COMPOSITIO MATHEMATICA

PETER BAXENDALE

Brownian motions in the diffeomorphism group I

Compositio Mathematica, tome 53, nº 1 (1984), p. 19-50
<http://www.numdam.org/item?id=CM_1984__53_1_19_0>

Compositio Mathematica 53 (1984) 19–50.

BROWNIAN MOTIONS IN THE DIFFEOMORPHISM GROUP I

Peter Baxendale

1. Introduction

Let M be a compact smooth manifold and let $\mathrm{Diff}^r(M)$ denote the group of C^r diffeomorphisms of M ($1 \leqslant r \leqslant \infty$). If $\mathrm{Diff}^r(M)$ is given the topology of uniform C^r convergence then it becomes a separable topological group which may be metrized in such a way as to become complete. The purpose of this paper is to characterize random processes $\{\phi_t\colon t \geqslant 0\}$ taking values in $\mathrm{Diff}^r(M)$ and satisfying

(i) independent increments on the left (i.e. if $0 \leqslant s < t \leqslant u < v$ then $\phi_t\phi_s^{-1}$ and $\phi_v\phi_u^{-1}$ are independent)

(ii) time homogeneous (i.e. the distribution of $\phi_t\phi_s^{-1}$ depends only on $t-s$)

(iii) continuous sample paths with probability one (w.p.1)

(iv) $\phi_0 = Id_M$ w.p.1.

We shall call any such process a *Brownian motion* in $\mathrm{Diff}^r(M)$.

More generally if $\mathrm{Diff}^r(M)$ is replaced by any topological group G we shall call a random process taking values in G and satisfying conditions (i)–(iv) a Brownian motion in G. For many G the classification of Brownian motions is known. If $G = \mathbb{R}$, any such process is of the form $aW_t + bt$ where W_t is the standard Brownian motion on $\mathbb{R}$ and a and b are constants. More generally if $G = \mathbb{R}^n$ then the process is of the form $\Sigma_{i=1}^N a_i W_t^i + bt$ where $W_t^1, \dots, W_t^n$ are n independent one-dimensional Brownian motions and $a_1, \dots, a_n$, b are fixed vectors in $\mathbb{R}^n$. This characterization may be extended to the case where G is any separable Fréchet space. (See Section 6.2 and the reference therein.)

In the case that G is a finite dimensional Lie group the processes have been classified by Hunt [10]. In fact Hunt deals with the more general case where the condition (iii) of continuous sample paths is weakened to that of sample paths which are right continuous with limits on the left. The classification is given in terms of the infinitesimal generator of the semigroup associated with the process. See also papers by Stroock and Varadhan [19] and Feinsilver [7] which use martingale techniques to extend Hunt's results to apply to non time homogeneous processes.

Our main result involves an adaptation of ideas contained in the paper by Hunt. The process $\{\phi_t\colon t \geqslant 0\}$ is determined up to equivalence by knowledge of the processes $\{\phi_t(\boldsymbol{x}):\ t \geq 0\}$ in M^n where $n \geq 1$, $\boldsymbol{x} = (x_1,\dots,x_n) \in M^n$ and $\phi_t(\boldsymbol{x}) = (\phi_t(x_1), \phi_t(x_2),\dots,\phi_t(x_n)) \in M^n$. Each process $\{\phi_t(\boldsymbol{x}):\ t \geq 0\}$ is a time homogeneous Markov process in M^n and the idea is to classify $\{\phi_t:\ t \geq 0\}$ via the infinitesimal generators of the n-point motions $\{\phi_t(\boldsymbol{x}):\ t \geq 0\}$, $\boldsymbol{x} \in M^n$, $n \geq 1$. We use the fact that the ϕ_t are C^r diffeomorphisms to prove that the n-point motions are diffusions with C^{r-2} coefficients. This should be compared with results of Harris [9] and Kunita [13] on Brownian motions of homeomorphisms where it is assumed that the n-point motions are diffusions with some specified degree of regularity. Formulae for the restrictions to C^2 functions of the infinitesimal generators of the n-point motions are given in Theorems 5.1 and 5.2 and uniqueness results are given in Theorems 5.3 and 5.4.

In the case $r \geq 3$ a Brownian motion in $\mathrm{Diff}^r(M)$ determines a real separable Hilbert space H continuously included in $C^{r-2}(TM)$ (or equivalently a covariance function $\alpha(x, y) \in T_xM \otimes T_yM$) and a vector field $X \in C^{r-3}(TM)$, and these specify it uniquely. H and X may be regarded as the information giving the diffusion and drift coefficients for a (Stratonovich) stochastic differential equation in M. Thus a link is achieved with the usual method of constructing Brownian motions in $\mathrm{Diff}^r(M)$ (or "stochastic flows of diffeomorphisms"), namely, as solutions of stochastic differential equations. See Section 6.3 for references. A result on the solutions of stochastic differential equations is used to show that any $H \subset C^{r+2}(TM)$ and $X \in C^{r+1}(TM)$ will do as the data for a Brownian motion in $\mathrm{Diff}^r(M)$, $r \geq 2$, and that for $r \geq 5$ all Brownian motions on $\mathrm{Diff}^r(M)$ occur as the solutions of stochastic differential equations.

The organisation of the paper is as follows. Some of the preliminary results are valid for general groups G. These are set out in Sections 2 and 3. In particular Section 3 exhibits the connection between one-parameter convolution semigroups $\{\mu_t:\ t > 0\}$ of Borel probability measures on G and time homogeneous processes with independent increments in G. In particular Theorem 3.1 gives criteria for the continuity or right continuity of sample paths in terms of the rate of convergence as $t \to 0$ of μ_t to the unit mass at the identity e in G. Thus from then on any result about a Brownian motion in $\mathrm{Diff}^r(M)$ can be thought of as a result about a one-parameter convolution semigroup $\{\mu_t:\ t \geq 0\}$ of Borel probability measures on $\mathrm{Diff}^r(M)$ satisfying $(1/t)\mu_t(\mathrm{Diff}^r(M) \setminus U) \to 0$ as $t \to 0$ for all neighbourhoods U of Id_M in $\mathrm{Diff}^r(M)$. Section 4 contains the main technical result of the paper. It concerns the infinitesimal generator of a particular semigroup of positive operators on the space of bounded continuous functions on a compact manifold. The results of Sections 2, 3 and 4 are then brought together to obtain the main results in Section 5. Section 6 provides an interpretation of H and X as coefficients in a stochastic differential equation, and also as the information to construct a Brownian motion $\{Z_t:\ t \geq 0\}$ in $C^{r-3}(TM)$, to be thought of as a "linearisation" of the original $\{\phi_t: t \geqslant 0\}$. Further details of the interplay between Brownian motions in $\mathrm{Diff}^r(M)$ and $C^r(TM)$ will be given in a forthcoming paper.

2. Polish groups

2.1. A *Polish* topological space F is one which is separable and can be metrised in such a way that it becomes complete. Let $\mathscr{B}(F)$ denote the Borel measurable subsets of F and $B(F)$ the Banach space of all bounded Borel measurable functions $F \to \mathbb{R}$ equipped with the supremum norm.

A *Polish group* G is a topological group which is a Polish topological space.

Examples

1. Any separable locally compact Hausdorff group.
2. $G = \mathrm{Homeo}(M) =$ homeomorphisms of M with compact-open topology, where M is compact metric or where M is locally compact, locally connected and 2nd countable. For details see Arens [1].
3. $G = \mathrm{Diff}^r(M) = C^r$ diffeomorphisms of a smooth finite-dimensional separable manifold M with the C^r compact-open topology $(1 \leqslant r \leqslant \infty)$.

As $\mathrm{Diff}^r(M)$ is our principal example we shall give some detail on it and some associated function spaces. Let M be a smooth m-dimensional separable manifold and let B be a separable Banach space. We let $C^r(M, B)$ denote the space of all C^r functions $f: M \to B$, equipped with the C^r compact-open topology, i.e. the topology of uniform C^r convergence on compact subsets of M. More precisely let $\{U_i\}$ be a countable cover of M by relatively compact open sets with $U_i \subset \bar{U}_i \subset V_i$ for charts (V_i, ϕ_i) for M. (Thus each V_i is open in M and each ϕ_i is a bijection of V_i onto an open set W_i in $\mathbb{R}^m$.) Consider the seminorms ρ_i on $C^r(M, B)$ defined by

$$\rho_i(f) = \sum_{k=0}^{r} \sup_{x \in W_i} \|D^k(f \circ \phi_i^{-1})(x)\|.$$

It can be shown that the topology generated by the seminorms ρ_i is independent of the choice of the U_i, V_i and ϕ_i; the topology obtained is the C^r compact-open topology. In fact the ρ_i generate a metric $\rho(f, g)$ which makes $C^r(M, B)$ a complete separable metric space. In particular if M is compact then $C^r(M, B)$ becomes a separable Banach space, with norm $\|f\|_r$ say.

Suppose now N is any other smooth finite-dimensional separable manifold. Then N may be embedded as a closed submanifold of some

Euclidean space $\mathbb{R}^k$, and we may consider $C^r(M, N) = \{ f \in C^r(M, \mathbb{R}^k)$: $f(M) \subset N \}$ with the relative topology inherited from $C^r(M, \mathbb{R}^k)$. It can be shown that the topological space $C^r(M, N)$ is independent of the choice of embedding. The topology is called the C^r compact-open topology. $C^r(M, N)$ is separable and complete with respect to the metric ρ inherited from $C^r(M, \mathbb{R}^k)$.

We may now define $\mathrm{Diff}^r(M) = \{ f \in C^r(M, M)$: f is a bijection and $f^{-1} \in C^r(M, M)\}$ with the relative topology from $C^r(M, M)$. The group multiplication is composition of diffeomorphisms, which is jointly continuous. (In fact the C^r compact open topology is designed so that

$$C^r(N, P) \times C^r(M, N) \to C^r(M, P)$$

$$(f, g) \mapsto f \circ g$$

is jointly continuous. See Franks [8] for a proof in the case where M is compact; also the discussion on topologies for function spaces in Kelley [12], p. 221.) It will follow that $\mathrm{Diff}^r(M)$ is a topological group in the C^r compact-open topology once we know that $f \to f^{-1}$ is continuous at Id_M. The same argument that works in case $G = \mathrm{Homeo}(M)$ shows that given any compact K in M there exists a neighbourhood U of Id_M in $\mathrm{Diff}^r(M)$ and a compact L in M such that $f \in U$ implies $f(L) \supset K$ and hence $f^{-1}(K) \subset L$. Thus knowledge of $f|_L$ gives knowledge of $f^{-1}|_K$. The fact that $f \to f^{-1}$ is continuous now follows from the equivalent result in the case that M is compact.

Finally if $\rho(f, g)$ is a complete metric on $C^r(M, M)$ define a metric $\bar{\rho}$ on $\mathrm{Diff}^r(M)$ by $\bar{\rho}(f, g) = \rho(f, g) + \rho(f^{-1}, g^{-1})$. Then the topology of $\bar{\rho}$ is still the C^r compact-open topology, but now $\mathrm{Diff}^r(M)$ is complete with respect to $\bar{\rho}$.

2.2. Any Polish group G may be metrized by a right-invariant metric d. It may or may not be complete with respect to d (or equivalently with respect to its right uniformity). However it is complete with respect to its two-sided uniformity. For details and further references see Kelley [12], pp. 210-212.

LEMMA 2.1: *Let d be a right-invariant metric on a Polish group G and let $(G, d) \overset{i}{\hookrightarrow} (\bar{G}, \bar{d})$ be a completion. Then*

(i) there is a multiplication on $\bar{G}$ extending the group multiplication of G and continuous with respect to $\bar{d}$,

(ii) if $g \in \bar{G} \setminus G$ and $h \in \bar{G}$ then $hg \in \bar{G} \setminus G$,

(iii) G is a G_δ in $\bar{G}$.

PROOF: Notice that (g_n) is a d-Cauchy sequence iff $g_n g_m^{-1} \to e$ as n, $m \to \infty$. Suppose (g_n) and (h_n) are d-Cauchy sequences. Thus $g_n g_m^{-1} \to e$

and $h_n h_m^{-1} \to e$ as $n, m \to \infty$. Let U be a neighbourhood of e. Then there exists a neighbourhood V of e such that $VVV^{-1} \subset U$. Choose N so that $n \geqslant N$ implies $g_n g_N^{-1} \in V$. Then choose M so that $m, n \geqslant M$ implies $h_n h_m^{-1} \in g_N^{-1} V g_N$. For $m, n \geqslant \max\{M, N\}$ we have

$$\begin{aligned}(g_n h_n)(g_m h_m)^{-1} &= (g_n g_N^{-1}) g_N (h_n h_m^{-1}) g_N^{-1} (g_m g_N^{-1})^{-1} \\ &\in VVV^{-1} \subset U.\end{aligned}$$

Therefore $(g_n h_n)$ is also a d-Cauchy sequence. Therefore we may define a multiplication in $\bar{G}$ by $(\lim g_n)(\lim h_n) = \lim g_n h_n$. It is clear that this is well defined and satisfies (i).

To prove (ii) it suffices to check that if $g, h \in \bar{G}$ and $hg = e$ then $g \in G$. So suppose $g = \lim g_n$ and $h = \lim h_n$ for d-Cauchy sequences (g_n) and (h_n). Then $g_n^{-1} g_m = (h_n g_n)^{-1} (h_n h_m^{-1})(h_m g_m) \to e$ as $n, m \to \infty$, so that both (g_n) and (g_n^{-1}) are d-Cauchy sequences. That is (g_n) is a Cauchy sequence in the two-sided uniformity and since G is complete in its two-sided uniformity, this implies $g \in G$ as required.

The result (iii) follows directly from the fact that G is a Polish space. □

As an example we can take $G = \text{Homeo}([0, 1])$ with the topology of uniform convergence. $\rho(f, g) = \sup_{0 \leqslant x \leqslant 1} |f(x) - g(x)|$ defines a right invariant metric ρ. Then $\bar{G}$ will be the closure of G as a subset of the Banach space $C([0, 1])$ of continuous real-valued function on [0,1] with the supremum norm.

3. Convolution semigroups of measures

3.1. Let $\{\mu_t\colon t > 0\}$ be a one-parameter convolution semigroup of Borel probability measures on a Polish group G. We may use translates of the measures μ_t to construct transition probabilities $P(t, g, \Gamma) = \mu_t(\Gamma g^{-1})$ of a Markov process $\{\phi_t\colon t \geqslant 0\}$ with values in G. The process has the following properties:

(i) independent increments on the left (i.e. if $0 \leqslant s < t \leqslant u < v$ then $\phi_v \phi_u^{-1}$ and $\phi_t \phi_s^{-1}$ are independent)

(ii) time homogeneous (i.e. the distribution of $\phi_t \phi_s^{-1}$ depends only on $t - s$).

Conversely if $\{\phi_t\colon t \geqslant 0\}$ is any stochastic process in G with the properties (i) and (ii) then letting μ_t = distribution of $\phi_t \phi_0^{-1}$ we obtain a one-parameter convolution semigroup $\{\mu_t\colon t > 0\}$. The convolution property $\mu_s * \mu_t = \mu_{t+s}$ corresponds exactly to the Chapman-Kolmogorov equation for transition probabilities.

It will be convenient to assume at this point that $\phi_0 = e$, the identity of G. This causes no real loss in generality.

The choice of $\mu_t(\Gamma g^{-1})$ rather than $\mu_t(g^{-1}\Gamma)$ implies that the independent increments are on the left rather than the right. The choice of left instead of right is fairly arbitrary at this stage though it is the more natural choice in the cases $G = \text{Homeo}(M)$ or $\text{Diff}^r(M)$.

THEOREM 3: *Let* $\{\mu_t: t > 0\}$ *and* $\{\phi_t: t \geqslant 0\}$ *be as above.*

(*i*) *The process* $\{\phi_t: t \geqslant 0\}$ *has a version with continuous sample paths if and only if* $(1/t)\mu_t(G \setminus U) \to 0$ *as* $t \to 0$ *for every neighbourhood* U *of* e.

(*ii*) *The process* $\{X_t: t \geqslant 0\}$ *has a version with sample paths continuous on right with limits on left if and only if* $\mu_t(G \setminus U) \to 0$ *as* $t \to 0$ *for every neighbourhood* U *of* e.

PROOF: Let d be a right invariant metric on G. It suffices to consider U of the form $U = \{g \in G: d(g, e) < \epsilon\}$

(i) only if: Let $\tau = \inf\{t: \phi_t \notin U\}$ = time of first exit of ϕ_t from U. Then $\mu_t(G \setminus U) = \mathbb{P}\{\phi_t \notin U\} \leqslant \mathbb{P}\{\tau \leqslant t\}$. Let $0 < t < 1$ and choose an integer n so that $1 \leqslant nt < 1 + t$. Then $n \geqslant 1/t$ so

$$
\begin{aligned}
(1 - \mathbb{P}\{\tau \leqslant t\})^{1/t} &= (\mathbb{P}\{t < \tau\})^{1/t} \\
&\geqslant (\mathbb{P}\{t < \tau\})^{n} \\
&= \mathbb{P}\left\{ \sup_{0 < s \leqslant t} \left\{ d\left(\phi_{kt+s}\phi_{kt}^{-1}, e\right)\right\} < \epsilon, 0 \leqslant k < n \right\} \\
&\geqslant \mathbb{P}\left\{ \sup_{\substack{0 \leqslant u < v \leqslant 2 \\ v - u \leqslant t}} \{ d(\phi_v, \phi_u)\} < \epsilon \right\} \\
&\to 1 \text{ as } t \to 0 \text{ since the process } \{\phi_s: s \geqslant 0\}
\end{aligned}
$$

has continuous sample paths. It follows that

$$
\frac{1}{t}\mu_t(G \setminus U) \leqslant \frac{1}{t}\mathbb{P}\{\tau \leqslant t\} \to 0 \text{ as } t \to 0.
$$

(ii) only if: This follows immediately from the right continuity of $\{\phi_t: t \geqslant 0\}$ at $t = 0$.

(i) if: Let $B(g, \epsilon) = \{h \in G: d(h, g) < \epsilon\}$, the open ball of radius ϵ and centre g. Then $B(g, \epsilon)g^{-1} = B(e, \epsilon)$ and it follows that

$$
\begin{aligned}
\frac{1}{t}P(t, g, G \setminus B(g, \epsilon)) &= \frac{1}{t}\mu_t(G - B(g, \epsilon)g^{-1}) \\
&= \frac{1}{t}\mu_t(G - B(e, \epsilon))
\end{aligned}
$$

which $\to 0$ as $t \to 0$ uniformly for $g \in G$. If G is complete with respect to the metric d the continuity of sample paths follows from the result of

Nelson [15]. However in general G is not complete with respect to d and an attempt on the estimate above with a complete but non right invariant metric is in danger of losing the uniformity in g of the estimate. Instead we work with the completion $(\bar{G}, \bar{d})$ of G in the metric d. Define Borel measures $\bar{\mu}_t$ on $\bar{G}$ by $\bar{\mu}_t(A) = \mu_t(A \cap G)$. Then we may apply the arguments above to obtain a Markov process $\{\bar{\phi}_t\colon t \geqslant 0\}$ taking values in the semigroup $\bar{G}$ with transition probabilities $\bar{P}(t, g, \Gamma) = \bar{\mu}_t\{h \in \bar{G}\colon hg \in \Gamma\}$. Let $\mathscr{F}_t$ denote the σ-algebra generated by $\{\bar{\phi}_s\colon 0 \leqslant s \leqslant t\}$. Since $\bar{G}$ is complete with respect to $\bar{d}$ we may suppose that $\{\bar{\phi}_t\colon t \geqslant 0\}$ has continuous sample paths. It remains to show that, with probability 1, $\bar{\phi}_t \in G$ for all $t \geqslant 0$. It is clear from the transition probabilities that for any fixed time t, $\mathbb{P}\{\bar{\phi}_t \in G\} = \bar{\mu}_t(G) = 1$. The idea is now that if a sample path leaves G at some time then (by (ii) of Lemma 2.1) it can never thereafter re-enter G.

Since G is a G_δ in $\bar{G}$, then $G = \bigcap_{n=1}^{\infty} G_n$ where each G_n is open in $\bar{G}$. It suffices to show that for each n, with probability 1, $\bar{\phi}_t \in G_n$ for all $t \geqslant 0$. Let τ be the first exit time of $\{\bar{\phi}_t\colon t \geqslant 0\}$ from G_n. Then τ is a Markov time and $\bar{\phi}_\tau \in \bar{G} \backslash G_n \subset \bar{G} \backslash G$ on $\{\tau < \infty\}$. If $\mathbb{P}\{\tau \leqslant t\} = 0$ for all $t \geqslant 0$ we are done. Otherwise, using the strong Markov property of the process $\{\bar{\phi}_t\colon t \geqslant 0\}$ and the result

$$\mathbb{P}\left\{\bar{\phi}_s \in \bar{G} \backslash G, 0 \leqslant s \leqslant t \mid \bar{\phi}_0 \in \bar{G} \backslash G_n\right\} = 1$$

(which follows from Lemma 2.1, part (ii)) we have

$$\begin{aligned} 0 &= \mathbb{P}\left\{\bar{\phi}_t \in \bar{G} \backslash G\right\} \\ &\geqslant \mathbb{P}\left\{\bar{\phi}_{\tau+s} \in \bar{G} \backslash G(0 \leqslant s \leqslant t), \tau \leqslant t\right\} \\ &= \mathbb{E}\left\{\mathbb{P}\left\{\bar{\phi}_{\tau+s} \in \bar{G} \backslash G(0 \leqslant s \leqslant t) | \mathscr{F}_\tau\right\} \chi_{\tau \leqslant t}\right\} \\ &= \mathbb{P}\{\tau \leqslant t\} \text{ as required.} \end{aligned}$$

(Here $\mathscr{F}_\tau$ denotes as usual the sub-σ-algebra generated by knowledge of the process $\bar{\phi}_t$ up to the Markov time τ.)

(ii) if: We use the same ideas with the stochastic continuity condition $\mu_t(G - U) \to 0$ to construct a strong Markov, right continuous (with limits on left) process $\{\bar{\phi}_t\colon t \geqslant 0\}$ in $\bar{G}$. See Dynkin [5], p. 127 and p. 116 for details. In this case $\{\tau > t\}$ may not be $\mathscr{F}_t$ measurable but it is $\mathscr{F}_t$-analytic or, in other words, Souslin-$\mathscr{F}_t$. See Rogers [17], p. 44, for details of the Souslin operation. Thus as the cost of passing to the completions of the $\mathscr{F}_t$ we obtain again the result that τ is a Markov time and the remainder of the proof proceeds as above. □

3.2. Let us suppose throughout this section that $\{\mu_t\colon t > 0\}$ is a convolution semigroup of Borel probability measures on G such that for all neighbourhoods U of e in G, $(1/t)\mu_t(G \setminus U) \to 0$ as $t \to 0$. Or equivalently, $\mu_t(\Gamma) = \mathbb{P}\{\phi_t \in \Gamma\}$ for some Brownian motion $\{\phi_t : t \geqslant 0\}$ in G.

LEMMA 3.1: *Given any neighbourhood U of e in G and $\epsilon > 0$, there exists $\delta > 0$ such that if $0 < t < \delta$ then $\mu_t(G \setminus U^n) \leqslant (t\epsilon)^n$, $n \geqslant 1$.*

PROOF Let τ_n = time of first exit of ϕ_t from U^n. Then, repeating the argument of Skorohod [18] and since $\{\phi_t : t \geqslant 0\}$ has continuous sample paths, we have

$$\begin{aligned}
\mathbb{P}\{\tau_n \leqslant t\} &= \mathbb{P}\{\phi_s \notin U^n \text{ for some } s \leqslant t\} \\
&\leqslant \mathbb{P}\Big\{\tau_{n-1} < t \text{ and } \phi_s \phi_{\tau_{n-1}}^{-1} \notin U \text{ for some } s \text{ with} \\
&\qquad \tau_{n-1} < s \leqslant t\Big\} \\
&\leqslant \mathbb{P}\{\tau_{n-1} < t\} \mathbb{P}\{\tau_1 < t\}
\end{aligned}$$

using the strong Markov property of the process at the Markov time τ_{n-1}. By induction we obtain

$$\mathbb{P}\{\tau_n \leqslant t\} \leqslant (\mathbb{P}\{\tau_1 < t\})^n.$$

Now the proof of "(i) only if" in Theorem 3.1 shows that $(1/t)\mathbb{P}\{\tau_1 < t\} \to 0$ as $t \to 0$. Take $\delta > 0$ so that $0 < t < \delta$ implies $(1/t)\mathbb{P}\{\tau_1 < t\} < \epsilon$. Then for $0 < t < \delta$

$$\begin{aligned}
\mu_t(G \setminus U^n) &= \mathbb{P}\{X_t \notin U^n\} \\
&\leqslant \mathbb{P}\{\tau_n \leqslant t\} \\
&\leqslant (\mathbb{P}\{\tau_1 < t\})^n \\
&\leqslant (t\epsilon)^n.
\end{aligned}$$

□

LEMMA 3.2: *Let d be a right invariant metric on G.*

(i) If $1 \leqslant p < \infty$,

$$\mathbb{E}(d(X_t, e))^p = \int (d(g, e))^p d\mu_t(g) \text{ is finite}$$

and tends to 0 as $t \to 0$ for each fixed p,

(ii) If $\alpha > 0$ there exists $\delta > 0$ such that for $0 < t < \delta$

$$\mathbb{E}(\exp\{\alpha d(X_t, e)\}) = \int \exp\{\alpha d(g, e)\} d\mu_t(g) < \infty$$

and tends to 1 as $t \to 0$ for each fixed p.

PROOF: For any monotone increasing f: $[0, \infty) \to \mathbb{R}$ with $f(0)=0$, any Borel measure ν on $[0, \infty)$ and any $\beta > 0$,

$$\begin{aligned}\int f(y)\,d\nu(y) &\leqslant \sum_{n=0}^{\infty} f((n+1)\beta)\,\nu((n\beta, (n+1)\beta)) \\ &\leqslant \sum_{n=0}^{\infty} \{f((n+1)\beta) - f(n\beta)\}\,\nu((n\beta, \infty)).\end{aligned}$$

(i) Take $f(y) = y^p$ and $\nu = \nu_t =$ distribution of $d(\phi_t, e)$. Fix $\beta > 0$ and $\epsilon = 1$. Taking $U = \{g \in G\colon d(g, e) < \beta\}$ in Lemma 3.1 we obtain $\delta > 0$ such that $0 < t < \delta$ implies that

$$\nu_t((n\beta, \infty)) = \mathbb{P}\{d(X_t, e) > n\beta\} \leqslant \mathbb{P}\{X_t \notin U^n\} < t^n.$$

Then if $t < \min(\delta, 1)$,

$$\begin{aligned}\mathbb{E}\big(d(\phi_t, e)^p\big) &= \int f(y)\,d\nu_t(y) \\ &\leqslant \sum_{n=0}^{\infty} \big((n+1)^p\beta^p - n^p\beta^p\big)t^n \\ &\leqslant p\beta^p \sum_{n=0}^{\infty} (n+1)^{p-1} t^n.\end{aligned}$$

Now $p\beta^p \sum_{n=0}^{\infty}(n+1)^{p-1}t^n < \infty$ for all $0 < t < 1$ and tends to $p\beta^p$ as $t \to 0$. Therefore $\mathbb{E}(d(\phi_t, e))^p$ is finite for $0 < t < \min(\delta, 1)$ and $\overline{\lim}_{t\to 0}\mathbb{E}(d(\phi_t, e))^p \leqslant p\beta^p$. But β is arbitrary, so $\mathbb{E}(d(\phi_t, e))^p \to 0$ as $t \to 0$. For general t, write $t = ru$ where $0 < u < \min(\delta, 1)$ and r is a positive integer. If $\psi_1, \dots, \psi_r$ are r independent random variables distributed as ϕ_u, then $\psi_1 \dots \psi_r$ has the same distribution as ϕ_t and

$$\begin{aligned}\mathbb{E}\big(d(\phi_t, e)\big)^p &= \mathbb{E}\big(d(\psi_1 \dots \psi_r, e)\big)^p \\ &\leqslant \mathbb{E}\left(\sum_{i=1}^{r} d(\psi_i \dots \psi_r, \psi_{i+1} \dots \psi_r)\right)^p \\ &= \mathbb{E}\left(\sum_{i=1}^{r} d(\psi_i, e)\right)^p \\ &\leqslant r^{p-1} \sum_{i=1}^{r} \mathbb{E}\big(d(\psi_i, e)\big)^p < \infty.\end{aligned}$$

(ii) We may apply the same technique to $f(y) = e^{\alpha y} - 1$. □

LEMMA 3.3: *Let $G \times B \to B$ be a continuous group action of G on a Banach space B such that for each $g \in G$, $b \mapsto gb$ is linear. Then for each $b \in B$ and $t > 0$, gb is μ_t integrable. Writing $P_t b = \int gb d\mu_t(g)$ we have $P_t \in L(B)$, $P_t P_s = P_{t+s}$, $\|P_t\| \leqslant ce^{dt}$ for some constants c, $d < \infty$ and $\|P_t b - b\| \to 0$ as $t \to 0$ for each fixed $b \in B$.*

PROOF: Since $(g, b) \mapsto gb$ is continuous there exists a neighbourhood U of e in G and $k > 1$ such that $g \in U$, $\|b\| \leqslant 1/k$ implies $\|gb\| \leqslant 1$. Therefore $g \in U$ implies $\|g\| \leqslant k$ (where $\|g\|$ denotes the operator norm on g as a bounded operator on B). Moreover $g \in U^n$ implies $\|g\| \leqslant k^n$. So for $b \in B$ and $t > 0$,

$$\int \|gb\| d\mu_t(g) \leqslant k\|b\|\mu_t(U) + \sum_{n=1}^{\infty} k^{n+1}\|b\|\mu_t(U^{n+1} \setminus U^n)$$

$$\leqslant k\|b\| + \sum_{n=1}^{\infty} (k^{n+1} - k^n)\|b\|\mu_t(G \setminus U^n).$$

By Lemma 3.1 there exists $\delta_0 > 0$ such that $0 < t < \delta_0$ implies $\mu_t(G \setminus U^n) < (t/2k)^n$. Then for $0 < t < \delta = \min(\delta_0, 1)$ we have

$$\int \|gb\| d\mu_t(g) \leqslant \left(k + (k-1) \sum_{n=1}^{\infty} k^n (t/2k)^n \right) \|b\|$$

$$\leqslant 2k\|b\|.$$

It follows that for $0 < t < \delta$, P_t is defined as a linear operator on B satisfying

$$\|P_t b\| \leqslant \int \|gb\| d\mu_t(g) \leqslant 2k\|b\|,$$

i.e. $\|P_t\| \leqslant 2k$. If $t \geqslant \delta$ we may write $t = t_1 + \ldots + t_r$ with $t_i < \delta$ each i. Then (writing μ_i for μ_{t_i}) we have

$$\int \|gb\| d\mu_t(g) = \int \|g_1 \ldots g_r b\| d\mu_1(g_1) \ldots d\mu_r(g_r)$$

$$\leqslant (2k)^r \|b\|$$

by repeated application of the result above. The exponential bound on $\|P_t\|$ follows immediately. The semigroup property $P_t P_s = P_{t+s}$ is an immediate consequence of the convolution property $\mu_t * \mu_s = \mu_{t+s}$.

Finally suppose $b \in B$ and $\epsilon > 0$ are given. There exists a neighbourhood U of e in G such that $g \in U$ implies $\|g\| \leqslant k$ (as above) and also

$\|gb - b\| < \epsilon$. Then for $g = g_1 \ldots g_n$ with $g_i \in U$, $i = 1, \ldots, n$ we have

$$gb - b = \sum_{i=1}^{n} g_1 \ldots g_{i-1}(g_i b - b),$$

so

$$\|gb - b\| \leqslant \sum_{i=1}^{n} k^{i-1}\epsilon = \frac{k^n - 1}{k - 1}\epsilon$$

(where we assume without loss of generality $k > 1$.) Then

$$\|P_t b - b\| \leqslant \int \|gb - b\| d\mu_t(g)$$

$$\leqslant \epsilon + \sum_{n=1}^{\infty} k^n \epsilon \mu_t(G \setminus U^n).$$

Arguing as before we obtain $\overline{\lim}_{t \to 0} \|P_t b - b\| \leqslant \epsilon$. But ϵ is arbitrary, so $\|P_t b - b\| \to 0$ as $t \to 0$.

The result fails if the condition on the $\{\mu_t\}$ is weakened to $(1/t)\mu_t(G \setminus U)$ bounded as $t \to 0$ for each neighbourhood U of e in G. For example take $G = B = \mathbb{R}$, $gb = e^g b$ and $\mu_t = e^{-t}\sum_{n=0}^{\infty} t^n (*^n \mu)/n!$ where the measure μ has support $[1, \infty)$ with $d\mu/dx = 1/x^2 (x \geqslant 1)$. Then if $U \subset (-\infty, 1)$, $(1/t)\mu_t(G - U) = (1/t)(1 - e^{-t}) \to 1$ and

$$\int gb\, d\mu_t(g) \geqslant be^{-t} t \int e^g d\mu(g) = \infty.$$

It also fails if the Banach space B is replaced by a Fréchet space F. For example take $G = \mathbb{R}$ acting by translation on $C(\mathbb{R})$, the continuous functions on $\mathbb{R}$ with the compact-open topology. Let $\mu_t = N(0, t)$, the Gaussian measure of mean 0 and variance t. If $f(x) = \exp(x^3)$ then $\int |(gf)(x)| d\mu_t(g) = \infty$ for all $x \in \mathbb{R}$, $t > 0$.

In general, it is not true that $\|P_t - I\| \to 0$ as $t \to 0$. This stronger result would follow from the stronger hypothesis that $G \to L(B)$ is continuous for the norm topology on $L(B)$.

4. Infinitesimal generators of semigroups

4.1. Let $\{P_t : t \geqslant 0\}$ be a one-parameter semigroup of bounded linear operators on a Banach space B with $P_0 = I$. We define the infinitesimal generator A of the semigroup by

$$Ab = \lim_{t \downarrow 0} \frac{1}{t}(P_t b - b),$$

i.e. A is the linear operator whose domain is the set $D(A) = \{b \in B: \lim_{t \downarrow 0}(1/t)(P_t b - b) \text{ exists}\}$, and, for $b \in D(A)$, Ab is given as above.

If $P_t b \to b$ as $t \to 0$ for all $b \in B$ we say $\{P_t: t \geqslant 0\} \subset L(B)$ is a semigroup of class (C_0) and it follows that

(i) $\|P_t\| \leqslant ce^{dt}$ for constants $c, d < \infty$

(ii) A is a closed operator and $D(A)$ is dense in B. (See Yosida [22], Chapter IX.) More generally if $\|P_t\|$ is bounded as $t \to 0$ then $B_0 = \{b \in B: P_t b \to b \text{ as } t \to 0\}$ is a closed subspace of B which is invariant under P_t for all $t > 0$. It follows from consideration of the semigroup $\{P_t|_{B_0}: t \geqslant 0\} \subset L(B_0)$ that A is a closed operator and $D(A)$ is a dense subspace of B_0.

4.2. Let M be a smooth compact m-dimensional manifold. We abbreviate $C(M, \mathbb{R})$ to $C(M)$ (with norm $\| \ \|$) and $C^r(M, \mathbb{R})$ to $C^r(M)$ (with norm $\| \ \|_r$). We recall that an operator P on $C(M)$ is said to be positive if $f(x) \geqslant 0$ all $x \in M$ implies $(Pf)(x) \geqslant 0$ all $x \in M$. A positive linear operator on $C(M)$ is automatically bounded with $\|P_t\| \leqslant \|P_t 1\|$ where $1 \in C(M)$ is given by $1(x) = 1$ all $x \in M$.

If $(u^1, \ldots, u^m)$ is a local coordinate system in M and $f \in C^2(M)$ we shall write $(\partial f)/\partial u^i)(x) = f_i(x)$ and $(\partial^2 f)/(\partial u^i \partial u^j) = f_{ij}(x)$.

THEOREM 4.1: *Suppose $\{P_t: t \geqslant 0\}$ is a family of positive linear operators on $C(M)$ satisfying*

(i) $P_t P_s = P_{t+s}$ for $t, s > 0$, $P_0 = I$.

(ii) $(1/t)((P_t f)(a) - f(a)) \to 0$ as $t \to 0$ whenever $f \in C(M)$ is zero in a neighbourhood of $a \in M$

(iii) there exists $r \geqslant 2$ such that $P_t(C^r(M)) \subset C^r(M)$ for all $t > 0$ and $\|P_t f - f\|_r \to 0$ as $t \to 0$ for each $f \in C^r(M)$.

Then $\{P_t: t \geqslant 0\}$ is a semigroup of class (C_0) with infinitesimal generator A such that $C^2(M) \subset D(A)$ and in any local coordinate system $(u^1, u^2, \ldots, u^m)$ in an open set U in M there exist C^{r-2} functions α^{ij} $(1 \leqslant i, j \leqslant m)$, β^i $(1 \leqslant i \leqslant m)$ and γ such that

$$(Af)(x) = \tfrac{1}{2}\sum_{i,j} \alpha^{ij}(x) f_{ij}(x) + \sum_i \beta^i(x) f_i(x) + \gamma(x) f(x)$$

for $f \in C^2(M)$, $x \in U$. Moreover for each x the matrix $[\alpha^{ij}(x)]$ is positive semi-definite.

PROOF:

Step 1. For any $f \in C(M)$ we have $-\|f\|1 \leqslant f \leqslant \|f\|1$. Therefore using the positivity of P_t and (iii) we have $\|P_t f\| \leqslant \|f\| \|P_t 1\| \leqslant \|f\| \|P_t 1\|_r \to \|f\| \|1\|_r$ as $t \to 0$. So $\|P_t\|$ remains bounded as $t \to 0$. Since P_t is continuous as a linear operator on $C(M)$ and since $P_t(C^r(M)) \subset C^r(M)$ it follows from the closed graph theorem that P_t is continuous as a linear operator on $C^r(M)$. Therefore $\{P_t\colon t \geqslant 0\}$ is a semigroup of class (C_0) when restricted to act as a family of operators on $C^r(M)$. If ${}^rA$ denotes the infinitesimal generator of the P_t acting on $C^r(M)$ then ${}^rA$ is a closed operator with domain $D({}^rA)$ dense in $C^r(M)$.

Step 2. Clearly $f \in D({}^rA)$ implies $f \in D(A)$. Therefore $D(A) \cap C^r(M)$ is C^r-dense in $C^r(M)$. But $C^r(M)$ is C^2-dense in $C^2(M)$ so $D(A) \cap C^r(M)$ is C^2-dense in $C^2(M)$, and a fortiori $D(A) \cap C^2(M)$ is C^2-dense in $C^2(M)$.

Notice also that $D(A)$ must be dense in $C(M)$, so $\{f \in C(M)\colon P_t f \to f$ as $t \to 0\}$ is dense in $C(M)$. But $\{f \in C(M)\colon P_t f \to f \text{ as } t \to 0\}$ is a closed subspace of $C(M)$ (since $\|P_t\|$ bounded as $t \to 0$) so we must have $P_t f \to f$ as $t \to 0$ for all $f \in C(M)$. So $\{P_t\colon t \geqslant 0\}$ is a semigroup of class (C_0) on $C(M)$.

Step 3. Fix $a \in M$ and fix a coordinate system $(u^1, u^2, \ldots, u^m)$ in a neighbourhood U of a. Suppose a has coordinates $(b^1, b^2, \ldots, b^m)$. Since $D({}^rA)$ is a C^r-dense subspace of $C^r(M)$ there exists $\phi \in D({}^rA)$ such that $\phi(a) = 0$, $\phi_{,i}(a) = 0$ $(1 \leqslant i \leqslant m)$, $\phi_{,ij}(a) = 2\delta_{ij}$ $(1 \leqslant i, j \leqslant m)$, the matrix $\varphi_{,ij}(x)$ is positive definite in some compact neighborhood K of a and $\varphi(x) > 0$ for $x \in M \setminus \operatorname{int} K$. Then $\varphi(x) > 0$ for $x \neq a$ and in particular $\phi(x) \sim \Sigma_i (u^i - b^i)^2$ as $x \to a$.

Now suppose $f \in C(M)$ and $f(x)/\phi(x) \to 0$ as $x \to a$. Given $\epsilon > 0$ we may write $f = f_1 + f_2$ where $f_1, f_2 \in C(M)$, f_1 is zero in some neighbourhood of a and $-\epsilon\phi(x) \leqslant f_2(x) \leqslant \epsilon\phi(x)$. Then $(1/t)(P_t f_1)(a) \to 0$ by (ii) and $(1/t)|(P_t f_2)(a)| \leqslant (1/t)\epsilon(P_t\phi)(a) \to \epsilon(A\phi)(a)$ as $t \to 0$. But $\epsilon > 0$ is arbitrary so $(1/t)(P_t f)(a) \to 0$ as $t \to 0$.

Step 4. Choose w, w^k $(1 \leqslant k \leqslant m)$ and w^{kl} $(1 \leqslant k \leqslant l \leqslant m)$ in $D({}^rA)$ so as to agree up to 2nd order at a with 1, $u^k - b^k$ $(1 \leqslant k \leqslant m)$ and $(u^k - b^k)(u^l - b^l)$ $(1 \leqslant k \leqslant l \leqslant m)$ respectively. Define $w^{kl} = w^{lk}$ for $k > l$. Then any $f \in C^2(M)$ may be written in the form

$$f(x) = f(a)w(x) + \sum_i f_{,i}(a)w^i(x) + \tfrac{1}{2}\sum_{i,j} f_{,ij}(a)w^{ij}(x) + h(x)$$

where $h \in C(M)$ and $h(x)/\phi(x) \to 0$ as $x \to a$. In particular if $f \in C^2(M) \cap D(A)$ we may apply the operation $g \mapsto \lim_{t \downarrow 0}((P_t g)(a) - g(a))/t$ to the expression above. We obtain

$$(Af)(a) = \gamma(a)f(a) + \sum_i \beta^i(a)f_{,i}(a) + \tfrac{1}{2}\sum_{i,j}\alpha^{ij}(a)f_{,ij}(a) \qquad (*)$$

where $\alpha^{ij}(a) = (Aw^{ij})(a)$, $\beta^i(a) = (Aw^i)(a)$, $\gamma(a) = (Aw)(a)$. We note that $(Aw)(a)$, $(Aw^i)(a)$ and $(Aw^{ij})(a)$ are independent of the choices made of w, w^i and w^{ij}. This may be checked by substituting any other choices into equation $(*)$.

Step 5. Clearly we can repeat the above analysis at any $x \in U$ to obtain functions $\alpha^{ij}(x)$, $\beta^i(x)$, $\gamma(x)$ from U to $\mathbb{R}$. For $x \in U$ we have

$$(Aw)(x) = \gamma(x)w(x) + \sum_i \beta^i(x)w_{,i}(x) + \sum_{i<j} \alpha^{ij}(x)w_{,ij}(x)$$
$$+ \tfrac{1}{2}\sum_i \alpha^{ii}(x)w_{,ii}(x)$$

$$(Aw^k)(x) = \gamma(x)w^k(x) + \sum_i \beta^i(x)w^k_{,i}(x) + \sum_{i<j} \alpha^{ij}(x)w^k_{,ij}(x)$$
$$+ \tfrac{1}{2}\sum_i \alpha^{ii}(x)w^k_{,ii}(x) \qquad (1 \leqslant k \leqslant m)$$

$$(Aw^{kl})(x) = \gamma(x)w^{kl}(x) + \sum_i \beta^i(x)w^{kl}_{,i}(x) + \sum_{i<j} \alpha^{ij}(x)w^{kl}_{,ij}(x)$$
$$+ \tfrac{1}{2}\sum_i \alpha^{ii}(x)w^{kl}_{,ii}(x) \qquad (1 \leqslant k \leqslant l \leqslant m).$$

Together we have $1 + m + \frac{1}{2}m(m+1)$ equations. The left hand sides are C^r functions of x; on the right hand sides the w, w^k, w^{kl} and their partial derivatives are C^{r-2} functions of x which we may regard as the coefficients in a linear transformation of $\gamma(x)$, $\beta^i(x)$ $(1 \leqslant i \leqslant m)$ and $\alpha^{ij}(x)$ $(1 \leqslant i \leqslant j \leqslant m)$ into $(Aw)(x)$, $(Aw^k)(x)$ $(1 \leqslant k \leqslant m)$ and $Aw^{kl}(x)$ $(1 \leqslant k \leqslant l \leqslant m)$. At $x = a$ this linear transformation is the identity, hence for x sufficiently near a it has an inverse with C^{r-2} coefficients. Thus for x sufficiently near a the functions $\gamma(x)$, $\beta^i(x)$, $\alpha^{ij}(x)$ are linear combinations of $(Aw)(x)$, $(Aw^k)(x)$, $(Aw^{kl})(x)$ where the coefficients in the linear combinations are C^{r-2} functions of x. Therefore γ, β^i and α^{ij} are C^{r-2} functions.

Step 6. Let $\tilde{A}$ be the operator on $C^2(M)$ defined, with respect to any local coordinate system, by

$$(\tilde{A}f)(x) = \tfrac{1}{2}\sum_{i,j} \alpha^{ij}(x)f_{,ij}(x) + \sum_i \beta^i(x)f_{,i}(x) + \gamma(x)f(x).$$

Then for $f \in C^2(M) \cap D(A)$ we have $Af = \tilde{A}f$. Let $f \in C^2(M)$. There exist f_n in $C^2(M) \cap D(A)$ such that $f_n \to f$ in $C^2(M)$. Then

$$Af_n = \tilde{A}f_n \to \tilde{A}f \text{ in } C(M).$$

But also $f_n \to f$ in $C(M)$ and A is a closed operator, so $f \in D(A)$. Thus we have shown that $C^2(M) \subset D(A)$.

Step 7. Given $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$, there exists $\psi \in C^2(M)$ such that $\psi(x) \geqslant 0$ for $x \in M$ and $\psi(x) = (\sum_{i=1}^m \lambda_i(x^i - a^i))^2$ in some neighbourhood of

a. Then $\psi \in D(A)$, $\psi(a)=0$ and $(P_t\psi)(a) \geqslant 0$, so that $(A\psi)(a) = \lim_{t\downarrow 0}((P_t\psi)(a)-\psi(a))/t \geqslant 0$. Also $\psi(a)=0$, $\psi_{,i}(a)=0$ and $\psi_{,ij}(a) = 2\lambda_i\lambda_j$, so that

$$\sum \alpha^{ij}(a)\lambda_i\lambda_j = (A\phi)(a) \geqslant 0.$$

This shows that for each a the matrix $[\alpha^{ij}(a)]_{i,j}$ is positive semidefinite. □

The main idea in the proof is in the use that is made of the denseness of the domain $D('A)$. The method used is an adaptation of techniques used by Hunt in [10].

For future reference we shall define

$$\alpha(x) = \sum_{i,j} \alpha^{ij}(x)\frac{\partial}{\partial u^i} \otimes \frac{\partial}{\partial u^j} \in T_xM \otimes T_xM.$$

It is easy to check that this defines $\alpha(x)$ independently of the coordinate system. If ξ, $\eta \in T_x^*M$ and if f, $g \in C^2(M)$ satisfy $df(x)=\xi$, $dg(x)=\eta$ then

$$A(fg)(x) - f(x)(Ag)(x) - g(x)(Af)(x) = \alpha(x)(\xi, \eta).$$

(This may be used as an equivalent coordinate-free definition of α.) Thus $\frac{1}{2}\alpha(x)$ is the symmetric bilinear form on T_x^*M whose corresponding quadratic form is the symbol of A. α is a C^{r-2} section of the bundle $TM \otimes TM$ over M.

4.3. We shall apply Theorem 4.1 in the case where $P_t1 = 1$ for all $t > 0$. In order to handle this case we have the following result. It helps to explain why we wrote $(1/t)((P_tf)(a) - f(a))$ rather than just $(1/t)((P_tf)(a))$ in condition (ii) of Theorem 4.1.

PROPOSITION 4.1: *In the situation of Theorem* 4.1 *the following are equivalent*:

(*i*) $P_t1 = 1$ *for all* $t \geqslant 0$.

(*ii*) *In condition* (*ii*) *of Theorem* 4.1, "*zero*" *may be replaced by* "*constant*".

(*iii*) $\gamma(x) = 0$ *for all* $x \in M$.

PROOF: It is clear that (i) ⇒ (ii) ⇒ (iii) (since $\gamma(x) = (A1)(x)$). Condition (iii) implies that given $\epsilon > 0$ there exists $\delta > 0$ such that if $0 < t < \delta$ then $\|P_t1 - 1\| < \epsilon t$,

$$\text{i.e. } 1 - \epsilon t < (P_t1)(x) < 1 + \epsilon t \quad \text{for all } x \in M.$$

Since P_s is a positive operator, if $0 < s,\ t < \delta$ we have

$$(1-\epsilon t)(1-\epsilon s) \leqslant (P_{t+s}1)(x) \leqslant (1+\epsilon t)(1+\epsilon s).$$

For any $t > 0$ we have for $n > t/\delta$

$$\left(1-\frac{\epsilon t}{n}\right)^n \leqslant (P_t 1)(x) \leqslant \left(1+\frac{\epsilon t}{n}\right)^n \quad \text{for all } x \in M.$$

Letting $n \to \infty$, we have

$$e^{-\epsilon} \leqslant (P_t 1)(x) \leqslant e^{\epsilon}.$$

But $\epsilon > 0$ is arbitrary so $P_t 1 = 1$. □

4.4. The statement and conclusions of the theorem would remain valid if the smooth manifold M was replaced by a C^r manifold. In this case, by looking at the way the coefficients α^{ij}, β^i and γ transform under a C^r change of coordinates, we see that the theoretically best possible result would be to assert that $\alpha^{ij} \in C^{r-1}(U)$, $\beta^i \in C^{r-2}(U)$ and $\gamma \in C^r(M)$. It can be shown that this is true if $1 \in D({}^rA)$ (and in particular if the conditions of Proposition 4.1 hold) and if we can choose the w^k of step 4 of the proof in such a way that $w^k w^l \in D({}^rA)$ for $1 \leqslant k,\ l \leqslant M$. For then we have

$$\begin{aligned} A(w^k w^l)(x) - w^k(x)(Aw^l)(x) - w^l(x)(Aw^k)(x) \\ + w^k(x)w^l(x)\gamma(x) \\ = \sum_{i,j} \alpha^{ij}(x) w_i^k(x) w_j^l(x). \end{aligned}$$

The argument used in step 5 of the proof can now be adapted to show that the α^{ij} $(1 \leqslant i, j \leqslant M)$ are C^{r-1} functions.

4.5. The question remains as to whether knowledge that $D(A) \supset C^2(M)$ and of $\tilde{A} = A|_{C^2(M)}$ is sufficient to determine A and hence the semigroup P_t uniquely. In this section we give some answers. Throughout $\{P_t\colon t \geqslant 0\}$ satisfy the conditions of Theorem 4.1.

LEMMA 4.1: *If $f \in D(A)$ has a maximum at a then $Af(a) \leqslant \gamma(a)f(a)$.*

PROOF: Let $g(x) = f(a) - f(x)$, then $g \in D(A)$, $g(a) = 0$ and $g(x) \geqslant 0$ for $x \in M$. Since the P_t are positive operators,

$$f(a)\gamma(a) - (Af)(a) = (Ag)(a) \geqslant 0.$$ □

COROLLARY 4.1: *If* $f \in D(A)$ *and* $Af = \lambda f$ *for some* $\lambda > \max_{x \in M} \gamma(x)$ *then* $f = 0$.

PROOF: Suppose not. Then without loss of generality f has a positive maximum at a, and we have

$$Af(a) \leqslant \gamma(a) f(a) < \lambda f(a) = Af(a). \qquad \square$$

PROPOSITION 4.2: $D({}^rA) = \{ f \in C^r(M) \colon Af \in C^r(M) \}$.

PROOF: Note that $C^r(M) \subset C^2(M) \subset D(A)$ so if $f \in C^r(M)$ then Af is defined. If $f \in D({}^rA)$ then $f \in C^r(M)$ and $Af = ({}^rA)f \in C^r(M)$. To prove the reverse inclusion we appeal to the Hille-Yosida semigroup theory. Since $\{P_t \colon t \geqslant 0\}$ acts as a semigroup of class (C_0) on $C^r(M)$ there exist $c, d < \infty$ such that $\|P_t\|_r \leqslant ce^{dt}$. It follows that for $\lambda > d$, $\lambda I - {}^rA$ is an injective mapping of $D({}^rA)$ onto $C^r(M)$ with inverse R_λ mapping $C^r(M)$ onto $D({}^rA)$. Suppose $f \in C^r(M)$ and $Af \in C^r(M)$. Choose λ such that $\lambda > d$ and $\lambda > \sup_{x \in M} \gamma(x)$. Let $g_\lambda = \lambda f - Af \in C^r(M)$ and $f_\lambda = R_\lambda(g_\lambda) \in D({}^rA)$. Then

$$\lambda f_\lambda - Af_\lambda = (\lambda I - {}^rA) f_\lambda = g_\lambda = \lambda f - Af,$$

so that $A(f - f_\lambda) = \lambda (f - f_\lambda)$, which implies $f = f_\lambda \in D({}^rA)$ as required. $\square$

COROLLARY 4.2: *The semigroup* $\{P_t \colon t \geqslant 0\}$ *is uniquely determined by* $A|_{C^2(M)}$.

PROOF: The proposition shows that ${}^rA$ is uniquely determined, so that the action of P_t on $C^r(M)$ is uniquely determined. But $C^r(M)$ is dense in $C(M)$ and each P_t is a bounded operator on $C(M)$ so P_t is uniquely determined. $\square$

This shows that $\{P_t \colon t \geqslant 0\}$ is uniquely determined by $A|_{C^2(M)}$ amongst the class of semigroups satisfying conditions (i)–(iii) of Theorem 4.1. It does not prove uniqueness amongst the class of all semigroups of bounded operators on $C(M)$. We obtain uniqueness amongst a somewhat larger class of semigroups with the following result.

PROPOSITION 4.3: *Let* $\{Q_t \colon t \geqslant 0\}$ *be a semigroup of class* (C_0) *on* $C(M)$ *with infinitesimal generator* B *such that* $C^2(M) \subset D(B)$ *and* $B|_{C^2(M)} = A|_{C^2(M)}$. *Suppose also* $Q_t(C^\infty(M)) \subset C^2(M)$ *for all* $t > 0$. *Then* $Q_t = P_t$ *for all* $t \geqslant 0$.

PROOF: Since Q_t and P_t are bounded operators on $C(M)$ it suffices to show $Q_t f = P_t f$ for $f \in C^\infty(M)$. So suppose $f \in C^\infty(M)$ and let $\lambda \geqslant \max_{x \in M} \gamma(x)$. Define $f(t, x) = e^{-\lambda t}((P_t f)(x) - (Q_t f)(x))$. Appealing to the theory of semigroups of class (C_0) (see Yosida [22], chapter IX, section 3) we see that f is jointly continuous in t and x, differentiable in t for each x, C^2 in x for each t, and

$$\frac{\partial f}{\partial t}(t,x) = -\lambda f(t,x) + e^{-\lambda t}((AP_t f)(x) - (BQ_t f)(x))$$

$$= ((A-\lambda)f)(t,x), \qquad (t>0)$$

$$f(0,x) = 0.$$

We must show $f(t,x) \equiv 0$. So suppose without loss of generality that $f(t,x) > 0$ for some (t,x). Then there exists an open interval $I = (a,b) \subset (0,\infty)$ such that $f(a,x) \leqslant 0$ for all $x \in M$ and $\max_{x\in M} f(t,x) > 0$ for $t \in I$. At any $(t,y) \in I \times M$ such that $f(t,y) = \max_{x\in M} f(t,x)$ we have

$$\frac{\partial f}{\partial t}(t,y) = (Af)(t,y) - \lambda f(t,y) \leqslant 0.$$

It follows, using a result of Hunt [10] section 1.8, that $\max_{x\in M} f(t,x) \leqslant \max_{x\in M} f(a,x)$ for $t \in I$, which gives the desired contradiction. □

We return to this question in Section 5.4 where the operators P_t arise from a Markov family and we require uniqueness of the Markov family. However we should draw attention to the example in Stroock and Varadhan [20], p. 170, wherein the operator

$$\tilde{A}f = \tfrac{1}{2}\min(|x|^{2\alpha}, 1)f''(x), \qquad f \in C_c^2(\mathbb{R})$$

for $0 < \alpha < 1$ is the restriction to $C_c^2(\mathbb{R})$ ($= C^2$ functions $\mathbb{R} \to \mathbb{R}$ with compact support) of the infinitesimal generators of uncountably many different semigroups of bounded operators on $C(\mathbb{R})$. It is clear that this example can exist equally well if $\mathbb{R}$ is replaced by $M = S^1$.

5. The main theorems

5.1. Let M be a smooth compact manifold of dimension m. For $n \geqslant 1$, $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in M^n$ and $\phi\colon M \to M$ let $\phi(\boldsymbol{x}) = (\phi(x_1), \phi(x_2), \ldots, \phi(x_n)) \in M^n$. If A is a differential operator on M and $f\colon M^n \to \mathbb{R}$ is a sufficiently differentiable function we shall let $A_i f$ denote the action of A on the ith variable only. More precisely if $\boldsymbol{a} \in M^n$ and $g(x_i) = f(a_1, \ldots, a_{i-1}, x_i, a_{i+1}, \ldots, a_n)$ then $(A_i f)(\boldsymbol{a}) = (Ag)(a_i)$. In particular if $f \in C^2(M^n)$ then $d_j f(\boldsymbol{x}) \in T^*_{x_j}M$ and $d_i d_j f(\boldsymbol{x}) \in T^*_{x_i}M \otimes T^*_{x_j}M$ for $i \neq j$. (Notice that, for $i \neq j$, $d_i d_j f(\boldsymbol{x})$ is defined just in terms of the differential structure of M; it is only the case $i = j$ which requires a connection on M.)

If Y is a continuous vector field on M we may think of it as a 1st order differential operator. If $f \in C^1(M^n)$ we define $Yf \in C(M^n)$ by $(Yf)(\boldsymbol{x}) = \sum_{i=1}^n (Y_i f)(\boldsymbol{x}) = \sum_{i=1}^n Y(x_i)(d_i f)(\boldsymbol{x})$. Alternatively let ϕ_t be a flow along the vector field Y defined in some neighbourhood of $\{0\} \times \{x_1, \dots, x_n\}$ in $\mathbb{R} \times M$. (We know that ϕ_t exists although it may not be unique if Y is not Lipschitz.) Then $\phi_t(\boldsymbol{x})$ is defined for t in a neighbourhood of 0. Let $(Yf)(\boldsymbol{x}) = (d/dt)(f(\phi_t(\boldsymbol{x})))|_{t=0}$. It is easy to check that these two definitions of Yf agree.

5.2. Suppose $\{\phi_t\colon t \geqslant 0\}$ is a Brownian motion in $\mathrm{Diff}^r(M)$, i.e. the process $\{\phi_t\colon t \geqslant 0\}$ satisfies conditions (i)–(iv) in the introduction. Our aim is to classify Brownian motions in $\mathrm{Diff}^r(M)$. We shall consider two Brownian motions equivalent if they induce the same probability measure on the space of continuous paths $[0, \infty) \to \mathrm{Diff}^r(M)$. This notion of equivalence coincides with the usual one of equivalence of random processes in terms of finite dimensional distributions because of the continuity in the time and space variables of a Brownian motion process.

For each $n \geqslant 1$ and $\boldsymbol{x} = (x_1, x_2, \dots, x_n) \in M^n$ we obtain a random process $\{\phi_t(\boldsymbol{x})\colon t \geqslant 0\}$ taking values in M^n, satisfying $\phi_0(\boldsymbol{x}) = \boldsymbol{x}$. Since $\phi_{t+s}(\boldsymbol{x}) = \phi_{t+s}\phi_t^{-1}(\phi_t(\boldsymbol{x}))$ it follows from the independence of $\phi_{t+s}\phi_t^{-1}$ and ϕ_t that each $\{\phi_t(\boldsymbol{x})\colon t \geqslant 0\}$ is a Markov process in M^n and that the family of processes $\{\phi_t(\boldsymbol{x})\colon t \geqslant 0\}$, $\boldsymbol{x} \in M^n$, is a Markov family of random processes for each fixed n. The original process $\{\phi_t\colon t \geqslant 0\}$ is time homogeneous; therefore the $\{\phi_t(\boldsymbol{x})\colon t \geqslant 0\}$ processes are all time homogeneous.

For fixed n let $P_n(t, \boldsymbol{x}, \Gamma)$, $t > 0$, $\boldsymbol{x} \in M^n$, $\Gamma \in \mathscr{B}(M^n)$, be the transition probability function for the Markov family $\{\phi_t(\boldsymbol{x})\colon t \geqslant 0\}$, $\boldsymbol{x} \in M^n$. Thus $P_n(t, \boldsymbol{x}, \Gamma) = \mathbb{P}\{\phi_t(\boldsymbol{x}) \in \Gamma\}$. We may now define $(P_{n,t}f)(\boldsymbol{x}) = \mathbb{E}(f(\phi_t(\boldsymbol{x}))) = \int_{M^n} f(y) P_n(t, \boldsymbol{x}, dy)$ for $f \in B(M^n)$. Then $P_{n,t}1 = 1$. Also $f(\boldsymbol{x}) \geqslant 0$, all $\boldsymbol{x} \in M^n$, implies $(P_{n,t}f)(\boldsymbol{x}) \geqslant 0$, all $\boldsymbol{x} \in M^n$. The Markov property of the process $\{\phi_t(\boldsymbol{x})\colon t \geqslant 0\}$ implies that $P_{n,t+s}f = P_{n,t}(P_{n,s}f)$ for all $s, t \geqslant 0, f \in B(M)$. In general a Markov process is characterised by its transition probabilities (together with the initial condition) and thus by the semigroup $\{P_{n,t}\colon t \geqslant 0\}$ of positive contraction operators on $B(M)$. In our case however $P_{n,t}(C(M^n)) \subset C(M^n)$ for all $t \geqslant 0$ (i.e. the processes $\{\phi_t(\boldsymbol{x})\colon t \geqslant 0\}$ are Feller processes) and we shall consider the family $\{P_{n,t}\colon t \geqslant 0\}$ as a semigroup of positive contraction operators on $C(M^n)$. Then knowledge of $P_{n,t}f$ for all $f \in C(M^n)$, $t > 0$ and $n \geqslant 1$ determines the transition probabilities for all of the $\{\phi_t(\boldsymbol{x})\colon t \geqslant 0\}$ processes and thereby determines up to equivalence the original process $\{\phi_t\colon t \geqslant 0\}$.

Let $A^{(n)}$ be the infinitesimal generator of the family $\{P_{n,t}\colon t \geqslant 0\}$ of bounded linear operators on $C(M^n)$ i.e. $A^{(n)}$ is the linear operator on $C(M^n)$ whose domain is the set $D(A^{(n)}) = \{f \in C(M^n)\colon \lim_{t \downarrow 0}(1/t)\mathbb{E}(f(\phi_t(\boldsymbol{x})) - f(\boldsymbol{x}))$ exists uniformly for $\boldsymbol{x} \in M^n\}$ and for $f \in D(A^{(n)})$, $A^{(n)}f(\boldsymbol{x}) = \lim_{t \downarrow 0}(1/t)\mathbb{E}(f(\phi_t(\boldsymbol{x})) - f(\boldsymbol{x}))$, $\boldsymbol{x} \in M^n$.

THEOREM 5.1: *Let* $r \geqslant 2$. *For* $n \geqslant 1$, $C^2(M^n) \subset D(A^{(n)})$. *Let* $A^{(1)} = A$. *Then* $A|_{C^2(M)}$ *is of the form described in Theorem* 4.1 *with* $\gamma \equiv 0$. *For* $n \geqslant 2$ *and* $f \in C^2(M^n)$,

$$(A^{(n)}f)(\boldsymbol{x}) = \sum_{i=1}^{n} (A_i f)(\boldsymbol{x}) + \tfrac{1}{2} \sum_{\substack{i,j=1 \\ i \neq j}}^{n} \alpha(x_i, x_j)(d_i d_j f)(\boldsymbol{x})$$

where $\alpha(x, y) \in T_x M \otimes T_y M$, $x, y \in M$, *and* α *is* $C^{r-2,r-2}$ *in the two variables* x *and* y. *Moreover* $\alpha(x, y)$ *is a positive semi-definite kernel and* $\alpha(x, x) = \alpha(x)$.

Note: Recall from the comment after Theorem 4.1 that $\frac{1}{2}\alpha(x)$ controls the matrix of second order coefficients of A. At various times we shall interpret $\alpha(x, y)$ as taking values in $T_x M \otimes T_y M$ or $L(T_x^*M, T_y M)$ or $\mathrm{BIL}(T_x^*M \times T_y^*M, \mathbb{R})$ or $L(T_x^*M \otimes T_y^*M, \mathbb{R})$. There are natural isomorphisms between all these spaces. Saying that α is positive semi-definite means that for any $\boldsymbol{x} = (x_1, \ldots, x_n) \in M^n$ and $\xi_i \in T_{x_i}^* M$, $1 \leqslant i \leqslant n$, then $\sum_{i,j=1}^n \alpha(x_i, x_j)(\xi_i, \xi_j) \geqslant 0$ and $\alpha(x_i, x_j)(\xi_i, \xi_j) = \alpha(x_j, x_i)(\xi_j, \xi_i)$ all i, j.

PROOF: Let $n \geqslant 1$ be temporarily fixed. We wish to apply Theorem 4.1 to the family $\{P_{n,t}: t \geqslant 0\}$ of positive linear operators on $C(M^n)$. (Thus the M of Theorem 4.1 is now M^n.) For $0 \leqslant s \leqslant r$ the map $\mathrm{Diff}^r(M) \times C^s(M^n) \to C^s(M^n)$

$$(\phi, f) \mapsto (\boldsymbol{x} \mapsto f(\phi(\boldsymbol{x})))$$

is continuous. Therefore we may apply Lemma 3.3. For $s = 0$ it shows that $\{P_{n,t}: t \geqslant 0\}$ is a semigroup of class (C_0) on $C(M^n)$, and in particular that $\{\phi_t(\boldsymbol{x}): t \geqslant 0\}$ is indeed a Feller process. For $s = r$ it gives us condition (iii) of Theorem 4.1. To prove condition (ii) suppose $f \in C(M^n)$ is zero in a neighbourhood V of $\boldsymbol{a} \in M^n$. Let $U = \{\phi \in \mathrm{Diff}^r(M): \varphi(\boldsymbol{a}) \in V\}$. Then U is a neighbourhood of Id_M in $\mathrm{Diff}^r(M)$ and

$$\begin{aligned} \left| \frac{1}{t}((P_{n,t}f)(\boldsymbol{a}) - f(\boldsymbol{a})) \right| &\leqslant \frac{1}{t} \|f\| \mathbb{P}\{\phi_t(\boldsymbol{x}) \notin V\} \\ &= \|f\| \frac{1}{t} \mathbb{P}\{\phi_t \notin U\} \\ &\to 0 \text{ as } t \to 0. \end{aligned}$$

It now follows from Theorem 4.1 and Proposition 4.1 that $C^2(M^n) \subset D(A^{(n)})$ and that $A^{(n)}|_{C^2(M^n)}$ is a second order differential operator with no zero order term.

Let $1 \leqslant p < n$. In the same way that the n-point motion $\{\phi_t(\boldsymbol{x}): t \geqslant 0\}$ is a quotient of the original process in $\mathrm{Diff}^r(M)$, then the p-point motion is a quotient of the n-point motion. This yields relations between the operators $A^{(n)}$ for various n. In particular if $f(x_1, x_2, \ldots, x_n) = g(x_{i_1}, \ldots, x_{i_p})$ for $1 \leqslant i_1 < i_2 < \ldots < i_p \leqslant n$ and $g \in C^2(M^p)$ then $(P_{n,t}f)(x_1, x_2, \ldots, x_n) = (P_{p,t}g)(x_{i_1}, \ldots, x_{i_p})$ and so $(A^{(n)}f)(x_1, x_2, \ldots, x_n) = (A^{(p)}g)(x_{i_1}, \ldots, x_{i_p})$. In particular if $f(x, y) = g(x)$ then $(A^{(2)}f)(x, y) = (Ag)(x)$.

Let $f, g \in C^r(M)$ and define $B(f, g)$ by $B(f, g)(x, y) = A^{(2)}(f(x)g(y)) - f(x)(Ag)(y) - g(y)(Af)(x)$. Then $B(f, g)$ is bilinear in f and g and in fact $B(f, g)(x, y)$ depends only on $df(x) \in T_x^*M$ and $dg(y) \in T_y^*M$. For let $a, b \in M$. Then

$$B(f, g)(a, b) = A^{(2)}((f(x) - f(a))(g(y) - g(b)))(a, b).$$

If we replace f by $\tilde{f}$ where $df(a) = d\tilde{f}(a)$ then $\tilde{f}(x) = f(x) + h_1(x)h_2(x) + k$ where k is a constant and $h_1, h_2 \in C^r(M)$ with $h_1(a) = h_2(a) = 0$. Therefore

$$\begin{aligned} B(\tilde{f}, g)(a, b) &= B(f, g)(a, b) \\ &\quad + A^{(2)}(h_1(x)h_2(x)(g(y) - g(b)))(a, b) \end{aligned}$$

and the last term is zero because it is a second order differential operator acting on a product of three functions which are all zero at (a, b). A similar argument works if we replace g by $\tilde{g}$ with $dg(b) = d\tilde{g}(b)$. We may now define $\alpha(x, y) \in T_xM \otimes T_yM$ by $\alpha(x, y)(\xi, \eta) = B(f, g)(x, y)$ whenever $df(x) = \xi \in T_x^*M$, $dg(y) = \eta \in T_y^*M$. Then $\alpha(x, y)(\xi, \eta) = \alpha(y, x)(\eta, \xi)$ and

$$\begin{aligned} A^{(2)}(f(x)g(y)) &= (Af)(x)g(y) + f(x)(Ag)(y) \\ &\quad + \alpha(x, y)(df(x), dg(y)). \end{aligned}$$

We know that $A^{(2)}|_{C^2(M^2)}$ is a 2nd order differential operator with no zero order term. Thus for $f \in C^2(M^2)$, we have

$$\begin{aligned} (A^{(2)}f)(x_1, x_2) &= (E_1 f)(x_1, x_2) + (E_2 f)(x_1, x_2) \\ &\quad + F(x_1, x_2)(d_1 d_2 f)(x_1, x_2) \end{aligned}$$

where E_1 and E_2 are 2nd order differential operators acting on the 1st and 2nd coordinates respectively, with coefficients depending possibly on both x_1 and x_2, and $F(x_1, x_2)$ takes values in $L(T_{x_1}^*M \otimes T_{x_2}^*M, \mathbb{R}) \cong T_{x_1}M \otimes T_{x_2}M$. By taking $f(x_1, x_2) = g(x_1)$ we obtain

$$(E_1 f)(x_1, x_2) = (Ag)(x_1) = (A_1 f)(x_1, x_2),$$

so that $E_1 = A_1$ and in particular the coefficients of E_1 depend only upon x_1. Similarly $E_2 = A_2$. Finally taking $f(x_1, x_2) = g(x_1)h(x_2)$ we see that $F(x_1, x_2) = \alpha(x_1, x_2)$.

More generally we may write

$$A^{(n)}f(x_1,\dots,x_n) = \sum_{i=1}^{n} (E_i f)(x_1,\dots,x_n) + \tfrac{1}{2}\sum_{i\neq j} F_{ij}(x_1,\dots,x_n)d_i d_j f(x_1,\dots,x_n).$$

By taking $f(x_1,\dots,x_n) = g(x_i)$ we obtain $E_i = A_i$. By taking $f(x_1\dots x_n) = g(x_i)h(x_j)$ we obtain $F_{ij}(x_1,\dots,x_n) = \alpha(x_i, x_j)$. Therefore we have obtained the general formula for $A^{(n)}$.

We must equate $\alpha(x)$ and $\alpha(x, x)$. Fix $a \in M$, $\xi, \eta \in T_x^*M$ and choose $g, h \in C^r(M)$ such that $dg(a) = \xi$, $dh(a) = \eta$. Let

$$\begin{aligned} f(x,y) &= (g(x)-g(y))(h(x)-h(y)) \\ &= g(x)h(x) - g(x)h(y) - h(x)g(y) + g(y)h(y). \end{aligned}$$

Then

$$\begin{aligned} A^{(2)}f(x,y) = A(gh)(x) &- (Ag)(x)h(y) - g(x)(Ah)(y) \\ &- \alpha(x,y)(dg(x), dh(y)) \\ &- (Ah)(x)g(y) - h(x)(Ag)(y) \\ &- \alpha(x,y)(dh(x), dg(y)) + A(gh)(y). \end{aligned}$$

Therefore

$$\begin{aligned} (A^{(2)}f)(a,a) &= 2\alpha(a)(\xi,\eta) - \alpha(a,a)(\xi,\eta) - \alpha(a,a)(\eta,\xi) \\ &= 2(\alpha(a) - \alpha(a,a))(\xi,\eta). \end{aligned}$$

But $f(x, x) = 0$ for all $x \in M$, so $(P_t f)(x, x) = \mathbb{E}(f(\phi_t(x), \phi_t(x))) = 0$ and in particular $(A^{(2)}f)(a, a) = 0$. Therefore $\alpha(a) = \alpha(a, a)$.

To show that $\alpha(x, y)$ is positive semi-definite, suppose $\boldsymbol{a} = (a_1,\dots,a_n)$ and $\xi_i \in T_{x_i}^*M$, $i = 1,\dots,n$. Choose $f_i \in C^r(M)$ such that $df_i(a_i) = \xi_i$. Consider $g(\boldsymbol{x}) = (\sum_{i=1}^n (f_i(x_i) - f_i(a_i)))^2$. Then $g(\boldsymbol{a}) = 0$ and $g(\boldsymbol{x}) \geqslant 0$ which implies $(A^{(n)}g)(\boldsymbol{a}) \geqslant 0$. Also $(A_i g)(\boldsymbol{a}) = \alpha(a_i)(\xi_i, \xi_i) = \alpha(a_i, a_i)(\xi_i, \xi_i)$ and $(d_i d_j g)(\boldsymbol{a}) = 2(\xi_i \otimes \xi_j)$ for $i \neq j$. Therefore

$$\begin{aligned} 0 \leqslant (A^{(n)}g)(\boldsymbol{a}) &= \sum_i (A_i g)(\boldsymbol{a}) + \tfrac{1}{2}\sum_{i\neq j}\alpha(x_i, x_j)d_i d_j g(\boldsymbol{a}) \\ &= \sum_{i,j}\alpha(a_i, a_j)(\xi_i, \xi_j). \end{aligned}$$

It follows from Theorem 4.1 that $\alpha(x, y)$ is jointly C^{r-2} in the x and y variables, i.e. if D_x and D_y denote differentiation with respect to the x and y variables respectively then $D_x^{r-2}D_y^{r-2}\alpha$ exists and is continuous (in any local coordinate system). Let $C^{r,r}(M^2)$ denote the space of all functions from M^2 to R which are jointly C^r in the first and second variables (in the sense above). Then $\mathrm{Diff}^r(M)$ acts on $C^{r,r}(M^2)$ in a continuous way, so that we may consider the family $\{P_t: t \geqslant 0\}$ as a semigroup of bounded linear operators on the space $C^{r,r}(M^2)$. We may now mimic the proof of Theorem 4.1 applied to the compact manifold M^2. The two degrees of differentiability which are lost in the coefficients of $A^{(2)}$ in step 5 now have the effect of converting $C^{r,r}$ functions into linear combinations of $C^{r,r-2}$, $C^{r-1,r-1}$ and $C^{r-2,r}$ functions. In particular $\alpha(x, y)$ which appears amongst the coefficients of $A^{(2)}$ is a $C^{r-2,r-2}$ function of x and y. □

One consequence of this result is that the $A^{(n)}|_{C^2(M^n)}$ for all $n \geqslant 1$ are determined by $A^{(2)}|_{C^2(M^2)}$ (corresponding to the 2-point motion). In particular they are determined by $A|_{C^2(M)}$ (i.e. the 1-point motion) and $\alpha(x, y)$ which may be thought of as a covariance for the 2-point motion. The connection between this observation and the similar result for Gaussian random processes will be pursued in Section 6.2.

5.3. For $x \in M$ and $\xi \in T_x^*M$ define the vector field $\alpha_{x,\xi}$ by $\alpha_{x,\xi}(y) = \alpha(x, y)(\xi) \in T_yM$, $y \in M$. Then $\alpha_{x,\xi} \in C^{r-2}(TM)$, the space of C^{r-2} sections of the tangent bundle TM. Associated with the positive semi-definite kernel α there is a real separable Hilbert space H consisting of sections of TM and characterised by

(i) $\alpha_{x,\xi} \in H$ for all $x \in M$, $\xi \in T_x^*M$

(ii) $\langle \alpha_{x,\xi}, Y\rangle = (Y(x), \xi)$ for all $Y \in H$, $x \in M$, $\xi \in T_x^*M$, where $(\ ,\ )$ denotes the natural pairing of T_xM and T_x^*M. H is called the reproducing kernel Hilbert space (RKHS) associated with α. For details of RKHS of sections of vector bundles see Baxendale [2], section 5.

Lemma 5.1: *$H \subset C^{r-2}(TM)$ and the inclusion is continuous.*

Proof Let $s = r - 2$. Recall that α is C^s in each variable. Let H_0 be the dense linear subspace of H consisting of finite linear combinations of $\alpha_{x,\xi}$, $x \in M$, $\xi \in T_x^*M$. Then $H_0 \subset C^s(TM)$ and it remains to show that this inclusion is continuous. Suppose first $s = 0$. We choose a Riemannian metric on M, giving an inner product on each tangent and cotangent space T_xM and T_x^*M. Then $C(TM)$ is a Banach space in the norm

$$\|X\|_{C(TM)} = \sup\{\|X(x)\|: x \in M\}$$

$$= \sup\left\{|(X(x), \xi)|: x \in M, \xi \in T_x^*M, \|\xi\| \leqslant 1\right\}.$$

Notice that $\|\alpha_{x,\xi}\|_H^2 = (\alpha_{x,\xi}(x), \xi) = \alpha(x, x)(\xi, \xi) \leqslant \|\alpha(x, x)\|\|\xi\|^2 \leqslant K^2$

for $x \in M$, $\xi \in T_x^*M$, $\|\xi\| \leqslant 1$ by continuity of α and compactness of M. Here $\|\alpha(x,y)\|$ denotes the operator norm on $\alpha(x,y) \in L(T_x^*M, T_yM)$. Therefore for $Y \in H_0$

$$\begin{aligned}
\|Y\|_{C(TM)} &= \sup\left\{ |(Y(x), \xi)| : x \in M, \xi \in T_x^*M, \|\xi\| \leqslant 1 \right\} \\
&= \sup\left\{ |\langle \alpha_{x,\xi}, Y \rangle| : x \in M, \xi \in T_x^*M, \|\xi\| \leqslant 1 \right\} \\
&\leqslant \sup\left\{ \|Y\|_H \, \|\alpha_{x,\xi}\|_H : x \in M, \xi \in T_x^*M, \|\xi\| \leqslant 1 \right\} \\
&\leqslant K \|Y\|_H,
\end{aligned}$$

which proves $H_0 \hookrightarrow C(M)$ is continuous, as required.

We reduce the general case to the case $s = 0$ by using s-jet bundles. See Palais [16] for definitions and notation. We shall use j_s^1 and j_s^2 for the s-jet extension maps applied to the 1st and 2nd variables respectively. Define $\alpha^s(x,y) = j_s^1 j_s^2(\alpha)(x,y) \in (J^sTM)_x \otimes (J^sTM)_y$. Then $\alpha^s(x,y)$ is a positive semi-definite kernel with corresponding *RKHS* H^s consisting of sections of $J^s(TM)$. By the result above we have a continuous inclusion $H^s \hookrightarrow C(J^sTM)$. For $\xi \in T_x^*M$, let $\bar{\xi} = (\xi, 0, 0, \ldots, 0) \in (J^sTM)_x^*$. Then

$$\begin{aligned}
j_s(\alpha_{x,\xi})(y) &= j_s(y \mapsto \alpha(x,y)(\xi)) \\
&= \left(j_s^1 j_s^2((x,y) \mapsto \alpha(x,y)) \right)(\bar{\xi}) \\
&= \alpha^s(x,y)(\bar{\xi}) \\
&= \alpha^s_{x,\bar{\xi}}(y).
\end{aligned}$$

Therefore $j_s(\alpha_{x,\xi}) \in H^s$. Also

$$\begin{aligned}
\left\| j_s\left(\sum_i \alpha_{x_i,\xi_i} \right) \right\|_{H^s}^2 &= \left\| \sum_i \alpha^s_{x_i,\bar{\xi}_i} \right\|_{H^s}^2 \\
&= \sum_{i,j} \alpha^s(x_i, x_j)\left(\bar{\xi}_i, \bar{\xi}_j \right) \\
&= \sum_{i,j} j_s^1 j_s^2 \alpha(x_i, x_j)\left(\bar{\xi}_i, \bar{\xi}_j \right) \\
&= \sum_{i,j} \alpha(x_i, x_j)(\xi_i, \xi_j) \\
&= \left\| \sum_i \alpha_{x_i,\xi_i} \right\|_H^2.
\end{aligned}$$

Therefore j_s: $H_0 \to H^s$ is an isometric embedding.

$$\begin{tikzcd} H_0 \arrow[r, hook] \arrow[d, "j_s"] & C^s(TM) \arrow[d, "j_s"] \\ H^s \arrow[r, hook, "\text{cont}'"] & C(J^sTM) \end{tikzcd}$$

The diagram commutes and the j_s on the right hand side is an isometry onto a closed subspace of $C(J^sTM)$. It follows that $H_0 \hookrightarrow C^s(TM)$ is continuous as required. □

Conversely to any real separable Hilbert space H continuously included in $C^{r-2}(TM)$ there exists a unique positive semi-definite kernel α which is C^{r-2} in each variable. For we may take (ii) above as a definition of $\alpha_{x,\xi}$ and then let

$$\alpha(x,y)(\xi) = (\alpha_{x,\xi})(y) \in T_yM.$$

An alternative way of recovering α from H is as follows. Let $(u^1, u^2, \dots, u^m)$ and $(v^1, v^2, \dots, v^m)$ be local coordinate systems near x and y respectively. If we write $\alpha_{x,i}$ for α_{x,du^i} and

$$Y(x) = \sum_{i=1}^{m} Y^i(x)\frac{\partial}{\partial u^i} \quad \text{then we have}$$

$$Y^i(x) = \big(Y(x), du^i\big) = \langle Y, \alpha_{x,i}\rangle, \qquad Y \in H.$$

Then

$$\begin{aligned} \operatorname{tr}_H\big(Y^i(x)Y^j(y)\big) &= \operatorname{tr}_H\big(\langle Y, \alpha_{x,i}\rangle\langle Y, \alpha_{y,j}\rangle\big) \\ &= \langle \alpha_{x,i}, \alpha_{y,j}\rangle \\ &= \big(\alpha_{x,i}(y), dv^j\big) \\ &= \alpha(x,y)(du^i, dv^j). \end{aligned}$$

Here tr_H denotes the trace on H of the quadratic form $Y \to Y^i(x)Y^j(y)$, $Y \in H$. We may rewrite this result in a coordinate free way as

$$\operatorname{tr}_H\big(Y(x) \otimes Y(y)\big) = \alpha(x,y).$$

The left hand side is now the trace on H of the $T_xM \otimes T_yM$ valued quadratic form $Y \to Y(x) \otimes Y(y)$, $Y \in H$. If we write $\alpha(x,y)(du^i, dv^j) =$

$\alpha^{ij}(x, y)$ then in the special case where $y = x$ (and $v^i = u^i$, $1 \leqslant i \leqslant m$) we have $\alpha^{ij}(x, x) = \alpha^{ij}(x) = \alpha(x)(du^i, du^j)$ where α and α^{ij} are the functions appearing in Theorem 4.1 and 5.1.

For $r \geqslant 3$ H is continuously included in $C^1(TM)$ so that the map $H \to R$ given by $Y \to Y^j_k(y) = \partial Y^j / \partial v^k(y)$ is continuous. Therefore there exists $\beta_{y,j,k}$, say, in H such that

$$\langle Y, \beta_{y,j,k} \rangle = Y^j_k(y) = \frac{\partial}{\partial v^k} \langle Y, \alpha_{y,j} \rangle.$$

Then

$$\begin{aligned}
\operatorname{tr}_H\big(Y^i(x) Y^j_k(y)\big) &= \operatorname{tr}_H\big(\langle Y, \alpha_{x,i}\rangle \langle Y, \beta_{y,j,k}\rangle\big) \\
&= \langle \alpha_{x,i}, \beta_{y,j,k} \rangle \\
&= \frac{\partial}{\partial v^k} \langle \alpha_{x,i}, \alpha_{y,j} \rangle \\
&= \frac{\partial}{\partial v^k} \big(\alpha^{ij}(x, y)\big).
\end{aligned}$$

If $Y \in H$ and $f \in C^2(M)$ we obtain

$$(Y(Yf))(x) = \sum_{i,j} Y^i(x) Y^j(x) f_{,ij}(x) + \sum_j \Big(\sum_i Y^i(x) Y^j_i(x)\Big) f_j(x).$$

Therefore if we define a second order differential operator B by

$$(Bf)(x) = \tfrac{1}{2} \operatorname{tr}_H\big((Y(Yf))(x)\big), \qquad f \in C^2(M)$$

we obtain

$$(Bf)(x) = \tfrac{1}{2} \sum_{i,j} \alpha^{ij}(x) f_{ij}(x) + \sum_j \Big(\tfrac{1}{2} \sum_i \frac{\partial}{\partial v^i} \alpha^{ij}(x, y)|_{y=x}\Big) f_j(x).$$

In particular the 2nd order coefficients of B agree with those of A and the 1st order coefficients are C^{r-3} functions of x.

Let $Xf(x) = (Af)(x) - (Bf)(x)$. Then X is a 1st order differential operator with no zero order term, i.e. X is a vector field on M. The first order terms of A and B are C^{r-2} and C^{r-3} respectively, so X is a C^{r-3} vector field, i.e. $X \in C^{r-3}(TM)$.

THEOREM 5.2: *Let $r \geqslant 3$. There exist a unique $X \in C^{r-3}(TM)$ and a unique real separable Hilbert space H continuously included in $C^{r-2}(TM)$ such that for $n \geqslant 1$ and $f \in C^2(M^n)$,*

$$(A^{(n)}f)(\mathbf{x}) = \tfrac{1}{2} \operatorname{tr}_H\big((Y(Yf))(\mathbf{x})\big) + (Xf)(\mathbf{x}).$$

PROOF: We have just seen that for $g \in C^2(M)$,

$$(A^{(1)}g)(x) = (Ag)(x) = \tfrac{1}{2}\mathrm{tr}_H((Y(Yg))(x)) + (Xg)(x).$$

For $i \neq j$

$$\big(Y(x_i) \otimes Y(x_j)\big)(d_i d_j f)(\boldsymbol{x}) = \big(Y_i(Y_j f)\big)(\boldsymbol{x}),$$

so that

$$\alpha(x_i, x_j)(d_i d_j f)(\boldsymbol{x}) = \mathrm{tr}_H\big(\big(Y_i(Y_j f)\big)(\boldsymbol{x})\big).$$

(Recall that $(Y_j f)(\boldsymbol{x})$ denotes Y as a differential operator acting in the jth variable of f.) Therefore

$$\begin{aligned}(A^{(n)}f)(\boldsymbol{x}) &= \sum_{i=1}^{n} (A_i f)(\boldsymbol{x}) + \tfrac{1}{2}\sum_{i\neq j} \alpha(x_i, x_j)(d_i d_j f)(\boldsymbol{x}) \\ &= \sum_{i=1}^{n} \{\tfrac{1}{2}\mathrm{tr}_H((Y_i(Y_i f))(\boldsymbol{x})) + (X_i f)(\boldsymbol{x})\} \\ &\quad + \sum_{i\neq j} \tfrac{1}{2}\mathrm{tr}_H\big(\big(Y_i(Y_j f)\big)(\boldsymbol{x})\big) \\ &= \tfrac{1}{2}\mathrm{tr}_H\left\{\left(\Big(\sum_i Y_i\Big)\Big(\sum_j Y_j f\Big)\right)(\boldsymbol{x})\right\} + \sum_i (X_i f)(\boldsymbol{x}) \\ &= \tfrac{1}{2}\mathrm{tr}_H((Y(Yf))(\boldsymbol{x})) + (Xf)(\boldsymbol{x}).\end{aligned}$$

The uniqueness is obvious as α, the kernel for H, and X can be defined directly in terms of $A^{(2)}$ and A. □

5.4. So far we have obtained Theorems 5.1 and 5.2 giving formulae for the restrictions to C^2 functions of the infinitesimal generators $A^{(n)}$ of the n-point motions $\{\phi_t(\boldsymbol{x})\colon t \geq 0\}$, $\boldsymbol{x} \in M^n$. The question remains as to whether this is sufficient to determine $\{\phi_t\colon t \geq 0\}$ uniquely (up to equivalence). We give two results dealing with different notions of uniqueness. Note that for $r \geq s \geq 1$ there is a natural continuous inclusion $\mathrm{Diff}^r(M) \subset \mathrm{Diff}^s(M)$.

THEOREM 5.3: *Let $r \geq s \geq 2$. Let $\{\phi_t\colon t \geq 0\}$ and $\{\psi_t\colon t \geq 0\}$ be Brownian motions in* $\mathrm{Diff}^r(M)$ *and* $\mathrm{Diff}^s(M)$ *respectively. Suppose that for each $n \geq 1$ the infinitesimal generators of the n-point motions of $\{\phi_t\colon t \geq 0\}$ and $\{\psi_t\colon t \geq 0\}$ coincide on $C^2(M^n)$. Then*

(i) $\psi_t \in \mathrm{Diff}^r(M)$ for all $t \geq 0$, w.p.1.

(ii) $\{\phi_t\colon t \geq 0\}$ and $\{\psi_t\colon t \geq 0\}$ are equivalent Brownian motions in $\mathrm{Diff}^r(M)$.

PROOF: Let $\{P_{n,t}: t \geqslant 0\}$ and $\{Q_{n,t}: t \geqslant 0\}$ be the semigroups corresponding to the n-point motions of $\{\phi_t: t \geqslant 0\}$ and $\{\psi_t: t \geqslant 0\}$ respectively. Then $Q_{n,t}(C^\infty(M^n)) \subset Q_{n,t}(C^s(M^n)) \subset C^s(M^n) \subset C^2(M^n)$, so we may apply Proposition 4.3 to deduce that $P_{n,t} = Q_{n,t}$ for all $t > 0$ as required. □

It follows that for $r \geqslant 3$ a Brownian motion $\{\phi_t: t \geqslant 0\}$ on $\mathrm{Diff}^r(M)$ is uniquely determined by H (continuously included in $C^{r-2}(TM)$) and X ($\in C^{r-3}(TM)$). We shall call the pair (H, X) the *data* of the Brownian motion $\{\phi_t: t \geqslant 0\}$. It may happen in the case $r = 2$ that the differential operator B is well-defined with continuous coefficients, for example if α is C^1 in one of its variables. In this case X is a well-defined continuous vector field and we shall again call (H, X) the data of the Brownian motion $\{\varphi_t: t \geqslant 0\}$. A way of constructing such examples is as follows. Let $\{\varphi_t: t \geqslant 0\}$ be a Brownian motion on $\mathrm{Diff}^\infty(M)$ with data (H, X) and differential operator B as above. Let g be a C^2, but not C^3, diffeomorphism of M and consider the new Brownian motion $\varphi_y = g^{-1} \circ \varphi_t \circ g$ which lies in $\mathrm{Diff}^2(M)$ but not in general in $\mathrm{Diff}^3(M)$. The process $\{\psi_t: t \geqslant 0\}$ has differential operator $f \to B(f \circ g^{-1}) \circ g$ which has C^1 second order and C^0 first order coefficients. In fact the data for the process $\{\psi_t: t \geqslant 0\}$ is (g^*H, g^*X) where for any vector field Y on M, $(g^*Y)(x) = (Dg(x))^{-1}Y(g(x))$ and the Hilbert space g^*H is the isometric image of H under g^*. g^*H has reproducing kernel $(x, y) \to ((Dg(x))^{-1} \otimes (Dg(y))^{-1})\alpha(g(x), g(y))$.

Up to now the Markov processes we have been considering have been Feller processes and we have considered the infinitesimal generators of semigroups acting on $C(M^n)$. More generally if $\{P_t: t \geqslant 0\}$ is the semigroup of positive contraction operators on $B(N)$ corresponding to a time homogeneous Markov process on a smooth compact manifold N we may define its infinitesimal generator $\bar{A}$ to be the operator on $B(N)$ with domain $D(\bar{A}) = \{f \in B(N): \lim_{t \downarrow 0}(1/t)(P_t f(x) - f(x))$ exists uniformly for $x \in M\}$ and given by

$$\bar{A}f(x) = \lim_{t \downarrow 0} \frac{1}{t}(P_t f(x) - f(x)) \qquad \text{if} \quad f \in D(\bar{A}).$$

If the process is in fact a Feller process with infinitesimal generator A (corresponding to P_t acting on $C(N)$) then $D(A) = C(N) \cap D(\bar{A})$ and $\bar{A}$ is an extension of A. Therefore we may talk without ambiguity of an infinitesimal generator having domain including $C^2(N)$ and of the restriction of the infinitesimal generator to $C^2(N)$.

THEOREM 5.4: *Let* $r \geqslant 4$. *Let* $\{\phi_t: t \geqslant 0\}$ *be a Brownian motion in* $\mathrm{Diff}^r(M)$. *Let* $\{\psi_t(x): t \geqslant 0, x \in M\}$ *be a random process with values in M such that for each* $n \geqslant 1$ *and* $\mathbf{x} \in M^n$,

(i) $\{\psi_t(\mathbf{x}): t \geqslant 0\}$ *is a time homogeneous Markov process in* M^n *with*

$\psi_0(\boldsymbol{x}) = \boldsymbol{x}$, *where* $\psi_t(\boldsymbol{x}) = (\psi_t(x_1), \ldots, \psi_t(x_n))$,

(ii) the infinitesimal generator of $\{\psi_t(\boldsymbol{x})\colon t \geqslant 0\}$ *has domain including* $C^2(M^n)$ *and its restriction to* $C^2(M^n)$ *agrees with the restriction of the infinitesimal generator of* $\{\phi_t(\boldsymbol{x})\colon t \geqslant 0\}$ *to* $C^2(M^n)$.

Then $\{\phi_t(x)\colon t \geqslant 0,\ x \in M\}$ *and* $\{\psi_t(x)\colon t \geqslant 0,\ x \in M\}$ *are equivalent processes.*

PROOF: Let $A^{(n)}$ be the infinitesimal generator of the n-point motions of $\{\phi_t\colon t \geqslant 0\}$. Then the coefficients of $A^{(n)}$ are C^{r-2} (and hence C^2) functions of $\boldsymbol{x}$. We may now apply Theorem 3.2.6 of Stroock and Varadhan [20] to see that the transition probability function $P_n(\cdot,\cdot,\cdot)$ is uniquely determined by $A^{(n)}|_{C^2(M^n)}$. Thus $\{\phi_t(\boldsymbol{x})\colon t \geqslant 0\}$ and $\{\psi_t(\boldsymbol{x})\colon t \geqslant 0\}$ have the same transition probability function for all $\boldsymbol{x}$, and so all the finite dimensional distributions of $\{\phi_t(x)\colon t \geqslant 0, x \in M\}$ and $\{\psi_t(x)\colon t \geqslant 0, x \in M\}$ agree. □

The power of Theorem 5.4 is that it gives uniqueness amongst a much larger class of processes than does Theorem 5.3.

6. Brownian motions in C^s(TM) and stochastic differential equations

6.1. We have seen that for $r \geqslant 3$ a Brownian motion $\{\phi_t\colon t \geqslant 0\}$ in $\mathrm{Diff}^r(M)$ is specified uniquely by a continuous inclusion of a separable real Hilbert space $H \subset C^{r-2}(TM)$ and a vector field $X \in C^{r-3}(TM)$. There are various questions which now arise.

1. Do every $H \subset C^{r-2}(TM)$ and $X \in C^{r-3}(TM)$ arise in this manner?
2. How do perturbations in the data (H, X) affect the process $\{\phi_t\colon t \geqslant 0\}$?
3. How can knowledge of H and X be used to give information about the process $\{\phi_t\colon t \geqslant 0\}$?

6.2. Suppose $s \geqslant 0$, H is a real separable Hilbert space continuously included in $C^{s+1}(TM)$ and $X \in C^s(TM)$. Then H is a RKHS with positive semi-definite kernel α, say. The induced inclusion $H \subset C^s(TM)$ is an abstract Wiener space, determining a mean-zero Gaussian measure ν on $C^s(TM)$. In particular for $x, y \in M$,

$$\int_{C^s(TM)} (Y(x) \otimes Y(y))\, d\nu(Y) = \mathrm{tr}_H(Y(x) \otimes Y(y))$$

$$= \alpha(x, y),$$

so that α is the covariance kernel for ν. Define Borel probability measures ν_t, $t > 0$, on $C^s(TM)$ by

$$\nu_t(\Gamma) = \nu(t^{-1/2}\Gamma), \qquad \Gamma \in \mathscr{B}(C^s(TM)).$$

Then each ν_t is Gaussian and $\nu_t * \nu_s = \nu_{t+s}$ for $s, t > 0$. Let $\{W_t\colon t \geqslant 0\}$

denote the Wiener process in $C^s(TM)$ corresponding to the inclusion $H \subset C^s(TM)$. Thus $\{W_t\colon t \geqslant 0\}$ has independent increments, is time homogeneous, has continuous sample paths, satisfies $W_0 = 0$ w.p.1, and $\mathbb{P}\{W_t \in \Gamma\} = \nu_t(\Gamma)$, $\Gamma \in \mathscr{B}(C^s(TM))$. Finally define $Z_t = W_t + tX$. Using the terminology of the introduction $\{W_t\colon t \geqslant 0\}$ and $\{Z_t\colon t \geqslant 0\}$ are Brownian motions in $C^s(TM)$.

In fact any Brownian motion in $C^s(TM)$ is of the form $Z_t = W_t + tX$ where $X \in C^s(TM)$ and $\{W_t\colon t \geqslant 0\}$ corresponds to some continuous inclusion $H \subset C^s(TM)$. H and X are determined uniquely by the Brownian motion. X can be any C^s vector field, whereas the inclusion $H \subset C^s(TM)$ must give an abstract Wiener space. The condition $H \subset C^{s+1}(TM)$ is thus sufficient but not necessary. See Baxendale [3], Section 2, for details.

We may restate our main result as follows. For $r \geqslant 3$, a Brownian motion $\{\phi_t\colon t \geqslant 0\}$ on $\mathrm{Diff}^r(M)$ is uniquely determined by a Brownian motion $\{Z_t = W_t + tX\colon t \geqslant 0\}$ in $C^{r-3}(TM)$. It may be useful to think of $\{Z_t\colon t \geqslant 0\}$ as the "linearisation" of the process $\{\phi_t\colon t \geqslant 0\}$ (regarding $C^r(TM)$ as the tangent space at Id_M to the manifold $\mathrm{Diff}^r(M)$ and ignoring for these purposes the loss of three orders of differentiability).

6.3. The first question in 6.1 can be reposed as follows. Which Brownian motions $\{Z_t\colon t \geqslant 0\}$ in $C^{r-3}(TM)$ come from a Brownian motion $\{\phi_t\colon t \geqslant 0\}$ in $\mathrm{Diff}^r(M)$? We may reverse the problem. Given a Brownian motion $\{Z_t\colon t \geqslant 0\}$ in $C^s(TM)$ can we construct a corresponding Brownian motion $\{\phi_t\colon t \geqslant 0\}$ in $\mathrm{Diff}^r(M)$ for some r? (It will be too much to hope that in general we may take $r = s + 3$.) There are (at least) two ways of constructing $\{\phi_t\colon t \geqslant 0\}$ from $\{Z_t\colon t \geqslant 0\}$.

One way is to obtain $\{\phi_t\colon t \geqslant 0\}$ as the solution of a stochastic differential equation. $\{W_t\colon t \geqslant 0\}$ provides an infinite-dimensional Wiener process and X will be a drift term. More precisely, consider

$$\left.\begin{aligned} d\phi_t(x) &= X(\phi_t(x))\,dt + \circ\, dW_t(\phi_t(x)) \\ \phi_0(x) &= x \end{aligned}\right\} \tag{1}$$

wnere $\circ\, dW_t$ denotes the Stratonovich stochastic differential, used to provide an invariant, connection-free, form of the stochastic differential equation. In the case that H is finite-dimensional we may write $W_t = \sum_{n=1}^{d} X_n W_t^n$ for d independent one-dimensional Wiener processes $(W_t^1, \ldots, W_t^d)$ and $\{X_1, \ldots, X_d\}$ a complete orthonormal basis for H. The equation (1) may now be written

$$\left.\begin{aligned} d\phi_t(x) &= X(\phi_t(x))\,dt + \sum_{n=1}^{d} X_n(\phi_t(x)) \circ dW_t^n \\ \phi_0(x) &= x. \end{aligned}\right\} \tag{2}$$

This equation has been much studied recently. See the article by Williams in [21] and references therein; also Elworthy [6] and Ikeda and Watanabe [11].

An alternative method is to use a version of the injection scheme of Gangolli and McKean (see [14]), using Φ: $C^s(TM) \to \mathrm{Diff}^s(M)$, $\Phi(X) =$ time 1 flow along X, to do the injecting.

Details of both of these methods will be given in a forthcoming paper, together with partial answers to questions 2 and 3. Some statements of results concerning question 3 may be found in [4]. Meanwhile we have the following.

THEOREM 6.1: *Suppose $s \geqslant 2$, $H \subset C^{s+1}(TM)$ is a continuous inclusion and $X \in C^s(TM)$. Then there is a Brownian motion $\{\psi_t\colon t \geqslant 0\}$ in* $\mathrm{Diff}^{s-1}(M)$ *satisfying* (1). *In particular the infinitesimal generators $B^{(n)}$ of the n-point motions of $\{\psi_t\colon t \geqslant 0\}$ satisfy $C^2(M^n) \subset D(B^{(n)})$ and for $f \in C^2(M^n)$,*

$$(B^{(n)}f)(x) = \tfrac{1}{2}\mathrm{tr}_H((Y(Yf))(x)) + (Xf)(x).$$

PROOF: See Elworthy [6], Chapter VIII, or the paper by Kunita in [21]. In fact the result of Elworthy shows that our statement is not best possible in terms of loss of differentiability. □

We many now answer question 1 as follows. Given $r \geqslant 2$, any continuous inclusion $H \subset C^{r+2}(TM)$ and $X \in C^{r+1}$ provide the data (H, X) for some Brownian motion $\{\phi_t\colon t \geqslant 0\}$ on $\mathrm{Diff}^r(M)$. In particular taking $r = \infty$ we obtain a bijection between (equivalence classes of) Brownian motions in $\mathrm{Diff}^\infty(M)$ and pairs (H, X) where $H \subset C^\infty(TM)$ is a continuous inclusion and $X \in X^\infty(TM)$.

It follows also from Theorem 6.1 that for $r \geqslant 5$ any Brownian motion $\{\phi_t\colon t \geqslant 0\}$ in $\mathrm{Diff}^r(M)$ arises as the solution of a stochastic differential equation. For if (H, X) are the data of $\{\phi_t\colon t \geqslant 0\}$ then $H \subset C^{r-2}(TM)$ and $X \in C^{r-3}(TM)$. We apply Theorem 6.1 with $s = r - 3 \geqslant 2$ to obtain $\{\psi_t\colon t \geqslant 0\}$ and use Theorem 5.4 to show that $\{\phi_t\colon t \geqslant 0\}$ and $\{\psi_t\colon t \geqslant 0\}$ are equivalent processes.

In fact in this case where (H, X) are the data of a Brownian motion $\{\phi_t\colon t \geqslant 0\}$ in $\mathrm{Diff}^r(M)$ a more careful analysis using the Itô version of the stochastic differential equation (1) may be made. It turns out that the Itô version has C^{r-2} coefficients so that $\{\psi_t\colon t \geqslant 0\}$ is a Brownian motion in $\mathrm{Diff}^{r-3}(M)$ and the result that $\{\phi_t\colon t \geqslant 0\}$ is the solution of a stochastic differential equation is valid for $r \geqslant 4$.

References

[1] R. ARENS: Topologies for homeomorphism groups. *Amer. J. Math.* 68 (1946) 593–610.
[2] P. BAXENDALE: Gaussian measures on function spaces. *Amer. J. Math.* 98 (1976) 891–952.

[3] P. BAXENDALE: Wiener processes on manifolds of maps. *Proc. Royal Soc. Edinburgh* 87A (1980) 127–152.

[4] P. BAXENDALE: Stochastic flows and Malliavin calculus. *Proc. 20th IEEE Conference on Decision and Control.* New York: IEEE (1981) pp. 595–599.

[5] E.B. DYNKIN: *Die Grundlagen der Theorie der Markoffshen Prozesse.* Berlin: Springer (1961).

[6] K.D. ELWORTHY: Stochastic differential equations on manifolds. London Mathematical Society, Lecture Note Series 70. Cambridge: Cambridge University Press (1982).

[7] P. FEINSILVER: Processes with independent increments on a Lie group. *Trans. Amer. Math. Soc.* 242 (1978) 73–121.

[8] J. FRANKS: Manifolds of C^r mappings and applications to differentiable dynamical systems. *Advances in Mathematics Supplementary Series* 4 (1979) 271–290.

[9] T.E. HARRIS: Brownian motions on the homeomorphisms of the plane. *Ann. Probability* 9 (1981) 232–254.

[10] G.A. HUNT: Semigroups of measures on Lie groups. *Trans. Amer. Math. Soc.* 81 (1956) 264–293.

[11] N. IKEDA and S. WATANABE: *Stochastic differential equations and diffusion processes.* Amsterdam: North-Holland; Tokyo: Kodansha (1981).

[12] J.L. KELLEY: *General topology.* Berlin: Springer (1955).

[13] H. KUNITA: Stochastic differential equations and stochastic flows of homeomorphisms, *Stochastic analysis and applications*, to appear.

[14] H.P. MCKEAN: *Stochastic integrals.* New York: Academic Press (1969).

[15] E. NELSON: An existence theorem for second order parabolic equations. *Trans. Amer. Math. Soc.* 88 (1958) 414–429.

[16] R.S. PALAIS: Foundations of global non-linear analysis. New York: Benjamin (1968).

[17] C.A. ROGERS: *Hausdorff measures.* Cambridge: Cambridge University Press (1970).

[18] A.V. SKOROHOD: A note on Gaussian measures in Banach space. *Theory Probab. Appl.* 15 (1970) 519–520.

[19] D.W. STROOCK and S.R.S. VARADHAN: Limit theorems for random walks on Lie groups. *Sankhyā Series A,* 35 (1973) 277–294.

[20] D.W. STROOCK and S.R.S. VARADHAN: *Multidimensional diffusion processes.* Berlin: Springer (1979).

[21] D. WILLIAMS (ed.): Stochastic integrals (Proceedings, LMS Durham symposium, 1980), *Lecture notes in mathematics 851.* Berlin: Springer (1981).

[22] K. YOSIDA: *Functional Analysis* (6th edn.). Berlin: Springer (1980).

(Oblatum 23-VIII-1982)

Department of Mathematics
University of Aberdeen
Aberdeen AB9 2TY
Scotland