

LOGARITHMIC SOBOLEV INEQUALITIES ON LIE GROUPS

BY

LEONARD GROSS¹

1. Introduction

The heat kernel on a manifold provides a natural analog for n dimensional Gauss measure. If Δ is the Laplace-Beltrami operator on a Riemannian manifold then the heat kernel $p_t(x, \cdot)$ is the measure on M given by $(e^{t\Delta}f)(x) = \int_M f(y)p_t(x, dy)$. $p_t(x, \cdot)$ reduces to Gauss measure centered at x in case $M = R^n$. Put $t = 1$, fix x and write $\mu(dy) = p_1(x, dy)$. The Dirichlet form operator L for μ is the self-adjoint operator on $L^2(\mu)$ defined by

$$(1.1) \quad (Lf, g)_{L^2(\mu)} = \int_M \text{grad } f(y) \cdot \text{grad } \bar{g}(y) \mu(dy)$$

for f and g in $C_c^\infty(M)$. In case $M = R^n$, L is the harmonic oscillator Hamiltonian in its ground state representation. Ever since E. Nelson [32] showed the usefulness to quantum field theory of operator bounds on $\exp(-sL)$ as an operator from $L^p(\mu)$ to $L^q(\mu)$, the boundedness and in particular the contractivity of this operator and similar operators has been explored with great intensity. A variety of techniques for exploring $\exp(-sL)$ as an operator from L^p to L^q for the harmonic oscillator Hamiltonian and for other second order elliptic operators have been investigated. Among them is the use of an equivalence between boundedness properties of $e^{-tL}: L^p \rightarrow L^q$ (hypercontractivity) and direct inequalities on the quadratic form of L itself. The latter have the form of logarithmic Sobolev inequalities [17] (see e.g. (3.7) below). A survey of these topics is given in [9] and a more recent bibliography is given in [18]. In the present paper it will be shown that a technique used in [17] for proving logarithmic Sobolev inequalities for Gauss measure on R^n goes over to the heat kernel measure on Lie groups.

Denote by W the space $C_0([0, T]; G)$ of continuous functions $g(\cdot)$ on $[0, T]$ with values in a connected Lie group G for which $g(0) = \text{identity}$. In Section

Received November 2, 1990.

1991 Mathematics Subject Classification. Primary 46E35, 58G32; Secondary 58D20, 60B15.

¹This work was partially supported by a grant from the National Science Foundation.

2 we discuss a notion of differentiability for functions on W which reduces, in case $G = R$, to a much studied notion of differentiability of functions on Wiener space. In fact W is a Banach space if $G = R$, and if H is the Cameron-Martin Hilbert space for Wiener measure on W then for a function $f: W \rightarrow R$ the derivative of f in H directions plays a central role in Gaussian integration theory, unlike the Frechet derivative which usually plays only a technical role in this context. R.H. Cameron [8] was the first to recognize that the derivative in H directions relates well to the Wiener integral and in fact allows an integration by parts formula. (Actually he considered only derivatives in W^* directions with W^* properly contained in H by the usual injection $W^* \hookrightarrow H^* = H \rightarrow W$. But the extension of his formula to H directions is immediate.) I.E. Segal [40, 41] emphasized and clarified the central role of H by dispensing with an ambient Banach space or other ambient vector space altogether. Directional derivatives in H directions are ubiquitous in Boson quantum field theory because they represent annihilation operators. For functions on a Banach space B with a given nondegenerate Gaussian measure μ and corresponding covariance Hilbert space H the H -derivative (i.e., gradient in H directions) is an intrinsic notion of differentiation relative to μ . The H -derivative was systematically exploited in [16] for regularity theorems in infinite dimensional potential theory and in many other works [2], [10], [12], [13], [14], [15], [22], [23], [24], [25], [26], [36], [37], [38] in the 1960's and early 1970's. More recently the closure of the H -derivative as a densely defined operator from $L^p(\mu)$ to $L^p(\mu) \otimes H$ has been intensely explored. The literature on the closed intrinsic derivative (i.e., closed H -derivative) is too vast to survey here. For extensive lists of some of the early work on these operators and their applications (e.g. to regularity of heat kernels for hypoelliptic operators) see the bibliographies of [1], [5], [7] as well as [20], [21], [27], [28], [30], [31], [45]. In Section 2 we describe the analogous intrinsic derivative (i.e., H -derivative) for functions on $W := C_0([0, 1]; G)$ when G is an arbitrary Lie group and Wiener measure is replaced by the G valued Brownian motion path space measure P on W . We do this by first defining an operation of "addition" of a finite energy Lie algebra valued function h with an element g of W . The operation $h, g \rightarrow h + g$ is defined on all of $H \times W$ by direct analytic means without reference to stochastic integrals or any probability measures. But in fact it is a thinly disguised version of ordinary translation $B \rightarrow h + B$ in the Lie algebra valued Wiener space when the Wiener path B is identified with g via an Ito-Stratonovich differential equation. In this paper we take an analytic viewpoint, avoiding direct use of stochastic differential equations (however, cf. [19]). Moreover since W is a group under pointwise multiplication there is another more natural notion of differentiation of functions on W and we shall explain the relation between these two derivatives (cf. Proposition 2.7).

In Section 3 we represent the G valued Brownian motion as a limit of random walks in G . We use the representation to derive a logarithmic

Sobolev inequality on W for P (cf. equation 3.6)) with best constant (§5). A logarithmic Sobolev inequality on the Lie group itself for the heat kernel measure follows from this (Section 4). But in Section 5 we show by example why we cannot expect our method to produce always the best Sobolev constant for the heat kernel measure on G . The example is motivated by the preceding random walk representation with the random walk taken as a simple random walk in the Lie algebra of the circle group. Specifically, if μ is the symmetric probability measure on R supported on the two point set $\{-1, 1\}$ we look at the discrete Dirichlet form operator N associated to the measure $\mu^3 \equiv \mu * \mu * \mu$ and describe the effect on the Sobolev constant and spectrum of N that occurs when one wraps R around S^1 , so as to identify the points -3 and 3 in the support of μ^3 .

It is a pleasure to acknowledge very useful discussions with G. Ben Arous, J. Deuschel, Z. Guo, A. Knapp, O. Rothaus, and D. Stroock.

2. Differential calculus on $C_0([0, T]; G)$

Let G be a connected Lie group and denote by \mathcal{L} its Lie algebra. We denote by W the set of continuous functions g from $[0, T]$ into G such that $g(0) = e$, the identity element of G . Throughout this paper we consider a fixed inner product (\cdot, \cdot) on \mathcal{L} with associated norm $|\cdot|$ which need not be invariant under the adjoint action of G except when specified. We denote by H the real Hilbert space consisting of absolutely continuous functions from $[0, 1]$ into \mathcal{L} such that

$$\|h\|^2 = \int_0^T |\dot{h}(t)|^2 dt < \infty \quad \text{and} \quad h(0) = 0.$$

In this section we shall describe an action $g \rightarrow h + g$ of H on W which reduces, in case G is the additive group R of real numbers, to translation by h : $(h + g)(t) = h(t) + g(t)$. This action, which underlies all of advanced calculus on Wiener space, will play a similar role for us for a general Lie group. If $c(\cdot)$ is a smooth curve in G then $c(t)^{-1}\dot{c}(t)$ is in $T_e(G)$, the tangent space to G at e . We shall generally omit the $*$, writing $c(t)^{-1}\dot{c}(t)$ instead. Moreover we identify the Lie algebra with $T_e(G)$ and if α is in $T_e(G)$ then we put

$$(\hat{\alpha}f)(x) = df(xe^{s\alpha})/ds|_{s=0}$$

for the corresponding left invariant vector field on G .

LEMMA 2.1. *Let h be in H . Then there is a unique function v in W such that the equations*

- (a) $v(0) = e$
- (b) $v(t)^{-1}\dot{v}(t) = \dot{h}(t) \quad 0 \leq t \leq T,$

hold in the following sense. For any smooth function $f: G \rightarrow R$ and $0 \leq a < b \leq T$

$$(2.1) \quad f(v(b)) - f(v(a)) = \int_a^b (h(t) \hat{f})(v(t)) dt.$$

Moreover if h_n converges in H to h then the corresponding solutions v_n converge uniformly on $[0, T]$ to v .

Let B be a ball centered at the origin of \mathcal{G} such that the exponential map is a diffeomorphism from $2B$ onto a neighborhood U of e in G . There is a number $\varepsilon > 0$ such that if $\|h\| < \varepsilon$ then $v(t)$ is in U for $0 \leq t \leq T$. For such an h put $\chi(h)(t) = \text{Log } v(t)$. Then for some $c > 0$, $h \rightarrow \chi(h)$ is a C^∞ diffeomorphism from $S \equiv \{h \in H: \|h\| < c\}$ into H and the Frechet derivative χ' of this map satisfies

$$(2.2) \quad \chi'(0) = I_H.$$

Proof. If α is in $T_e(G)$ then $\hat{\alpha}$ is represented on $2B$ in exponential coordinates by a vector field $y \rightarrow \beta(y)\alpha$ where the linear map $\beta(y): \mathcal{G} \rightarrow \mathcal{G}$ is defined by

$$\beta(y)\alpha = d \text{Log}((\exp y)(\exp t\alpha)) / dt|_{t=0}.$$

$\beta(y)$ is C^∞ on $2B$, has bounded derivatives on B and clearly $\beta(0) = I_{\mathcal{G}}$. A standard proof of the existence theorem proceeds by showing that once one has established the existence of a solution v up to time t_0 one may continue the solution into the neighborhood $v(t_0)\exp B$ by solving the integral equation

$$(2.3) \quad w(t) = \int_{t_0}^t \beta(w(s)) \dot{h}(s) ds, \quad |t - t_0| < \delta$$

where $v(t) = v(t_0)\exp w(t)$ and δ depends on h and β but not on t_0 . The continuity of $h \rightarrow v$ from H (in fact also with respect to $\|\dot{h}\|_{L^1}$) to W (with uniform convergence) is also standard. Now if $\|\beta(y)\| \leq M$ for all $y \in B$ then $|w(t)| \leq M \int_0^t |\dot{h}(s)| ds$. So if B has radius a and $MT^{1/2}\|h\| \leq a$ then the solution $w(t)$ to (2.3), with $t_0 = 0$, remains in B up to time T , and we may therefore use just one coordinate patch up to time T for such small h . Moreover standard O.D.E. methods show that $w(t)$ is a C^∞ function of \dot{h} in L^1 norm, hence in H norm uniformly in t . Thus for small h , $h \rightarrow \beta(w(t))$ is C^∞ in h into $C([0, T]; \text{Hom}(\mathcal{G}, \mathcal{G}))$. But $\chi(h)(t) = w(t)$ and, denoting time derivative with a dot, (2.3) gives $\dot{\chi}(h)(t) = \beta(w(t))\dot{h}(t)$ which is a "product" of a smooth, uniformly bounded multiplication operator on H with \dot{h} itself. $h \rightarrow \chi(h)$ is therefore C^∞ into H for small h . We may compute the Frechet

derivative by the formula

$$\begin{aligned} \partial\chi(h + rk)/\partial r|_{r=0}(t) &= \beta'(w(t)) \\ &\times \langle \partial\chi(h + rk)/\partial r \rangle_{r=0} \dot{h}(t) + \beta(w(t)) \dot{k}(t). \end{aligned}$$

But $w \equiv 0$ if $h = 0$. Thus at $h = 0$ we have $\chi'(0)\langle k \rangle(t) = \beta(0)\dot{k}(t) = \dot{k}(t)$. Hence $\chi'(0) = I_H$. By the inverse function theorem χ is a diffeomorphism on a neighborhood of zero in H .

DEFINITION 2.2. Let g be in W and h in H . Let v be the solution to

$$(2.4) \quad v(t)^{-1}\dot{v}(t) = (\text{Ad } g(t))\dot{h}(t) \quad v(0) = e$$

as in Lemma 2.1. Put $u(t) = v(t)g(t)$, $0 \leq t \leq T$ and write $h + g = u$. Then $h + g$ is in W .

Remark 2.3. For each t , $\text{Ad } g(t)$ is a bounded operator on \mathcal{S} . Since the image of $g(\cdot)$ is compact for g in W $|\text{Ad } g(t)|_{\mathcal{S} \rightarrow \mathcal{S}}$ is uniformly bounded on $[0, T]$. Hence the function $t \rightarrow (\text{Ad } g(t))\dot{h}(t)$ is in L^2 when h is in H . Thus Lemma 2.1 is applicable to (2.4) and $h + g$ is therefore well defined.

THEOREM 2.4. The map $h, g \rightarrow h + g$ from $H \times W$ to W has the following properties:

- (a) $0 + g = g$.
- (b) $(h_1 + h_2) + g = h_1 + (h_2 + g)$ for $h_1, h_2 \in H, g \in W$.
- (c) If $g \in W$ and $h_1 + g = h_2 + g$ then $h_1 = h_2$.
- (d) $h + g$ is continuous as a function from $H \times W$ to W .
- (e) If $g \in W$ then $\{h + g: h \in C^\infty([0, T]; \mathcal{S})\}$ is dense in W .
- (f) If g is piecewise C^1 , h is in H , and $u = h + g$ then u is absolutely continuous and is the unique solution to the initial value problem

$$(2.5) \quad u(t)^{-1}\dot{u}(t) = g(t)^{-1}\dot{g}(t) + \dot{h}(t) \text{ a.e. } u(0) = e.$$

- (g) If $\dot{h} = 0$ on (a, b) then $(h + g)(a)^{-1}(h + g)(b) = g(a)^{-1}g(b)$ for all g in W .

Proof. (a) If $h = 0$ then $(\text{Ad } g(t))\dot{h}(t) = 0$. The unique solution to (2.4) is $v(t) = e$. Hence $(0 + g)(t) = e \cdot g(t) = g(t)$.

(b) Let $u_1 = h_1 + g$. If $v_1(t)^{-1}\dot{v}_1(t) = (\text{Ad } g(t))\dot{h}_1(t)$ and $v_1(0) = e$ then $u_1(t) = v_1(t)g(t)$. Let v_2 be the solution to

$$v_2(t)^{-1}\dot{v}_2(t) = (\text{Ad } u_1(t))\dot{h}_2(t) \quad \text{with } v_2(0) = e.$$

Then $(h_2 + (h_1 + g))(t) = v_2(t)u_1(t)$. Put $v(t) = v_2(t)v_1(t)$. Then for almost all t ,

$$v(t)^{-1}\dot{v}(t) = v_1(t)^{-1}v_2(t)^{-1}(v_2(t) * \dot{v}_1(t) + \dot{v}_2(t)v_1(t) *)$$

where $v_1(t)_*$ on the right in the last term means the differential of right multiplication acting on the tangent vector $\dot{v}_2(t)$. Thus for a.e. t ,

$$\begin{aligned} v(t)^{-1}\dot{v}(t) &= v_1(t)^{-1}\dot{v}_1(t) + v_1(t)^{-1}v_2(t)^{-1}\dot{v}_2(t)v_1(t)_* \\ &= (\text{Ad } g(t))\dot{h}_1(t) + v_1(t)^{-1}\{\text{Ad}(v_1(t)g(t))\dot{h}_2(t)\}v_1(t)_* \\ &= (\text{Ad } g(t))\dot{h}_1(t) + (\text{Ad } g(t))\dot{h}_2(t) \end{aligned}$$

since $(\text{Ad } a) = (L_a)_*(R_{a^{-1}})_*$ on \mathcal{S} where L_a and $R_{a^{-1}}$ are left and right multiplication. Hence $v(t)^{-1}\dot{v}(t) = (\text{Ad } g(t))(\dot{h}_1(t) + \dot{h}_2(t))$ while $v(0) = e$. Thus

$$\begin{aligned} ((h_1 + h_2) + g)(t) &= v(t)g(t) = v_2(t)v_1(t)g(t) = v_2(t)u_1(t) \\ &= (h_2 + (h_1 + g))(t). \end{aligned}$$

This proves (b). We emphasize that the $+$ sign has two different meanings in (b). $h_1 + h_2$ is the sum in the vector space H .

(c) Suppose $h_1 + g = h_2 + g$. Write $(h_i + g)(t) = v_i(t)g(t)$, $i = 1, 2$. Then $v_1 = v_2$. So

$$(\text{Ad } g(t))\dot{h}_1(t) = v_1(t)^{-1}\dot{v}_1(t) = v_2(t)^{-1}\dot{v}_2(t) = (\text{Ad } g(t))\dot{h}_2(t) \text{ a.e.}$$

Hence $\dot{h}_1 = \dot{h}_2$ a.e.

(e) Choose a left invariant metric ρ on G . The topology on W is determined by the metric

$$\text{dist}(f, g) = \sup\{\rho(f(t), g(t)) : 0 \leq t \leq T\}.$$

Fix f and g in W and $\varepsilon > 0$. By choosing local charts in G along f and g one can find f_0 and g_0 in $W \cap C^\infty([0, T]; G)$ such that $\text{dist}(f, f_0) < \varepsilon$ and $\text{dist}(g, g_0) < \varepsilon$. Let

$$k(t) = f_0(t)g_0(t)^{-1}$$

and

$$\dot{h}_0(t) = (\text{Ad } g(t)^{-1})(k(t)^{-1}\dot{k}(t)).$$

Then h_0 is in $C([0, T], \mathcal{S})$, but may not be smooth. Since

$$k(t)^{-1}\dot{k}(t) = (\text{Ad } g(t))\dot{h}_0(t),$$

we have

$$(h_0 + g)(t) = k(t)g(t) = f_0(t)(g_0(t)^{-1}g(t)).$$

Thus

$$\begin{aligned} \text{dist}(h_0 + g, f) &\leq \text{dist}(h_0 + g, f_0) + \varepsilon = \text{dist}(f_0 \cdot (g_0^{-1} \cdot g), f_0) + \varepsilon \\ &= \text{dist}(g, g_0) + \varepsilon \leq 2\varepsilon. \end{aligned}$$

Since $C^\infty([0, T]; \mathcal{S})$ is dense in H we may, by d), find h in $C^\infty([0, T]; \mathcal{S})$ close to h_0 such that $\text{dist}(h + g, h_0 + g) < \varepsilon$. Then $\text{dist}(h + g, f) < 3\varepsilon$.

(d) Suppose that h_n converges to h in H while g_n converges to g in the topology of W (i.e., uniformly on $[0, T]$). Then $\text{Ad } g_n(t) - \text{Ad } g(t)$ converges to zero in operator norm on \mathcal{S} uniformly in t . Hence

$$\begin{aligned} &\left(\int_0^T |(\text{Ad } g_n(t))\dot{h}_n(t) - (\text{Ad } g(t))\dot{h}(t)|^2 dt \right)^{1/2} \\ &\leq \left(\int_0^T |(\text{Ad } g_n(t) - \text{Ad } g(t))\dot{h}_n(t)|^2 dt \right)^{1/2} \\ &\quad + \left(\int_0^T |\text{Ad } g(t)(\dot{h}_n(t) - \dot{h}(t))|^2 dt \right)^{1/2} \\ &\leq \sup_t \|\text{Ad } g_n(t) - \text{Ad } g(t)\| \|h_n\|_H + \sup_t \|\text{Ad } g(t)\| \|h_n - h\|_H. \end{aligned}$$

But since $g(\cdot)$ has compact range $\sup\{\|\text{Ad } g(t)\|: 0 \leq t \leq T\}$ is finite. So the function $(\text{Ad } g_n(t))\dot{h}_n(t)$ converges to $(\text{Ad } g(t))\dot{h}(t)$ in H . By Lemma 2.1 the solutions to

$$v_n(t)^{-1}\dot{v}_n(t) = (\text{Ad } g_n(t))\dot{h}_n(t)$$

converge uniformly to v . The functions $(h_n + g_n)(t) \equiv v_n(t)g_n(t)$ also converge to $v(t)g(t) \equiv (h + g)(t)$ in W because if V is a neighborhood of e in G then, since range g is compact, there are neighborhoods U and U' of e such that $x'g(t)^{-1}xg(t)$ is in V for all t if x' is in U' and x is in U . Thus if n is so large that $g_n(t)^{-1}g(t) \in U'$ for all t and $v_n(t)^{-1}v(t) \in U$ for all t we

have

$$(v_n(t)g_n(t))^{-1}v(t)g(t) = (g_n(t)^{-1}g(t))g(t)^{-1}(v_n(t)^{-1}v_n(t))g(t)$$

is in V for all t in $[0, T]$.

(f) If v is defined by (2.4) then $u(t) \equiv v(t)g(t)$ is a product of two absolutely continuous functions on $[0, T]$ and is therefore absolutely continuous. Moreover

$$\begin{aligned} u(t)^{-1}\dot{u}(t) &= g(t)^{-1}v(t)^{-1}d(v(t)g(t))/dt \\ &= g(t)^{-1}v(t)^{-1}(\dot{v}(t)g(t) + v(t)\dot{g}(t)) \\ &= g(t)^{-1}\{(\text{Ad } g(t))\dot{h}(t)\}g(t) + g(t)^{-1}\dot{g}(t) \\ &= \dot{h}(t) + g(t)^{-1}\dot{g}(t). \end{aligned}$$

We have used, as in part (b), the identity $(\text{Ad } a) = (L_a)_*(R_{a^{-1}})_*$ and the notation $\xi a = (R_a)_*\xi$ for $\xi \in \mathcal{S}$ and $a \in G$.

(g) Write

$$(h + g)(t) = v(t)g(t) \quad \text{where } v(t)^{-1}\dot{v}(t) = (\text{Ad } g(t))\dot{h}(t) \text{ a.e.}$$

as in Lemma 2.1. Since $(\text{Ad } g(t))\dot{h}(t) = 0$ on (a, b) it follows from (2.1) that $f(v(b)) - f(v(a)) = 0$ for all f in $C^\infty(G)$. Hence $v(b) = v(a)$. This proves (g).

DEFINITION 2.5. If f is a function from W to R and h is in H we put

$$(\partial_h f)(g) = df(sh + g)/ds|_{s=0}, \quad h \in H, g \in W$$

if the derivative exists. We say f has a gradient at g if there is a vector $(\nabla f)(g)$ in H such that $(\partial_h f)(g) = ((\nabla f)(g), h)_H$ for all h in H . We say f is in $C^1(W)$ if $(\nabla f)(g)$ exists for each g in W and $\nabla f: W \rightarrow H$ is continuous.

Notation 2.6. For g in W and h in H let

$$(g \cdot h)(s) = \int_0^s (\text{Ad } g(\sigma))\dot{h}(\sigma) d\sigma.$$

Note that $g \cdot h$ is again in H because $\{\text{Ad } g(\sigma)\}_{0 \leq \sigma \leq T}$ is a uniformly bounded family of operators on \mathcal{S} for each g in W . W is a group under pointwise multiplication. In the next proposition we relate the group structure of W to the preceding notion of gradient.

PROPOSITION 2.7. *Let*

$$(e^h g)(s) = (\exp h(s))g(s)$$

for h in H and g in W . Let $f: W \rightarrow R$ and fix g in W . Then the functions $h \rightarrow f(h + g)$ and $h \rightarrow f(e^{g \cdot h} g)$ are both Frechet differentiable at $h = 0$ if either is. Moreover they have the same Frechet derivative. In particular if u is in $C^\infty(G^k)$ and f has the form

$$(2.6) \quad f(g) = u(g(t_1), g(t_1)^{-1}g(t_2), g(t_2)^{-1}g(t_3), \dots, g(t_{k-1})^{-1}g(t_k))$$

where $0 < t_1 < \dots < t_k \leq T$ then $(\partial_h f)(g)$ exists for all h in H and g in W and is given by

$$(2.7) \quad (\partial_h f)(g) = \sum_{i=1}^k u_i(g(t_1), \dots, g(t_{k-1})^{-1}g(t_k)) \times \left\langle \int_{t_{i-1}}^{t_i} (\text{Ad } g(t_i)^{-1}g(s)) \dot{h}(s) ds \right\rangle$$

where

$$u_i(x_1, \dots, x_k) \langle \xi \rangle = du(x_1, \dots, x_i e^{s\xi}, x_{i+1}, \dots, x_k) / ds \Big|_{s=0}$$

for ξ in \mathcal{G} . Moreover $(\nabla f)(g)$ exists for all g and

$$(2.8) \quad |(\nabla f)(g)|^2 = \sum_{i=1}^k \int_{t_{i-1}}^{t_i} \left| (\text{Ad } g(t_i)^{-1}g(s))^{\text{tr}} u_i \right|_{\mathcal{G}^*}^2 ds,$$

where $u_i = u_i(g(t_1), \dots, g(t_{k-1})^{-1}g(t_k))$ is in the dual space \mathcal{G}^* and the superscript tr denotes the transposed operator.

Proof. By Definition 2.2 we may write

$$(h + g)(s) = [\exp \chi(g \cdot h)(s)]g(s)$$

where χ is defined in Lemma 2.1 for h with small norm. Let $\varphi(h) = f(e^h g)$ and $\psi(h) = f(e^{\chi(h)} g)$ where $(e^{\chi(h)} g)(s) = [\exp \chi(h)(s)]g(s)$. Then $\psi(h) = \varphi(\chi(h))$. By Lemma 2.1, φ and ψ have the same Frechet derivative at $h = 0$. Hence so do the functions $h \rightarrow \varphi(g \cdot h) = f(e^{g \cdot h} g)$ and $h \rightarrow \psi(g \cdot h) = f(h + g)$ since the map $h \rightarrow g \cdot h$ is a bounded linear map from H into H with inverse $h \rightarrow g^{-1} \cdot h$ where $g^{-1}(s) = g(s)^{-1}$. We now use this to verify (2.7).

For real r we have

$$\begin{aligned} &(e^{rg \cdot h}g)(t_{i-1})^{-1}(e^{rg \cdot h}g)(t_i) \\ &= g(t_{i-1})^{-1}e^{-r(g \cdot h)(t_{i-1})}e^{r(g \cdot h)(t_i)}g(t_i) \\ &= g(t_{i-1})^{-1}g(t_i)\exp\left[-r \operatorname{Ad} g(t_i)^{-1}(g \cdot h)(t_{i-1})\right] \\ &\quad \cdot \exp\left[r \operatorname{Ad} g(t_i)^{-1}(g \cdot h)(t_i)\right]. \end{aligned}$$

If we use the identity to compute $f(e^{g \cdot h}g)$ with f given by (2.6) we see that $h \rightarrow f(e^{g \cdot h}g)$ is a Frechet differentiable function of h at $h = 0$. Moreover for X and Y in \mathcal{S} the tangent vector to the curve $\exp(-rX)\exp(rY)$ is $Y - X$ at $r = 0$. Hence at $r = 0$,

$$\begin{aligned} &df(e^{rg \cdot h}g)/dr \\ &= \sum_{i=1}^k u_i(g(t_1), \dots) \left\langle (\operatorname{Ad} g(t_i)^{-1})\{(g \cdot h)(t_i) - (g \cdot h)(t_{i-1})\} \right\rangle \end{aligned}$$

which is (2.7). Since the right side of (2.7) is a linear functional of h given in each interval $(t_{i-1}, t_i]$ by integration against a continuous function into $\mathcal{S}^*(\partial_h f)(g)$ is, for each g in W , a continuous linear functional of h on H whose norm is given by (2.8).

It is technically useful to have available functions on W with bounded gradient. Unfortunately the simplest kinds of functions on W may never have this property as the following corollary shows.

COROLLARY 2.8. *Assume that the dual of the adjoint representation of G has the property that every non zero orbit is unbounded. In other words, $\{\operatorname{Ad}(x)^{\operatorname{tr}}\eta: x \in G\}$ is an unbounded set in \mathcal{S}^* for each non-zero η in \mathcal{S}^* . Let v be in $C^\infty(G^k)$ and let $0 < t_1 < t_2 < \dots < t_k \leq T$. Put*

$$(2.10) \quad f(g) = v(g(t_1), g(t_2), \dots, g(t_k)).$$

Then $|(\nabla f)(g)|^2$ is unbounded on W unless v is constant.

Remark 2.9. If $G = SL(2, R)$ then the hypothesis of Corollary 2.8 holds because the image of the adjoint representation of G is $SO(2, 1)$ which is isomorphic to its contragredient representation.

Proof of Corollary 2.8. The map $S: G^k \rightarrow G^k$ defined by

$$S(x_1, \dots, x_k) = (x_1, x_1x_2, x_1x_2x_3, \dots, x_1x_2 \cdots x_k)$$

is a diffeomorphism with inverse

$$S^{-1}(y_1, \dots, y_k) = (y_1, y_1^{-1}y_2, y_2^{-1}y_3, \dots, y_{k-1}^{-1}y_k).$$

Let $u = v \circ S$. Then u is also in $C^\infty(G^k)$ and f is given by (2.6). By Proposition 2.7, $(\nabla f)(g)$ exists for all g and (2.8) holds. Suppose that v and hence u is not constant. Then there is a point x_1, \dots, x_k in G^k such that $u_i(x_1, \dots, x_k) \neq 0$ for some i . Consider all functions g in W for which $g(t_1) = x_1$ and $g(t_j)^{-1}g(t_j) = x_j$ for $j = 2, \dots, k$. Such functions are restricted only at the points t_j and by the requirement of continuity in between. Since $(\text{Ad } g(t_i)^{-1})^{\text{tr}}u_i$ is not zero $(\text{Ad } x)^{\text{tr}}(\text{Ad } g(t_i)^{-1})^{\text{tr}}u_i$ is unbounded as a function of x on G . Thus

$$\int_{t_{i-1}}^{t_i} \left| (\text{Ad } g(s))^{\text{tr}} (\text{Ad } g(t_i)^{-1})^{\text{tr}} u_i \right|_{\mathcal{L}^*}^2 ds$$

can be made arbitrarily large by choosing $g(s)$ suitably on (t_{i-1}, t_i) .

In order to construct bounded functions with bounded gradients on W the function defined in the following Lemma will be useful.

LEMMA 2.10. *Let*

$$(2.11) \quad \varphi(g) = \int_0^T \text{trace}\{(\text{Ad } g(t))^*(\text{Ad } g(t))\} dt.$$

Then $(\nabla\varphi)(g)$ exists, is continuous from W into H^ and*

$$(2.12) \quad |(\nabla\varphi)(g)| \leq 2M\varphi(g)^{3/2}$$

where M is the operator norm of the map $\eta \rightarrow \text{ad } \eta$ from \mathcal{L} to operators on \mathcal{L} and the superscript $$ denotes the adjoint relative to the given inner product on \mathcal{L} .*

Proof. Let $u(x) = \text{trace}\{(\text{Ad } x)^*\text{Ad}(x)\}$ for x in G . Then

$$\begin{aligned} u'(x)\langle \xi \rangle &:= du(xe^{s\xi})/ds|_{s=0} \\ &= \text{trace}\{(\text{Ad } x)^*(\text{Ad } x)\text{ad } \xi\} + \text{trace}\{(\text{ad } \xi)^*(\text{Ad } x)^*\text{Ad } x\} \\ &= 2 \text{trace}\{(\text{Ad } x)^*(\text{Ad } x)\text{ad } \xi\}. \end{aligned}$$

Hence by (2.7) and the identity $\text{Ad } x \text{ ad } \xi = \text{ad}((\text{Ad } x)\xi)\text{Ad } x$ we have

$$\begin{aligned} \partial_h u(g(t)) &= 2 \text{ trace} \left\{ (\text{Ad } g(t))^* (\text{Ad } g(t)) \text{ad} \int_0^t \text{Ad } g(t)^{-1} (\text{Ad } g(s)) \dot{h}(s) ds \right\} \\ &= 2 \text{ trace} \left\{ (\text{Ad } g(t))^* \left(\text{ad} \int_0^t (\text{Ad } g(s)) \dot{h}(s) ds \right) \text{Ad } g(t) \right\}. \end{aligned}$$

As a function of g from W into H^* this is clearly continuous and is bounded on bounded sets in W uniformly in t . Hence $u((sh + g)(t))$ is continuously differentiable in s for each t with a uniformly bounded (in t) derivative in $|s| \leq 1$. Thus we may interchange the derivative and integral in φ to get

$$(\partial_h \varphi)(g) = \int_0^T \partial_h u(g(t)) dt.$$

The operator norm $\| \cdot \|$ satisfies $\| \text{ad } \eta \| \leq M \|\eta\|$ by assumption and is of course majorized by the Hilbert-Schmidt norm. Hence

$$\begin{aligned} |(\partial_h \varphi)(g)| &\leq \int_0^T |\partial_h u(g(t))| dt \\ &\leq \int_0^T 2 \text{ trace} \{ (\text{Ad } g(t))^* \text{Ad } g(t) \} \left\| \text{ad} \int_0^t (\text{Ad } g(s)) \dot{h}(s) ds \right\| dt \\ &\leq 2 \int_0^T \text{ trace} \{ (\text{Ad } g(t))^* \text{Ad } g(t) \} M \left| \int_0^t (\text{Ad } g(s)) \dot{h}(s) ds \right| dt \\ &\leq 2 \int_0^T \text{ trace} \{ (\text{Ad } g(t))^* \text{Ad } g(t) \} M \int_0^T |(\text{Ad } g(s)) \dot{h}(s)| ds dt \\ &\leq 2 \varphi(g) M \int_0^T |(\text{Ad } g(s)) \dot{h}(s)| ds \\ &\leq 2 \varphi(g) M \left(\int_0^T \text{ trace} \{ (\text{Ad } g(s))^* \text{Ad } g(s) \} ds \right)^{1/2} \\ &\quad \times \left(\int_0^T |\dot{h}(s)|^2 ds \right)^{1/2} \\ &= 2 \varphi(g) M \varphi(g)^{1/2} \|\dot{h}\| \end{aligned}$$

which proves (2.12).

COROLLARY 2.11. *Let u and v be in $C_c^1(G^k)$. Define $f: W \rightarrow R$ by (2.6) or (2.10) and define φ by (2.11). Let $\varepsilon > 0$ and put*

$$F_\varepsilon(g) = f(g)(1 + \varepsilon\varphi(g))^{-1}.$$

Then F_ε has a bounded continuous gradient on W .

Proof. Suppose f is given by (2.6). Then

$$(2.13) \quad \begin{aligned} \nabla F_\varepsilon(g) &= (\nabla f)(g)(1 + \varepsilon\varphi(g))^{-1} \\ &\quad - \varepsilon f(g)(1 + \varepsilon\varphi(g))^{-2}(\nabla\varphi)(g). \end{aligned}$$

By (2.8) we have

$$(2.14) \quad \begin{aligned} |\nabla f(g)|^2 &\leq \sum_{i=1}^k \int_{t_{i-1}}^{t_i} \|(\text{Ad } g(s))^{\text{tr}}\|_{\mathcal{G}^* \rightarrow \mathcal{G}^*}^2 ds \left| (\text{Ad } g(t_i)^{-1})^{\text{tr}} u_i \right|_{\mathcal{G}^*}^2 \\ &\leq \int_0^T \|\text{Ad } g(s)\|_{\mathcal{G} \rightarrow \mathcal{G}}^2 ds \max_i \left| (\text{Ad } g(t_i)^{-1})^{\text{tr}} u_i \right|_{\mathcal{G}^*}^2 \\ &\leq C\varphi(g) \end{aligned}$$

where

$$C = \sup_{i,x} \left| (\text{Ad } x_i^{-1})^{\text{tr}} u_i(x_1, \dots, x_k) \right|_{\mathcal{G}^*}$$

which is finite because $u \in C_c^1(G^k)$. Thus by Lemma 2.10,

$$\begin{aligned} |(\nabla F_\varepsilon)(g)| &\leq C^{1/2}\varphi(g)^{1/2}(1 + \varepsilon\varphi(g))^{-1} \\ &\quad + \varepsilon|f(g)|2M\varphi(g)^{3/2}(1 + \varepsilon\varphi(g))^{-2}. \end{aligned}$$

Both terms are bounded. The continuity of $g \rightarrow (\nabla F_\varepsilon)(g)$ from W to H^* follows from Lemma 2.10 and (2.7).

In case f is given by (2.10) then it is also given by (2.6) where $u = v \circ S$ and S is the diffeomorphism of G^k defined in the proof of Corollary 2.8. Since u is then in $C_c^1(G^k)$ the previous case applies.

Next we define polygonal functions in W . Choose a partition $0 = t_0 < t_1 < t_2 < \dots < t_k \leq T$ and let y_1, \dots, y_k be in \mathcal{S} . Define

$$(2.15) \quad \psi(y_1, \dots, y_k)(t) = \begin{cases} (\exp y_1)(\exp y_2) \cdots (\exp y_{r-1}) \left(\exp \frac{t - t_{r-1}}{t_r - t_{r-1}} y_r \right) & \text{if } t_{r-1} \leq t \leq t_r \\ \text{constant} & \text{if } t_k \leq t \leq T. \end{cases}$$

Then $\psi(y_1, \dots, y_k)(\cdot)$ is continuous and has bounded derivatives of all orders on each interval (t_{r-1}, t_r) . In particular it is absolutely continuous on $[0, T]$.

PROPOSITION 2.12. *Suppose that $f: W \rightarrow R$ has a gradient $\nabla f(g)$ for all g in W . Define ψ by (2.15) and put $v = f \circ \psi: \mathcal{S}^k \rightarrow R$. Let ξ_1, \dots, ξ_d be an O.N. of \mathcal{S} . Then the partial derivative*

$$(\partial_{r,j} v)(y_1, \dots, y_k) = dv(y_1, \dots, y_r + s\xi_j, y_{r+1}, \dots, y_k) / ds \Big|_{s=0}$$

exists and is given by

$$(2.16) \quad \partial_{r,j} v = (\partial_{h_{r,j}} f) \circ \psi$$

where $h_{r,j}$ is the element of H determined by

$$(2.17) \quad h_{r,j}(t) = \begin{cases} (t_r - t_{r-1})^{-1} \xi_j & \text{if } t_{r-1} < t < t_r \\ 0 & \text{otherwise.} \end{cases}$$

Moreover if $t_r = r/n$, $r = 1, \dots, k$ then

$$(2.18) \quad n^{-1} |(\nabla v)(y_1, \dots, y_k)|^2 \leq |(\nabla f)(\psi(y_1, \dots, y_k))|^2$$

where ∇v is the gradient on the Euclidean space \mathcal{S}^k .

Proof. Fix y_1, \dots, y_k in \mathcal{S}^k and write $g = \psi(y_1, \dots, y_k)$. Put $g_s = sh_{r,j} + g$. Since g is piecewise C^1 , by Theorem 2.4(f) we have

$$(2.19) \quad g_s(t)^{-1} \dot{g}_s(t) = g(t)^{-1} \dot{g}(t) + s\dot{h}_{r,j}(t) \quad \text{for a.e. } t.$$

But (2.15) shows that $\dot{g}(t) = g(t)_*(t_r - t_{r-1})^{-1} y_r$ for $t_{r-1} < t < t_r$. Hence $g(t)^{-1} \dot{g}(t) = (t_r - t_{r-1})^{-1} y_r$ on the r th interval for each r and in particular for the r in the statement of the proposition. Thus $g_s(t)^{-1} \dot{g}_s(t) = g(t)^{-1} \dot{g}(t)$

on each of these open intervals except the r th interval where it takes the constant value $(t_r - t_{r-1})^{-1}(y_r + s\xi_j)$. The unique solution to (2.19) is therefore

$$g_s(t) = \psi(y_1, \dots, y_r + s\xi_j, y_{r+1}, \dots, y_k)(t).$$

Hence

$$(2.20) \quad sh_{r,j} + \psi(y_1, \dots, y_k) = \psi(y_1, \dots, y_r + s\xi_j, \dots, y_k).$$

Thus

$$df(sh_{r,j} + \psi(y_1, \dots, y_k))/ds = df(\psi(y_1, \dots, y_r + s\xi_j, \dots, y_k))/ds$$

which proves (2.16). Finally, if $t_r = r/n$ for $r = 1, \dots, k$ then

$$\|h_{r,j}\|^2 = \int_{(r-1)/n}^{r/n} |(t_r - t_{r-1})^{-1}\xi_j|^2 dt = n.$$

Hence the functions $n^{-1/2}h_{r,j}$ form an O.N. set in H . Thus, writing $y = (y_1, \dots, y_k)$ in \mathcal{S}^k , we have

$$\begin{aligned} |\nabla v(y)|^2 &= \sum_{r=1}^k \sum_{j=1}^d |(\partial_{r,j}v)(y)|^2 \\ &= \sum_{r=1}^k \sum_{j=1}^d ((\nabla f)(\psi(y)), h_{r,j})^2 \\ &= n \sum_{r=1}^k \sum_{j=1}^d ((\nabla f)(\psi(y)), n^{-1/2}h_{r,j})^2 \\ &\leq n|(\nabla f)(\psi(y))|^2. \end{aligned}$$

3. Logarithmic Sobolev inequality on path space

Let k be the integer part of nT . Define $\psi_n: \mathcal{S}^k \rightarrow W$ by

$$(3.1) \quad \begin{aligned} \psi_n(y_1, \dots, y_k)(t) &= \exp[y_1] \cdots \exp[y_{r-1}] \\ &\quad \times \exp[n(t - (r-1)/n)y_r] \\ &\quad \text{if } (r-1)/n \leq t \leq r/n \end{aligned}$$

for $r = 1, \dots, k$ and define $\psi_n(y_1, \dots, y_k)(t)$ to be constant on $[k/n, T]$. Let

λ be a probability measure on \mathcal{S} such that

- (a) $\int_{\mathcal{S}} |y|^2 \lambda(dy) < \infty$
- (b) $\int_{\mathcal{S}} y \lambda(dy) = 0$
- (c) $\int_{\mathcal{S}} (\xi_i, y)(\xi_j, y) \lambda(dy) = \delta_{ij}$

where ξ_1, \dots, ξ_d is an O.N. basis of \mathcal{S} .

Let $\lambda_n(E) = \lambda(n^{1/2}E)$ for Borel sets $E \subseteq \mathcal{S}$ and put $\lambda_n^k = \lambda_n \times \dots \times \lambda_n$ on \mathcal{S}^k . Define, for a Borel set B in W ,

$$(3.2) \quad P_n(B) = \lambda_n^k(\psi_n^{-1}B).$$

Then P_n is a probability measure on W .

PROPOSITION 3.1. *The sequence of measures P_n converges weakly on W to the probability measure P corresponding to the diffusion process on G with initial value $X(0) = e$ and infinitesimal generator $(1/2)\sum_{j=1}^d \hat{\xi}_j^2$.*

Proof. The proof consists largely of verifying the hypothesis of the central limit theorem of Stroock and Varadhan [43, Theorem 2.4] using largely standard techniques. Put

$$(3.3) \quad \chi_n(y_1, \dots, y_k)(t) = \exp[y_1] \cdots \exp[y_r] \quad \text{if } (r - 1)/n \leq t < r/n$$

for $r = 1, \dots, k$ and define χ on $[k/n, T]$ so as to be constant on $[(k - 1)/n, T]$. Then $\chi_n(y_1, \dots, y_k)(\cdot)$ is piecewise constant, and lies in the space Ω_d of functions from $[0, T]$ into G which are right continuous on $[0, T]$ and left continuous at T and have only discontinuities of the first kind. The Skorohod topology on Ω_d may be defined with the aid of a left invariant metric ρ on G in the usual way [6, 35]. For a Borel measurable set B in Ω_d define

$$R_n(B) = \lambda_n^k(\chi_n^{-1}(B)).$$

The central limit theorem of Stroock and Varadhan [43, Theorem 2.4] asserts that R_n converges weakly to P on Ω_d under hypotheses which we will now verify. We adhere to the notation of [43] in the rest of this proof.

Let $\mu_n = \lambda_n \circ \exp^{-1}$ be the probability measure on G induced from λ_n by the exponential map. With $B_r = \{y \in \mathcal{S} : |y| < r\}$ we choose a so that \exp is a diffeomorphism of B_{2a} onto a neighborhood N of e in G and put $V = \exp B_a$. We use exponential coordinates on N defined by $\varphi_j(\exp \sum_{i=1}^d s_i \xi_i) = s_j$. The independent G valued random variables on $(\mathcal{S}^k, \lambda_n^k)$, $X_{n,j} \equiv \exp y_j$, have a common distribution $\alpha_{n,j} = \mu_n$, $j = 1, \dots, k$. We need to verify the Lindeberg condition, covariance and mean conditions

[43, (2.6), (2.22), (2.23)]. But in fact if $U = \exp B_r$ is a neighborhood of e then

$$\begin{aligned} \mu_n(U^c) &\leq \lambda_n(|y| \geq r) = \lambda(|z| \geq n^{1/2}r) \leq r^{-2}n^{-1} \int_{|z| \geq n^{1/2}r} |z|^2 \lambda(dz) \\ &= o(n^{-1}). \end{aligned}$$

Hence $n\mu_n(U^c) \rightarrow 0$, which verifies the Lindeberg condition. The mean $g_{n,j}$ in V defined by [43, (2.8)] satisfies

$$\begin{aligned} |\varphi(g_{n,j})| &= \left| \int_V \varphi(x) \alpha_{n,j}(dx) \right| = \left| \int_{\mathcal{G}} \varphi(\exp y) C_V(\exp y) \lambda_n(dy) \right| \\ &= \left| \int_{B_a} y \lambda_n(dy) + \int_{\mathcal{G}-B_a} \varphi(\exp y) C_V(\exp y) \lambda_n(dy) \right| \\ &\leq \left| \int_{B_a} y \lambda_n(dy) \right| + a \lambda_n(B_a^c) \end{aligned}$$

because $|\varphi(x)C_V(x)| \leq a$. But $\lambda_n(B_a^c) = o(n^{-1})$. Since $\int y \lambda_n(dy) = 0$ we have

$$\begin{aligned} \left| \int_{B_a} y \lambda_n(dy) \right| &= \left| - \int_{B_a^c} y \lambda_n(dy) \right| \\ &= \left| n^{-1/2} \int_{|z| \geq n^{1/2}a} z \lambda(dz) \right| \leq a^{-1}n^{-1} \int_{|z| \geq n^{1/2}a} |z|^2 \lambda(dz) \\ &= o(n^{-1}). \end{aligned}$$

Hence $|\varphi(g_{n,j})| = \varepsilon_n/n$ with $\varepsilon_n \rightarrow 0$ as $n \rightarrow \infty$. If $y_n = \varphi(g_{n,j})$ (which is independent of j) then $g_{n,1} \cdots g_{n,r} = \exp(ry_n)$ and since $|ry_n| = (r/n)\varepsilon_n \leq T\varepsilon_n$ the cumulative means $h_n(t)$ (cf. [43, (2.14)]) converge to e uniformly on $[0, T]$.

Since

$$\begin{aligned} &\int_V \varphi_p(x) \varphi_q(x) \mu_n(dx) \\ &= \int_{B_a} (\xi_p, y)(\xi_q, y) \lambda_n(dy) + \int_{B_a^c} \varphi_p(\exp y) \varphi_q(\exp y) C_V(\exp y) \lambda_n(dy) \\ &= n^{-1} \int_{|z| < n^{1/2}a} (\xi_p, z)(\xi_q, z) \lambda(dz) + o(n^{-1}) \\ &= n^{-1} \left[\delta_{pq} - \int_{|z| \geq n^{1/2}a} (\xi_p, z)(\xi_q, z) \lambda(dz) \right] + o(n^{-1}) \\ &= n^{-1} \delta_{pq} + o(n^{-1}) \end{aligned}$$

and since $\varphi_p(g_{n,j}) = o(n^{-1})$, the cumulative variance $A^{(n)}(t)$, cf. [43, (2.15)] converges uniformly on $[0, T]$ to t times the identity matrix.

The central limit theorem of Stroock and Varadhan [43, Theorem 2.4] now shows that R_n converges weakly to P . P is supported on W and we want to show now that the measures P_n (on W) converge in the weak sense on W to P . Let $V_\varepsilon = \exp B_\varepsilon$ and put

$$A_\varepsilon = \left\{ (y_1, \dots, y_k) \in \mathcal{G}^k : \chi_n(y_1, \dots, y_k)(t)^{-1} \psi_n(y_1, \dots, y_k)(t) \in V_\varepsilon \right. \\ \left. \text{for } 0 \leq t \leq T \right\}.$$

By (3.1) and (3.3) the ratio $\chi_n(y)(t)^{-1} \psi_n(y)(t)$ is $\exp[n(t - r/n)y_r]$ if $(r - 1)/n \leq t < r/n$ and is e on $[k/n, T]$. But

$$|n(t - r/n)y_r| \leq |y_r| \quad \text{if } (r - 1)/n \leq t < r/n.$$

Hence

$$\lambda_n^k(A_\varepsilon^c) \leq \lambda_n^k \left(\sup_{r=1, \dots, k} |y_r| \geq \varepsilon \right) \leq \sum_{r=1}^k \lambda_n^k(|y_r| \geq \varepsilon) \\ \leq Tn\lambda_n(|y| \geq \varepsilon)$$

which goes to zero by the Lindeberg condition. Thus

$$(3.4) \quad \lambda_n^k(A_\varepsilon^c) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

So if $u: G^m \rightarrow R$ is bounded and uniformly continuous and if

$$|u(x_1, \dots, x_m) - u(x'_1, \dots, x'_m)| < \delta$$

whenever all $x_i^{-1}x'_i \in V_\varepsilon$ then for any $0 < t_1 < \dots < t_m \leq T$,

$$(3.5) \quad \left| \int_{\Omega_d} u(g(t_1), \dots, g(t_m))R_n(dg) - \int_W u(g(t_1), \dots, g(t_m))P_n(dg) \right| \\ = \left| \int_{\mathcal{G}^k} [u(\chi_n(y_1, \dots, y_k)(t_1), \dots, \chi_n(t_m)) \right. \\ \left. - u(\psi_n(y_1, \dots, y_k)(t_1), \dots, \psi_n(t_m))] \lambda_n^k(dy) \right| \\ \leq \int_{A_\varepsilon} |\dots| \lambda_n^k(dy) + 2\|u\|_\infty \lambda_n^k(A_\varepsilon^c) \\ \leq \delta + 2\|u\|_\infty \lambda_n^k(A_\varepsilon^c).$$

The $\limsup_{n \rightarrow \infty}$ of the left side of (3.5) is therefore $\leq \delta$ for any $\delta > 0$ and hence is zero. But

$$\int_{\Omega_d} uR_n \rightarrow \int_W u(g(t_1), \dots, (g(t_m)))P(dg)$$

for any t_1, \dots, t_m because P is supported on W . Thus

$$\int_W u(g(t_1), \dots, g(t_m))P_n(dg) \rightarrow \int_W u(\dots)P(g).$$

It remains to show that $\{P_n\}$ is tight on W . For this it suffices to show that, regarded as measures on Ω_d , $\{P_n\}_{n=1}^\infty$ is tight. For if K is a compact set in Ω_d with $P_n(K) \geq 1 - \delta$ then, since W is closed in Ω_d and has the relative topology, $K \cap W$ is compact in W and $P_n(K \cap W) = P_n(K) \geq 1 - \delta$. To show tightness of $\{P_n\}_{n=1}^\infty$ on Ω_d we use the already established [43] tightness of $\{R_n\}_{n=1}^\infty$ on Ω_d . If ρ is a left invariant metric on G the modulus of continuity relevant for Ω_d may be defined as

$$w'_g(\delta) = \inf_{\{t_i\}} \sup_i \sup \{ \rho(g(t), g(s)) : t_{i-1} \leq s < t \leq t_i \}$$

where the inf is taken over partitions $0 = t_0 < t_1 < \dots < t_r = T$ with $t_i - t_{i-1} > \delta, i = 1, \dots, r$ and g is an arbitrary G valued function. Note that if g_1 and g_2 are in Ω_d and

$$\rho(g_1, g_2) \equiv \sup_t \rho(g_1(t), g_2(t)) \leq \varepsilon$$

then

$$w'_{g_1}(\delta) \leq w'_{g_2}(\delta) + 2\varepsilon.$$

Thus for fixed y_1, \dots, y_k and $\delta > 0, w'_{\chi_n}(\delta) \geq \varepsilon$ if $w'_{\psi_n}(\delta) \geq 3\varepsilon$ and $\rho(\psi_n, \chi_n) \leq \varepsilon$. Hence

$$\lambda_n^k \{ w'_{\psi_n}(\delta) \geq 3\varepsilon, \rho(\psi_n, \chi_n) \leq \varepsilon \} \leq \lambda_n^k \{ w'_{\chi_n}(\delta) \geq \varepsilon \}$$

and therefore

$$\lambda_n^k \{ w'_{\psi_n}(\delta) \geq 3\varepsilon \} \leq \lambda_n^k \{ w'_{\chi_n}(\delta) \geq \varepsilon \} + \lambda_n^k \{ \rho(\psi_n, \chi_n) > \varepsilon \}.$$

That is

$$P_n \{ w'_g(\delta) \geq 3\varepsilon \} \leq R_n \{ w'_g(\delta) \geq \varepsilon \} + \lambda_n^k \{ \rho(\psi_n, \chi_n) > \varepsilon \}.$$

Given $\varepsilon > 0$ and $\eta > 0$ there exists n_0 and $\delta > 0$ such that $R_n\{w'_g(\delta) \geq \varepsilon\} < \eta$ for $n \geq n_0$ by a standard tightness theorem (cf. [6, Theorem 15.2]). Since $\lambda_n^k\{\rho(\psi, \chi) > \varepsilon\} \rightarrow 0$ by (3.4), there is an n_1 such that $P_n(w'_g(\delta) \geq 3\varepsilon) < 2\eta$ for $n \geq n_1$. Since $P_n\{g(0) = e\} = 1$, the same tightness theorem establishes the tightness of $\{P_n\}_{n=1}^\infty$ on Ω_d .

Example 3.2 (Bernoulli measure). If λ assigns mass 2^{-d} to each of the points $\sum_{j=1}^d \varepsilon_j \xi_j$ where $d = \dim \mathcal{S}$ and each $\varepsilon_j = \pm 1$ then the conditions (a), (b), (c) for Proposition 3.1 hold. This will be useful for understanding the loss of best constants that we will explore in Section 5.

Example 3.3 (Gauss measure). $\lambda(dz) = (2\pi)^{-d/2} e^{-|z|^2/2} dz$. In this case a stronger version of Proposition 3.1 has been proved by McKean [29] and shows that P is the measure on W induced by “wrapping a \mathcal{S} valued Wiener process around G ”. For a manifold a similar theorem has been shown by Elworthy [11]. See also the method of Baxendale [3].

THEOREM 3.4. *Let P be the distribution on W of the diffusion process X in G with infinitesimal generator $(1/2)\sum_{j=1}^d \hat{\xi}_j^2$ and with $X(0) = e$. Let $f: W \rightarrow \mathbb{R}$ be bounded and continuous and have a bounded continuous gradient on W . Then*

$$(3.6) \quad \int_W f(g)^2 \ln|f(g)| P(dg) \leq \int_W |(\nabla f)(g)|^2 P(dg) + \|f\|_{L^2(P)}^2 \ln\|f\|_{L^2(P)}.$$

Proof. Denote by λ Gauss measure as in Example 3.3. We may apply Proposition 3.1 to construct P . We start with the known [17] logarithmic Sobolev inequality for the Gauss measure λ^k on \mathcal{G}^k .

$$(3.7) \quad \int_{\mathcal{G}^k} |u(z)|^2 \ln|u(z)| \lambda^k(dz) \leq \int_{\mathcal{G}^k} |(\nabla u)(z)|^2 \lambda^k(dz) + \|u\|_{L^2(\lambda^k)}^2 \ln\|u\|_{L^2(\lambda^k)}$$

where $z = (z_1, \dots, z_k)$ is in \mathcal{G}^k and ∇U is the gradient of the C^1 function u . Put $\lambda_n(E) = \lambda(n^{1/2}E)$ and let $u(z) = v(n^{-1/2}z)$. Then $(\nabla u)(z) = n^{-1/2}(\nabla v)(n^{-1/2}z)$ by the chain rule. We change variables to $y = n^{-1/2}z$ in

(3.7) to get

$$(3.8) \quad \int_{\mathcal{G}^k} |v(y)|^2 \ln|v(y)| \lambda_n^k(dy) \leq n^{-1} \int_{\mathcal{G}^k} |(\nabla v)(y)|^2 \lambda_n^k(dy) + \|v\|_{L^2(\lambda_n^k)}^2 \ln \|v\|_{L^2(\lambda_n^k)}.$$

With ψ_n defined by (3.1) put $v(y) = f(\psi_n(y))$. By the definition (3.2) of P_n and by (2.18) we have

$$(3.9) \quad \int_W |f(g)|^2 \ln|f(g)| P_n(dg) \leq \int_W |(\nabla f)(g)|^2 P_n(dg) + \|f\|_{L^2(P_n)}^2 \ln \|f\|_{L^2(P_n)}.$$

Since the integrands are all bounded and continuous on W we may let $n \rightarrow \infty$ in (3.9) to get (3.6).

We want now to extend the class of functions for which (3.6) holds to some unbounded functions.

LEMMA 3.5. *Let M be the norm of the linear map $y \rightarrow \text{ad } y$ from \mathcal{G} to operators on \mathcal{G} in operator norm. Put $\gamma = 2M^2$. Then for $0 \leq t \leq T$ we have*

$$(3.10) \quad \int_W |(\text{Ad } g(t))\xi|^2 P(dg) \leq e^{\gamma t} |\xi|^2 \quad \text{for } \xi \text{ in } \mathcal{G},$$

and

$$(3.11) \quad \int_W \left| (\text{Ad } g(b)^{-1} g(a))^{\text{tr}} \eta \right|^2 P(dg) \leq e^{\gamma(b-a)} |\eta|^2$$

for $\eta \in \mathcal{G}^*, 0 \leq a < b \leq T$.

Moreover if φ is defined by (2.11) then

$$(3.12) \quad \int_W \varphi(g) P(dg) < \infty.$$

Proof. Two proofs of a related integrability theorem are given in [34], which would yield a proof. We give another short proof based on the random walk representation of P given in Proposition 3.1. Let λ be Gauss measure on \mathcal{G} as in the proof of Theorem 3.4. Since $\|\text{Ad exp } sy\| = \|e^{\text{ad } sy}\| \leq e^{M|sy|}$

the function

$$u(s) := \int_{\mathcal{G}} |e^{s \operatorname{ad} y \xi}|^2 \lambda(dy)$$

extends to an entire function of s . Clearly $u(0) = |\xi|^2$. Moreover

$$u'(s) = \int 2((\operatorname{ad} y) e^{s \operatorname{ad} y \xi}, e^{s \operatorname{ad} y \xi}) \lambda(dy)$$

and

$$u''(s) = \int \left\{ 2((\operatorname{ad} y)^2 e^{s \operatorname{ad} y \xi}, e^{s \operatorname{ad} y \xi}) + 2|(\operatorname{ad} y) e^{s \operatorname{ad} y \xi}|^2 \right\} \lambda(dy).$$

Thus $u'(0) = 0$ since the integrand in $u'(0)$ is odd. Similarly $u'''(0) = 0$. Moreover

$$u''(0) \leq \int 4M^2 |y|^2 |\xi| \lambda(dy) = 4M^2 |\xi|^2.$$

Hence $u(s) = (1 + as^2 + s^4 b(s, \xi)) |\xi|^2$ where $|a| \leq 2M^2$ and $b(s, \xi)$ is bounded for $|s| \leq 1$. Thus if $0 \leq \alpha \leq 1$ then

$$\begin{aligned} \int |e^{\operatorname{ad} \alpha y \xi}|^2 \lambda_n(dy) &= \int |e^{n^{-1/2} \alpha \operatorname{ad} z \xi}|^2 \lambda(dz) = u(n^{-1/2} \alpha) \\ &= (1 + a\alpha^2/n + O(n^{-2})) |\xi|^2. \end{aligned}$$

Hence

$$(3.13) \quad \int |e^{\operatorname{ad} \alpha y \xi}|^2 \lambda_n(dy) \leq \beta_n |\xi|^2$$

where $\beta_n = 1 + \gamma/n + O(n^{-2})$. If $(r - 1)/n \leq t < r/n$ put $\alpha = n(t - (r - 1)/n)$. Then $0 \leq \alpha \leq 1$ and by (3.2) and repeated use of (3.13) we have

$$\begin{aligned} &\int_W |(\operatorname{Ad} g(t)) \xi|^2 P_n(dg) \\ &= \int_{\mathcal{G}^r} |e^{\operatorname{ad} y_1} \dots e^{\operatorname{ad} y_{r-1}} e^{\operatorname{ad} \alpha y_r} \xi|^2 \lambda_n(dy_1) \dots \lambda_n(dy_r) \\ &\leq \beta_n^r |\xi|^2. \end{aligned}$$

Note that β_n^r converges to $e^{\gamma t}$ as $n \rightarrow \infty$. Let

$$\chi(g) = |(\text{Ad } g(t))\xi|^2 \quad \text{and} \quad \chi_m(g) = \min(m, \chi(g)).$$

χ_m is bounded and continuous on W . Hence

$$\begin{aligned} \int \chi(g)P(dg) &= \lim_{m \rightarrow \infty} \int \chi_m P(dg) = \lim_m \lim_n \int \chi_m P_n \\ &\leq \limsup_n \int \chi P_n \leq \limsup_n \beta_n^r |\xi|^2 = e^{\gamma t} |\xi|^2. \end{aligned}$$

The proof of (3.11) is similar but with the factors $(e^{-\text{ad } y_j})^{\text{tr}}$ occurring in reverse order. This time one integrates first w.r.t. y_j for the largest j . Finally (3.12) follows by replacing ξ by ξ_j , summing over j to get the trace and then integrating over $[0, T]$. The integrand is continuous in g and t and positive so that the P and dt integrals can be interchanged.

LEMMA 3.6. *Let u and v be in $C_c^1(G^k)$. Define $f: W \rightarrow R$ by (2.6) or (2.10). Then all integrals in (3.6) are finite and (3.6) holds.*

Proof. Just as in the proof of Corollary 2.11 it suffices to consider only the case in which f is given by (2.6). Define $F_\varepsilon(g)$ as in Corollary 2.11. Since F_ε and its gradient are bounded and continuous (3.6) holds for F_ε . As $\varepsilon \downarrow 0$ the nongradient terms converge to $\int |f|^2 \ln|f| dP$ and $\|f\|_2^2 \ln\|f\|_2$ respectively. By (2.14) we have $|\nabla f(g)|^2 \leq C\varphi(g)$ which is integrable by Lemma 3.5. Hence the first term in (2.13) converges in $L^2(P)$ to ∇f as $\varepsilon \downarrow 0$. The second term in (2.13) converges pointwise to zero and is dominated by $2M|f(g)|_\varepsilon \varphi(g)(1 + \varepsilon \varphi(g))^{-2} \varphi(g)^{1/2}$ by Lemma 2.10. This is bounded by $2M|f(g)|\varphi(g)^{1/2}$ which is in $L^2(P)$ by Lemma 3.5.

Remark 3.7. Even though u and its derivatives are bounded in the preceding lemma $\nabla f(g)$ may be unbounded as we noted in Corollary 2.8.

LEMMA 3.8. *Let u and v be bounded functions in $C^1(G^k)$. Define $f: W \rightarrow R$ by (2.6) or (2.10). Then (3.6) holds.*

Proof. It suffices as before to consider only f given by (2.6). We must remove the compact support condition of Lemma 3.6. Let U be a neighborhood of e in G with compact closure such that $U^{-1} = U$. Let $\beta \in C_c^\infty(G)$ with $\text{supt } \beta \subset U$ such that $\beta \geq 0$ and $\int_G \beta(z) dz = 1$ where dz is left invari-

ant Haar measure on G . Let

$$\beta_n(x) = \int_{U^n} \beta(y^{-1}x) dy.$$

One verifies easily that (a) $\beta_n \in C_c^\infty(G)$, (b) $\beta_n(x) = 1$ on U^{n-1} , (c) $0 \leq \beta_n(x)/1$ on G as $n \rightarrow \infty$ and (d) there is a constant C such that $|\beta'_n(x)\langle \xi \rangle| \leq C|\xi|$ for all x in G and all n . Now let

$$u_n(x_1, \dots, x_k) = u(x_1, \dots, x_k) \prod_{i=1}^k \beta_n(x_i).$$

Let

$$f_n(g) = u_n(g(t_1), g(t_1)^{-1}g(t_2), \dots)$$

and put

$$B_n(g) = \prod_{i=1}^k \beta_n(g(t_{i-1})^{-1}g(t_i)).$$

Then u_n is in $C_c^1(G^k)$ and therefore by Lemma 3.6 f_n satisfies (3.6). Moreover

$$f_n(g) = f(g)B_n(g)$$

and

$$(\nabla f_n)(g) = (\nabla f)(g)B_n(g) + f(g)\nabla B_n(g).$$

Since $0 \leq B_n(g)/1$ for each g as $n \rightarrow \infty$ and f is bounded we shall obtain (3.6) for f in the limit as $n \rightarrow \infty$ if we only show that

$$\int_W |(\nabla B_n)(g)|^2 P(dg) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

But by (b) and (d) above, $|\beta'_n(x)\langle \xi \rangle| \leq C\chi_n(x)|\xi|$ where $\chi_n(x) = 0$ on U^{n-1} and is one otherwise. Hence by (2.16),

$$\begin{aligned} |(\nabla B_n)(g)|^2 &\leq \sum_{i=1}^k C \int_{t_{i-1}}^{t_i} \chi_n(g(t_{i-1})^{-1}g(t_i)) \|\text{Ad } g(t_i)^{-1}g(s)\|_{\text{op}}^2 ds \\ &\leq C \sum_{i=1}^k \int_{t_{i-1}}^{t_i} \chi_n(g(t_{i-1})^{-1}g(t_i)) \varphi(s, t_i, g) ds \end{aligned}$$

where $\varphi(a, b, g)$ denotes the square of the Hilbert-Schmidt norm of $\text{Ad } g(b)^{-1}g(a)$. Hence

$$\int_W |(\nabla B_n(g))|^2 P(dg) \leq C \sum_{i=1}^k \int_{t_{i-1}}^{t_i} \left[\int_W \chi_n(g(t_{i-1})^{-1}g(t_i)) \varphi(s, t_i, g) P(dg) \right] ds.$$

Since $\chi_n(g(t_{i-1})^{-1}g(t_i)) \rightarrow 0$ for each g in W , (3.11) shows that the P integral on the right goes to zero for each s and boundedly in s . Hence $\int |(\nabla B_n)(g)|^2 P(dg) \rightarrow 0$ as $n \rightarrow \infty$.

COROLLARY 3.9. *Let u and v be in $C^1(G^k)$. Define $f: W \rightarrow R$ by (2.6) or (2.10). Then (3.6) holds.*

Proof. As usual we need to consider only the case (2.6). Put

$$u_n(x) = n \tan^{-1}(n^{-1}u(x)).$$

Then u_n is bounded, and by Lemma 3.8 the corresponding function

$$f_n(g) = n \tan^{-1}(n^{-1}f(g))$$

satisfies (3.6). Moreover $|f_n(g)|$ increases to $|f(g)|$ and $|\nabla f_n(g)| = (1 + n^{-2}f(g)^2)^{-1}|\nabla f(g)|$ increases to $|\nabla f(g)|$. Since $t^2 \ln|t|$ is bounded below on $(0, \infty)$ we may take the limit on n to get (3.6) for f .

Remark 3.10. Just as we derived (3.6) by starting with the known logarithmic Sobolev inequality (3.7) for Gauss measure on R^n it is also possible to derive (3.6) by starting with the known finite difference version of (3.7) which is in fact the most elementary form of a logarithmic Sobolev inequality. One starts with the purely atomic Bernoulli measure of Example 3.2. We will sketch how this may be done because it will be useful in understanding the failure of our method to produce best constants.

Let $\Omega = \{1, -1\}$ and let μ be the measure on Ω which assigns weight $1/2$ to each point. If $v: \Omega \rightarrow R$ put $Av = (1/2)(v(1) - v(-1))$. We regard A as an operator on $L^2(\Omega, \mu)$. Then the inequality $\int_{\Omega} v^2 \ln|v| d\mu \leq \int (Av)^2 d\mu + \|v\|_2^2 \ln\|v\|_2$ is the simplest logarithmic Sobolev inequality [17]. It extends to product spaces [17] yielding

$$(3.14) \quad \int_{\Omega^m} v^2 \ln|v| d\mu^m \leq \int \sum_j (A_j v)^2 d\mu^m + \|v\|_2^2 \ln\|v\|_2$$

where $(A_j v)(s_1, \dots, s_m) = \frac{1}{2}(v(s)|_{s_j=1} - v(s)|_{s_j=-1})$ and $s = (s_1, \dots, s_n)$. Now put $m = kd$ with $d = \dim \mathcal{G}$ and write $s = \{s_{r,j}\}_{r=1, j=1}^{k,d}$. We embed Ω^{kd} into \mathcal{G}^k by mapping s to

$$y := \bigoplus_{r=1}^k \sum_{j=1}^d s_{r,j} \xi_j / n^{1/2}$$

with $k = [n\pi]$ as in (3.1). Then μ^{kd} clearly induces the measure λ_n^k on \mathcal{G}^k where λ is the Bernoulli measure of Example 3.2. For each fixed n we identify Ω^{kd} with its image in \mathcal{G}^k henceforth. With ψ_n defined in (3.1) and $v = f \circ \psi_n$ for some bounded function f in $C^1(W)$ the inequality (3.14) reads

$$\begin{aligned} (3.15) \quad & \int_{\mathcal{G}^k} v(y)^2 \ln|v(y)| \lambda_n^k(dy) \\ & \leq \int_{\mathcal{G}^k} \sum_{r=1}^k \sum_{j=1}^d (A_{r,j} v(y))^2 \lambda_n^k(dy) + \|v\|_2^2 \ln \|v\|_2 \end{aligned}$$

where

$$(3.16) \quad (A_{r,j} v)(y) = \frac{1}{2}(v(y)|_{y_{r,j}=n^{-1/2}} - v(y)|_{y_{r,j}=-n^{-1/2}})$$

and $y = \bigoplus_{r=1}^k \sum_{j=1}^d y_{r,j} \xi_j$ is in Ω^{kd} . The first and last terms in (3.15) converge to $\int_W f^2 \ln|f| dP$ and $\|f\|_{L^2(P)}^2 \ln \|f\|_{L^2(P)}$ respectively by Proposition 3.1. We shall show in the next proposition that the finite difference terms converge to at most $\int |\nabla f(g)|^2 P(dg)$ under suitable conditions on f , giving another derivation of (3.6).

In the following proposition the conditions on f are satisfied by functions of the form (2.6) when u has compact support in G and G is compact times abelian. This case will be the only one of interest to us.

PROPOSITION 3.11. *Assume that $f: W \rightarrow R$ is bounded and in $C^1(W)$. Define the kernel $\nabla f(g)(t)$ by*

$$(\partial_h f)(g) = \int_0^T (\nabla f(g)(t), \dot{h}(t))_{\mathcal{G}} dt,$$

and put $\|\nabla f(g)\|_\infty = \sup\{|\nabla f(g)(t)|_{\mathcal{G}}: 0 \leq t \leq T\}$. Suppose there is a constant C_1 such that $\|\nabla f(g)\|_\infty \leq C_1$ for all g in W and that the map $g \rightarrow \nabla f(g)(\cdot)$ is

continuous from W into $L^\infty([0, T]; \mathcal{S})$. Then

$$(3.17) \quad \limsup_{n \rightarrow \infty} \int_{\mathcal{S}^k} \sum_{r=1}^k \sum_{j=1}^d (A_{r,j}(f \circ \psi_n))^2 \lambda_n^k \leq \int_W |(\nabla f)(g)|^2 P(dg).$$

Proof. Write $a = n^{-1/2}$. If u is in $C^1(R)$ then

$$\begin{aligned} (u(a) - u(-a))/2 &= \frac{1}{2} \int_{-a}^a u'(t) dt = au'(a) + \frac{1}{2} \int_{-a}^a (u'(t) - u'(a)) dt \\ &= au'(a) + R(a) \end{aligned}$$

where $|R(a)| \leq a \sup\{|u'(t) - u'(a)|; |t| \leq a\}$. Similarly we may approximate $(u(a) - u(-a))/2$ by $au'(-a)$ with a similar estimate on the error. Thus

$$(u(a) - u(-a))/2 = au'(\pm a) + R(\pm a)$$

with

$$|R(\pm a)| \leq a \sup\{|u'(t) - u'(\pm a)|; |t| \leq a\}.$$

Put $u = f \circ \psi_n$. By (3.16) we therefore may write, for y in Ω^{kd} ,

$$(A_{r,j}v)(y) = a\partial v(y)/\partial y_{rj} + R_{r,j}(y)$$

where

$$|R_{r,j}(y)| \leq a \sup\{|\partial v(y')/\partial y_{r,j} - \partial v(y)/\partial y_{rj}|\}$$

and the supremum is taken over those y' in \mathcal{S}^k which agree with y except in the (r, j) th coordinate while $|y'_{r,j} - y_{r,j}| \leq 2a$. Now if $h_{r,j}(t) = n\xi_j$ for $(r-1)/n \leq t < r/n$ and is zero otherwise we put $\partial_{r,j}f(g) = n^{-1/2}\partial_{h_{r,j}}f(g)$. Since $\{n^{-1/2}h_{r,j}\}_{r=1,j=1}^{kd}$ is an O.N. set in H we have $\sum_{r,j}(\partial_{r,j}f(g))^2 \leq |\nabla f(g)|^2$ and in fact as $n \rightarrow \infty$ the left side converges to the right side because the projection Q_n onto $\text{span}\{h_{r,j}\}_{r,j}$ converges strongly to the identity operator on H as $n \rightarrow \infty$. In view of (2.16) we have

$$(A_{r,j}v)(y) = (\partial_{r,j}f)(\psi_n(y)) + R_{r,j}(y)$$

with

$$(3.18) \quad |R_{r,j}(y)| \leq \sup\{(\partial_{r,j}f)(\psi_n(y')) - (\partial_{r,j}f)(\psi_n(y))\},$$

the supremum being over the same set of y' as before. Now

$$\begin{aligned} (\partial_{r,j}f)(g) &= \int_0^T (\nabla f(g)(t), n^{-1/2} \dot{h}_{r,j}(t))_{\mathcal{S}} dt \\ &= n^{1/2} \int_{(r-1)/n}^{r/n} (\nabla f(g)(t), \xi_j) dt. \end{aligned}$$

So if g and g' are in W then

$$(3.19) \quad |\partial_{r,j}f(g)| \leq n^{-1/2} C_1$$

and

$$(3.20) \quad |(\partial_{r,j}f)(g') - (\partial_{r,j}f)(g)| \leq n^{-1/2} \|\nabla f(g') - \nabla f(g)\|_{\infty}.$$

Hence $|R_{r,j}(y)| \leq 2C_1 n^{-1/2}$ by the first of these inequalities. But

(3.21)

$$\begin{aligned} &\sum_{r,j} (A_{r,j}v)(y)^2 \\ &= \sum_{r,j} (\partial_{r,j}f(\psi_n(y)) + R_{r,j}(y))^2 \\ &\leq \sum_{r,j} (\partial_{r,j}f(\psi_n(y))^2 + 2|(\nabla f)(\psi_n(y))| \left(\sum_{r,j} R_{r,j}(y)^2 \right)^{1/2} + \sum_{r,j} R_{r,j}(y)^2 \\ &\leq |(\nabla f)(\psi_n(y))|^2 + 2|\nabla f(\psi_n(y))| (4C_1^2 kd/n)^{1/2} + (4C_1^2 kd/n). \end{aligned}$$

Since $k/n \leq T$ the left side of (3.21) is uniformly bounded in n and y by a constant C_2 . Given $\varepsilon > 0$ there is a compact set C in W such that $P_n(C) \geq 1 - \varepsilon$ for all n because the weakly convergent sequence P_n is tight. There is a compact set α in G such that $g(t)$ is in α for $0 \leq t \leq T$ whenever g is in C . Since $g \rightarrow \nabla f(g)$ is continuous on W to $L^\infty([0, T], \mathcal{S})$ and C is compact there is a neighborhood V of e in G such that $\|\nabla f(g') - \nabla f(g)\|_{\infty} < \varepsilon$ whenever g is in C and $g(t)^{-1}g'(t)$ is in V for all t in $[0, T]$. There is moreover a neighborhood V_1 of e such that

$$(h^{-1}k)^{-1}V_1(h^{-1}k) \subset V$$

for all h and k in α . There is an integer n_0 such that $\exp z_1 \exp z_2$ is in V_1 whenever $\max(|z_1|, |z_2|) \leq 2(d/n_0)^{1/2}$. Thus if $n \geq n_0$, y is in Ω^{kd} , y' differs from y only in the (r, j) th coordinate as before, with $|y'_{r,j} - y_{r,j}| \leq 2n^{-1/2}$,

and if $g = \psi_n(y)$, $g' = \psi_n(y')$ and g is in C then we see from (3.1) that $g(t) = g'(t)$ for $t \leq (r - 1)/n$, and $g(r/n)^{-1}g(t) = g'(r/n)^{-1}g'(t)$ for $t \geq r/n$, so that

$$g(t)^{-1}g'(t) = (g(r/n)^{-1}g(t))^{-1}(\exp y_r)(\exp y'_r)(g(r/n))^{-1}g(t)$$

is in V for $t \geq r/n$, and that $g(t)^{-1}g'(t)$ is in V for $(r - 1)/n \leq t \leq r/n$ also. Hence $\|\nabla f(\psi_n(y')) - \nabla f(\psi_n(y))\|_\infty < \varepsilon$. By (3.18) and (3.20) $R_{r,j}(y) \leq \varepsilon n^{-1/2}$ if $\psi(y)$ is in C and $n \geq n_0$. Apply this estimate to the second line of (3.21) to get

$$\begin{aligned} (3.22) \quad & \sum_{r,j} (A_{r,j^v})(y)^2 \\ & \leq \sum_{r,j} (\partial_{r,j} f)(\psi_n(y))^2 + 2|(\nabla f)(\psi_n(y))|(\varepsilon^2 Td)^{1/2} + \varepsilon^2 Td \\ & \leq \sum_{r,j} |(\partial_{r,j} f)(\psi_n(y))|^2 + \varepsilon C_3 \end{aligned}$$

if $\psi_n(y)$ is in C . Thus

$$\begin{aligned} \int_{\mathcal{G}^k} \sum_{r,j} (A_{r,j^v})(y)^2 \lambda_n^k(dy) & \leq \int_C \{|\nabla f(g)|^2 + \varepsilon C_3\} P_n(dg) + C_2 \varepsilon \\ & \leq \int_W |\nabla f(g)|^2 P_n(dg) + \varepsilon C_4 \end{aligned}$$

for some constants C_3 and C_4 if $n \geq n_0$. This proves Proposition 3.11. However it might be useful to point out that if one lets n run through the sequence $n = 2^m$ then $\sum_{r,j} |\partial_{r,j} f(g)|^2$ increases to $|\nabla f(g)|^2$ hence the convergence is uniform on the compact set C . By reversing the inequality (3.22) along with the sign in front of ε it then follows easily that (3.17) becomes equality and $\limsup_{n \rightarrow \infty}$ may be replaced by $\lim_{n=2^m \rightarrow \infty}$.

4. Logarithmic Sobolev inequalities on G and quotient spaces

Now we shall derive from Corollary 3.9 a logarithmic Sobolev inequality for functions on G itself with respect to the heat kernel measure. We fix $t > 0$. Let μ_t be the distribution of $g(t)$ with respect to P . Then convolution by μ_t gives the solution to the heat equation on G for the operator $\frac{1}{2} \sum_{j=1}^d \hat{\xi}_j^2$. In view of the integrability result of (3.11) there is a positive operator valued

function $A(t, x)$ which operates on \mathcal{G}^* and which is defined by the equation

$$(4.1) \quad A(t, x) = \int_W \int_0^t (\text{Ad } g(t)^{-1} g(s))^{\text{tr}*} \times (\text{Ad } g(t)^{-1} g(s))^{\text{tr}} ds P(dg | g(t) = x).$$

Note that the integrand is strictly positive and so therefore is $A(t, x)$ for a.e. $x [\mu_t]$. Explicitly $A(t, x)$ is determined a.e. $[\mu_t]$ by the equation

$$(4.2) \quad \int_G (A(t, x)\eta(x), \eta(x))_{\mathcal{G}^* \mu_t} (dx) = \int_W \int_0^t \left| (\text{Ad } g(t)^{-1} g(s))^{\text{tr}} \eta(g(t)) \right|^2 ds P(dg)$$

where $\eta: G \rightarrow \mathcal{G}^*$ runs over arbitrary bounded measurable functions. As usual the superscript tr denotes transpose. In particular

$$\int_G \text{trace } A(t, x) \mu_t(dx) < \infty.$$

THEOREM 4.1. *Fix $t > 0$. Let u be in $C^1(G)$. Then*

$$(4.3) \quad \int_G u(x)^2 \ln|u(x)| \mu_t(dx) \leq \int_G (A(t, x)u'(x), u'(x)) \mu_t(dx) + \|u\|_{L^2(\mu_t)}^2 \ln \|u\|_{L^2(\mu_t)}$$

where $u'(x)\langle \xi \rangle = du(x \exp s\xi) / ds$ at $s = 0$.

Proof. Put $f(g) = u(g(t))$. By (2.8) and the equation (4.2), which clearly extends to arbitrary measurable η by the monotone convergence theorem,

$$(4.4) \quad \int_W |(\nabla f)(g)|^2 P(dg) = \int_W \int_0^t \left| (\text{Ad } g(t)^{-1} g(s))^{\text{tr}} u'(g(t)) \right|^2 ds P(dg) = \int_G (A(t, x)u'(x), u'(x))_{\mathcal{G}^* \mu_t} (dx).$$

By Corollary 3.9, (3.6) holds and this is (4.3).

THEOREM 4.2. *Assume that $\text{Ad } G$ acts orthogonally on \mathcal{G} with respect to the given inner product. That is, $|(\text{Ad } x)\xi| = |\xi|$ for x in G and ξ in \mathcal{G} . Then for u in $C^1(G)$,*

$$(4.5) \quad \int_G u(x)^2 \ln|u(x)| \mu_t(dx) \leq t \int_G |u'(x)|^2 \mu_t(dx) + \|u\|_{L^2(\mu_t)}^2 \ln \|u\|_{L^2(\mu_t)}.$$

Proof. Since $(\text{Ad } x)^t$ is also orthogonal the right side of the first equality in (4.4) is

$$\int_W \int_0^t |u'(g(s))|^2 ds P(dg) = t \int_G |u'(x)|^2 \mu_t(dx).$$

Thus $A(t, x) = tI_{\mathcal{G}^*}$. (4.5) now follows from (4.3).

Remark 4.3. The hypothesis of Corollary 4.2 is very restrictive. The only groups whose adjoint action is orthogonal for some inner product are compact \times abelian.

Let K be a closed subgroup of G with Lie algebra $\mathcal{K} \subset \mathcal{G}$. Write $M = G/K = \{Kx: x \in G\}$ for the manifold of right K cosets. Let $\mathcal{K}_x = (\text{Ad } x^{-1})\mathcal{K}$ for x in G and denote by \mathcal{K}_x^0 the annihilator of \mathcal{K}_x in \mathcal{G}^* . If u is in $C^1(G)$ and is K invariant; i.e., $u(kx) = u(x)$ for k in K and x in G then for any ξ in \mathcal{K}_x we have $u(x \exp s\xi) = u(\{\exp s(\text{Ad } x)\xi\}x) = u(x)$ because $(\text{Ad } x)\xi$ is in \mathcal{K} . Hence $u'(x)\langle \xi \rangle = 0$. Therefore $u'(x)$ is in \mathcal{K}_x^0 . Since $\mathcal{K}_{kx} = \mathcal{K}_x$ for k in K we may identify \mathcal{K}_x^0 with the dual space to the tangent space $T_y(M)$ if $y = \pi(x)$ and we write $\mathcal{K}_{[x]}$ and $\mathcal{K}_{[x]}^0$ to emphasize dependence only on $\pi(x)$. If v is in $C^1(M)$ then its derivative $v'(y)$ takes its value in \mathcal{K}_y^0 and may be computed as $v'(y) = u'(x)$ where $u = v \circ \pi$ and $y = \pi(x)$.

We summarize in the following lemma a well known fact.

LEMMA 4.4. *Let $p(x)$ be a strictly positive continuous function on G such that $\int_G p(x) dx = 1$ where dx is right Haar measure on G . Let $\mu(dx) = p(x) dx$ and let $\nu = \mu \circ \pi^{-1}$ be the induced probability measure on M . For each point y in M there is a unique measure $\gamma(y, \cdot)$ on $\pi^{-1}(y)$ such that for any bounded continuous function u on $G \int_{\pi^{-1}(y)} u(z) \gamma(y, dz)$ is continuous in y and*

$$(4.6) \quad \int_G u(x) p(x) dx = \int_M \left(\int_{\pi^{-1}(y)} u(z) \gamma(y, dz) \right) \nu(dy).$$

We apply this Lemma to μ_t , which has a smooth density p_t with respect to Haar measure on G . We write $\nu_t = \mu_t \circ \pi^{-1}$ for the induced measure on M and γ_t in place of γ .

Define $B(t, y): \mathcal{K}_y^0 \rightarrow \mathcal{K}_y^0$ for y in M by

$$(4.7) \quad (B(t, y)\eta, \eta)_{\mathcal{S}^*} = \int_{\pi^{-1}y} (A(t, z)\eta, \eta)\gamma_t(y, dz) \quad \eta \in \mathcal{K}_y^0.$$

The integrand is nonnegative and $B(t, y)$ exists for almost all y [ν_t] since

$$(4.8) \quad \int_M (B(t, y)\eta(y), \eta(y))_{\mathcal{S}^*} \nu_t(dy) \\ = \int_G (A(t, x)\eta(x), \eta(x))\mu_t(dx) < \infty$$

in case η is a left K invariant continuous function on G with compact support with $\eta(x) \in \mathcal{K}_x^0$. That is, η is a continuous section of $T^*(M)$ with compact support.

COROLLARY 4.5. *Let v be in $C^1(M)$. Then*

$$(4.9) \quad \int_M v(y)^2 \ln|v(y)|\nu_t(dy) \leq \int_M (B(t, y)v'(y), v'(y))\nu_t(dy) \\ + \|v\|_{L^2(\nu_t)}^2 \ln\|v\|_{L^2(\nu_t)}.$$

Proof. Let $u(x) = v(\pi(x))$. Then u is in $C^1(G)$ and (4.3) holds. The first and third terms of (4.3) agree with those of (4.9). Moreover since u is constant on right K cosets and $u'(x)$ is in \mathcal{K}_x^0 for all x we have by (4.6) and (4.8),

$$\int_G (A(t, x)u'(x), u'(x))\mu_t(dx) \\ = \int_M \int_{\pi^{-1}y} (A(t, z)v'(y), v'(y))\gamma_t(y, dz)\nu_t(dx) \\ = \int_M (B(t, y)v'(y), v'(y))\nu_t(dy).$$

Remark 4.6. In case the given inner product on \mathcal{S} is $\text{Ad } G$ invariant then $A(t, x) = tI_{\mathcal{S}^*}$ as we noted in Corollary 4.2. In this case $B(y) = t\gamma(y, \pi^{-1}y)I_{\mathcal{K}_y^0}$.

5. Best constants

We will show that the constant one in front of the gradient term in (3.6) is the smallest possible. On the other hand we will see by example that the coefficients of the gradient term in (4.3) and (4.5) are unlikely to be best possible in some cases.

THEOREM 5.1. *Define P as in Theorem 3.4. Suppose that c is a constant such that*

$$(5.1) \quad \int_W f(g)^2 \ln|f(g)|P(dg) \leq c \int_W |(\nabla f)(g)|^2 P(dg) + \|f\|_{L^2(P)}^2 \ln \|f\|_{L^2(P)}$$

for all bounded continuous real valued functions f with bounded continuous gradient. Then $c \geq 1$.

LEMMA 5.2. *Let v be in $C_c(\mathcal{G})$. Suppose that the exponential map is a diffeomorphism of $B_{2a} \equiv \{y \in \mathcal{G} : |y| < 2a\}$ onto a neighborhood U of e in G and that $\varphi(\exp y) = y$ for y in B_{2a} . Let $u(t, x) = v(t^{-1/2}\varphi(x))$ if x is in U and define $u(t, x)$ to be zero otherwise. Then*

$$(5.2) \quad \lim_{t \downarrow 0} \int_W u(t, g(t))P(dg) = \int_{\mathcal{G}} v(y)\lambda(dy)$$

where λ is Gauss measure on \mathcal{G} with mean zero and variance one, given in Example 3.3.

Proof. Let ξ_1, \dots, ξ_d be an O.N. basis of \mathcal{G} . Define $\beta(y): \mathcal{G} \rightarrow \mathcal{G}$ as in Lemma 2.1 for $|y| < 2a$ and denote by $\beta_{j,k}(y)$ its matrix: $\beta(y)\xi_k = \sum_j \beta_{j,k}(y)\xi_j$. Since $\beta(0) = I_{\mathcal{G}}$ we have $\beta_{j,k}(0) = \delta_{jk}$. The left invariant vector field $\hat{\xi}_k$ is represented over U in canonical coordinates by $\sum_j \beta_{j,k}(y)\partial/\partial y_j$. Put

$$a_{ij}(y) = \sum_{k=1}^d \beta_{i,k}(y)\beta_{j,k}(y)$$

and

$$b_i(y) = \frac{1}{2} \sum_{j,k} \beta_{j,k}(y)\partial_j \beta_{i,k}(y).$$

Then a_{ij} and b_i are in $C^\infty(B_{2a})$ and $a_{ij}(0) = \delta_{ij}$. Let $\zeta \in C_c^\infty(\mathcal{G})$ with

$0 \leq \zeta \leq 1$, $\zeta = 1$ on B_a and $\zeta = 0$ outside $B_{3a/2}$. Let $\hat{a}_{ij}(y)$ be $a_{ij}(y)\zeta(y) + \delta_{ij}(1 - \zeta(y))$ on B_{2a} and equal to δ_{ij} off B_{2a} . Then \hat{a}_{ij} is bounded and smooth on \mathcal{S} and uniformly elliptic and $\hat{a}_{ij}(y) = a_{ij}(y)$ for $|y| \leq a$. Extend b_i similarly to a bounded C^∞ function \hat{b}_i on \mathcal{S} which coincides with b_i on B_a . For $\varepsilon \geq 0$ define

$$(5.3) \quad L^\varepsilon = \frac{1}{2} \sum_{i,j} \hat{a}_{ij}(\varepsilon^{1/2}y) \partial^2 / \partial y_i \partial y_j + \varepsilon^{1/2} \sum_i b_i(\varepsilon^{1/2}y) \partial / \partial y_i.$$

Then the operator $L \equiv \frac{1}{2} \sum_{k=1}^d (\hat{\xi}_k)^2$ is correctly given in canonical coordinates over B_a by the operator L^1 , as one computes readily from the previous definitions.

Let X be a Markov process on $[0, \infty)$ with state space \mathcal{S} , with continuous paths and infinitesimal generator L^1 . Define $X_\varepsilon(t) = \varepsilon^{-1/2}X(\varepsilon t)$ for $\varepsilon > 0$. Then X_ε is again a Markov process and one computes readily that its infinitesimal generator is L^ε . Now as $\varepsilon \downarrow 0$ we see from (5.3) that the coefficients of L^ε converge uniformly on bounded sets in \mathcal{S} to the corresponding coefficients of L^0 which is just the operator $\frac{1}{2} \sum_{i=1}^d \partial^2 / \partial y_i^2$. Hence if X_0 denotes the corresponding Brownian motion in \mathcal{S} beginning at the origin and if we begin X at the origin also then by [44, Theorem 11.1.4] the processes X_ε converge in distribution to X_0 . In particular if $v \in C_c(\mathcal{S})$ then $E(v(X_\varepsilon(1)))$ converges to $E(v(X_0(1)))$. That is,

$$\lim_{\varepsilon \downarrow 0} E(v(\varepsilon^{-1/2}X(\varepsilon))) = \int_{\mathcal{S}} v(y) \lambda(dy).$$

But since X has continuous sample paths and starts at the origin we have

$$\begin{aligned} & |E(v(t^{-1/2}X(t))) - E(v(t^{-1/2}X(t)), |X(s)| \leq a \text{ for } 0 \leq s \leq t)| \\ & \leq (\sup |v(y)|) E(|X(s)| > a \text{ for some } s \text{ in } [0, t]) \\ & \rightarrow 0 \text{ as } t \downarrow 0. \end{aligned}$$

Let L_D^1 be the operator L over B_a with Dirichlet boundary conditions. Then for any continuous bounded function f on B_a we have

$$E(f(X(t)), |X(s)| \leq a, 0 \leq s \leq t) = (e^{tL_D^1} f)(0).$$

Hence

$$(5.4) \quad \lim_{t \downarrow 0} (e^{tL_D^1} v(t^{-1/2} \cdot))(0) = \int_{\mathcal{S}} v(y) \lambda(dy).$$

Under the exponential map the operator $e^{tL_D^1}$ goes over to the operator e^{tL_D}

acting on $C(\bar{V})$ where \bar{V} is the closure in G of $\exp B_a$. Since $u(t, x) = v(t^{-1/2}\varphi(x))$ for x in \bar{V} the left side of (5.4) is

$$\lim_{t \downarrow 0} (e^{tL_D}u(t, \cdot))(e)$$

which is given in terms of the measure P by

$$\lim_{t \downarrow 0} E_P(u(t, g(t)), g(s) \in \bar{V}, 0 \leq s \leq t).$$

Since u is bounded and the paths $g(\cdot)$ are continuous this limit is the same as $\lim_{t \downarrow 0} E_P(u(t, g(t)))$ by the same argument used above for the process X . This proves (5.2).

LEMMA 5.3. *Define $A(t, x)$ by (4.1). Then*

$$(5.5) \quad \lim_{t \downarrow 0} \int_G \|t^{-1}A(t, x) - I\| \mu_t(dx) = 0$$

for any norm on operators on \mathcal{S}^* .

Proof. The space of symmetric operators on \mathcal{S}^* is a finite dimensional real inner product space with respect to the inner product trace AB and is spanned by the set of all one dimensional projections by the spectral theorem. Hence there is a finite set η_1, \dots, η_k of unit vectors in \mathcal{S}^* such that the corresponding set P_{η_j} of one dimensional projections span this space. Since $\text{trace}(P_{\eta}B) = (B\eta, \eta)_{\mathcal{S}^*}$, $\|B\| := \sum_{j=1}^k |(B\eta_j, \eta_j)_{\mathcal{S}^*}|$ is a norm on symmetric operators and is equivalent to any other one. It suffices therefore to show that for any vector η in \mathcal{S}^* ,

$$\lim_{t \downarrow 0} \int_G |((t^{-1}A(t, x) - I)\eta, \eta)| \mu_t(dx) = 0.$$

Choose a sequence t_n which converges to zero and put

$$f_n(g) = t^{-1} \int_0^t \left| \left(\text{Ad}(g(t)^{-1}g(s)) \right)^{\text{tr}} \eta \right|^2 ds$$

with $t = t_n$. Put $f(g) = (I\eta, \eta) = 1$. In view of (4.1) we must prove that $\int_W |f_n(g) - f(g)| P(dg) \rightarrow 0$. Note that $0 \leq f_n(g) \rightarrow f(g)$ for each g because $g(\cdot)$ is continuous. Moreover

$$\int f_n(g) P(dg) \leq t_n^{-1} \int_0^{t_n} e^{\gamma(t_n-s)} ds$$

by inequality (3.11). Thus

$$\limsup \int f_n(g)P(dg) \leq 1 = \int f(g)P(dg).$$

But if $0 \leq a < b$ then $|a - b| = a - b + 2|a - b|$. Hence if

$$S_n = \{g: f_n(g) < f(g)\}$$

then

$$\int |f_n - f| = \int f_n - f + 2 \int (f(g) - f_n(g))C_{S_n}(g).$$

The second integrand is positive, dominated by f and goes to zero pointwise. Hence the second integral goes to zero. But as

$$\limsup \int (f_n - f) \leq 0$$

the lemma follows.

Proof of Theorem 5.1. Assume that (5.1) holds. Then Theorem 4.1 holds also with the constant c appearing in front of the gradient term. Let v be in $C_c^\infty(\mathcal{G})$ and define $u(t, x)$ as in Lemma 5.2. Note that if $v(y) = 0$ for $|y| > r$ and t_0 is so small that $t_0^{1/2}r < a$ then for $0 < t \leq t_0$, $v(t^{-1/2}y) = 0$ if $|y| > a$. Consequently $u(t, \cdot)$ is in $C_c^\infty(G)$ if $0 < t \leq t_0$ and is supported in $V := \exp B_a$. Thus by (4.3) we have

$$\begin{aligned} (5.6) \quad & \int_G u(t, x)^2 \ln|u(t, x)|\mu_t(dx) \\ & \leq c \int_G (A(t, x)u'(t, x), u'(t, x))\mu_t(dx) \\ & \quad + \|u(t, \cdot)\|_{L^2(\mu_t)}^2 \ln\|u(t, \cdot)\|_{L^2(\mu_t)}, \end{aligned}$$

if $t \leq t_0$. By Lemma 5.2 the non gradient terms converge as $t \downarrow 0$ to

$$\int_{\mathcal{G}} v(y)^2 \ln|v(y)|\lambda(dy) \quad \text{and} \quad \|v\|_{L^2(\lambda)}^2 \ln\|v\|_{L^2(\lambda)}$$

respectively. We now consider the gradient term as $t \downarrow 0$. For $t \leq t_0$ we may compute $u'(t, x)$ in local coordinates using the notation of the proof of

Lemma 5.2. Thus

$$\begin{aligned} \hat{\xi}_k u(t, \exp y) &= \sum_j \beta_{j,k}(y) (\partial/\partial y_j) v(t^{-1/2}y) \\ &= t^{-1/2} \sum_j \beta_{j,k}(y) (\partial v/\partial y_j)(t^{-1/2}y) \end{aligned}$$

for $|y| < 2a$. If $\{\eta_k\}_{k=1}^d$ denotes the basis of \mathcal{S}^* dual to ξ_1, \dots, ξ_d put

$$w(t, y) = \sum_{j,k} \beta_{j,k}(y) (\partial v/\partial y_j)(t^{-1/2}y) \eta_k.$$

Then $u'(t, \exp y)$ is $t^{-1/2}w(t, y)$ for $|y| < 2a$ and is zero if $|y| > a$. Furthermore putting $(\nabla v)(t^{-1/2}y) = \sum_k (\partial v/\partial y_k)(t^{-1/2}y) \eta_k$ we see that $|w(t, y) - (\nabla v)(t^{-1/2}y)|$ goes to zero uniformly in y as $t \downarrow 0$ because both terms are zero if $t^{-1/2}|y| > r$ while for $|y| \leq t^{1/2}r$, $\beta_{j,k}(y) - \delta_{j,k}$ is small if t is small, since $\beta_{j,k}(0) = \delta_{j,k}$, as noted in the proof of Lemma 5.2. Clearly $|(\nabla v)(t^{-1/2}y)|$ and hence $|w(t, y)|$ are both bounded on $(0, t_0] \times \mathcal{S}$. Therefore writing $\varphi(\exp y) = y$ for $|y| < 2a$ we have

$$\begin{aligned} &\int_G (A(t, x)u'(t, x), u'(t, x)) \mu_t(dx) \\ &= \int_V (A(t, x)t^{-1/2}w(t, \varphi(x)), t^{-1/2}w(t, \varphi(x))) \mu_t(dx) \\ &= \int_V ((t^{-1}A(t, x) - I)w(t, \varphi(x)), w(t, \varphi(x))) \mu_t(dx) \\ &\quad + \int_V |w(t, \varphi(x))|^2 \mu_t(dx). \end{aligned}$$

The first integral on the right goes to zero as $t \downarrow 0$ by Lemma 5.3 because $w(t, \varphi(x))$ is uniformly bounded. The second integrand may be replaced by $|(\nabla v)(t^{-1/2}\varphi(x))|^2$ in the limit $t \downarrow 0$ because the difference goes to zero uniformly for x in V . But by Lemma 5.2,

$$\lim_{t \downarrow 0} \int_V |(\nabla v)(t^{-1/2}\varphi(x))|^2 \mu_t(dx) = \int_{\mathcal{S}} |(\nabla v)(y)|^2 \lambda(dy).$$

Thus in the limit (5.6) becomes

$$\int_{\mathcal{S}} |v(y)|^2 \ln|v(y)| \lambda(dy) \leq c \int_{\mathcal{S}} |(\nabla v)(y)|^2 \lambda(dy) + \|v\|^2 \ln\|v\|.$$

But it is known [17, Remark 3.4] that the smallest value of c for which the

Gaussian logarithmic Sobolev inequality holds is $c = 1$. This proves the theorem.

COROLLARY 5.4. *Assume the hypothesis of Theorem 4.2 and assume also that the inequality (4.5) holds with the coefficient t in front of the derivative term replaced by a constant $c(t)$ which is at most t . Then $\lim_{t \downarrow 0} c(t)/t = 1$.*

Proof. In the notation of the proof of Theorem 5.1 put $u(x) = u(t, x)$ in the assumed inequality to get

$$\int_G |u(t, x)|^2 \ln|u(t, x)| \mu_t(dx) \leq c(t) \int_G |u'(t, x)|^2 \mu_t(dx) + \|u\|_t^2 \ln\|u\|_t$$

where $\|u\|_t$ denotes the $L^2(\mu_t)$ norm. Since $A(t, x) = tI_{\mathcal{G}^*}$, the proof of Theorem 5.1 shows that

$$\lim_{t \downarrow 0} t \int |u'(t, x)|^2 \mu_t(dx) = \int |\nabla v(y)|^2 \lambda(dy).$$

Hence

$$\begin{aligned} & \int_{\mathcal{G}} |v(y)|^2 \ln|v(y)| \lambda(dy) \\ & \leq (\liminf c(t)/t) \int_{\mathcal{G}} |\nabla v(y)|^2 \lambda(dy) + \|v\|^2 \ln\|v\| \end{aligned}$$

for all v in $C_c^\infty(\mathcal{G})$. As noted in the proof of the theorem this implies that $\liminf c(t)/t \geq 1$. But $c(t)/t \leq 1$ by assumption.

Remark 5.5. The preceding corollary is of interest only if G is a compact group because if $G = \text{compact times } R^m$ then the validity of (4.5) with t replaced by $c(t)$ for functions which depend only on the R^m coordinates forces $c(t) \geq t$ since μ_t has a Gaussian factor.

Remark 5.6. The measure P on W is induced by “wrapping a \mathcal{G} valued Wiener process around G ” (cf. McKean [29] and Elworthy [11]) as we noted in Example 3.3. Indeed the basic Theorem 3.4 may be regarded informally as following from the pulling back of the terms in (3.6) to the path space of an ordinary \mathcal{G} valued Wiener process on which logarithmic Sobolev inequalities are known. However the discontinuity of this map leads to technical difficulties, some of which have been studied in [7]. The proof of Theorem 5.1 is based on the fact that for very small t the Wiener process does not wrap very far around G . But for a fixed $t > 0$ the Wiener process may wrap around G many times (e.g., if G is compact) with the result that the constants in (4.3)

and (4.5) will in some cases not be the smallest possible. Indeed if G is compact and t is large then μ_t will be a very good approximation to Haar measure and the large constant t in (4.5) is hardly likely to be best possible. The next example illustrates the mechanism by which the best constant in front of the gradient term in (4.5) may be reduced because the exponential map is not one to one. We shall first give the example, in which $G = S^1$, and then contrast it with the case $G = R$.

Example 5.7 (Periodic Bernoulli algebra). We take $G = S^1$ and study in detail the discrete approximation to the heat kernel on S^1 described in Remark 3.10. We take $T = 1$ and study the process only at time $t = 1$. Moreover we analyze only the case $n = 3$, which gives a very simple but interesting “approximation” to the heat kernel on S^1 . Let $\Omega = \{1, -1\}$ as in Remark 3.10. Identify the Lie algebra \mathcal{L} of the circle with $\sqrt{-1}R$. The exponential map is $i\theta \rightarrow e^{i\theta}$. We take for a basis of \mathcal{L} the element $\xi = i\pi 3^{-1/2}$. In the notation of Section 3 we wish to take $k = n = 3$. The measure $\lambda_n^k = \lambda_3^3$ on \mathcal{L}^3 is then supported on the points

$$y = (s_1\xi \oplus s_2\xi \oplus s_3\xi)/3^{1/2}$$

where each $s_r = \pm 1$ for $r = 1, 2, 3$ in accordance with Remark 3.10. λ_3^3 assigns equal weight to each of these eight points. At time $t = 1$ the map $\psi \equiv \psi_3$ is given (cf. 3.1)) by

$$\psi_3(y)(1) = \exp(s_1\xi/3^{1/2})\exp(s_2\xi/3^{1/2})\exp(s_3\xi/3^{1/2}) = e^{i\pi(s_1+s_2+s_3)/3}.$$

Let $x = s_1 + s_2 + s_3$. Then x is a function on Ω^3 which takes all values in the set $S = \{3, 1, -1, -3\}$. If φ is a function on S^1 then $y \rightarrow \varphi(\psi(y)(1))$ is a function on Ω^3 which depends only on x : $f(x) = \varphi(e^{i\pi x/3})$. Moreover f is a periodic function on S in the sense that $f(3) = f(-3)$. Let \mathcal{A} be the four dimensional algebra of functions on Ω^3 generated by x . Let \mathcal{A}_p be the three dimensional subalgebra of \mathcal{A} consisting of the functions $f(x)$ which are periodic. I.e., $f(3) = f(-3)$. We write $E(f)$ for the expectation of f over (Ω^3, μ^3) and $(f, g) = E(fg)$. The Dirichlet form operators D and N associated to the periodic and nonperiodic cases respectively are the operators on the real Hilbert space \mathcal{A}_p and \mathcal{A} defined by the equations

$$(5.7) \quad (Df, g) = \sum_{r=1}^3 (A_r f, A_r g), \quad f, g \in \mathcal{A}_p$$

and $(Nf, g) = \sum_{r=1}^3 (A_r f, A_r g)$, $f, g \in \mathcal{A}$. The operator N is very well understood with respect to its spectrum and hypercontractivity properties. Just as D is the $n = 3$ approximation to the Dirichlet form operator for the

heat kernel on the circle so also N is the $n = 3$ approximation to the Dirichlet form operator for the heat kernel (Gauss measure) on the line. In the next proposition we analyze D and then we shall contrast N and D with respect to their spectral and hypercontractivity properties.

- PROPOSITION 5.8. (a) spectrum $D = \{0, 2, 2\frac{1}{2}\}$
 (b) *There is a constant c such that*

$$(5.8) \quad E(f^2 \ln|f|) \leq c(Df, f) + \|f\|_2^2 \ln\|f\|_2, \quad f \in \mathcal{A}_P.$$

The smallest such constant lies in the interval (.532, .602).

- (c) *The operator e^{-tD} satisfies*

$$(5.9) \quad \|e^{-tD}\|_{L^2 \rightarrow L^4} \leq 1 \quad \text{if and only if} \quad e^{-2t} \leq e^{-2t_0}$$

where $e^{-2t_0} = .55$ (approximately).

Proof. We shall merely sketch the tedious computational proof with a view toward explaining the origin of the above constants. Let $u = x^2 - 3$ and $v = (x^2 - 9)x$. Then $1, u, v$ form an orthogonal basis for \mathcal{A}_P and from (5.7) one can compute that $D1 = 0, Du = 2u$ and $Dv = (2\frac{1}{2})v$. This proves (a). To prove (c) it is convenient to compute L^4 norms in terms of $1, u$ and v and their products, for which the following algebraic relations are useful and straightforward to verify:

$$\begin{aligned} x^2 &= 3 + 2(s_1s_2 + s_2s_3 + s_3s_1), & x^3 &= 7x + 6s_1s_2s_3, & x^4 &= 10x^2 - 9, \\ u^2 &= 4(u + 3) = 4x^2, & v^2 &= 8(6 - u) = 8(9 - x^2), & uv &= -2v. \end{aligned}$$

Moreover $E(x^m) = 0$ if m is odd while $E(x^2) = 3$. $E(u) = E(v) = E(uv) = 0$ while $E(u^2) = 12$ and $E(v^2) = 48$. Let $f(x) = 1 + au + bv$ where a and b are real constants. Then

$$e^{-tD}f = 1 + e^{-2t}au + e^{-(2\frac{1}{2})t}bv.$$

We wish to find the smallest positive number t for which

$$\|e^{-tD}f\|_4^4 \leq \|f\|_2^4$$

for all real a and b . Put $\alpha = e^{-t/2}$. A straightforward computation shows that

$$\|f\|_2^4 - \|e^{-tD}f\|_4^4 = 24a^2A + 48 \cdot 16b^4B + 96b^2C$$

where

$$A = (6 - 14\alpha^{16})a^2 - 8\alpha^{12}a + (1 - 3\alpha^8), \quad B = 3 - 4\alpha^{20}$$

and

$$C = 12(1 - \alpha^{18})a^2 + 12\alpha^{14}a + (1 - 3\alpha^{10}).$$

If we put $b = 0$ we see that we must have $A \geq 0$ for all real a . This can happen if and only if the discriminant of this quadratic form is nonpositive. Put $s = \alpha^8$ and let $p(s) = 13s^3 - 7s^2 - 9s + 3$. The condition of nonpositive discriminant is $p(s) \geq 0$. Standard analysis of this cubic polynomial shows that $p(s)$ has exactly one zero on the interval $[0, 1)$ which is the only range of interest to us. Denote this zero by s_0 . Then for $0 \leq s < 1$ we have $p(s) \geq 0$ iff $0 \leq s \leq s_0$. Since $p(.3) > 0$ and $p(.31) < 0$ we have $.3 < s_0 < .31$. Define t_0 by $e^{-4t_0} = s_0$. Then $(.3)^{1/2} < e^{-2t_0} < (.31)^{1/2}$. Hence $.5477 < e^{-2t_0} < .557$. That is, $e^{-2t_0} = .55$, approximately. We have shown that $A \geq 0$ for all a if and only if $e^{-2t} \leq e^{-2t_0}$. To complete the proof of (c) assume $\alpha^8 \leq s_0$. Then $\alpha^8 \leq .31$. One can compute then that $B \geq 0$ while the discriminant of C , which is $48\{3\alpha^{10} + \alpha^{18} - 1\}$ is negative. This proves (c). To prove (b) we first use the interpolation argument from [17, Example 2] which shows that if $\|e^{-t_0 D}\|_{L^2 \rightarrow L^4} \leq 1$ then (5.8) holds for some constant $c \leq 2t_0$. On the other hand if (5.8) holds for some constant c then the argument in [17, Theorem 6] may be applied because D is a (discrete) Dirichlet form operator on an algebra of functions and [17, Theorem 3, Case 2] may be used. From this theorem we may conclude that $\|e^{-tD}\|_{L^2 \rightarrow L^4} \leq 1$ if $e^{-2t/c} \leq 1/3$. But since $\|e^{-tD}\|_{L^2 \rightarrow L^4} > 1$ if $t < t_0$ it follows that $e^{-2t_0/c} \geq 1/3$. Hence $c \geq 2t_0/\ln 3$. Thus the smallest constant c for which (5.9) holds satisfies $2t_0/\ln 3 \leq c_0 \leq 2t_0$. But since $\ln(.31)^{-1/2} < 2t_0 < \ln(.3)^{-1/2}$ we have

$$(-1/2)(\ln .31)/\ln 3 < c_0 < -(1/2)\ln .3,$$

which implies the assertion (b) and proves Proposition 5.6.

Remark 5.9. The example illustrates three points. Suppose that ν is a probability measure on a Lie group G and L is a self-adjoint operator on $L^2(\nu)$ with $C_c^2(G)$ as a core such that

$$(Lf, g)_{L^2(\nu)} = \sum_{j=1}^d \int_G (\hat{\xi}_j f)(x) (\hat{\xi}_j g)(x) \nu(dx)$$

for f and g in C_c^2 . Suppose further that there is a constant c such that

$$(5.10) \quad \int f^2 \ln|f| \, d\nu \leq c(Lf, f) + \|f\|_2^2 \ln\|f\|_2$$

and that c_0 is the smallest such constant. The mass gap of L is by definition the infimum m of the spectrum of L in the subspace orthogonal to the constant functions. One has always $mc_0 \geq 1$. (See e.g. [39], [42]. One need only put $f(x) = e^{sg(x)}$, $g \in C_c^2(G)$ in (5.10) and compare second order terms in s .) Moreover $mc_0 = 1$ for Gauss measure on R as well as for the discrete Dirichlet form operator N defined after equation (5.7). N is a well understood operator for all n [17, 4] with spectrum $N = \{0, 1, 2, \dots, n-1\}$ and $c_0 = 1$. The Gaussian case on the line follows from this via the central limit theorem argument of Remark 3.10 above, which was used in [17]. Thus the Gaussian case on R reflects the discrete case and in view of Remark 3.10 and Proposition 3.11 we may expect that the behavior of c_0 for the circle reflects Example 5.5. Since the heat kernel measure for S^1 is the image of Gauss measure on R under the exponential map it is easy to see that the mass gap for L on S^1 is strictly larger than that for L on R because the second lowest eigenfunction for L on R is the linear coordinate function on R and this is not periodic. But it is not a priori clear that the Sobolev coefficient c_0 for the heat kernel measure on S^1 is strictly less than that for R . It is no bigger, in any case, by Corollary 4.2. Example 5.7 shows that in the discrete approximation ($n = 3$) to the heat kernel on the circle (a) the mass gap increases ($m = 1$ for N but $m = 2$ for D) (b) the Sobolev constant decreases ($c_0 = 1$ for N but c_0 lies in $(.532, .602)$ for D). In particular $mc_0 > 1$ for D . This also follows from Rothaus' Lemma [39, second Lemma p. 105] since $E(u^3) = 48 \neq 0$. Finally part (c) in Proposition 5.8 shows that e^{-tD} is a more regularizing operator than e^{-tN} in the sense that e^{-tD} is a contraction from L^2 to L^4 for smaller values of t since $\|e^{-tN}\|_{L^2 \rightarrow L^4} \leq 1$ if and only if $e^{-2t} \leq 1/3$ [17]. This is clearly related to and perhaps presages the fact that on the circle e^{-tL} is bounded from L^2 to L^4 (even to L^∞) for all t whereas on the line e^{-tL} is unbounded from L^2 to L^4 if $e^{-2t} > 1/3$ [33].

REFERENCES

1. H. AIRAULT and P. MALLIAVIN, *Integration geometrique sur l'espace de Wiener*, Bull. Sci. Math., vol. 112 (1988), pp. 3–52.
2. J. ALVAREZ, *The Riesz decomposition theorem for distributions on a Wiener space*, Cornell Ph.D. Thesis, 1973 (unpublished).
3. P. BAXENDALE, *Wiener processes on manifolds of maps*, Proc. Roy. Soc. Edinburgh, vol. 87A (1980), pp. 127–152.
4. W. BECKNER, *Inequalities in Fourier Analysis*, Ann. of Math., vol. 102 (1975), pp. 159–182.
5. Y.M. BEREZANSKI and Y.G. KONDRATEV, *Spectral methods in infinite dimensional analysis*. Scientific Thought, Kiev, 1988.
6. P. BILLINGSLEY, *Convergence of probability measures*, Wiley, New York, 1968.

7. J.M. BISMUT, *Large deviations and the Malliavin calculus*, Birkhäuser, Boston, 1984.
8. R.H. CAMERON, *The first variation of an indefinite Wiener integral*, Proc. Amer. Math. Soc., vol. 2 (1951), 914–924.
9. E.B. DAVIES, L. GROSS and B. SIMON, *Hypercontractivity: a bibliographic review*, Proc. Hoegh-Krohn Memorial Conference, S. Albeverio (Editor), to appear.
10. C. ELSON, *An extension of Weyl's Lemma to infinite dimensions*, Trans. Amer. Math. Soc., vol. 194 (1974), pp. 301–324.
11. K.D. ELWORTHY, *Stochastic differential equations on manifolds*, London Math. Soc. Lecture Note Series 70, Cambridge Univ. Press, Cambridge, England, 1982.
12. V. GOODMAN, *Quasi-differentiable functions on Banach spaces*, Proc. Amer. Math. Soc., vol. 30 (1971), pp. 367–370.
13. ———, *A divergence theorem for Hilbert space*, Trans. Amer. Math. Soc., vol. 164 (1972), pp. 411–426.
14. ———, *Harmonic functions on Hilbert space*, J. Funct. Anal., vol. 10 (1972), pp. 451–470.
15. ———, *A Liouville theorem for abstract Wiener spaces*, Amer. J. Math., vol. 45 (1973), pp. 215–220.
16. L. GROSS, *Potential theory on Hilbert space*, J. Funct. Anal., vol. 1 (1967), pp. 123–181.
17. ———, *Logarithmic Sobolev inequalities*, Amer. J. Math., vol. 97 (1975), pp. 1061–1083.
18. ———, “Logarithmic Sobolev inequalities for the heat kernel on a Lie group” in *White noise analysis, mathematics and applications*, 1989 Bielefeld Conference, Editors T. Hida, Kuo, Potthoff, Streit, World Scientific, New Jersey, 1990, pp. 108–130.
19. ———, *Logarithmic Sobolev inequalities on loop groups*, J. Funct. Anal., vol. 102 (1991), pp. 268–313.
20. M. KREE, *Propriété de trace en dimension infinie, d'espaces du type Sobolev*, Bull. Soc. Math. France, vol. 105 (1977), pp. 141–163.
21. M. KREE and P. KREE, *Continuity of la divergence dans les espaces de Sobolev relatifs a l'espace de Wiener*, Note C.R.A.S., vol. 296 (1983), pp. 833–836.
22. H.H. KUO, *Integration by parts for abstract Wiener measures*, Duke Math. J., vol. 41 (1974), pp. 373–379.
23. ———, *On Gross differentiation on Banach spaces*, Pacific J. Math., vol. 59 (1975), pp. 135–145.
24. ———, *Gaussian measures in Banach spaces*, Lecture Notes in Math., vol. 463, Springer-Verlag, New York, 1975.
25. ———, *Potential theory associated with Ornstein-Uhlenbeck process*, J. Funct. Anal., vol. 21 (1976), pp. 63–75.
26. ———, *Ornstein-Uhlenbeck process on a Riemann-Wiener manifold*, Proc. Symposium on Stochastic Differential Equations, Kyoto 1976, K. Ito (Editor), pp. 187–193.
27. P. MALLIAVIN, *Stochastic calculus of variations and hypoelliptic operators*, Proc. Symposium on Stochastic Differential Equations, Kyoto 1976, K. Ito (Editor), Kinokuniya-Wiley, Tokyo, 1978, pp. 195–263.
28. ———, *Geometrie differentielle stochastique*, Presses de l'Univ. de Montreal, Montreal, Canada, 1978.
29. H.P. MCKEAN, JR., *Stochastic integrals*, Academic Press, New York, 1969.
30. PAUL MEYER, *Some analytical results on the Ornstein-Uhlenbeck semigroup in infinitely many dimensions*, IFIP Working Conference on Theory and Applications of Random Fields, Bangalore, Jan. 4–9, 1982 preprint (10 pages)
31. ———, *Transformations de Riesz pour les lois Gaussiennes*, Seminaire de Probabilités XVIII, Lecture Notes in Math., vol. 1059, Springer-Verlag, New York, 1984, pp. 179–193.
32. E. NELSON, “A quartic interaction in two dimensions” in *Mathematical theory of elementary particles*, (R. Goodman and I. Segal, Editors) M.I.T. Press, Cambridge, Mass., 1966, pp. 69–73.
33. ———, *The free Markoff field*, J. Funct. Anal. vol. 12 (1973), pp. 211–227.

34. G.C. PAPANICOLAOU and S.R.S. VARADHAN, *A limit theorem with strong mixing in Banach space and two applications to stochastic differential equations*, *Comm. Pure Appl. Math.*, vol. 26 (1973), pp. 497–524.
35. K.R. PARTHASARATHY, *Probability measures on metric spaces*, Academic Press, New York, 1967.
36. M.A. PIECH, *A fundamental solution of the parabolic equation on Hilbert space*, *J. Funct. Anal.*, vol. 3 (1969), pp. 85–114.
37. _____, *Diffusion semigroups on abstract Wiener space*, *Trans. Amer. Math. Soc.*, vol. 166 (1972), pp. 411–430.
38. _____, *The Ornstein-Uhlenbeck semigroup in an infinite dimensional L^2 setting*, *J. Funct. Anal.*, vol. 18 (1975), pp. 271–285.
39. O.S. ROTHBAUS, *Diffusion on compact Riemannian manifolds and logarithmic Sobolev inequalities*, *J. Funct. Anal.*, vol. 42 (1981), pp. 102–109.
40. I.E. SEGAL, *Tensor algebras over Hilbert spaces*, *Trans. Amer. Math. Soc.*, vol. 81 (1956), pp. 106–134.
41. _____, *Distributions in Hilbert space and canonical systems of operators*, *Trans. Amer. Math. Soc.*, vol. 88 (1958), pp. 12–41.
42. B. SIMON, *A remark on Nelson's best hypercontractive estimates*, *Proc. Amer. Math. Soc.*, vol. 55 (1976), pp. 376–378.
43. D.W. STROOCK and S.R.S. VARADHAN, *Limit theorems for random walks on Lie groups*, *Sankhyā*, vol. 35 (1973), pp. 277–294.
44. _____, *Multidimensional diffusion processes*, Springer-Verlag, New York, 1979.
45. S. WATANABE, *Lectures on stochastic differential equations and Malliavin calculus*, Tata Institute of Fundamental Research, Springer-Verlag, New York, 1984.

CORNELL UNIVERSITY
ITHACA, NEW YORK