

## ESTIMATES OF DERIVATIVES OF THE HEAT KERNEL ON A COMPACT RIEMANNIAN MANIFOLD

ELTON P. HSU

(Communicated by Stanley Sawyer)

ABSTRACT. We give global estimates on the covariant derivatives of the heat kernel on a compact Riemannian manifold on a fixed finite time interval. The proof is based on analyzing the behavior of the heat kernel along Riemannian Brownian bridge.

### 1. INTRODUCTION

Let  $M$  be a compact Riemannian manifold of dimension  $n$  and  $p(T, x, y)$  the heat kernel on  $M$ . The present work concerns with estimates of derivatives of the heat kernel. Let  $\nabla^N \log p(T, x, y)$  be the  $N$ th covariant derivative of the logarithm of the heat kernel with respect to its first space variable. The study of such estimates is motivated by various problems involving Brownian bridge on  $M$  where one needs to control the behavior of the process at the terminal time. Estimates of this kind were obtained previously by Sheu [6]. In our present setting, his results can be restated as follows. For all  $(T, x, y) \in (0, 1] \times M \times M$ ,

$$|\nabla \log p(T, x, y)| \leq C_1 \left\{ \frac{d(x, y)}{T} + \frac{1}{\sqrt{T}} \right\},$$
$$|\nabla^N \log p(T, x, y)| \leq C_N \left\{ \frac{d(x, y)}{T} + \frac{1}{\sqrt{T}} \right\}^N \left\{ \frac{d(x, y)}{\sqrt{T}} + 1 \right\}^{2(N-2)},$$

where  $C_N$  is a constant depending on  $N$  and the manifold  $M$ . One expects that for  $N \geq 3$  an estimate without the last factor should be the one with the correct order of magnitude. The purpose of this work is to prove such an inequality.

From Sheu's work we find two basic observations. First, from stochastic control theory it is more natural to work with derivatives of  $\log p(T, x, y)$  than with those of  $p(T, x, y)$  itself. Second, the optimal control is attained at Brownian bridge. These two observations lead us to consider directly the process  $\log p(T - t, \gamma_t, y)$ , where  $\{\gamma_t, 0 \leq t \leq T\}$  is a Brownian bridge from  $x$  to  $y$  in time  $T$ . Using repeatedly Itô's formula on the process, we express  $\nabla^N \log p(T, x, y)$  in terms of lower derivatives and obtain the desired estimates by induction, much in the same way as was done in Sheu [6]. It seems that our more intrinsic approach to the problem resulted in

---

Received by the editors March 24, 1997 and, in revised form, December 22, 1997 and February 26, 1998.

1991 *Mathematics Subject Classification*. Primary 58G32; Secondary 60J60.

The research was supported in part by NSF grant 9406888-DMS.

a precise tracking of the induction step, thus allowing us to obtain estimates with the correct order of magnitude. Our result can be stated as follows.

**Theorem 1.1.** *For each  $N$ , there is a constant  $C_N$  depending on  $N$  and  $M$  such that for all  $(T, x, y) \in (0, 1] \times M \times M$*

$$|\nabla^N \log p(T, x, y)| \leq C_N \left\{ \frac{d(x, y)}{T} + \frac{1}{\sqrt{T}} \right\}^N.$$

The following corollary is immediate.

**Corollary 1.2.** *For each  $N$ , there is a constant  $D_N$  depending on  $N$  and  $M$  such that for all  $(T, x, y) \in (0, 1] \times M \times M$*

$$|\nabla^N p(T, x, y)| \leq D_N \left\{ \frac{d(x, y)}{T} + \frac{1}{\sqrt{T}} \right\}^N p(T, x, y).$$

More recent discussions on the cases  $N \leq 2$  can be found in Hamilton [2], Malliavin and Stroock [4], Stroock [7], and Stroock and Zeitouni [9]. We emphasize that the estimate stated in the theorem holds for all  $(T, x, y) \in [0, 1] \times M \times M$ . The argument in Malliavin and Stroock [4] shows that the estimate of the form stated in the theorem is in general best possible. However, if  $x, y$  are kept a positive distance away from the cut locus, the term  $1/\sqrt{T}$  can be dispensed with. For a detailed discussion on this case, see Norris [5].

We are happy to acknowledge that Theorem 1.1 was obtained independently in a recent work of Stroock and Turetsky [8] by a different method.

Finally I thank Professors Stanley Sawyer, Wilfrid Kendall, and an anonymous reviewer for their helpful comments on the work.

## 2. PROOF OF THE THEOREM

In the course of the proof, the letter  $C$  will denote a constant depending on the index  $N$  and the manifold  $M$  whose value may differ from one appearance to another.

Let  $O(M)$  be the orthonormal frame bundle of  $M$  and  $\pi : O(M) \rightarrow M$  the canonical projection. We use  $H_i, 1 \leq i \leq n$ , to denote the canonical horizontal vector fields and  $\Omega_{ij}^*, 1 \leq i, j \leq n$ , the canonical vertical vector fields on  $O(M)$ . We denote by  $\Omega$  the  $o(n)$ -valued curvature form on  $O(M)$ . By the structure equations, we have the following commutation relations:

$$(1) \quad [H_i, H_j] = \Omega_{ij}^*, \quad [H_i, \Omega_{jk}^*] = \Omega_{jk}^{il} H_l, \quad [\Omega_{ij}^*, \Omega_{kl}^*] = c_{ij,kl}^{ab} \Omega_{ab}^*,$$

where  $c_{ij,kl}^{ab}$  are the structure constants of  $o(n)$ , whose explicit values we do not need. If  $I = \{i_1, \dots, i_l\}$  is a multi-index with length  $|I| = l$ , then we use the abbreviation  $H_I J = H_{i_1} \cdots H_{i_l} J$  for a function  $J$  on  $O(M)$ .

Let  $J(t, u) = \log p(t, \pi u, y)$ , the lift of  $\log p(t, x, y)$  to  $O(M)$ . Then it satisfies the equation

$$(2) \quad \partial_t J(T - t, u) + \frac{1}{2} \Delta^H J(T - t, u) + \frac{1}{2} |\nabla^H J(T - t, u)|^2 = 0,$$

where  $\Delta^H = \sum_{i=1}^n H_i^2$  is Bochner's Laplacian on  $O(M)$  and

$$\nabla^H J = \{H_1 J, \dots, H_n J\}$$

is the horizontal gradient of  $J$ .

Let  $\{\gamma_t\}$  be a Brownian bridge from  $x$  to  $y$  in time  $T$  and  $\{u_t\}$  its horizontal lift with initial value  $u_o$ , where  $u_o$  is an orthonormal frame over  $x$ . It is well known that there is a Brownian motion  $\{b_t\}$  such that

$$(3) \quad du_t = H_{u_t} \circ db_t + \nabla^H J(T-t, u_t) dt.$$

Using (3), and Itô's formula we have (with  $J(T-t, u_t)$  abbreviated as  $J$ ):

$$dH_I J = \langle \nabla^H H_I J, db_t \rangle + \left\{ \partial_t H_I J + \frac{1}{2} \Delta^H H_I J + \langle \nabla^H H_I J, \nabla^H J \rangle \right\} dt.$$

Inserting the equation for  $J$  in (2) we have

$$(4) \quad dH_I J = \langle \nabla^H H_I J, db_t \rangle + \{F_I + G_I\} dt,$$

where

$$F_I = \frac{1}{2} [\Delta^H, H_I] J,$$

and

$$G_I = \langle \nabla^H H_I J, \nabla^H J \rangle - \frac{1}{2} H_I \langle \nabla^H J, \nabla^H J \rangle.$$

Consider first the cases  $|I| = 0$  and  $1$ . For  $I$  of length  $0$  we have  $F_I = 0$  and  $G_I = \frac{1}{2} |\nabla^H J|^2$ . Integrating (4) from  $0$  to  $T/2$  and taking expectation we have

$$E \int_0^{T/2} |\nabla^H J(T-t, u_t)|^2 dt = 2EJ(T/2, u_{T/2}) - 2J(T, u_o).$$

From Bellanche [1] or Li and Yau [3], there is a constant  $C$  such that

$$\frac{C^{-1}}{t^{n/2}} e^{-d(x,y)^2/Ct} \leq p(t, x, y) \leq \frac{C}{t^{n/2}}.$$

Hence we have immediately

$$E \int_0^{T/2} |\nabla^H J(T-s, u_s)|^2 ds \leq C \left\{ \frac{d(x, y)^2}{T} + 1 \right\}.$$

For the sake of simplicity we set

$$Q = \frac{d(x, y)}{T} + \frac{1}{\sqrt{T}}.$$

The above estimate can be written as

$$(5) \quad E \int_0^{T/2} |\nabla^H J(T-t, u_t)|^2 ds \leq CTQ^2.$$

We now apply (4) to an index  $I = \{i\}$  of length  $1$ . In this case using (1) and noting that  $\Omega_{ij}^* J = 0$  we see that  $F_I$  is a linear combination of  $H_j J$  and  $G_I = 0$ . Therefore  $|F_I| \leq C |\nabla^H J|$ . Integrating (4) from  $0$  to  $t$  and taking expectation, we have

$$H_i J(T, u_o) = E H_i J(T-t, u_t) - E \int_0^t F_I dt.$$

Integrating from  $0$  to  $T/2$  we have

$$\frac{T}{2} H_i J(T, u_o) = E \int_0^{T/2} H_i J(T-t, u_t) dt - E \int_0^{T/2} \left( \frac{T}{2} - t \right) F_I dt.$$

Using the Cauchy-Schwarz inequality and (5) we have  $|H_i J(T, u_o)| \leq CQ$ , which proves the case  $N = 1$  of the main theorem.

Now it is clear how we should proceed inductively for the general case. The induction hypothesis has two parts. First,

$$(6) \quad E \int_0^{T/2} |\nabla^H J(T-t, u_t)|^2 dt \leq CQ^2,$$

and for all  $I$  such that  $2 \leq |I| \leq N$ ,

$$(7) \quad E \int_0^{T/2} |H_I J(T-t, u_t)|^2 dt \leq CQ^{2(|I|-1)}.$$

Second, for all  $I$  such that  $|I| \leq N$ ,

$$(8) \quad |H_I J(T, u_o)| \leq CQ^{|I|}.$$

Note that (6)–(8) are supposed to hold uniformly for all  $(T, u_o) \in (0, 1] \times O(M)$ . We have already proved the initial step  $N = 1$ . Suppose the above inequalities hold for  $N$  and let  $I$  be an index of length  $N$ . Integrating (4) from 0 to  $T/2$  we have

$$\begin{aligned} \int_0^{T/2} \langle \nabla^H H_I J(T-t, u_t), db_t \rangle &= H_I J(T/2, u_{T/2}) - H_I J(T, u_o) \\ &\quad - \int_0^{T/2} \{F_I + G_I\} dt. \end{aligned}$$

Squaring and taking expected value, we have

$$\begin{aligned} (9) \quad & E \int_0^{T/2} |\nabla^H H_I J(T-t, u_t)|^2 dt \\ & \leq C \left\{ E \int_0^{T/2} |F_I + G_I| dt \right\}^2 \\ & \quad + CE |H_I J(T/2, u_{T/2})|^2 + C |H_I J(T, u_o)|^2. \end{aligned}$$

From (8) we have

$$(10) \quad E |H_I J(T/2, u_{T/2})|^2 + |H_I J(T, u_o)|^2 \leq CQ^{2N}.$$

Using the commutation relations (1) we can write  $[\Delta^H, H_I] J$  as a linear combination of  $H_L J$  with  $|L| = |I|$ . Hence for  $|I| = N = 1$  we have by (6)

$$E \int_0^{T/2} |F_I|^2 dt \leq CQ^2 = CQ^{2N}$$

and for  $N > 1$  we have by (7)

$$E \int_0^{T/2} |F_I|^2 dt \leq CQ^{2(N-1)} \leq CQ^{2N}.$$

Hence we always have

$$E \int_0^{T/2} |F_I|^2 dt \leq CQ^{2N}.$$

For the terms involving  $G_I$ , we observe first that  $G_I = 0$  if  $N = |I| = 1$  and if  $N \geq 2$ , then by the commutation relations (1)  $G_I$  is a linear combination of the terms of

the form  $\langle H_K J, H_L J \rangle$  with  $2 \leq |K| \leq N$ ,  $|L| = N + 2 - |K|$  or  $|K| = N - 1$ ,  $|L| = 1$ . In the first case we have  $|L| \geq 2$  and  $|K| \geq 2$ , and by (7)

$$\begin{aligned} & \left\{ E \int_0^{T/2} |\langle H_K J, H_L J \rangle| dt \right\}^2 \\ & \leq E \left\{ \int_0^{T/2} |H_K|^2 dt \right\} E \left\{ \int_0^{T/2} |H_L|^2 dt \right\} \\ & \leq CQ^{2(|K|-1)} Q^{2(|L|-1)} \\ & = CQ^{2N}. \end{aligned}$$

In the second case we have  $|L| = 1$ ,  $N \geq 2$ , and using both (6) and (7) we obtain the same bound  $CQ^{2N}$ . Therefore we have

$$\left\{ E \int_0^{T/2} |G_I| dt \right\} \leq CQ^N.$$

Thus we have shown that (6) and (7) for all indices of length less than  $|I|$  imply

$$(11) \quad E \int_0^{T/2} |F_I + G_I| dt \leq CQ^{2|I|}.$$

From (9), (10), and (11) we have

$$E \int_0^{T/2} |\nabla^H H_I J(T - t, u_t)| dt \leq CQ^N.$$

This proves (7) for all indices of length  $N + 1$ .

To prove (8) for indices of length  $N + 1$ , we assume that  $I$  is such an index. Integrating (4) from 0 to  $t$  and taking expectation, we have

$$H_I J(T, u_o) = E H_I J(T - t, u_t) - E \int_0^t \{F_I + G_I\} ds.$$

Integrating from 0 to  $T/2$ , we have

$$(12) \quad \begin{aligned} \frac{T}{2} H_I J(T, u_o) &= E \int_0^{T/2} H_I J(T - t, u_t) dt \\ &+ E \int_0^{T/2} \left( \frac{T}{2} - t \right) \{F_I + G_I\} dt. \end{aligned}$$

Since  $|I| = N + 1$ , the first term on the right-hand side can be estimated by (7) and we obtain

$$(13) \quad \left| E \int_0^{T/2} H_I J(T - t, u_t) dt \right| \leq C\sqrt{T}Q^N \leq CTQ^{N+1}.$$

The argument leading to (11) shows that

$$E \int_0^{T/2} |F_I + G_I| dt \leq CQ^{N+1}.$$

Hence

$$(14) \quad E \int_0^{T/2} \left( \frac{T}{2} - t \right) |F_I + G_I| dt \leq CTQ^{N+1}.$$

It follows from (12), (13), and (14) that  $|H_I J(T, u_o)| \leq CQ^{N+1}$ , and the theorem is proved.

## REFERENCES

1. Bellanche, C., Le comportement asymptotique de  $p(t, x, y)$ , *Asterisque*, **84-85** (1981), ed. by R. Azencott.
2. Hamilton, R., A matrix Harnack estimate for the heat equation, *Comm. Anal. & Geom.*, **1** (1993), 113–126. MR **94g**:58215
3. Li, P. and Yau, S.-T., On the parabolic kernel of the Schrödinger operator, *Acta Math.*, **156** (1986), 153–201. MR **87f**:58156
4. Malliavin, P. and Stroock, D. W., Short time behavior of the heat kernel and its logarithmic derivatives, *J. of Diff. Geom.*, **44**, No. 3 (1996), 550–570. MR **98c**:58164
5. Norris, J. R., Path integral formulae for heat kernels and their derivatives, *Prob. Theory and Related Fields*, **94** (1993), 525–541. MR **94c**:58224
6. Sheu, S.-J., Some estimates of the transition density function of a nondegenerate diffusion Markov process, *Annals of Probability*, **19**, No. 2 (1991), 538–561. MR **93a**:60087
7. Stroock, D. W., An estimate on the Hessian of the heat kernel, *Itô's Stochastic Calculus and Probability Theory*, edited by N. Ikeda, S. Watanabe, M. Fukushima, and H. Kunita, Springer (1996). MR **97m**:58212
8. Stroock, D. W., and Turetsky, J., Upper bounds on derivatives of the logarithm of the heat kernel, preprint (1996).
9. Stroock, D. W. and Zeitouni, O., Variations on a theme by Bismut, *Astérisque*, **236** (1996), 291–301. MR **97m**:58213

DEPARTMENT OF MATHEMATICS, NORTHWESTERN UNIVERSITY, EVANSTON, ILLINOIS 60208  
E-mail address: [elton@math.nwu.edu](mailto:elton@math.nwu.edu)