VARIATIONS ON A THEME BY BISMUT

Daniel W. Stroock and Ofer Zeitouni

M.I.T. and Technion

Abstract. Let M be a compact, connected, Riemannian manifold of dimension d, let $\{P_t : t > 0\}$ denote the Markov semigroups on $C(M)$ determined by $\frac{1}{2}\Delta$, and let $p_t(x,y)$ denote the kernel (with respect to the Riemannian volume measure) for the operator P_t. (The existence of this kernel as a positive, smooth function is well-known, see e.g. [D].) Bismut's celebrated formula, presented in [B], equates $\nabla \log\bigl(p_t(\,\cdot\,,y)\bigr)$ with certain stochastic integrals (see (20) below.) Various derivations of this formula and its extensions can be found in [AM], [EL] and [N]. In this note, we give a quick derivation of Bismut's and related formulae by lifting considerations to the bundle of orthonormal frames, using Bochner's identity, and applying a little elementary stochastic analysis. Some consequences of these identities are then explored. In particular, after deriving a standard logarithmic Sobolev inequality, we present (see (26)) a sharp pointwise estimate on the logarithmic derivative of the heat kernel in terms of known estimates on the heat kernel itself.

§1 Bismut's Formula and Variations

Let $\mathcal{O}(M)$ denote the bundle of orthonormal frames associated to M, equipped with the Lévi–Civita connection. (Throughout, we will take our basic reference for differential geometry to be the book [BC]. In particular, see Chapter 7 for an explanation of $\mathcal{O}(M)$.) The advantage gained by moving considerations to $\mathcal{O}(M)$ is that many differential geometric quantities resemble their classical analogs. For example, if $(\mathbf{e}_1,\dots,\mathbf{e}_d)$ denotes the standard orthonormal basis in $\mathbb{R}^d$ and $\mathfrak{E}_1,\dots,\mathfrak{E}_d$ are the corresponding basic vector fields on $\mathcal{O}(M)$ (i.e., $\mathfrak{E}_k$ is the horizontal vector field on $\mathcal{O}(M)$ for which $d\pi\mathfrak{E}_k(\mathfrak{f}) = \mathfrak{f}\mathbf{e}_k$ at each $\mathfrak{f} \in \mathcal{O}(M)$), then we can define the gradient $\mathfrak{f} \in \mathcal{O}(M) \longmapsto \nabla_{\mathfrak{f}}\varphi \in \mathbb{R}^d$ for $\varphi \in C^1(M)$ so that

$$\nabla\varphi(\mathfrak{f}) = \nabla_{\mathfrak{f}}\varphi = \sum_1^d \mathfrak{E}_k(\mathfrak{f})(\varphi\circ\pi)\mathbf{e}_k, \tag{1}$$

where $\pi : \mathcal{O}(M) \longrightarrow M$ denotes the fiber map. Similarly, if, for $F \in C^2\bigl(\mathcal{O}(M)\bigr)$,

$$\Delta F = \sum_{k=1}^d \mathfrak{E}_k^2 F, \tag{2}$$

then

$$\Delta\varphi \equiv \Delta(\varphi\circ\pi), \quad \varphi \in C^2(M), \tag{3}$$

is well-defined as a function on M and, in fact, gives the action of the standard Laplacian (Laplace–Beltrami operator) on φ.

Both authors gratefully acknowledge support from NSF grant 9302709–DMS during the period of this research.

Typeset by $\mathcal{A}_{\mathcal{M}}\mathcal{S}$-TEX

Next, let $\phi : T\big(\mathcal{O}(M)\big) \longrightarrow o(d)$ (the Lie algebra of $d \times d$ skew symmetric matrices) denote the connection 1-form determined by the Lévi–Civita connection (cf. §5.2 in [BC]). That is, for any $\mathfrak{f} \in \mathcal{O}(M)$ and $X_{\mathfrak{f}} \in T_{\mathfrak{f}}\big(\mathcal{O}(M)\big)$, $\phi(X_{\mathfrak{f}})$ is determined so that

$$\mathbf{H}X_{\mathfrak{f}} \equiv X_{\mathfrak{f}} - \lambda\big(\phi(X_{\mathfrak{f}})\big)$$

is the horizontal component of $X_{\mathfrak{f}}$, where, for $A \in o(d)$, $\lambda(A) \in T\big(\mathcal{O}(M)\big)$ is the vertical vector field such that

$$\lambda(A)F(\mathfrak{f}) = \frac{d}{ds}F\big(R_{e^{sA}}\mathfrak{f}\big)\Big|_{s=0}, \quad \mathfrak{f} \in \mathcal{O}(M),$$

and $R_{\mathcal{O}} : \mathcal{O}(M) \longrightarrow \mathcal{O}(M)$ is the natural right action given by $R_{\mathcal{O}}\mathfrak{f}\mathbf{v} = \mathfrak{f}\mathcal{O}\mathbf{v}$ for $\mathcal{O} \in \mathrm{O}(d)$, $\mathfrak{f} \in \mathcal{O}(M)$, and $\mathbf{v} \in \mathbb{R}^d$. Then, the curvature 2-form $\Phi : T\big(\mathcal{O}(M)\big) \times T\big(\mathcal{O}(M)\big) \longrightarrow o(d)$ is defined to be the horizontal part of the exterior derivative $d\phi$ of ϕ:

$$\Phi(X_{\mathfrak{f}}, Y_{\mathfrak{f}}) = d\phi\big(\mathbf{H}X_{\mathfrak{f}}, \mathbf{H}Y_{\mathfrak{f}}\big), \quad X_{\mathfrak{f}}, Y_{\mathfrak{f}} \in T_{\mathfrak{f}}\big(\mathcal{O}(M)\big). \tag{4}$$

As a consequence of the fact that the Lévi–Civita connection is torsion free and the second structural equation (cf. Theorem 4 in §6.2 of [BC]), one finds (cf. §5.3 of [BC]) that the commutator of $\mathfrak{E}_k$ with $\mathfrak{E}_\ell$ is vertical and is given by

$$\big[\mathfrak{E}_k, \mathfrak{E}_\ell\big](\mathfrak{f}) = -\lambda\big(\Phi_{k,\ell}(\mathfrak{f})\big), \quad \text{where } \Phi_{k,\ell} \equiv \Phi\big(\mathfrak{E}_k, \mathfrak{E}_\ell\big). \tag{5}$$

In particular, for $\varphi \in C^2(M)$,

$$\mathfrak{E}_k\mathfrak{E}_\ell(\varphi \circ \pi) = \mathfrak{E}_\ell\mathfrak{E}_k(\varphi \circ \pi),$$

and, for $\varphi \in C^3(M)$,

$$\begin{aligned}\mathfrak{E}_k^2\mathfrak{E}_\ell(\varphi \circ \pi) &= \mathfrak{E}_\ell\mathfrak{E}_k^2(\varphi \circ \pi) - \lambda\big(\Phi_{k,\ell}\big)\mathfrak{E}_k(\varphi \circ \pi)\\ &= \mathfrak{E}_\ell\mathfrak{E}_k^2(\varphi \circ \pi) - \sum_{j=1}^d \big(\Phi_{k,\ell}\mathbf{e}_k, \mathbf{e}_j\big)_{\mathbb{R}^d}\mathfrak{E}_j(\varphi \circ \pi).\end{aligned}$$

Hence, after summing with respect of k, we arrive at the Bochner identity

$$\Delta\nabla\varphi = \nabla\Delta\varphi + \mathrm{Ric}\nabla\varphi, \tag{6}$$

where $\mathrm{Ric} : \mathcal{O}(M) \longrightarrow \mathbb{R}^d \otimes \mathbb{R}^d$ is the Ricci curvature (symmetric) matrix

$$\mathrm{Ric}_{i,j} = -\sum_{k=1}^d \big(\Phi_{k,i}\mathbf{e}_k, \mathbf{e}_j\big)_{\mathbb{R}^d}. \tag{7}$$

Bochner's identity is the starting point for a great deal of analysis on M. To wit, let $\{P_t : t > 0\}$ denote the Markov semigroups on $C(M)$ determined by $\frac{1}{2}\Delta$. Then, as an application of (6), we find that

$$\frac{d}{dt}\nabla P_t\varphi = \frac{1}{2}\Delta\nabla P_t\varphi - \frac{1}{2}\mathrm{Ric}\nabla P_t\varphi, \tag{8}$$

where the action of Δ on an $\mathbb{R}^d$-valued function is component by component; and, from (8), one has

$$\frac{d}{dt}\big|\nabla P_t\varphi\big|^2 = \big(\nabla P_t\varphi, \Delta\nabla P_t\varphi\big)_{\mathbb{R}^d} - \big(\nabla P_t\varphi, \mathrm{Ric}\nabla P_t\varphi\big)_{\mathbb{R}^d}.$$

At the same time, an easy computation leads to

$$\Delta\big|\nabla P_t\varphi\big|^2 = 2\big(\nabla P_t\varphi, \Delta\nabla P_t\varphi\big)_{\mathbb{R}^d} + 2\big\|\mathrm{Hess}\big(P_t\varphi\big)\big\|_{\mathrm{H.S.}}^2,$$

where $\mathrm{Hess}f \equiv \big(\big(\mathfrak{E}_k\mathfrak{E}_\ell f\big)\big)$ is the Hessian matrix of $f \in C^2(M)$ and $\|\cdot\|_{\mathrm{H.S.}}$ is the standard Hilbert–Schmidt norm for $d \times d$ matrices. Hence, we find that

$$\begin{aligned}\frac{d}{dt}\big|\nabla P_t\varphi\big|^2 &= \frac{1}{2}\Delta\big|\nabla P_t\varphi\big|^2 - \big\|\mathrm{Hess}\big(P_t\varphi\big)\big\|_{\mathrm{H.S.}}^2 - \big(\nabla P_t\varphi, \mathrm{Ric}\nabla P_t\varphi\big)_{\mathbb{R}^d}\\ &\le \frac{1}{2}\Delta\big|P_t\varphi\big|^2 - \alpha\big|\nabla P_t\varphi\big|^2,\end{aligned}$$

where

$$\alpha \equiv \inf\big\{\big(\mathbf{e}, \mathrm{Ric}(\mathfrak{f})\mathbf{e}\big)_{\mathbb{R}^d} : \mathfrak{f} \in \mathcal{O}(M) \text{ and } |\mathbf{e}| = 1\big\}. \tag{9}$$

In particular, for $T \in (0,\infty)$,

$$\frac{d}{dt}P_{T-t}\Big(\big|\nabla P_t\varphi\big|^2\Big) \le -\alpha P_{T-t}\Big(\big|\nabla P_t\varphi\big|^2\Big), \quad t \in (0,T),$$

and so

$$\big|\nabla P_T\varphi\big|^2 \le e^{-\alpha T}P_T\big(|\nabla\varphi|^2\big), \quad T \in (0,\infty). \tag{10}$$

The estimate in (10) is very useful as it stands. For example, when $\alpha > 0$, it leads immediately to the well known fact that the spectral gap for Δ as an operator on $L^2(M)$ is at least α. However, as Bismut [B] noticed, (8) can be effectively combined with elementary probability theory to replace estimates like (10) with intriguing equalities. To see this, let $(\mathfrak{W}, \mathcal{B}_{\mathfrak{W}}, \mu)$ be the standard Wiener space of $\mathbb{R}^d$-valued paths and, for each $\mathfrak{f} \in \mathcal{O}(M)$, use $\mathfrak{F}_{\mathfrak{f}} : [0,\infty) \times \mathfrak{W} \longrightarrow \mathcal{O}(M)$ to denote the progressively measurable solution to the Stratonovich stochastic differential equation

$$d\mathfrak{F}_{\mathfrak{f}}(t,\mathbf{w}) = \sum_1^d \mathfrak{E}_k\big(\mathfrak{F}_{\mathfrak{f}}(t,\mathbf{w})\big) \circ d\mathbf{w}(t)_k \quad \text{with } \mathfrak{F}_{\mathfrak{f}}(0,\mathbf{w}) = \mathfrak{f}. \tag{11}$$

Next, define $A_{\mathfrak{f}} : [0,\infty) \times \mathfrak{W} \longrightarrow \mathbb{R}^d \otimes \mathbb{R}^d$ by the integral equation

$$A_{\mathfrak{f}}(t,\mathbf{w}) = \mathbf{I} - \frac{1}{2}\int_0^t A_{\mathfrak{f}}(\tau,\mathbf{w})\mathrm{Ric}\big(\mathfrak{F}_{\mathfrak{f}}(\tau,\mathbf{w})\big)\,d\tau, \quad t \in [0,\infty). \tag{12}$$

Then, from (8) and Itô's formula, one finds that, for each $T \in (0,\infty)$,

$$M(t,\mathbf{w}) = A_{\mathfrak{f}}(t\wedge T,\mathbf{w})\big[\nabla P_{T-t\wedge T}\varphi\big]\big(\mathfrak{F}_{\mathfrak{f}}(t\wedge T,\mathbf{w})\big) \tag{13}$$

is an $\mathbb{R}^d$-valued martingale. In particular, this means that

$$\big[\nabla P_T\varphi\big](\mathfrak{f}) = \mathbb{E}\Big[A_{\mathfrak{f}}(T)\nabla\varphi\big(\mathfrak{F}_{\mathfrak{f}}(T)\big)\Big]. \tag{14}$$

Since it is obvious that (cf. (9))

$$\big\|A_{\mathfrak{f}}(T,\mathbf{w})\big\|_{\mathrm{H.S.}}^2 \le e^{-\alpha T}, \quad (T,\mathbf{w}) \in [0,\infty) \times \mathfrak{W}, \tag{15}$$

(14) represents a considerable sharpening of (10). For example, from (14) and (15), we know that

$$|\nabla P_T\varphi| \leq e^{-\frac{\alpha T}{2}} P_T\big(|\nabla\varphi|\big). \tag{16}$$

(Notice that although $\nabla\psi$ is defined only on $\mathcal{O}(M)$, $|\nabla\psi|$ is well-defined on M itself.) To see why (16) represents an improvement on (10), we follow the reasoning of D. Bakry and M. Emery [BM] to derive from it the logarithmic Sobolev inequality

$$P_T\big(\varphi\log\varphi\big) - P_T\varphi\log P_T\varphi \leq \left(\frac{1}{2}\int_0^T e^{-\alpha t}\,dt\right) P_T\left(\frac{|\nabla\varphi|^2}{\varphi}\right), \quad \varphi \in C^1\big(M;(0,\infty)\big). \tag{17}$$

Indeed, note that,

$$\begin{aligned} 2\frac{d}{dt}P_t\big(P_{T-t}\varphi\log P_{T-t}\varphi\big) &= P_t\left(\frac{\big|\nabla P_{T-t}\varphi\big|^2}{P_{T-t}\varphi}\right) \\ &\leq e^{\alpha(t-T)}P_t\left(\frac{\big(P_{T-t}|\nabla\varphi|\big)^2}{P_{T-t}\varphi}\right) \leq e^{\alpha(t-T)}P_T\left(\frac{|\nabla\varphi|^2}{\varphi}\right), \end{aligned}$$

where we have used (16) to get the first inequality and the Markov property and

$$\begin{aligned} \big(P_{T-t}|\nabla\varphi|\big)^2 &= 4\Big(P_{T-t}\big(\varphi^{\frac{1}{2}}\nabla\varphi^{\frac{1}{2}}\big)\Big)^2 \\ &\leq 4\big(P_{T-t}\varphi\big)\Big(P_{T-t}\big|\nabla\varphi^{\frac{1}{2}}\big|^2\Big) = \big(P_{T-t}\varphi\big)\left(P_{T-t}\frac{|\nabla\varphi|^2}{\varphi}\right) \end{aligned}$$

to get the second.

In addition to (17), (14) leads immediately to a remarkable identity, which, because it was discovered[1] originally by Bismut, we call *Bismut's formula*. Namely, by applying Itô's formula to first (cf. (13)) $t \in [0,T] \longmapsto tM(t)$ and then $t \in [0,T] \longmapsto P_{T-t}\varphi\big(\mathfrak{F}_{\mathfrak{f}}(t)\big)$, we see that

$$\begin{aligned} T\mathbb{E}\Big[A_{\mathfrak{f}}(T)\nabla\varphi\big(\mathfrak{F}_{\mathfrak{f}}(T)\big)\Big] &= \mathbb{E}\left[\int_0^T A_{\mathfrak{f}}(t)\nabla P_{T-t}\varphi\big(\mathfrak{F}_{\mathfrak{f}}(t)\big)\,dt\right] \\ &= \mathbb{E}\left[\int_0^T A_{\mathfrak{f}}(t)\,d\mathbf{w}(t)\int_0^T \big[\nabla P_{T-t}\varphi\big]\big(\mathfrak{F}_{\mathfrak{f}}(t)\big)\,d\mathbf{w}(t)\right] \\ &= \mathbb{E}\left[\int_0^T A_{\mathfrak{f}}(t)\,d\mathbf{w}(t)\Big(\varphi\circ\pi\big(\mathfrak{F}_{\mathfrak{f}}(T)\big) - \big[P_T\varphi\big](\mathfrak{f})\Big)\right] \\ &= \mathbb{E}\left[\int_0^T A_{\mathfrak{f}}(t)\,d\mathbf{w}(t)\varphi\circ\pi\big(\mathfrak{F}_{\mathfrak{f}}(T)\big)\right], \end{aligned}$$

where all the stochastic integrals here are taken in the sense of Itô. (More generally, the notation $d\mathbf{w}(t)$, as opposed to "$\circ d\mathbf{w}(t)$", will be used to indicate Itô, as opposed to Stratonovich, stochastic integration.) Thus, in conjunction with (14), we arrive at Bismut's formula

$$\big[\nabla P_T\varphi\big](\mathfrak{f}) = T^{-1}\mathbb{E}\left[\int_0^T A_{\mathfrak{f}}(t)\,d\mathbf{w}(t)\varphi\circ\pi\big(\mathfrak{F}_{\mathfrak{f}}(T)\big)\right]. \tag{18}$$

[1] Actually, after deriving his formula by a quite different line of reasoning, Bismut [B] offers a second derivation which, even if it is not identical, is closely related to the one which we give here.

Before examining (18) further, we remark that essentially the same line of reasoning leads to a related formula. Namely, let $\{\mathfrak{P} : t > 0\}$ denote the Markov semigroup given by

$$\mathfrak{P}_t\Psi(\mathfrak{f}) = \mathbb{E}\Big[\Psi\big(\mathfrak{F}_{\mathfrak{f}}(t)\big)\Big], \quad \Psi \in C\big(\mathcal{O}(M)\big).$$

Clearly, $\big[P_t\varphi\big] \circ \pi = \mathfrak{P}_t(\varphi \circ \pi)$ for $\varphi \in C(M)$, but what is perhaps less obvious is that

$$\nabla P_T\varphi(\mathfrak{f}) = \big[\mathfrak{P}_T\nabla\varphi\big](\mathfrak{f}) - \frac{1}{2}\mathbb{E}\left[\int_0^T \mathrm{Ric}\big(\mathfrak{F}_{\mathfrak{f}}(t)\big)\, d\mathbf{w}(t)\, \varphi \circ \pi\big(\mathfrak{F}_{\mathfrak{f}}(T)\big)\right]. \tag{19}$$

To derive (19), first note that, from (6),

$$2\frac{d}{dt}\Big[\mathfrak{P}_t\big(\nabla P_{T-t}\varphi\big)\Big] = \mathfrak{P}_t\Big(\mathrm{Ric}\nabla P_{T-t}\varphi\Big), \quad t \in (0,T),$$

and therefore

$$\begin{aligned}
2\big[\mathfrak{P}_T\big(\nabla\varphi\big)\big](\mathfrak{f}) - 2\big[\nabla P_T\varphi\big](\mathfrak{f}) &= \mathbb{E}\left[\int_0^T \mathrm{Ric}\big(\mathfrak{F}_{\mathfrak{f}}(t)\big)\big[\nabla P_{T-t}\varphi\big(\mathfrak{F}_{\mathfrak{f}}(t)\big)\big]\right] \\
&= \mathbb{E}\left[\int_0^T \mathrm{Ric}\big(\mathfrak{F}_{\mathfrak{f}}(t)\big)\, d\mathbf{w}(t) \int_0^T \big[\nabla P_{T-t}\varphi\big]\big(\mathfrak{F}_{\mathfrak{f}}(t)\big)\, d\mathbf{w}(t)\right] \\
&= \mathbb{E}\left[\int_0^T \mathrm{Ric}\big(\mathfrak{F}_{\mathfrak{f}}(t)\big)\, d\mathbf{w}(t)\, \varphi \circ \pi\big(\mathfrak{F}_{\mathfrak{f}}(T)\big)\right].
\end{aligned}$$

§2 Estimates and Applications

We conclude this note with an examination of the potential applications of Bismut's formula (18). To begin with, we follow Bismut by converting (18) into the statement that

$$\big[\nabla \log\big(p_T(\,\cdot\,,y)\big)\big](\mathfrak{f}) = \frac{\big[\nabla p_T(\,\cdot\,,y)\big](\mathfrak{f})}{p_T\big(\pi(\mathfrak{f}),y\big)} = T^{-1}\mathbb{E}\left[\int_0^T A_{\mathfrak{f}}(t)\, d\mathbf{w}(t)\,\middle|\, \pi\big(\mathfrak{F}_{\mathfrak{f}}(T)\big) = y\right]. \tag{20}$$

where $p_t(x,y)$ denotes the kernel (with respect to the Riemannian volume measure) for the operator P_t. (The existence of this kernel as a positive, smooth function is well-known, see e.g. [D].) Indeed, as soon as one shows that the conditional expectation value on the right makes sense and admits a version which is continuous in $y \in M$, there is no question that (20) is simply a dramatic re-interpretation of (18). For this purpose, first observe that there is no problem about the interpretation of $\int_0^T A_{\mathfrak{f}}(t,\mathbf{w})\, d\mathbf{w}(t)$ under the conditional measure. Namely, although this integral arose as an Itô integral which is defined only up to a set of μ-measure 0, it makes perfectly good sense as a classical Riemann–Stieltjes integral for each $\mathbf{w} \in \mathfrak{W}$. In particular, this means that there is no question about the meaning of the right hand side and no doubt that (20) holds for $P\big(T, \pi(\mathfrak{f}), \,\cdot\,\big)$-almost every $y \in M$, where $P(T,x,\,\cdot\,)$ is the transition probability function whose density is $p_T(x,\,\cdot\,)$.

Now, let $T \in (0,\infty)$ be given, set $T_n = (1-2^{-n})T$, and define

$$G_{n,T}(\mathfrak{f},y) = \mathbb{E}\left[\int_0^{T_n} A_{\mathfrak{f}}(t)\, d\mathbf{w}(t)\, F_{n,T}\Big(\mathfrak{f}, \pi\big(\mathfrak{F}_{\mathfrak{f}}(T_n)\big), y\Big)\right] \quad \text{where } F_{n,T}(\mathfrak{f},\xi,y) \equiv \frac{p_{2^{-n}T}\big(\xi,y\big)}{p_T\big(\pi(\mathfrak{f}),y\big)}.$$

We know from (18) that

$$T\,\frac{\big[\nabla P_T\varphi\big](\mathfrak{f})}{p_T\big(\pi(\mathfrak{f}),y\big)} = \lim_{n\to\infty}\int_M G_{n,T}(\mathfrak{f},y)\varphi(y)\, dy.$$

Thus, our problem comes down to showing that there exists a $G_T(\mathfrak{f},\,\cdot\,) \in C\big(M;\mathbb{R}^d\big)$ to which $\big\{G_{n,T}(\mathfrak{f},\,\cdot\,)\big\}_1^\infty$ converges uniformly. But, since

$$G_{n,T}(\mathfrak{f},y) - G_{n-1,T}(\mathfrak{f},y) = \mathbb{E}\left[\int_{T_{n-1}}^{T_n} A_{\mathfrak{f}}(t)\,d\mathbf{w}(t) F_{n,T}\Big(\mathfrak{f},\pi\big(\mathfrak{F}_{\mathfrak{f}}(T_n)\big),y\Big)\right], \quad n \geq 1,$$

our problem comes down to estimating the quantities

$$B_{n,T}(\mathfrak{f},y) \equiv \left|\mathbb{E}\left[\int_{T_{n-1}}^{T_n} A_{\mathfrak{f}}(t)\,d\mathbf{w}(t) F_{n,T}\Big(\mathfrak{f},\pi\big(\mathfrak{F}_{\mathfrak{f}}(T_n)\big),y\Big)\right]\right|.$$

However, by standard (cf. Chapter 5 of [D]) estimates,

$$M(T) \equiv \sup_{t\in(0,T]}\ \sup_{\xi,\eta\in M} (2\pi t)^{\frac{d}{2}} p_t(\xi,\eta) < \infty \quad \text{and} \quad \epsilon(T) \equiv \inf_{\xi,\eta\in M} p_T(\xi,\eta) > 0,$$

while, by Hölder's and Burkholder's inequalities,

$$\begin{aligned} B_{n,T}(\mathfrak{f},y) &\leq \mathbb{E}\left[\left|\int_{T_{n-1}}^{T_n} A_{\mathfrak{f}}(t)\,d\mathbf{w}(t)\right|^{d+1}\right]^{\frac{1}{d+1}} \mathbb{E}\Big[F_{n,T}\Big(\mathfrak{f},\pi\big(\mathfrak{F}_{\mathfrak{f}}(T_n)\big),y\Big)^{1+\frac{1}{d}}\Big]^{\frac{d}{d+1}} \\ &\leq de^{\frac{|\alpha|T}{2}}\big(T2^{-n}\big)^{\frac{1}{2}}\big\|F_{n,T}(\mathfrak{f},\,\cdot\,,y)\big\|_{\mathrm{u}}^{\frac{1}{d+1}} \leq de^{\frac{|\alpha|T}{2}}\left(\frac{M(T)}{\sqrt{T}\,\epsilon(T)}\right)^{\frac{1}{d+1}} 2^{-\frac{n}{2(d+1)}}, \end{aligned}$$

which is more than enough to justify (20).

Obviously, the preceding argument is extremely crude and leads to far from optimal estimates. In order to remedy this situation, we return again to (18) and consider general $\varphi \in C\big(M;(0,\infty)\big)$. Next, recall (see, for example, Lemma 3.2.13 in [DS]) the application Jensen's inequality which says that, for any probability measure μ and non-negative $f \in L^1(\mu)$ with integral 1,

$$\int \psi f\,d\mu \leq \int f\log f\,d\mu + \log\left[\int e^{\psi}\,d\mu\right], \quad \text{for all measurable } \psi \text{ with } \psi f \in L^1(\mu),$$

take

$$\psi = \lambda\int_0^T \big(\mathbf{e}, A_{\mathfrak{f}}(t)\,d\mathbf{w}(t)\big)_{\mathbb{R}^d} \quad \text{and} \quad f = \frac{\varphi\circ\pi\big(\mathfrak{F}_{\mathfrak{f}}(T)\big)}{P_T\varphi\big(\pi(\mathfrak{f})\big)},$$

and conclude from (18) that, for every $\lambda > 0$ and $\mathbf{e} \in S^{d-1}$,

$$\lambda T\frac{\big(\mathbf{e},\nabla P_T\varphi(\mathfrak{f})\big)_{\mathbb{R}^d}}{P_T\varphi\big(\pi(\mathfrak{f})\big)} \leq h_T\big(\pi(\mathfrak{f}),\varphi\big) + \log\mathbb{E}\left[\exp\left(\lambda\int_0^T\big(\mathbf{e},A_{\mathfrak{f}}(t)\,d\mathbf{w}(t)\big)_{\mathbb{R}^d}\right)\right],$$

where

$$h_T(x,\varphi) \equiv \int_M \frac{\varphi(\xi)}{P_T\varphi(x)}\log\frac{\varphi(\xi)}{P_T\varphi(x)}\,p_T(x,\xi)\,d\xi = \frac{P_T\big(\varphi\log\varphi\big)(x) - P_T\varphi(x)\log P_T\varphi(x)}{P_T\varphi(x)}. \tag{21}$$

Finally, observe that

$$\mathbb{E}\left[\exp\left(\lambda\int_0^T\big(\mathbf{e},A_{\mathfrak{f}}(t)\,d\mathbf{w}(t)\big)_{\mathbb{R}^d} - \frac{\lambda^2}{2}\int_0^T\big|A_{\mathfrak{f}}(t)^\top\mathbf{e}\big|^2\,dt\right)\right] = 1,$$

and therefore, by (15),

$$\log \mathbb{E}\left[\exp\left(\lambda \int_0^T \left(\mathbf{e}, A_{\mathfrak{f}}(t)\,d\mathbf{w}(t)\right)_{\mathbb{R}^d}\right)\right] \le \frac{\lambda^2}{2}\int_0^T e^{-\alpha t}\,dt.$$

Hence, after minimizing with respect to $\lambda > 0$, we arrive at

$$\frac{\left|\nabla P_T\varphi\right|(x)}{P_T\varphi(x)} \le T^{-1}\sqrt{2h_T(x,\varphi)E_\alpha(T)}, \quad \text{where } E_\alpha(t) \equiv \int_0^t e^{-\alpha\tau}\,d\tau. \tag{22}$$

The estimate in (22) has several potentially interesting features. For one thing, it is a complement to the logarithmic Sobolev inequality in (17). Indeed, (17), the second part of (21), and (22) yield:

$$\frac{\left|\nabla P_T\varphi\right|^2(x)}{P_T\varphi(x)} \le \frac{2E_\alpha(T)}{T^2}\Big(\big(P_T\varphi\log\varphi\big)(x) - P_T\varphi(x)\log P_T\varphi(x)\Big) \le \frac{E_\alpha(T)^2}{T^2}P_T\left(\frac{|\nabla\varphi|^2}{\varphi}\right)(x). \tag{23}$$

In addition, (22) enables one to pass from estimates on $p_T(x,y)$ to estimates for $|\nabla p_T(\,\cdot\,,y)|(x)$. Namely, by taking $\varphi = p_{\epsilon T}(\,\cdot\,,y)$ in (22), we obtain

$$\left|\nabla \log p_{(1+\epsilon)T}(\,\cdot\,,y)\right|(x) = \frac{\left|\nabla p_{(1+\epsilon)T}(\,\cdot\,,y)\right|(x)}{p_{(1+\epsilon)T}(x,y)} \le T^{-1}\sqrt{2E_\alpha(T)H_{\epsilon,T}(x,y)}, \tag{24}$$

where

$$H_{\epsilon,T}(x,y) \equiv \int_M \frac{p_{\epsilon T}(\xi,y)}{p_{(1+\epsilon)T}(x,y)}\log\frac{p_{\epsilon T}(\xi,y)}{p_{(1+\epsilon)T}(x,y)}\,p_T(x,\xi)\,d\xi \le \log\frac{\left\|p_{\epsilon T}(\,\cdot\,,y)\right\|_{\mathrm{u}}}{p_{(1+\epsilon)T}(x,y)}. \tag{25}$$

To test that (24) is reasonably sharp, we consider the case when (cf. (9)) $\alpha \ge 0$ and use some of the beautiful estimates given by Cheeger, Li, and Yau in [CY] and [LY]. Namely, in that case, (cf. Proposition 5.5.1 and Theorem 5.5.11 in [D]), on the one hand, there exists a universal $a_d \in (0,\infty)$ such that

$$(2\pi t)^{\frac{d}{2}}\|p_t(\,\cdot\,,y)\|_{\mathrm{u}} \le a_d\frac{(2\pi t)^{\frac{d}{2}}}{\mathrm{Vol}\big(y,\sqrt{t}\big)} \le a_d\frac{(2\pi T)^{\frac{d}{2}}}{\mathrm{Vol}\big(y,\sqrt{T}\big)}, \quad 0 < t \le T,$$

where $\mathrm{Vol}(y,r)$ is the Riemannian volume of the Riemannian ball of radius r around y. On the other hand, $\alpha \ge 0$ implies (cf. Theorem 5.6.1 in [D]) that

$$(2\pi T)^{\frac{d}{2}}p_T\big(\pi(\mathfrak{f}),y\big) \ge \exp\left[-\frac{\mathrm{dist}^2\big(\pi(\mathfrak{f}),y\big)}{2T}\right],$$

where distance is measured in the Riemannian metric. Hence, after putting these together with (24), we find that, when $\alpha \ge 0$:

$$\left|\nabla\log p_{(1+\epsilon)T}(\,\cdot\,,y)\right|(x) \le \frac{\mathrm{dist}(x,y)}{(1+\epsilon)^{\frac{1}{2}}T} + T^{-\frac{1}{2}}\sqrt{2\log a_d + d\log\frac{1+\epsilon}{\epsilon} + 2\log\frac{\big(2\pi\epsilon T\big)^{\frac{d}{2}}}{\mathrm{Vol}\big(y,\sqrt{\epsilon T}\big)}}, \tag{26}$$

which is surprisingly close to what one knows to be true in the classical, Euclidean setting.

Discouraging Observation: Experience in such matters makes one suspect that estimates like (22) (equivalently, the first inequality in (23)) are easier to derive by a direct analytic argument than they are by way of a probabilistic formula like Bismut's. Unfortunately for stochastic analysis, the one here is no exception. Indeed, recall (cf. the derivation of (17)) that

$$2\Big(P_T(\varphi\log\varphi) - P_T\varphi\log P_T\varphi\Big) = \int_0^T P_{T-t}\left(\frac{|\nabla\varphi_t|^2}{\varphi_t}\right)dt,$$

where $\varphi_t \equiv P_t\varphi$. Next, integrate by parts to get

$$2\Big(P_T(\varphi\log\varphi) - P_T\varphi\log P_T\varphi\Big) = E_{-\alpha}(T)\frac{\left|\nabla P_T\varphi\right|^2}{P_T\varphi} + \int_0^T E_{-\alpha}(t)P_{T-t}\big(F_\alpha(t)\big)\,dt,$$

where

$$F_\alpha(t) \equiv \frac{1}{2}\Delta\left(\frac{|\nabla\varphi_t|^2}{\varphi_t}\right) - \frac{d}{dt}\left(\frac{|\nabla\varphi_t|^2}{\varphi_t}\right) - \alpha\frac{|\nabla\varphi_t|^2}{\varphi_t}.$$

Since, by elementary computation and the equality preceding (9),

$$\begin{aligned} F_\alpha(t) = {}& \frac{\|\mathrm{Hess}(\varphi_t)\|^2_{\mathrm{H.S.}}}{\varphi_t} + \frac{\big(\nabla\varphi_t, \mathrm{Ric}\,\nabla\varphi_t\big)_{\mathbb{R}^d}}{\varphi_t} - \frac{2\big(\nabla\varphi_t, \mathrm{Hess}(\varphi_t)\,\nabla\varphi_t\big)_{\mathbb{R}^d}}{\varphi_t^2} + \frac{|\nabla\varphi_t|^4}{\varphi_t^3} - \alpha\frac{|\nabla\varphi_t|^2}{\varphi_t} \\ &\geq \frac{1}{\varphi_t}\left|\frac{\mathrm{Hess}(\varphi_t)\nabla\varphi_t}{|\nabla\varphi_t|} - \frac{|\nabla\varphi_t|\nabla\varphi_t}{\varphi_t}\right|^2 \geq 0, \end{aligned}$$

we now get that

$$E_{-\alpha}(T)\frac{\left|\nabla P_T\varphi\right|^2}{P_T\varphi} \leq 2\Big(P_T(\varphi\log\varphi) - P_T\varphi\log P_T\varphi\Big). \tag{27}$$

Finally, after an application of Jensen's inequality, one sees that (27) is actually a little sharper than the first inequality in (23).

Of course, with twenty-twenty hindsight, one sees how to amend Bismut's formula so that (27) comes out of a probabilistic argument. Namely, exactly the same sort of calculation which led to (18) shows that

$$\big[\nabla P_T\varphi\big](\mathfrak{f}) = E_{-\frac{\alpha}{2}}(T)^{-1}\mathbb{E}\left[\int_0^T e^{\frac{\alpha t}{2}}A_{\mathfrak{f}}(t)\,d\mathbf{w}(t)\varphi\circ\pi\big(\mathfrak{F}_{\mathfrak{f}}(T)\big)\right],$$

which leads, in turn, to

$$\frac{\left|\nabla P_T\varphi\right|(x)}{P_T\varphi(x)} \leq E_{-\frac{\alpha}{2}}(T)^{-1}\sqrt{2Th_T(x,\varphi)}. \tag{28}$$

One may derive some gratification from the observation that, again by Jensen's inequality, (28) is a little better than (27). However, the improvement is hardly cause for celebration. *Sic semper fors.*

References

[AM] Airault, H. & Malliavin, P., *Semi-martingales with values in Euclidean vector bundles and Ocone's formula on a Riemannian manifold* (to appearProceedings of the 1993 AMS conference at Cornell).

[B] Bismut, J.M., *Large Deviations and the Malliavin Calculus*, Progress in Math. #45, Birkhäuser, Cambridge USA, 1984.

[BC] Bishop, R. & Crittenden, R., *Geometry of Manifolds*, Pure and Appl. Math. #15, Academic Press, San Diego, 1964.

[BM] Bakry, D. and Emery, M., *Hypercontractivité de semi-groupes des diffusions*, C.R. Acad. Sci. Paris t. 299 Serie 1 (1984), 775–777.

[CY] Cheeger, J. & Yau, S.T., *A lower bound for the heat kernel*, Comm. Pure Appl. Math. **34** (1981), 465–480.

[D] Davies, E.B., *Heat Kernels and Spectral Theory*, Cambridge Tracts in Math. #92, Cambridge Univ. Press, Cambridge U.K. & NY, USA, 1989.

[DS] Deuschel, J.D. & Stroock, D., *Large Deviations*, Pure and Appl. Math. #137, Academic Press, San Diego USA, 1989.

[EL] Elworthy, K.D. & Li, X.-M., *Formulae for the derivative of heat semigroups*, JFA **125** (1994), 252–286.

[LY] Li, P. & Yau, S.T., *On the parabolic kernel of the Schrödinger operator*, Acta Math. **156** (1986), 153–201.

[N] Norris, J., *Path integral formulae for heat kernels and derivatives*, Prob. Th. Rel. Fields **94** (1993), 525–541.

MIT, rm. 2-272, Cambridge, MA 02139, USA and Dept. Elec. Eng., Technion, Haifa, Israel
E-mail address: dws@math.mit.edu and zeitouni@ee.technion.ac.il