transformation $\tilde{\mathcal{O}}$ on $\tilde{H}$, and a non-zero, compact, linear map $K : H \to \tilde{H}$ such that $K \circ \mathcal{O}^\top = \tilde{\mathcal{O}} \circ K$.*

Proof. First suppose that $\tilde{H}$, $\tilde{\mathcal{O}}$, and K exist, and set $B = K^\top K$. Then B is a non-zero, compact, non-negative definite operator on H, and so B admits an eigenvalue $\lambda \in (0, \infty)$ whose eigenspace $L = \{h \in H : Bh = \lambda h\}$ has $\dim(L) \in \mathbb{Z}^+$. Moreover, because B commutes with $\mathcal{O}$, if $h \in L$, then $B \circ \mathcal{O}h = \mathcal{O} \circ Bh = \lambda \mathcal{O}h$, and so $\mathcal{O}h \in L$. Hence, L is a non-trivial, finite-dimensional, $\mathcal{O}$-invariant subspace.

Conversely, suppose that $\mathcal{O}$ admits a non-trivial, finite-dimensional invariant subspace L. Then there exist $(\xi_1, \xi_2) \in \mathbb{R}^2$ with $\xi_1^2 + \xi_2^2 = 1$ and $(g_1, g_2) \in H^2$ with $\|g_1\|_H^2 + \|g_g\|_H^2 = 1$ such that $\mathcal{O}g_1 = \xi_1 g_1 - \xi_2 g_2$ and $\mathcal{O}g_2 = \xi_2 g_1 + \xi_1 g_2$. Moreover, if $\xi_2 = 0$, then we may and will assume that $g_2 = 0$, and if $\xi_2 \neq 0$, then neither g_1 nor g_2 is 0 and $(g_1, g_2)_H = 0$. Now define $K : H \to H$ by

$$Kh = \begin{cases} (h, g_1)_H g_1 & \text{if } \xi_2 = 0, \\ (h, g_2)_H g_1 + (h, g_1)_H g_2 & \text{if } \xi_2 \neq 0. \end{cases}$$

Then K is a non-zero, compact, linear map, and $K \circ \mathcal{O}^\top = \mathcal{O} \circ K$. Hence, we can take $\tilde{H} = H$ and $\tilde{\mathcal{O}} = \mathcal{O}$. ∎

Our goal is to use Lemma 4 to prove the following slight strengthening of Segal's Theorem.

THEOREM 5. *Suppose that H is a separable, real Hilbert space and that $(H, \Theta, \mathcal{W}_H)$ is an abstract Wiener space. Given an orthogonal transformation $\mathcal{O}$ on H, $\mathcal{O}$ is atomic if and only if there is a finite-dimensional, $T_{\mathcal{O}}$-invariant subspace of $L^2(\mathcal{W}_H;\mathbb{R})$ containing non-constant functions. Thus, the following are equivalent:*

(1) *$\mathcal{O}$ is non-atomic.*
(2) *$T_{\mathcal{O}}$ is weakly mixing in the sense that for any $f \in L^2(\mathcal{W}_H;\mathbb{R})$ with f having mean-value 0,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} |(f \circ T_{\mathcal{O}}^n, g)_{L^2(\mathcal{W}_H;\mathbb{R})}| = 0$$

for all $g \in L^2(\mathcal{W}_H;\mathbb{R})$.
(3) *$T_{\mathcal{O}}$ is $\mathcal{W}_H$-ergodic.*

That the non-existence of a finite-dimensional, $T_{\mathcal{O}}$-invariant subspace with non-constant elements implies weak mixing is a familiar fact in classical ergodic theory. See, for example, page 39 of [2]. Thus, everything comes down to proving the first assertion.

If $\mathcal{O}$ admits a non-trivial, finite-dimensional subspace L, set $\tilde{L} = \{\mathcal{I}(h) : h \in L\}$. Then $\tilde{L}$ has the same dimension as L and L contains non-constant elements. In addition, because L is also $\mathcal{O}^\top$-invariant and $\mathcal{I}(h)\circ T_{\mathcal{O}} = \mathcal{I}(\mathcal{O}^\top h)$, it is clear that $\tilde{L}$ is $T_{\mathcal{O}}$-invariant. Hence, what remains to be shown is that $\mathcal{O}$ is atomic whenever $T_{\mathcal{O}}$ admits a finite-dimensional subspace containing non-constant functions, and it is in the proof of this part that we will employ Lemma 4.

In order to construct the quantities required to apply Lemma 4, we will need to recall Wiener's decomposition of $L^2(\mathcal{W};\mathbb{R})$ into subspaces $Z^{(n)}$, $n \in \mathbb{N}$, of homogeneous chaos. Namely, set

$$H_n(x) = (-1)^n e^{x^2/2} \partial_x^n e^{-x^2/2}, \qquad n \in \mathbb{N} \text{ and } x \in \mathbb{R},$$

and define $H_\mu : \mathbb{R}^\mathbb{N} \to \mathbb{R}$ for $\mu \in \mathbb{N}^\mathbb{N}$ with $\|\mu\| \equiv \sum_{k=0}^\infty \mu_k < \infty$ so that

$$H_\mu(\mathbf{x}) = \prod_{k=0}^\infty H_{\mu_k}(x_k) \qquad \text{for } \mathbf{x} = (x_0, \dots, x_k, \dots) \in \mathbb{R}^\mathbb{N}.$$

Given an orthonormal basis $E = \{e_k : k \geq 0\}$ for H, define $\mathbf{I}^E : \Theta \to \mathbb{R}^\mathbb{N}$ so that $\mathbf{I}^E(\theta)_k = \mathcal{I}(e_k)(\theta)$, $k \in \mathbb{N}$, and set $\mathcal{H}_\mu^E = H_\mu \circ \mathbf{I}^E$. Then $Z^{(n)}$ is the closure in $L^2(\mathcal{W};\mathbb{R})$ of $\operatorname{span}(\{\mathcal{H}_\mu^E : \|\mu\| = n\})$. The basic facts about the $Z^{(n)}$'s are that they are independent of the choice of orthonormal basis E, they are mutually orthogonal, and their span Z is dense in $L^2(\mathcal{W};\mathbb{R})$. All these properties are elementary applications of the generating function

$$e^{\alpha x-\alpha^2/2}=\sum_{m=0}^{\infty}\frac{\alpha^m}{m!}\,H_m(x),\qquad \alpha\in\mathbb{C}.$$

In particular, each $Z^{(n)}$ is $T_{\mathcal{O}}$-invariant.

Now suppose that L is a finite-dimensional, $T_{\mathcal{O}}$-invariant subspace of $L^2(\mathcal{W}_H;\mathbb{R})$ containing non-constant elements. By the preceding, we may and will assume that $L\subseteq Z^{(m)}$ for some $m\in\mathbb{Z}^+$. In addition, because L is finite-dimensional, we know that there exist $\xi_1,\xi_2\in\mathbb{R}^2$ and $G_1,G_2\in L^2(\mathcal{W}_H;\mathbb{R})$ such that $\xi_1^2+\xi_2^2=1=\|G_1\|^2_{L^2(\mathcal{W}_H;\mathbb{R})}+\|G_2\|^2_{L^2(\mathcal{W}_H;\mathbb{R})}$, $\big(G_1,G_2\big)_{L^2(\mathcal{W}_H;\mathbb{R})}=0$, and

$$\begin{pmatrix}G_1\circ T_{\mathcal{O}}\\ G_2\circ T_{\mathcal{O}}\end{pmatrix}=R\begin{pmatrix}G_1\\ G_2\end{pmatrix},$$

where $R:L^2(\mathcal{W}_H;\mathbb{R}^2)\to L^2(\mathcal{W}_H;\mathbb{R}^2)$ is given by

$$R\begin{pmatrix}F_1\\ F_2\end{pmatrix}=\begin{pmatrix}\xi_1F_1-\xi_2F_2\\ \xi_2F_1+\xi_1F_2\end{pmatrix}.$$

Now set $\tilde{H}=L^2(\mathcal{W}_H;\mathbb{R}^2)$, and define $\tilde{\mathcal{O}}$ on $\tilde{H}$ by

$$\tilde{\mathcal{O}}\begin{pmatrix}F_1\\ F_2\end{pmatrix}=R^{-1}\begin{pmatrix}F_1\circ T_{\mathcal{O}}\\ F_2\circ T_{\mathcal{O}}\end{pmatrix}=\begin{pmatrix}\xi_1F_1+\xi_2F_2\\ -\xi_2F_1+\xi_1F_2\end{pmatrix}\circ T_{\mathcal{O}}.$$

Clearly, $\tilde{\mathcal{O}}$ is an orthogonal transformation on $\tilde{H}$, and

$$\tilde{\mathcal{O}}\begin{pmatrix}G_1\\ G_2\end{pmatrix}=\begin{pmatrix}G_1\\ G_2\end{pmatrix}. \tag{2}$$

Before introducing the operator $K:H\to\tilde{H}$, we need to discuss the gradient operator $D:Z\to L^2(\mathcal{W}_H;H)$, which is determined by

$$D\mathcal{H}^E_\mu=\sum_{k=1}^{\infty}\mu^E_k\mathcal{H}^E_{\mu^k}e_k,$$

where $(\mu^k)_l\equiv(\mu_l-\delta_{k,l})^+$. Because $\|\mathcal{H}^E_\mu\|^2_{L^2(\mathcal{W}_H;\mathbb{R})}=\mu!$, it is easy to check that

$$(D\Phi_1,D\Phi_2)^2_{L^2(\mathcal{W}_H;H)}=n(\Phi_1,\Phi_2)^2_{L^2(\mathcal{W}_H;\mathbb{R})}\quad\text{for }\Phi_1,\Phi_2\in Z^{(n)}. \tag{3}$$

In addition, the definition of D as an operator is independent of the choice of orthonormal basis E. In particular, by taking $\{\lambda_k:k\geq 0\}\subseteq\Theta^*$ so that $\{h_{\lambda_k}:k\geq 0\}$ is an orthonormal basis in H, one can show that

$$(h,D\Phi(\theta))_H=\frac{d}{dt}\Phi(\theta+th)\Big|_{t=0}$$

when $\Phi(\theta)$ is a polynomial function of $\{\langle \theta, \lambda_k \rangle : k \geq 0\}$. Starting with such Φ's and using the preceding together with the obvious estimate coming from (3), one arrives at

$$(4) \qquad (h, D(\Phi \circ T_{\mathcal{O}}))_H = (\mathcal{O}h, D\Phi)_H \circ T_{\mathcal{O}} \quad \text{for } \Phi \in Z.$$

We now define $K : H \to \tilde{H}$ so that

$$Kh = \begin{pmatrix} D_h G_1 \\ D_h G_2 \end{pmatrix},$$

where we have introduced the notation $D_h F \equiv \big(h, DF\big)_H$ for $F \in Z$. We will show that K satisfies the hypotheses in Lemma 4 relative to $\mathcal{O}$ on H and $\tilde{\mathcal{O}}$ on $\tilde{H}$.

We begin by checking that K is a non-zero, compact operator. For this purpose, choose an orthonormal basis $\{e_k : k \geq 0\}$ for H, and observe that, by (3),

$$\sum_{k=0}^{\infty} \|Ke_k\|_{\tilde{H}}^2 = m,$$

which means that K is a non-zero, Hilbert–Schmidt, and therefore compact, operator. Thus, all that remains is to check that $K \circ \mathcal{O}^\top = \tilde{\mathcal{O}} \circ K$, or equivalently, that $K = \tilde{\mathcal{O}} \circ K \circ \mathcal{O}$. To this end, use (2) and (4) to check that

$$Kh = R^{-1} \begin{pmatrix} D_h(G_1 \circ T_{\mathcal{O}}) \\ D_h(G_2 \circ T_{\mathcal{O}}) \end{pmatrix} = R^{-1} \begin{pmatrix} (D_{\mathcal{O}h} G_1) \circ T_{\mathcal{O}} \\ (D_{\mathcal{O}h} G_2) \circ T_{\mathcal{O}} \end{pmatrix} = \tilde{\mathcal{O}} K \mathcal{O} h.$$

Before moving on, it should be observed that there is an easier result which, in most practical situations, is not only sufficient to check ergodicity but also opens the possibility of giving quantitative information.

THEOREM 6. *Let $(H, \Theta, \mathcal{W}_H)$ be an abstract Wiener space. If $\mathcal{O}$ is an orthogonal transformation on H with the property that, for every $(g, h) \in H^2$, $\varliminf_{n\to\infty} (\mathcal{O}^n g, h)_H = 0$, then $T_{\mathcal{O}}$ admits no finite-dimensional, invariant subspace containing non-constant elements.*

If one wants to use Theorem 5, this result is obvious. Indeed, if $\mathcal{O}$ satisfies the stated condition, then it certainly cannot have a non-trivial, finite-dimensional invariant subspace. However, there is a more quantitative way to prove it. Namely, the non-existence assertion follows immediately from the mixing property

$$\varliminf_{n\to\infty} |\mathbb{E}^{\mathcal{W}_H}[(\Phi \circ T_{\mathcal{O}}^n)\Phi]| = 0$$

for all $\Phi \in L^2(\mathcal{W}_H; \mathbb{R})$ with mean value 0. In fact, if $\{e_k : k \geq 1\}$ is an orthonormal basis for H, then it suffices to check this mixing property when

$$\Phi(\theta) = f([\mathcal{I}(e_1)](\theta), \ldots, [\mathcal{I}(e_N)](\theta))$$

for some $N \in \mathbb{Z}^+$ and bounded, Borel measurable $f : \mathbb{R}^N \to \mathbb{R}$. The reason why it is sufficient to check it for such Φ's is that, because $T_{\mathcal{O}}$ is $\mathcal{W}_H$-measure preserving, the set of Φ's for which the desired property holds is closed in $L^2(\mathcal{W}_H; \mathbb{R})$. Hence, if we start with any $\Phi \in L^2(\mathcal{W}_H; \mathbb{R})$ with mean value 0, we can first approximate it in $L^2(\mathcal{W}_H; \mathbb{R})$ by bounded functions with mean value 0 and then condition these bounded approximates with respect to $\sigma\big(\{\mathcal{I}(e_1), \ldots, \mathcal{I}(e_N)\}\big)$ to give them the required form.

Now suppose that $\Phi = f\big(\mathcal{I}(e_1), \ldots, \mathcal{I}(e_N)\big)$ for some N and bounded, measurable f. Then

$$\mathbb{E}^{\mathcal{W}_H}[\Phi \circ T_{\mathcal{O}}^n \Phi] = \iint_{\mathbb{R}^N \times \mathbb{R}^N} f(\mathbf{x}) f(\mathbf{y})\, \gamma_{\mathbf{0},\mathbf{C}_n}(d\mathbf{x} \times d\mathbf{y}),$$

where $\gamma_{\mathbf{0},\mathbf{C}_n}$ is the Gauss measure on $\mathbb{R}^N \times \mathbb{R}^N$ with mean $\mathbf{0}$ and covariance

$$\mathbf{C}_n = \begin{pmatrix} \mathbf{I} & \mathbf{B}_n \\ \mathbf{B}_n^\top & \mathbf{I} \end{pmatrix} \qquad \text{with } \mathbf{B}_n = (((e_k, \mathcal{O}^n e_l)_H))_{1 \le k,l \le N},$$

where the block structure corresponds to $\mathbb{R}^N \times \mathbb{R}^N$. Finally, by our hypothesis about $\mathcal{O}$, we can find a subsequence $\{n_m : m \ge 0\}$ such that $\lim_{m\to\infty} \mathbf{B}_{n_m} = \mathbf{0}$, from which it is clear that $\gamma_{\mathbf{0},\mathbf{C}_{n_m}}$ tends to $\gamma_{\mathbf{0},\mathbf{I}} \times \gamma_{\mathbf{0},\mathbf{I}}$ in variation and therefore

$$\lim_{m\to\infty} \mathbb{E}^{\mathcal{W}_H}[(\Phi \circ T_{\mathcal{O}}^{n_m})\Phi] = \mathbb{E}^{\mathcal{W}_H}[\Phi]^2 = 0.$$

As the preceding makes clear, the more one knows about the rate at which $(\mathcal{O}^n g, h)_H$ tends to 0, the more one will know about the strength of the mixing property for $T_{\mathcal{O}}$; and from this one can get estimates on the $L^2(\mathcal{W}_H; \mathbb{R})$ rate at which ergodic sums are converging.

3. Wiener isomorphisms. Given a pair of abstract Wiener spaces $(H, \Theta, \mathcal{W}_H)$ and $(H', \Theta', \mathcal{W}_{H'})$, we will say that a map $\Phi : \Theta \to \Theta'$ is a *Wiener isomorphism* if

(a) Φ is Borel measurable and $\mathcal{W}_{H'} = \Phi_* \mathcal{W}_H$.
(b) Φ is almost surely invertible in the sense that there is a Borel measurable $\Phi^{-1} : \Theta' \to \Theta$ such that $\Phi^{-1} \circ \Phi(\theta) = \theta$ and $\Phi \circ \Phi^{-1}(\theta') = \theta'$ for, respectively, $\mathcal{W}_H$-almost all $\theta \in \Theta$ and $\mathcal{W}_{H'}$-almost all $\theta' \in \Theta'$.
(c) For $\mathcal{W}_H^2$-almost every $(\theta_1, \theta_2) \in \Theta^2$,

$$\Phi\bigg(\frac{\theta_1 + \theta_2}{\sqrt{2}}\bigg) = \frac{\Phi(\theta_1) + \Phi(\theta_2)}{\sqrt{2}}.$$

Notice that if $\Psi' : \Theta \to \Theta'$ is Borel measurable and Ψ is $\mathcal{W}_H$-almost surely equal to a Borel measurable $\Phi : \Theta \to \Theta'$ which satisfies the additivity condition in (c), then Ψ also satisfies that condition. Indeed, $\Psi(2^{-1/2}(\theta_1 + \theta_2))$

$= 2^{-1/2}(\Psi(\theta_1)+\Psi(\theta_2))$ for $(\theta_1,\theta_2) \in (A\times A)\cap B\cap C$, where $A = \{\theta : \Phi(\theta) = \Psi(\theta)\}$ and

$$B = \{(\theta_1,\theta_2) : \Psi(2^{-1/2}(\theta_1+\theta_2)) = 2^{1/2}(\Psi(\theta_1)+\Psi(\theta_2))\},$$
$$C = \{(\theta_1,\theta_2) : \Psi(2^{-1/2}(\theta_1+\theta_2)) = \Phi(2^{-1/2}(\theta_1+\theta_2))\}.$$

Obviously, $A\times A$ and B have $\mathcal{W}_H^2$-measure 1, and because $\mathcal{W}_H$ is the $\mathcal{W}_{H^2}$-distribution of $(\theta_1,\theta_2) \rightsquigarrow 2^{-1/2}(\theta_1+\theta_2)$, so does C. Hence, if Φ is a Wiener isomorphism, then Ψ is also. Similarly, if $\Phi : \Theta \to \Theta'$ is a Wiener isomorphism and Φ^{-1} is an almost sure inverse of Φ, then Φ^{-1} is also a Wiener isomorphism. Since Φ^{-1} is certainly measure preserving, to check this it suffices to show that Φ^{-1} has the additivity property. For this purpose, take $A = \{\theta\in\Theta : \Phi^{-1}\circ\Phi(\theta) = \theta\}$, $A' = \{\theta'\in\Theta' : \Phi\circ\Phi^{-1}(\theta') = \theta'\}$ and

$$B = \{(\theta_1,\theta_2)\in\Theta^2 : \Phi(2^{-1/2}(\theta_1+\theta_2)) = 2^{-1/2}(\Phi(\theta_1)+\Phi(\theta_2))\},$$
$$C = \{(\theta_1'\theta_2')\in A'\times A' : (\Phi^{-1}(\theta_1'),\Phi^{-1}(\theta_2'))\in(A\times A)\cap B\}.$$

Then $\mathcal{W}_{H'}^2(C) = 1$ and, for $(\theta_1',\theta_2')\in C$,

$$\frac{\Phi^{-1}(\theta_1')+\Phi^{-1}(\theta_2')}{\sqrt{2}} = \Phi^{-1}\circ\Phi\left(\frac{\Phi^{-1}(\theta_1')+\Phi^{-1}(\theta_2')}{\sqrt{2}}\right) = \Phi^{-1}\left(\frac{\theta_1'+\theta_2'}{\sqrt{2}}\right).$$

The goal here is to prove the following theorem.

THEOREM 7. *Suppose that $\mathcal{O}$ and $\mathcal{O}'$ are orthogonal transformations on, respectively, H and H'. Then there exists a Wiener isomorphism $\Phi : \Theta \to \Theta'$ such that $\Phi\circ T_{\mathcal{O}} = T_{\mathcal{O}'}\circ\Phi$ $\mathcal{W}_H$-almost surely if and only if there is a unitary, linear map $U : H \to H'$ such that $U\mathcal{O} = \mathcal{O}'U$. In particular, $\Phi : \Theta\to\Theta$ is a Wiener isomorphism if and only if $\Phi = T_{\mathcal{O}}$ for some orthogonal transformation $\mathcal{O}$ on H.*

There are two steps in the proof of Theorem 7, the first of which is taken in the following lemma.

LEMMA 8. *Suppose that $U : H\to H'$ is a linear, unitary map. Then there exists a $\mathcal{W}_H$-almost surely unique, Borel measurable $T_U : \Theta\to\Theta'$ such that $\mathcal{I}(h')\circ T_U = \mathcal{I}(U^\top h')$ $\mathcal{W}_H$-almost surely for all $h'\in H'$. Moreover, T_U is a Wiener isomorphism and $T_{U^\top}$ is an almost sure inverse of T_U. Finally, $T_U\circ T_{\mathcal{O}} = T_{\mathcal{O}'}\circ T_U$ $\mathcal{W}_H$-almost surely if and only if $U\mathcal{O} = \mathcal{O}'U$.*

Proof. Since, for any Borel measurable maps Φ and Ψ from Θ to Θ', $\Phi = \Psi$ $\mathcal{W}_H$-almost surely if and only if $\mathcal{I}(h')\circ\Phi = \mathcal{I}(h')\circ\Psi$ $\mathcal{W}_H$-almost surely for each $h'\in H'$, uniqueness of T_U is clear. To prove existence, choose an orthonormal basis $(e_1,\dots,e_n,\dots)$ for H. Then $\{\mathcal{I}(e_n) : n\ge 0\}$ is a sequence of independent, standard normal random variables under $\mathcal{W}_H$, $(Ue_1,\dots,Ue_n,\dots)$ is an orthonormal basis for H', and so (cf. §2 in [4]) the series $\sum_{n=0}^\infty \mathcal{I}(e_n)Ue_n$ is $\mathcal{W}_H$-almost surely convergent in Θ.

Thus, we can define a Borel measurable $T_U : \Theta \to \Theta'$ so that, for $\mathcal{W}_H$-almost all $\theta \in \Theta$, $T_U(\theta) = \sum_{n=0}^{\infty} \mathcal{I}(e_n)(\theta)Ue_n$, where the series converges in Θ'. To check that this choice of T_U has the right property, first observe that $\mathcal{W}_{H'}$ is the $\mathcal{W}_H$-distribution of $\sum_{n=0}^{\infty} \mathcal{I}(e_n)(\theta)Ue_n$ and therefore that $(T_U)_*\mathcal{W}_H = \mathcal{W}_{H'}$. Hence, $h' \in H' \mapsto \mathcal{I}(h') \circ T_U \in L^2(\mathcal{W}_H;\mathbb{R})$ and $h' \in H' \mapsto \mathcal{I}(U^\top h') \in L^2(\mathcal{W}_H;\mathbb{R})$ are both isometric, and they are therefore equal if they agree on $\{h'_{\lambda'} : \lambda' \in (\Theta')^*\}$. But, for $\mathcal{W}_H$-almost all θ,

$$\begin{aligned}\mathcal{I}(h'_{\lambda'}) \circ T_U(\theta) = \langle T_U(\theta), \lambda'\rangle &= \sum_{n=0}^{\infty} \mathcal{I}(e_n)\langle Ue_n, \lambda'\rangle \\ &= \sum_{n=0}^{\infty} \mathcal{I}(e_n)(\theta)(U^\top h'_{\lambda'}, e_n)_H = \mathcal{I}(h'_{\lambda'})(\theta).\end{aligned}$$

To see that T_U is a Wiener isomorphism, we must still check that it is almost surely invertible and that it has the additivity property in (c). But

$$\mathcal{I}(h')\circ T_U \circ T_{U^\top} = \mathcal{I}(U^\top h')\circ T_{U^\top} = \mathcal{I}(UU^\top h') = \mathcal{I}(h') \quad \mathcal{W}_H\text{-almost surely,}$$

and so not only is T_U almost surely invertible but $T_{U^\top}$ is an almost sure inverse. As for the additivity property, note that we can choose the e_n's entering the preceding definition of T_U so that $e_n = h_{\lambda_n}$ with $\lambda_n \in \Theta^*$ for each $n \in \mathbb{N}$. Now let A be the set of θ for which $S(\theta) \equiv \sum_{n=0} \langle \theta, \lambda_n\rangle Ue_n$ converges, and take $S(\theta) = 0$ for $\theta \notin A$. Then S is a legitimate choice for T_U, and we know that $2^{-1/2}(\theta_1 + \theta_2) \in A$ and $S(2^{-1/2}(\theta_1 + \theta_2)) = 2^{-1/2}(S(\theta_1) + S(\theta_2))$ for $(\theta_1, \theta_2) \in A \times A$. Hence, since $A \times A$ has $\mathcal{W}_H^2$-measure 1, we have completed the proof that T_U is a Wiener isomorphism and that $T_{U^\top}$ is an almost sure inverse.

Finally, $\mathcal{O}'U = U\mathcal{O}$ if and only if for each $h' \in H'$,

$$\begin{aligned}\mathcal{I}(h') \circ T_U \circ T_{\mathcal{O}} &= \mathcal{I}(U^\top h') \circ T_{\mathcal{O}} = \mathcal{I}((U\mathcal{O})^\top h') \\ &= \mathcal{I}((\mathcal{O}'U)^\top h') = \mathcal{I}((\mathcal{O}')^\top h') \circ T_U = \mathcal{I}(h') \circ T_{\mathcal{O}'} \circ T_U\end{aligned}$$

$\mathcal{W}_H$-almost surely. Hence $\mathcal{O}'U = U\mathcal{O} \Leftrightarrow T_{\mathcal{O}'} \circ T_U = T_U \circ T_{\mathcal{O}}$ $\mathcal{W}_H$-almost surely. ∎

The second step in the proof of Theorem 7 is to show that if $\Phi : \Theta \to \Theta'$ is a Wiener isomorphism, then there exists a unique unitary, linear $U : H \to H'$ such that $\Phi = T_U$ $\mathcal{W}_H$-almost surely. Again the uniqueness is essentially trivial, and so we will concentrate on existence.

Lemma 9. *If $f \in L^2(\mathcal{W}_H;\mathbb{R})$, then $f = \mathcal{I}(h)$ for some $h \in H$ if and only if $(2^{-1/2}(\theta_1 + \theta_2)) = 2^{-1/2}(f(\theta_1) + f(\theta_2))$ $\mathcal{W}_H^2$-almost surely.*

Proof. The "if" assertion is easy. Indeed, again using the fact that $\mathcal{W}_H$ is the $\mathcal{W}_H^2$-distribution of $(\theta_1, \theta_2) \rightsquigarrow 2^{-1/2}(\theta_1 + \theta_2)$, one can easily check that it suffices to check the additivity property for $\mathcal{I}(h)$ when $h = h_\lambda$ for some $\lambda \in \Theta^*$, in which case it is trivial.

To prove the converse, what we have to check is that the additivity property implies that $f \in Z^{(1)}$. Thus, let f_n denote the projection of f onto $Z^{(n)}$. We want to show that $f_n = 0$ when $n \neq 1$, and the key to doing so is to show that each f_n inherits the additivity property from f. For this purpose, set $F(\theta_1, \theta_2) = f(2^{-1/2}(\theta_1 + \theta_2))$. Then $F \in L^2(\mathcal{W}_H^2; \mathbb{R})$ and $F = \sum_{n=0}^\infty F_n$ in $L^2(\mathcal{W}_H^2; \mathbb{R})$, where $F_n(\theta_1, \theta_2) = f_n(2^{-1/2}(\theta_1 + \theta_2))$. At the same time, we know that

$$F(\theta_1, \theta_2) = 2^{-1/2} \sum_{n=0}^\infty (f_n(\theta_1) + f_n(\theta_2))$$

in $L^2(\mathcal{W}_H^2; \mathbb{R})$, and clearly each summand in the preceding is an element of $Z^{(n)}(H^2)$, the space of nth order homogeneous chaos for the abstract Wiener space $(H^2, \Theta^2, \mathcal{W}_H^2)$. Hence, we will know that the additivity property holds for f_n once we show that $F_n \in Z^{(n)}(H^2)$. That is, we must check that if $g \in Z^{(n)}$ and $G(\theta_1, \theta_2) = g(2^{-1/2}(\theta_1 + \theta_2))$, then $G \in Z^{(n)}(H^2)$. After a little thought, one realizes that checking this last statement reduces to showing that if H_n is the nth Hermite polynomial described following the statement of Theorem 5, then $H_n(2^{-1/2}(x_1 + x_2))$ is a linear combination of $H_m(x_1)H_{n-m}(x_2)$, $0 \le m \le n$. Finally, using the generating function for the H_n's, it is easy to check that

$$H_n(2^{-1/2}(x_1 + x_2)) = 2^{-n/2} \sum_{m=0}^n H_m(x_1) H_{n-m}(x_2).$$

In view of the preceding, we will be done once we show that if $f \in Z^{(n)}$ has the additivity property, then $f = 0$ unless $n = 1$. This is obvious when $n = 0$. Thus, assume that $n \ge 2$. Define F from f as in the preceding paragraph, and observe that $((h, -h), DF)_{H^2} = 0$ $\mathcal{W}_H^2$-almost surely for each $h \in H$. At the same time, by the additivity property,

$$0 = 2^{1/2}((h, -h), DF(\theta_1, \theta_2))_{H^2} = (h, Df(\theta_1))_H - (h, Df(\theta_2))_H$$

for $\mathcal{W}_H^2$-almost all (θ_1, θ_2). Hence, since the terms on the right are independent under $\mathcal{W}_H^2$, the only way that this could hold is if each of them is $\mathcal{W}_H^2$-almost surely constant. Hence, we now know Df is $\mathcal{W}_H$-almost surely constant, which, because $n \ge 2$, is possible only if $f = 0$. ∎

The following completes the proof of Theorem 7.

THEOREM 10. *$\Phi : \Theta \to \Theta'$ is a Wiener isomorphism if and only if there is a unitary, linear $U : H \to H'$ such that $\Phi = T_U$ $\mathcal{W}_H$-almost surely.*

Proof. We need only address the "only if" part. To this end, let $\lambda' \in \Theta^*$ be given, and set $f = \mathcal{I}(h'_{\lambda'}) \circ \Phi$. Then f satisfies the additivity property in Lemma 11, and therefore, by that lemma, there exists a unique $F'h \in H'$

such that $\mathcal{I}(h'_{\lambda'})\circ\Phi = \mathcal{I}(F'h'_{\lambda'})$. Moreover, F' is linear and isometric. Hence F' admits a unique extension as an isometry from H' into H, and it is easy to check that $\mathcal{I}(h')\circ\Phi = \mathcal{I}(F'h')$ $\mathcal{W}_H$-almost surely continues to hold for all $h'\in H'$. Next, apply the same argument to Φ^{-1} and thereby produce a linear isometry $F : H \to H'$ such that $\mathcal{I}(h)\circ\Phi^{-1} = \mathcal{I}(Fh)$ $\mathcal{W}_{H'}$-almost surely for each $h\in H$. Finally, note that

$$\mathcal{I}(h) = \mathcal{I}(h)\circ\Phi^{-1}\circ\Phi = \mathcal{I}(Fh)\circ\Phi = \mathcal{I}(F'Fh) \qquad \mathcal{W}_H\text{-almost surely for each } h\in H,$$

and similarly, $\mathcal{I}(h') = \mathcal{I}(FF'h')$ $\mathcal{W}_{H'}$-almost surely for each $h'\in H'$, and from these conclude that F is unitary. Thus, we can take $U = F$. ∎

Remark. It should be emphasized that Theorem 7 does not solve the problem of determining when $T_{\mathcal{O}}$ on Θ is measure theoretically isomorphic to $T_{\mathcal{O}'}$ on Θ'. That is, when there is an almost surely invertible, measure preserving $\Phi : \Theta \to \Theta'$ such that $T_{\mathcal{O}'}\circ\Phi = \Phi\circ\Phi$ almost surely. It only determines when there is such a Φ which is a Wiener isomorphism.

4. Some examples. The Hilbert space for Wiener's own abstract Wiener space, the one corresponding to Brownian motion, is the space $H^1(\mathbb{R}^N)$ of absolutely continuous paths $h : [0,\infty)\to\mathbb{R}^N$ with $h(0)=0$ and square integrable derivative $\dot h$, the norm $\|h\|_{H^1(\mathbb{R}^N)}$ being $\|\dot h\|_{L^2([0,\infty);\mathbb{R}^N)}$. Wiener's renowned theorem shows that in this case the Banach space can be taken to be the space $\Theta(\mathbb{R}^N)$ of continuous $\theta : [0,\infty)\to\mathbb{R}^N$ with $\theta(0) = 0 = \lim_{t\to\infty} t^{-1}|\theta(t)|$ endowed with the norm $\|\theta\|_{\Theta(\mathbb{R}^N)} = \sup_{t\ge 0}(1+t)^{-1}|\theta(t)|$.

Among the orthogonal transformations on $H^1(\mathbb{R}^N)$, perhaps the most important to Brownian motion *aficionados* are the natural rescaling maps. That is, given $\alpha\in(0,\infty)$, consider the transformation $\mathcal{O}_\alpha$ on $H^1(\mathbb{R}^N)$ determined by $\mathcal{O}_\alpha h(t) = \alpha^{-1/2}h(\alpha t)$. Obviously, $\mathcal{O}_\alpha$ is an orthogonal transformation on $H^1(\mathbb{R}^N)$. In addition, it extends trivially as a continuous map S_α on $\Theta(\mathbb{R}^N)$ into itself, known to probabilists as the *Brownian scaling transformation.* Thus, by the Remark following Corollary 2, we may and will take $T_{\mathcal{O}_\alpha} = S_\alpha$. Finally, for each $\alpha\in(0,\infty)\setminus\{1\}$, $S_\alpha^n g$ tends weakly to 0 in $H^1(\mathbb{R}^N)$ for all $g\in H(\mathbb{R}^N)$. Perhaps the easiest way to see this is to first note that it suffices to prove it for a dense set of g's. Thus, suppose that $\dot g\in C_{\mathrm{c}}^\infty((0,\infty);\mathbb{R}^N)$ and observe that, for all $h\in H^1(\mathbb{R}^N)$,

$$(\mathcal{O}_\alpha^n g, h)_H = \int\limits_{(0,\infty)} \alpha^{1/2}\dot g(\alpha t)\dot h(t)\,dt \to 0$$

as α tends to either 0 or ∞. Since $\mathcal{O}_\alpha^n = \mathcal{O}_{\alpha^n}$, there is nothing more to do.

As a consequence, we can now apply Birkhoff's individual ergodic theorem to see that, for any $\alpha\in(0,\infty)\setminus\{1\}$ and any $\mathcal{W}_{H^1(\mathbb{R}^N)}$-integrable

$F : \Theta(\mathbb{R}^N) \to \mathbb{R}$,

$$\frac{1}{n}\sum_{m=0}^{n-1} F \circ S_{\alpha^m} \to \int_{\Theta(\mathbb{R}^N)} F\, d\mathcal{W}_{H^1(\mathbb{R}^N)}$$

both $\mathcal{W}_{H^1(\mathbb{R}^N)}$-almost surely and in $L^1(\mathcal{W}_{H^1(\mathbb{R}^N)};\mathbb{R})$. Alternatively, by taking advantage of the obvious fact that $\{S_{e^t} : t \in [0,\infty)\}$ is a semigroup, one can use the continuous time version of Birkhoff's theorem to show that

$$\frac{1}{\log t}\int_1^t \tau^{-1} F \circ S_\tau \, d\tau \to \int_{\Theta(\mathbb{R}^N)} F\, d\mathcal{W}_{H^1(\mathbb{R}^N)}$$

both $\mathcal{W}_{H^1(\mathbb{R}^N)}$-almost surely and in $L^1(\mathcal{W}_{H^1(\mathbb{R}^N)};\mathbb{R})$. Unfortunately, at least from the standpoint of Segal's theory, the conclusion just drawn can be seen (cf. §5) as an application of a more familiar, and far better understood, ergodic theorem.

Here is a second example in the same setting, one which I do not know how to handle by any other approach. Assume that N is even and therefore that there exist non-singular, skew-symmetric transformations A on $\mathbb{R}^N$. Given such an A, define $\mathcal{O}_A$ on $H^1(\mathbb{R}^N)$ so that

$$\mathcal{O}_A h(t) = \int_0^t e^{\tau A}\dot{h}(\tau)\, d\tau, \qquad t \in [0,\infty).$$

Again it is obvious that $\mathcal{O}_A$ is an orthogonal transformation. Further, Riemann–Stieltjes integration provides a continuous extension of $\mathcal{O}_A$. Namely, because $\tau \rightsquigarrow e^{\tau A}$ is smooth and $\theta \in \Theta(\mathbb{R}^N)$ is continuous, $\tau \rightsquigarrow e^{\tau A}$ is locally Riemann–Stieltjes integrable with respect to every $\theta \in \Theta(\mathbb{R}^N)$, and it is an elementary exercise to show that the map I_A given by

$$I_A\theta(t) = \int_0^t e^{\tau A}\, d\theta(\tau), \qquad t \in [0,\infty),$$

is a continuous on $\Theta(\mathbb{R}^N)$ into itself. Finally, because $I_A^n = I_{nA}$ and there are orthonormal coordinates in which the matrix representation of $e^{\tau A}$ consists of $N/2$ 2×2 blocks along the diagonal, each of which has the form

$$\begin{pmatrix} \cos\tau\alpha & -\sin\tau\alpha \\ \sin\tau\alpha & \cos\tau\alpha \end{pmatrix}$$

for some $\alpha \neq 0$, it follows that $(\mathcal{O}_A^n g, h)_{H^1(\mathbb{R}^N)} \to 0$ by the Riemann–Lebesgue Lemma. Hence, just as before, if $F \in L^1(\mathcal{W}_{H^1(\mathbb{R}^N)};\mathbb{R})$, then

$$\frac{1}{n}\sum_{m=0}^{n-1} F \circ I_{mA} \to \int_{\Theta(\mathbb{R}^N)} F\, d\mathcal{W}_{H^1(\mathbb{R}^N)}$$

and

$$\frac{1}{t}\int_0^t F \circ I_{\tau A}\,d\tau \to \int_{\Theta(\mathbb{R}^N)} F\,d\mathcal{W}_{H^1(\mathbb{R}^N)},$$

both $\mathcal{W}_{H^1(\mathbb{R}^N)}$-almost surely and in $L^1(\mathcal{W}_{H^1(\mathbb{R}^N)};\mathbb{R})$.

5. A large deviations result. As I said in §4, the Brownian scaling example there can be understood as an application of a more familiar result. Namely, define $\Psi : \Theta(\mathbb{R}^N) \to C(\mathbb{R};\mathbb{R}^N)$ so that $[\Psi(\theta)](t) = e^{-t/2}\theta(e^t)$, set $\Theta^U(\mathbb{R}^N) = \Psi(\Theta(\mathbb{R}^N))$, and define the norm $\|\cdot\|_{\Theta^U(\mathbb{R}^N)}$ so that Ψ becomes an isometry. Then the pushforward of $\mathcal{W}_{H^1(\mathbb{R}^N)}$ under Ψ is the distribution $\mathcal{U}^{(N)}$ of the stationary Ornstein–Uhlenbeck process. To be precise, $(H^U(\mathbb{R}^N), \Theta^U(\mathbb{R}^N), \mathcal{U}^{(N)})$ is an abstract Wiener space when $H^U(\mathbb{R}^N)$ is the space of absolutely continuous $h \in L^2(\mathbb{R};\mathbb{R}^N)$ with

$$\|h\|_{H^U(\mathbb{R}^N)} \equiv \sqrt{\|\dot{h}\|^2_{L^2(\mathbb{R};\mathbb{R}^N)} + \tfrac{1}{4}\|h\|^2_{L^2(\mathbb{R};\mathbb{R}^N)}} < \infty.$$

Furthermore, it is easy to check that $\Psi \circ S_{e^\tau} = \Sigma_\tau \circ \Pi$, where Σ_τ is the natural time-shift transformation on $C(\mathbb{R};\mathbb{R}^N)$. Hence, the distribution of

$$\left\{\frac{1}{n}\sum_{m=0}^{n-1} F \circ S_\alpha^n : n \ge 1\right\}$$

under $\mathcal{W}_{H^1(\mathbb{R}^N)}$ is the same as that of

$$\left\{\frac{1}{n}\sum_{m=0}^{n-1} (F \circ \Psi^{-1}) \circ \Sigma_{n\log\alpha} : n \ge 1\right\}$$

under $\mathcal{U}^{(N)}$, and the same line of reasoning applies in the continuous time setting.

Since, for each $\tau \ne 0$, the time-shift transformation Σ_τ is ergodic for $\mathcal{U}^{(N)}$, the ergodic result proved in §4 for the Brownian scaling transformations follows immediately. In addition, a lot is known about the ergodic properties of the time-shift transformation group under $\mathcal{U}^{(N)}$. For example, consider the associated empirical measure given by

$$R_T^U(\cdot,\theta^U) \equiv \frac{1}{2T}\int_{-T}^{T} \delta_{\Sigma_\tau\theta^U}\,d\tau, \qquad (T,\theta^U) \in (0,\infty)\times\Theta^U(\mathbb{R}^N).$$

Then, as a consequence of ergodicity and the Individual Ergodic Theorem, one knows that, as $T \to \infty$, $R_T^U(\cdot,\theta^U)$ tends weakly to $\mathcal{U}^{(N)}$ for $\mathcal{U}^{(N)}$-almost every $\theta^U \in \Theta^U(\mathbb{R}^N)$. Moreover, one has (cf. §5.4 and Exercise 6.2.47 in [5]) estimates on the rate at which this convergence is taking place. To be precise, given an interval $I \subseteq \mathbb{R}$, let $\mathcal{B}_I^U$ be the σ-algebra over $\Theta^U(\mathbb{R}^N)$ generated by the maps $\theta^U \rightsquigarrow \theta^U(t)$ as t runs over I, and given a $\mathbb{P} \in \mathbf{M}_1(\Theta^U(\mathbb{R}^N))$, define $H_I^U(\mathbb{P})$ to be the entropy of $\mathbb{P}\restriction\mathcal{B}_I^U$ relative to $\mathcal{U}^{(N)}\restriction\mathcal{B}_I^U$. That is, if $\mathbb{P}\restriction\mathcal{B}_I^U \ll \mathcal{U}^{(N)}\restriction\mathcal{B}_I^U$ and F_I is its Radon–Nikodym derivative, then

$$H_I^U(\mathbb{P}) = \int F_I \log F_I \, d\mathcal{U}^{(N)} = \int \log F_I \, d\mathbb{P},$$

and $H_I^U(\mathbb{P}) = \infty$ otherwise. If $\mathbb{P}$ is time-shift invariant, then the limit

$$\lim_{T\to\infty} \frac{H^U_{[-T,T]}(\mathbb{P})}{2T}$$

exists in $[0,\infty]$. Thus, we can define the specific entropy $H^U(\mathbb{P})$ of $\mathbb{P}$ relative to $\mathcal{U}^{(N)}$ to be this limit when $\mathbb{P}$ is time-shift invariant and to be $+\infty$ otherwise, and if we adopt this definition, one can show that the large deviations of $\{R_T : T > 0\}$ are governed by H^U in the sense that, for any Borel measurable $A \subseteq \mathbf{M}_1(\Theta^U(\mathbb{R}^N))$,

$$\begin{aligned}
-\inf_{\mathbb{P}\in\mathring{A}} H^U(\mathbb{P}) &\le \varliminf \frac{1}{2T}\log(\mathcal{U}^{(N)}(R_T^U \in A)) \\
&\le \varlimsup \frac{1}{2T}\log(\mathcal{U}^{(N)}(R_T^U \in A)) \le -\inf_{\mathbb{P}\in\overline{A}} H^U(\mathbb{P}),
\end{aligned}$$

where $\mathring{A}$ and $\overline{A}$ are, respectively, the interior and closure of A with respect to the weak topology.

To transfer this large deviations estimate back to the setting of the Brownian scaling transformations, define the empirical process $\{R_T : T > 0\}$ by

$$R_T(\cdot,\theta) \equiv \frac{1}{T}\int_0^T \delta_{S_{e^\tau}\theta}\, d\tau \qquad \text{for } (T,\theta) \in (0,\infty)\times\Theta(\mathbb{R}^N).$$

Then the preceding result says that the large deviations of $\{R_T : T > 0\}$ under $\mathcal{W}^{(N)}$ are governed by $H^U \circ \Psi_*$. That is, for any Borel measurable $A \subseteq \mathbf{M}_1(\Theta(\mathbb{R}^N))$,

$$\begin{aligned}
-\inf_{\mathbb{P}\in\mathring{A}} H^U(\Psi_*\mathbb{P}) &\le \varliminf \frac{1}{2T}\log(\mathcal{W}^{(N)}(R_T \in A)) \\
&\le \varlimsup \frac{1}{2T}\log(\mathcal{W}^{(N)}(R_T \in A)) \le -\inf_{\mathbb{P}\in\overline{A}} H^U(\Psi_*\mathbb{P}).
\end{aligned}$$

Further, because, for any $\mathbb{P} \in \mathbf{M}_1(\Theta(\mathbb{R}^N))$ and $T > 0$,

$$\Psi_*(\mathbb{P}\restriction\mathcal{B}_{[e^{-T},e^T]}) = (\Psi_*\mathbb{P})\restriction\mathcal{B}^U_{[-T,T]},$$

where $\mathcal{B}_I$ is the σ-algebra over $\Theta(\mathbb{R}^N)$ generated by the maps $\theta \rightsquigarrow \theta(t)$ for $t \in I \subseteq [0,\infty)$, it is easy to see that $\mathbb{P}\restriction\mathcal{B}_{[e^{-T},e^T]} \ll \mathcal{W}^{(N)}\restriction\mathcal{B}_{[e^{-T},e^T]}$ if and only if $(\Psi_*\mathbb{P})\lceil\mathcal{B}^U_{[-T,T]} \ll \mathcal{U}^{(N)}\lceil\mathcal{B}^U_{[-T,T]}$ and that

$$H^U_{[-T,T]}(\Psi_*\mathbb{P}) = \int F_T \log F_T \, d\mathcal{W}^{(N)} = \int \log F_T \, d\mathbb{P}$$
$$\text{if } d(\mathbb{P}\lceil\mathcal{B}_{[e^{-T},e^T]}) = F_T d(\mathcal{W}^{(N)}\lceil\mathcal{B}_{[e^{-T},e^T]}).$$

Hence, if we define $H^{\mathcal{W}}(\mathbb{P})$ for $\mathbb{P} \in \mathbf{M}_1(\Theta(\mathbb{R}^N))$ so that

$$H^{\mathcal{W}}(\mathbb{P}) = \lim_{T\to\infty} \frac{1}{2T} \int \log\left(\frac{d(\mathbb{P}\lceil\mathcal{B}_{[e^{-T},e^T]})}{d(\mathcal{W}^{(N)}\lceil\mathcal{B}_{[e^{-T},e^T]})}\right) d\mathbb{P}$$

when $\mathbb{P}$ is invariant under the Brownian scaling transformations $\{S_\alpha : \alpha \in [0,\infty)\}$ and $\mathbb{P}\lceil\mathcal{B}_{[e^{-T},e^T]} \ll \mathcal{W}^{(N)}\lceil\mathcal{B}_{[e^{-T},e^T]}$ for all $T > 0$ and $H^W(\mathbb{P}) = +\infty$ otherwise, then

$$(5) \qquad -\inf_{\mathbb{P}\in\mathring{A}} H^W(\mathbb{P}) \le \varliminf \frac{1}{2T}\log(\mathcal{W}^{(N)}(R_T \in A))$$
$$\le \varlimsup \frac{1}{2T}\log(\mathcal{W}^{(N)}(R_T \in A)) \le -\inf_{\mathbb{P}\in\overline{A}} H^W(\mathbb{P}).$$

It is instructive to calculate $H^W(\mathbb{P})$ in a simple case. Thus, let $N = 1$, and consider perturbations of Brownian motion by a deterministic path ψ. That is, take $\mathbb{P}$ to be the distribution of $\theta \rightsquigarrow \theta + \psi$ under $\mathcal{W} = \mathcal{W}^{(1)}$. In order for $\mathbb{P}$ to be invariant under Brownian scaling, it is clear that ψ must be homogeneous of order $1/2$. Thus, consider $\psi(t) = t^{1/2}$. Because $\dot\psi$ is not square integrable at 0, $\mathbb{P}\lceil\mathcal{B}_{[0,T]}$ fails to be absolutely continuous with respect to $\mathcal{W}\lceil\mathcal{B}_{[0,T]}$ for any $T > 0$. Nonetheless, for each $T > 0$, $\mathbb{P}\lceil\mathcal{B}_{[e^{-T},e^T]} \ll \mathcal{W}\lceil\mathcal{B}_{[e^{-T},e^T]}$. In fact, the corresponding Radon–Nikodym derivative is given by

$$F_T(\theta) = \exp\left(\int_0^\infty \dot h_T(t)\, d\theta(t) - \frac{1}{2}\, \|h_T\|^2_{H^1(\mathbb{R})}\right),$$

where

$$h_T(t) = \begin{cases} e^{T/2}t & \text{if } t \in [0, e^{-T}], \\ \sqrt{t \wedge e^T} & \text{if } t \ge e^{-T} \end{cases}$$

and the integral is taken in the sense of Riemann–Stieltjes. To see this, one need only note that $\theta(t) + t^{1/2} = \theta(t) + h_T(t)$ for $e^{-T} \le t \le e^T$ and that F_T is $\mathcal{B}_{[e^{-T},e^T]}$-measurable, and apply the Cameron–Martin formula (cf. [4]) to check that F_T is the Radon–Nikodym derivative with respect to $\mathcal{W}$ of the translate of $\mathcal{W}$ by h_T. Furthermore,

$$\int \log F_T \, d\mathbb{P} = \int \log F_T(\theta + h_T)\, d\mathcal{W} = \frac{\|h_T\|^2_{H^1(\mathbb{R})}}{2} = \frac{1}{2} + \frac{T}{8}.$$

Hence, $H^W(\mathbb{P}) = 1/16$ for this $\mathbb{P}$.

REFERENCES

[1] L. Gross, *Abstract Wiener spaces*, in: Proc. 5th Berkeley Sympos. on Probability and Statistics, Vol. 2, Part 1, Univ. of California Press, 1967, 31–42.
[2] P. Halmos, *Lectures on Ergodic Theory*, Chelsea, 1956.
[3] I. M. Segal, *Ergodic subgroups of the orthogonal group on a real Hilbert space*, Ann. of Math. 66 (1957), 297–303.
[4] D. Stroock, *Abstract Wiener space, revisited*, Comm. Stoch. Anal. 2 (2008), 145–151.
[5] D. Stroock and J.-D. Deuschel, *Large Deviations*, Academic Press, 1989; republished by AMS Chelsea, 2000.

Daniel W. Stroock
M.I.T. 2-272
Cambridge, MA 02139, U.S.A.
E-mail: dws@math.mit.edu

Received 3 February 2009;
revised 27 June 2009 (5155)