RECENT RESULTS ON THE SHORT-TIME
GEOMETRY OF RANDOM HEAT KERNELS

RICHARD B. SOWERS

ABSTRACT. We give some recent results concerning the short-time behavior of the random heat kernel associated with the stochastic partial differential equation $du = \frac{1}{2}\Delta u\, dt + (\sigma, \nabla u) \circ dW_t$ on some Riemannian manifold M. Here Δ is the Laplace-Beltrami operator, σ is some vector field, ∇ is the gradient operator, and $\circ\, dW_t$ denotes Stratonovich integration against a standard Wiener process. These results show how classical short-time asymptotics of deterministic heat kernels must be corrected to account for the random term; an exponential term must be added.

0. Introduction

This research announcement reports some recent results concerning the short-time geometry of random heat kernels of the form

$$(0.1) \qquad dp_y = \frac{1}{2}\Delta p_y\, dt + (\sigma, \nabla p_y) \circ dW_t, \quad t > 0; \qquad \lim_{t \to 0} p_y(t, \cdot) = \delta_y.$$

The spatial variable takes values on a C^∞, compact, connected, and d-dimensional Riemannian manifold M with inner product $(\cdot, \cdot)$, Δ and ∇ are respectively the standard Laplace-Beltrami and gradient operators, and σ is some vector field. Here W is a standard Wiener process, which is defined on some underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and $\circ$ denotes Stratonovich

integration. Finally, δ_y denotes the Dirac measure which has mass at y, where y is any fixed point in M. The equation (0.1) is an example of a *stochastic partial differential equation*, or SPDE; SPDE's are currently an object of much study (see [17]), particularly in some areas of applied mathematics. We might let $\sigma\dot{\xi}$ model a rough or turbulent velocity field; as the turbulence increases, an equation like (0.1) might be the appropriate limit (see [16]). Alternately, (0.1) describes the evolution of the unnormalized density for a certain nonlinear filtering problem. Our interest here is the *short-time asymptotics* of p_y (i.e., as $t \to 0$). The short-time asymptotics of a deterministic counterpart of (0.1) are well-known—consider the PDE on M given by

$$(0.2) \qquad \frac{\partial q_y}{\partial t} = \frac{1}{2}\Delta q_y + (\sigma, \nabla q_y)\dot{\zeta}(t), \quad t > 0; \qquad \lim_{t \to 0} q_y(t, \cdot) = \delta_y,$$

where ζ is some C^1 function on $[0, \infty)$. The classical short-time asymptotics of q_y are

$$(0.3) \quad q_y(t, x)$$
$$= \exp\left[-\frac{d^2(x, y)}{2t} - \int_0^t \left(\dot{\gamma}_y^{t,x}(s), \sigma(\gamma_y^{t,x}(s))\right)\dot{\zeta}(s)\,ds + O(t)\right]$$
$$\cdot \Theta_y(x)(2\pi t)^{-d/2}$$

for x sufficiently close to y. In this expression, $d(\cdot, \cdot)$ is the Riemannian distance function, $\gamma_y^{t,x}$ is the geodesic running from y to x in time t, which we assume to be unique, and Θ_y is related to the Riemannian volume element at y. A very abbreviated list of references concerning (0.3) is: [2], [4], [6], [7], [8], [10], [11], [13], [14], and [20]. The exponential rate of decay, i.e., $-d^2(x, y)/(2t)$, was isolated by Varadhan in [19], and $\int_0^t \left(\dot{\gamma}_y^{t,x}(s), \sigma(\gamma_y^{t,x}(s))\right)\dot{\zeta}(s)\,ds$, the so-called *work term* (since it is the work of the time-varying vector field $\sigma\dot{\zeta}$ along the geodesic $\gamma_y^{t,x}$), was isolated in [14]. The formula (0.3), in particular when $\sigma \equiv 0$, has led to deep results in differential geometry (see [3]). A quite natural question is to ask how results such as (0.3) may be translated to the SPDE (0.1). This is the essence of [18], of which this is a brief summary. A corresponding study of a related problem can be found in [22].

In order to point out what is unusual in this problem, consider a simple example. Namely, consider the SPDE on $\mathbb{R}$ given by

$$dp_y = \frac{1}{2}\frac{\partial^2 p_y}{\partial x^2}\,dt + c\frac{\partial p_y}{\partial x} \circ dW_t, \quad t > 0; \qquad \lim_{t \to 0} p_y(t, \cdot) = \delta_y$$

where c is any nonzero constant. This of course is the SPDE (0.1) when $M = \mathbb{R}$ (with the usual atlas), the Riemannian inner product field is $(\cdot, \cdot) = dx \otimes dx$, and the vector field σ is $\sigma = c\frac{\partial}{\partial x}$ (we shall understand tangent vectors as derivations). This does not exactly fit the requirements we have specified for (0.1) since $\mathbb{R}$ is not compact, but we shall proceed anyways. It is easily checked that the solution of this SPDE is explicitly

$$(0.4) \quad p_y^{can}(t, x) \triangleq \exp\left[-\frac{|x - y|^2}{2t} - \left(\frac{x - y}{t}\right)(cW_t) - \frac{(cW_t)^2}{2t}\right] \Big/ \sqrt{2\pi t},$$

for all $t > 0$ and all x in $\mathbb{R}$, the superscript can indicating that this is a "canonical" case. The reader familiar with filtering theory may be more comfortable with the representation

$$p_y^{can}(t, x) = \exp\left[-\frac{|x - (y - cW_t)|^2}{2t}\right] \Big/ \sqrt{2\pi t}, \qquad t > 0, \, x \in \mathbb{R}.$$

We may from this understand $p_y^{can}(t, \cdot)$ as the conditional density of $y - cW_t + \tilde{W}_t$ given $\{W_s : s \geq 0\}$, where $\tilde{W}$ is any standard Wiener process independent of W.

In this simple situation, a complete analysis of the short-time asymptotics of p_y^{can} is possible, thanks to well-known and classical results on Brownian motions. Via the law of the iterated logarithm [12, Theorem 1.9.23, and 15, Theorem II.1.9], the exact small-time asymptotics of the second and third exponential terms of (0.4) are

$$(0.5) \quad \left(\frac{x - y}{t}\right)(cW_t) = O\left(\sqrt{t^{-1}\ln\ln t^{-1}}\right); \qquad \frac{(cW_t)^2}{2t} = O\left(\ln\ln t^{-1}\right),$$

with the first asymptotic holding only, of course, when $x \neq y$. Thus all of the terms in the exponent in (0.4) blow up almost-surely, but with decreasing rates from left to right.

Let's now try to fit the formula (0.4) into the framework given by (0.3). Obviously, $|x - y|^2 = d^2(x, y)$, where $d(\cdot, \cdot)$ is the Riemannian distance function on $\mathbb{R}$ defined by the metric $(\cdot, \cdot) = dx \otimes dx$. Also,

$$\left(\frac{x - y}{t}\right)(cW_t) = \int_0^t \left(\dot{\gamma}_y^{t,x}(s), \sigma(\gamma_y^{t,x}(s))\right) \circ dW_s$$

where $\{\gamma_y^{t,x}(s) : 0 \leq s \leq t\}$ is again the unique geodesic running from y to x in time t; in this case, $\gamma_y^{t,x}(s) \triangleq y + \frac{s}{t}(x - y)$ for $0 \leq s \leq t$. Thus a

more qualitatively valuable representation of (0.4) is

(0.6)
$$p_y^{can}(t,x) = \frac{\exp\left[-\frac{d^2(x,y)}{2t} - \int_0^t \left(\dot{\gamma}_y^{t,x}(s), \sigma(\gamma_y^{t,x}(s))\right) \circ dW_s - \frac{(cW_t)^2}{2t}\right]}{\sqrt{2\pi t}}$$

for each $t > 0$ and x and y in $\mathbb{R}$. Comparing this to (0.3), we find the usual distance and work terms. From (0.5), however, we see that the third exponential term in (0.6) *blows up* ($\mathbb{P}$-a.s.), whereas in (0.3) the third exponential term was negligible. Thus, for small time, there are *more* non-negligible exponential terms in the random heat kernel (0.1) than in the deterministic heat kernel (0.2).

The explicit expression (0.4), along with the asymptotics of (0.5), also points out a potential problem in technique. A general way of studying short-time asymptotics of heat kernels is by rescaling time. Essentially, rescaling here would replace W_t in (0.4) by $t^{1/2}\eta$, where η is any zero-mean unit-variance Gaussian random variable. Then for each $t > 0$ and x and y in $\mathbb{R}$, we have the following equality of *law*:

$$p_y^{can}(t,x) \overset{(\mathcal{L})}{=} \exp\left[-\frac{|x-y|^2}{2t} - \left(\frac{x-y}{t^{1/2}}\right)(c\eta) - \frac{(c\eta)^2}{2}\right] \Big/ \sqrt{2\pi t},$$

which does not at all reflect the second asymptotic statement of (0.5), which forms the focus of our interest. Thus, as in the study of Brownian motion, we must distinguish between the *almost-sure* asymptotics of p_y^{can} and the asymptotics of the *law* of p_y^{can}. To be explicit, our aim is a study of its *almost-sure* asymptotics.

1. The main result

The main result is given by Theorem 1 below. Preparatory to stating this theorem, we need some notation. Let $\overline{\nabla}$ denote the Levi-Civita connection and let R_{op} and R respectively denote the curvature operator and curvature tensor. To remove any confusion about signs, we use the definition $R_{op}(X,Y)Z \overset{\triangle}{=} \overline{\nabla}_X\overline{\nabla}_Y Z - \overline{\nabla}_Y\overline{\nabla}_X Z - \overline{\nabla}_{[X,Y]}Z$ for any vector fields X, Y, and Z on M, where $[\cdot,\cdot]$ denotes the Lie Bracket of vector fields. Finally, for each $y \in M$, let $\mathrm{Cut}(y)$ denote the cut-locus of y.

We now define a certain stochastic differential equation (SDE). Let $TM \oplus TM$ be the Whitney sum of two copies of TM; i.e., $TM \oplus TM$ is the vector bundle over M for which the fibre over x, for each $x \in M$, is the Cartesian product $T_xM \times T_xM$. Fix $t > 0$, $y \in M$, and $x \in M \sim \mathrm{Cut}(y)$, and let $\{\gamma_y^{t,x}(s) : 0 \le s \le t\}$ be the unique geodesic which runs from y to x

in time t. Let $\{(\Xi_y^{1,t,x}(s), \Xi_y^{2,t,x}(s)) : 0 \leq s \leq t\}$ be the solution of the $TM \oplus TM$-valued SDE

$$
\begin{pmatrix} D_s\Xi_y^{1,t,x} \\ D_s\Xi_y^{2,t,x} \end{pmatrix} = \begin{pmatrix} 0 & I \\ R_{op}\left(\dot{\gamma}_y^{t,x}(s), \cdot\right)\dot{\gamma}_y^{t,x}(s) & 0 \end{pmatrix} \begin{pmatrix} \Xi_y^{1,t,x}(s) \\ \Xi_y^{2,t,x}(s) \end{pmatrix} ds
$$

(1.1)
$$
+ \begin{pmatrix} -\sigma(\gamma_y^{t,x}(s)) \\ \overline{\nabla}^*_{\dot{\gamma}_y^{t,x}(s)}\sigma \end{pmatrix} \circ dW_s,
$$

$$
\Xi_y^{1,t,x}(0) = 0_y,
$$

$$
\Xi_y^{1,t,x}(t) = 0_x,
$$

$$
0 \leq s \leq t.
$$

Here $\overline{\nabla}^*$ is the adjoint of $\overline{\nabla}$; i.e., $(\overline{\nabla}^*_Y\sigma, X) = (\overline{\nabla}_X\sigma, Y)$ for all vector fields X and Y. Secondly, 0_y and 0_x are respectively the zero elements of T_yM and T_xM. This equation describes two *random vector fields* along the *deterministic curve* $\gamma_y^{t,x}$. The form of this equation is simple. The operator

$$
\begin{pmatrix} X \\ Y \end{pmatrix} \mapsto \begin{pmatrix} Y \\ R_{op}\left(\dot{\gamma}_y^{t,x}(s), X\right)\dot{\gamma}_y^{t,x}(s) \end{pmatrix}
$$

which is the effect of the first term on the right of (1.1), is clearly linear for each $0 \leq s \leq t$. The $\circ\, dW_t$ term on the right of (1.1) does not depend on $\Xi_y^{1,t,x}$ or $\Xi_y^{2,t,x}$; thus the evolution of the pair $(\Xi_y^{1,t,x}, \Xi_y^{2,t,x})$ is governed by a *forced linear* SDE. The terminal condition on $\Xi_y^{1,t,x}$ makes the stochastic calculus associated with this SDE anticipative and therefore potentially problematic, but the simple linear form of (1.1) allows one to solve (1.1) using a simple variation of parameters formula and in the process resolve such anticipativity issues. Geometrically, the linear part of (1.1) is exactly the equation for *Jacobi fields*; thus one can interpret (1.1) in some sense as a forced Jacobi field (the reason for this will become clear at the end of Section 2).

We now can use the solution of (1.1) to define the third term in the exponential expansion of p_y. For each $t > 0$, $y \in M$ and $x \in M \sim \mathrm{Cut}(y)$, define

(1.2) $\quad \mathfrak{G}_y(t, x; W)$

$$
\stackrel{\triangle}{=} -\frac{1}{2}\int_0^t \|\Xi_y^{2,t,x}(s)\|^2\, ds - \frac{1}{2}\int_0^t R\left(\Xi_y^{1,t,x}(s), \dot{\gamma}_y^{t,x}(s), \Xi_y^{1,t,x}(s), \dot{\gamma}_y^{t,x}(s)\right)\, ds
$$

$$
- \int_0^t \left(\overline{\nabla}^*_{\dot{\gamma}_y^{t,x}(s)}\sigma, \Xi_y^{1,t,x}(s)\right) \circ dW_s.
$$

Similarly to (1.1), the stochastic calculus associated with the third expression on the right is complicated by some anticipativity problems (since $\Xi_y^{1,t,x}$ is anticipative); however, the above comment that we can solve (1.1) by a variation of parameters formula also leads to a meaningful definition of (1.2). Geometrically, (1.2) is similar to the *index form* in differential geometry—the second derivative of the energy functional of curves (again, the reason for this will be made clearer at the end of Section 2).

We finally have the main result:

Theorem 1. *We have that for each $\varepsilon \in (0, 1/2)$,*

(1.3)
$$p_y(t,x) = \exp\left[-\frac{d^2(x,y)}{2t} - \int_0^t \left(\dot{\gamma}_y^{t,x}(s), \sigma(\gamma_y^{t,x}(s)) \right) \circ dW_s + \mathfrak{G}_y(t,x;W) \right.$$
$$\left. + O(t^{1/2-\varepsilon}) \right] \Theta_y(x)(2\pi t)^{-d/2}$$

$\mathbb{P}$-*a.s. as t tends to zero, for $y \in M$ and $x \in M \sim \mathrm{Cut}(y)$.*

This is the counterpart of (0.3).

2. Heuristics

Where do the formulæ of Section 1 come from? Here we give a brief outline.

Let's start with a classical way of focussing on the short-time behavior of heat kernels—rescaling time. For every $\varepsilon > 0$, define

(2.1)
$$p_y^\varepsilon(t,x) \overset{\triangle}{=} p_y(t\varepsilon, x). \qquad t > 0,\ x \in M$$

Then p_y^ε satisfies the SPDE

$$dp_y^\varepsilon = \frac{\varepsilon}{2}\Delta p_y^\varepsilon\, dt + \sqrt{\varepsilon}(\sigma, \nabla p_y^\varepsilon) \circ dW_t^\varepsilon, \quad t > 0; \qquad \lim_{t \to 0} p_y^\varepsilon(t, \cdot) = \delta_y.$$

Here W^ε is the Wiener process defined by

(2.2)
$$W_t^\varepsilon \overset{\triangle}{=} \varepsilon^{-1/2} W_{t\varepsilon}. \qquad t > 0$$

Ultimately, we will want to vary ε. We will, however, want to be careful when we do this, since the dependence of W^ε upon ε is very irregular due to the irregularity of Brownian motion. To somewhat separate the different

effects of ε in (2.1), let us now take a fixed Wiener process B on $(\Omega, \mathcal{F}, \mathbb{P})$ which does not depend on $\varepsilon > 0$. Then we can study the SPDE

$$(2.3) \quad d\tilde{p}_y^\varepsilon = \frac{\varepsilon}{2} \Delta \tilde{p}_y^\varepsilon \, dt + \sqrt{\varepsilon}(\sigma, \nabla \tilde{p}_y^\varepsilon) \circ dB_t, \quad t > 0; \qquad \lim_{t \to 0} \tilde{p}_y^\varepsilon(t, \cdot) = \delta_y$$

and, assuming that the solution of this SPDE exists and is unique, we have that $p_y^\varepsilon \overset{(\mathcal{L})}{=} \tilde{p}_y^\varepsilon$. Note that since the scaling properties of Wiener processes are statements about the *laws* of Wiener processes, we have lost the possibility of directly studying the *almost-sure* behavior of p_y; we can only study the *law* of p_y (recall the comments at the end of Section 0). However, we shall for the present ignore this obstacle—we here are searching only for heuristic arguments.

Next we replace B in (2.3) with a smooth function b and consider the solution of

$$dq_y^{b,\varepsilon} = \frac{\varepsilon}{2} \Delta q_y^{b,\varepsilon} \, dt + \sqrt{\varepsilon}(\sigma, \nabla q_y^{b,\varepsilon})\dot{b}(t), \quad t > 0; \qquad \lim_{t \to 0} q_y^{b,\varepsilon}(t, \cdot) = \delta_y;$$

this is similar to (0.2), so at least we are in known territory. We do not really care here whether b is random or deterministic; we wish only to find some *dependencies* of the behavior of $q_y^{b,\varepsilon}$ upon b; in particular, we will try to identify some function $\Gamma_y : M \times (0, 1) \times C^1([0, \infty); \mathbb{R}) \to \mathbb{R}$ such that as ε tends to zero, $q_y^{b,\varepsilon}(1, x) \asymp \exp[\Gamma_y(x; \varepsilon, b)]$. Here "$\asymp$" denotes logarithmic equivalence. By the celebrated result of Wong and Zakai [21], if we take b to be random and an approximation of B, then $q_y^{b,\varepsilon}$ should approximate $\tilde{p}_y^\varepsilon$, at least in law, and thus hopefully we will have in some sense that

$$(2.4) \qquad\qquad \tilde{p}_y^\varepsilon(1, x) \asymp \exp[\Gamma_y(x; \varepsilon, B)].$$

Of course to carry out any analysis of this, the dependence of Γ_y upon B should be sufficiently simple. If (2.4) is true, then we next may hope that in some sense $p_y^\varepsilon(1, x) \asymp \exp[\Gamma_y(x; \varepsilon, W^\varepsilon)]$ and, since $p_y(t, x) = p_y^t(1, x)$ for all $t > 0$ and $x \in M$, that finally in some sense

$$(2.5) \qquad\qquad p_y(t, x) \asymp \exp[\Gamma_y(x; t, W^t)].$$

Recall now that $q_y^{b,\varepsilon}$ can be interpreted as the density of a certain M-valued diffusion. The theory of *large deviations* (see [5]) now suggests that the asymptotics of $q_y^{b,\varepsilon}$ as ε tends to zero are

$$(2.6) \qquad\qquad q_y^{b,\varepsilon}(1, x) \asymp \exp\left[-\frac{1}{\varepsilon} H_y(x; \varepsilon, b)\right]$$

for any $t > 0$, x in M, and b in $C^1([0, \infty); \mathbb{R})$, where for each $\varepsilon \in (-1, 1)$

$$(2.7) \qquad H_y(x; \varepsilon, b) \triangleq \inf_{\substack{\varphi \in C^1([0,1]; M) \\ \varphi(0) = y, \varphi(1) = x}} \frac{1}{2} \int_0^1 \|\dot{\varphi}(s) + \sqrt{\varepsilon}\sigma(\varphi(s))\dot{b}(s)\|^2 \, ds$$

(see [9], [13]). This is slightly different from the classical result of Varadhan [19]. The classical result follows from this by realizing that the effect of $\sqrt{\varepsilon}\sigma(\varphi(s))\dot{b}(s)$ becomes negligible as ε becomes small. One might say that here we are stopping at an intermediate step in the proof of the classical result. The motivation of (2.7) is the Azencott approach to large deviations (see [1]); $q_y^{b,\varepsilon}$ can be understood as describing the law of a process given by transforming a small Wiener process, where this transformation itself depends on ε. Note also that we are considering curves which go from y to x; usually one considers curves from x to y, which would change the "+" sign to a "$-$" sign; it is more natural here to integrate in the same direction that b evolves. The asymptotics (2.6) and the "action functional" of (2.7) suggest that the desired asymptotics of p_y should be given by taking

$$(2.8) \qquad \Gamma_y(x; \varepsilon, b) \approx -\frac{1}{\varepsilon} H_y(x; \varepsilon, b)$$

and realizing the program outlined above. Note, however, a serious problem: to proceed to the next step, we would have to replace b in (2.7) with a trajectory of the Wiener process B; this would formally give us

$$H_y(x; \varepsilon, B) = \inf_{\substack{\varphi \in C^1([0,1]; M) \\ \varphi_0 = y, \varphi_1 = x}} \frac{1}{2} \int_0^1 \|\dot{\varphi}_s + \sqrt{\varepsilon}\sigma(\varphi_s) \circ \dot{B}_s\|^2 \, ds$$

at which point we are faced with the problem that Wiener processes are nowhere differentiable and this expression is not easily made rigorous.

The resolution of this difficulty lay in *expanding H_y in powers of $\sqrt{\varepsilon}$*; we write

$$H_y(x; \varepsilon, b) \approx H_y(x; 0, b) + \sqrt{\varepsilon}\dot{H}_y(x; 0, b) + \frac{\varepsilon}{2}\ddot{H}_y(x; 0, b)$$

where

$$\dot{H}_y(x; \varepsilon, b) \triangleq \frac{\partial H_y}{\partial \sqrt{\varepsilon}}(x; \varepsilon, b); \qquad \ddot{H}_y(x; \varepsilon, b) \triangleq \frac{\partial^2 H_y}{\partial (\sqrt{\varepsilon})^2}(x; \varepsilon, b)$$

for all $\varepsilon \in (-1, 1)$. Then instead of (2.8), we can define $\Gamma_y(x; \varepsilon, b)$ as

$$\Gamma_y(x; \varepsilon, b) \triangleq -\frac{1}{\varepsilon}H_y(x; 0, b) - \varepsilon^{-1/2}\dot{H}_y(x; 0, b) - \frac{1}{2}\ddot{H}_y(x; 0, b).$$

Since we are still only interested in heuristics, we do not need to verify here that H_y is indeed twice-differentiable with respect to $\sqrt{\varepsilon}$ (issues of dependence of calculus of variation problems upon parameters are often delicate).

The expressions for $H_y(x; 0, b)$, $\dot{H}_y(x; 0, b)$, and $\ddot{H}_y(x; 0, b)$ follow from standard calculus of variations calculations. We will give only some representative parts of these calculations. We have that

$$H_y(x; 0, b) = \frac{d^2(x, y)}{2}; \qquad \dot{H}_y(x; 0, b) = \int_0^1 \left(\sigma(\gamma_y^{1,x}(s)), \dot{\gamma}_y^{1,x}(s) \right) \dot{b}(s)\, ds.$$

The term $\ddot{H}_y(x; 0, b)$ similarly has an explicit formula. Note that since $\ddot{H}_y(x; 0, b)$ can be understood as the second-derivative of a functional of $\dot{b}$ in the direction of $\dot{b}$, $\ddot{H}_y(x; 0, b)$ is a quadratic functional of $\dot{b}$. Thus

$$\Gamma_y(x; \varepsilon, b)$$
$$= -\frac{1}{\varepsilon} \left\{ \frac{d^2(x, y)}{2} + \sqrt{\varepsilon} \int_0^1 \left(\sigma(\gamma_y^{1,x}(s)), \dot{\gamma}_y^{1,x}(s) \right) \dot{b}(s)\, ds + \frac{\varepsilon}{2} \ddot{H}_y(x; 0, b) \right\}.$$

Now let b be a smooth approximation of the Wiener process B. We get that in some sense

$$\Gamma_y(x; \varepsilon, B) = -\frac{d^2(x, y)}{2\varepsilon} - \frac{1}{\sqrt{\varepsilon}} \int_0^1 \left(\sigma(\gamma_y^{1,x}(s)), \dot{\gamma}_y^{1,x}(s) \right) \circ dB_s - \frac{1}{2} \ddot{H}_y(x; 0, B).$$

By our previous comments about $\ddot{H}_y$, we see that the term $\ddot{H}_y(x; 0, B)$ can be understood as a second-order integral in B, which of course can be rigorously understood. Finally substituting W^t for B, as required by (2.5) and then undoing the scaling of (2.2), we get that in some sense

$$\Gamma_y(x; t, W^t)$$
$$= -\frac{d^2(x, y)}{2t} - \frac{1}{\sqrt{t}} \int_0^1 \left(\sigma(\gamma_y^{1,x}(s)), \dot{\gamma}_y^{1,x}(s) \right) \circ dW_s^t - \frac{1}{2} \ddot{H}_y(x; 0, W^t)$$
$$= -\frac{d^2(x, y)}{2t} - \int_0^t \left(\sigma(\gamma_y^{t,x}(s)), \dot{\gamma}_y^{t,x}(s) \right) \circ dW_s - \frac{1}{2} \ddot{H}_y(x; 0, W^t).$$

Since the term $\ddot{H}_y(x; 0, W^t)$ is a second-order integral in W^t, the scaling of (2.2) can be reversed as in the work term. We get exactly (1.1), (1.2), and (1.3). Reviewing our arguments, we also can now explain more clearly the similarity of (1.1) and (1.2), respectively, to Jacobi fields and the index form. Jacobi fields and the index form arise in studying the second variation of the energy of curves; (2.7) can be viewed as a perturbation of the energy functional, and the third exponential term $\mathfrak{G}_y$ in (1.3) is essentially given by the second variation of H_y in $\sqrt{\varepsilon}$.

3. Completion

Thus far we have a general guess at the short-time asymptotics of the solution of (0.1), namely

$$\hat{P}_y(t,x) = \exp\left[-\frac{d^2(x,y)}{2t} - \int_0^t \left(\dot{\gamma}_y^{t,x}(s), \sigma(\gamma_y^{t,x}(s))\right) \circ dW_s \right.$$

$$\left. + \mathfrak{G}_y(t,x;W)\right]\Theta_y(x)(2\pi t)^{-d/2},$$

$$t > 0, \ y \in M, \ x \in M \sim \mathrm{Cut}(y).$$

What remains is to show that this guess is correct. It turns out that some of the technicalities of anticipative stochastic calculus preclude standard parametrix-type calculations (or at least they make such calculations much more demanding). What we can do, however, is to write an SPDE for $\hat{P}_y$. The completion of the proof comes from using SPDE methods to study the error between p_y and $\hat{P}_y$. These calculations, which are carried out in [18], are somewhat lengthy.

To get a better feel for the complexity of things, let's see why standard parametrix calculations are inapplicable. We can extend the definition of $\hat{P}_y$ by setting

$$\underline{\hat{P}}(t,x;s,y)$$

$$\stackrel{\triangle}{=} \exp\left[-\frac{d^2(x,y)}{2(t-s)} - \int_s^t \left(\dot{\gamma}_y^{t-s,x}(r-s), \sigma(\gamma_y^{t-s,x}(r-s))\right) \circ dW_r \right.$$

$$\left. + \mathfrak{G}_y(t-s,x;W\circ\theta_s)\right]\Theta_y(x)(2\pi(t-s))^{-d/2},$$

$$t > s \geq 0, \ y \in M, \ x \in M \sim \mathrm{Cut}(y)$$

where for each $s \geq 0$, $W \circ \theta_s$ is the Wiener process defined by $(W \circ \theta_s)(r) \stackrel{\triangle}{=} W_{r+s} - W_s$ for all $r \geq 0$. Then for each $s \geq 0$ and $y \in M$, $\{\underline{\hat{P}}(t,x;s,y) : t > s, \ x \in M \sim \mathrm{Cut}(y)\}$ is the natural approximation for the solution of the SPDE

$$dp_{s,y} = \frac{1}{2}\Delta p_{s,y}\, dt + (\sigma, \nabla p_{s,y}) \circ dW_t, \quad t > 0; \qquad \lim_{t \to s} p_{s,y}(t,\cdot) = \delta_y.$$

The parametrix method then suggests that we look for the solution of (0.1) in the form

$$(3.1) \quad p_y(t,x) = \hat{P}_y(t,x) + \int_0^t \int_M \underline{\hat{P}}(t,x;s,z)g_y^0(s,z)\mathfrak{V}(dz)\, ds$$

$$+ \int_0^t \int_M \underline{\hat{P}}(t,x;s,z)g_y^1(s,z)\mathfrak{V}(dz) \circ dW_s, \qquad t \geq 0, \ x \in M$$

for some functions g_y^0 and g_y^1, where $\mathfrak{V}$ is the Riemannian volume measure on $(M, \mathcal{B}(M))$ (here we have ignored problems arising from cut-loci). It turns out that the g_y^i's should be adapted to the filtration generated by W; however, for any $0 \leq s \leq t$, the function $\underline{\hat{P}}(t, x; s, \cdot)$ is measurable with respect to the *future* sigma-field $\sigma\{W_r - W_s : s \leq r \leq t\}$; this puts the parametrix procedure suggested by (3.1) outside the framework of classical Ito-based stochastic calculus. Unless there are some delicate integrability calculations one could use, this more or less precludes the whole parametrix approach.

There are, fortunately, several ways to resolve this difficulty. In [18], we follow an approach which appears to give final results which very closely parallel those of deterministic kernels like q_y of (0.2) (see Theorem 2 below). We find an SPDE for $\hat{P}_y$. This allows us to write down an SPDE for the difference

$$(3.2) \qquad\qquad p_y - \hat{P}_y.$$

Suppose that, purely by SPDE methods, we can show that this SPDE (for the difference random field (3.2)) has a solution r_y which is in some sense "small". Then the random field $\hat{P}_y + r_y$ satisfies the same SPDE (0.1) as p_y, so necessarily $p_y = \hat{P}_y + r_y$; since r_y is small, we would have shown that $\hat{P}_y$ is indeed a good approximation of p_y.

Actually, the arguments of [18] are even a bit more complex. As a preliminary to the arguments just outlined, we carry out a Minakshisundaram-Pleijel-type expansion. This involves defining

$$(3.3) \quad \hat{P}_y^*(t, x) \triangleq \hat{P}_y(t, x) \left\{ 1 + \alpha_y^0(t, x) + \int_{s=0}^t \alpha_y^1(t, x; s) \circ dW_s \right.$$

$$+ \int_{s_1=0}^t \int_{s_2=0}^{s_1} \alpha_y^2(t, x; s_1, s_2) \circ dW_{s_2} \circ dW_{s_1} + \cdots$$

$$\left. + \int_{s_1=0}^t \int_{s_2=0}^{s_1} \cdots \int_{s_k=0}^{s_{k-1}} \alpha_y^k(t, x; s_1, s_2, \ldots s_k) \circ dW_{s_k} \circ \cdots \circ dW_{s_2} \circ dW_{s_1} \right\},$$

$$t > 0, \ x \in M \sim \text{Cut}(y)$$

for some well-chosen positive integer k and some collection of deterministic integrands $\alpha_y^0, \alpha_y^1, \ldots \alpha_y^k$. These α_y^k's are some form of "polynomials", which may be thought of as following some generalized Minakshisundaram-Pleijel recursion. Replacing $\hat{P}_y$ by $\hat{P}_y^*$ in (3.2) and the succeeding arguments, we can more easily show that the difference random field of (3.2) is small. In the end, we get

Theorem 2. *We have that for any integer $l \geq 1$, if k is large enough in* (3.3)*, there is a constant $K > 0$ such that*

$$p_y(t, x) = \hat{P}_y^*(t, x) + \hat{P}_y(t, x)O(t^l) + O(e^{-K/t})$$

$\mathbb{P}$-*a.s. as t tends to zero, for $y \in M$ and $x \in M \sim \mathrm{Cut}(y)$.*

References

1. R. Azencott, *Grandes Dèviations et Applications*, École d'Éte de Probabilités de Saint-Flour VIII, 1978 (P. L. Hennequin, ed.), Lecture Notes in Mathematics 774, Springer, New York, 1980.
2. R. Azencott, et. al., *Géodésiques et diffusions en temps petit*, Astérisque 84–85, Sociéte Mathématique de France, 1981.
3. N. Berline, E. Getzler and M. Vergne, *Heat Kernels and Dirac Operators*, Springer Verlag, New York, 1992.
4. J.-M. Bismut, *Large Deviations and the Malliavin Calculus*, Birkhäuser, Boston, 1984.
5. J.-D. Deuschel and D. W. Stroock, *Large Deviations*, Academic Press, New York, 1989.
6. K.D. Elworthy, *Stochastic Differential Equations on Manifolds*, London Mathematical Society Lecture Note Series 70, Cambridge University Press, London, 1982.
7. D. Elworthy, *Geometric Aspects of Diffusions on Manifolds*, École d'Éte de Probabilités de Saint-Flour XV–XVII, 1985–87 (P. L. Hennequin, ed.), Lecture Notes in Mathematics 1362, Springer, New York, 1988, pp. 277–421.
8. D. Elworthy and A. Truman, *Classical mechanics, the diffusion (heat) equation and the Schrödinger equation on a Riemannian manifold*, J. Math. Phys. **22** (1981), 2144–2166.
9. M. I. Freidlin and A. D. Wentzell, *Random Perturbations of Dynamical Systems*, Springer, New York, 1984.
10. N. Ikeda, *Probabilistic Methods in the study of Asymptotics*, École d'Éte de Probabilités de Saint-Flour XVIII, 1988 (P. L. Hennequin, ed.), Lecture Notes in Mathematics 1427, Springer, New York, 1990, pp. 197–325.
11. Y. Kannai, *Off diagonal short time asymptotics for fundamental solutions of diffusion equations*, Comm. in Partial Differential Equations **2** (1977), 781–830.
12. I. Karatzas and S. Shreve, *Brownian Motion and Stochastic Calculus*, Springer, New York, 1987.
13. Yu. I. Kifer, *On the asymptotics of the transition density of processes with small diffusion*, Prob. Theory and Appl. **21** (1976), 513–522.
14. S.A. Molchanov, *Diffusion processes and Riemannian geometry*, Russian Math Surveys **30** (1975), 1–63.
15. D. Revuz, and M. Yor, *Continuous Martingales and Brownian Motion*, Springer, New York, 1991.
16. B. L. Rozovskii, *On the mathematical theory of a hydromagnetic dynamo in a random flow*, Soviet Math. Dokl. **35** (1987), 462–465.
17. B. L. Rozovskii, *Stochastic Evolution Systems: Linear Theory and Applications to Non-linear Filtering*, Kluwer Academic Publishers, Boston, 1990.
18. R. B. Sowers, *Short-Time Geometry of Random Heat Kernels*, submitted.

19. S.R.S. Varadhan, *On the behavior of the fundamental solution of the heat equation with variable coefficients*, Comm. Pure Appl. Math. **20** (1967), 432–455.
20. K.D. Watling, *Formulae for the heat kernel of an elliptic operator exhibiting small-time asymptotics*, Stochastic Mechanics and Stochastic Processes (A. Truman and I.M. Davies, eds.), Lecture Notes in Mathematics 1325, Springer, New York, 1988, pp. 167–180.
21. E. Wong and M. Zakai, *On the relation between ordinary and stochastic differential equations*, Intern. J. Engng. Sci. **3** (1965), 213–229.
22. Huilong Zhang, *Diffusion en temps petit: Application a l'equation de Zakai*, preprint.

Department of Mathematics, Northwestern University, Evanston, Illinois, 60208

E-mail address: sowers@math.nwu.edu