JOURNAL OF FUNCTIONAL ANALYSIS **135**, 39–49 (1996)
Article No. 0002

A Functional Form of the Isoperimetric Inequality for the Gaussian Measure*

S. Bobkov

Department of Mathematics, Syktyvkar University, 167001 Syktyvkar, Russia; and Center for Stochastic Processes, Department of Statistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599

Received July 27, 1993

Let g be a smooth function on $\mathbb{R}^n$ with values in $[0, 1]$. Using the isoperimetric property of the Gaussian measure, it is proved that $\phi(\Phi^{-1}(\mathbf{E}g)) - \mathbf{E}\phi(\Phi^{-1}(g)) \leqslant \mathbf{E}\,|\nabla g|$. Conversely, this inequality implies the isoperimetric property of the Gaussian measure. © 1996 Academic Press, Inc.

The isoperimetric property of the Gaussian measure states ([3, 16]) that for any Borel measurable set $A \subset \mathbb{R}^n$ of measure $\gamma_n(A) = p$ and for all $h > 0$,

$$\gamma_n(A^h) \geqslant \Phi(\Phi^{-1}(p) + h). \tag{1}$$

Here γ_n is the standard Gaussian measure in $\mathbb{R}^n$, of density $d\gamma_n(x) = \prod_{k=1}^n \phi(x_k)\, dx_k$, $x = (x_1, ..., x_n) \in \mathbb{R}^n$, $\phi(x_k) = 1/\sqrt{2\pi} \exp(-x_k^2/2)$, Φ^{-1} is the inverse of the distribution function Φ of γ_1, and $A^h = \{x \in \mathbb{R}^n$: $|x - a| < h$ for some $a \in A\}$ denotes the open h-neighborhood of A. (1) becomes identity for all half-spaces A of measure p.

In these notes we suggest an equivalent analytic form for (1) involving a relation between smooth functions and their derivatives. Relations of such type are well-known for Lebesgue measure (see e.g. [12], Section 3); the Sobolev unequality, for example, provides an equivalent form for the isoperimetric property of balls in the Euclidean space. There is a number of inequalities for the Gaussian measure like Poincaré-type or logarithmic Sobolev-type inequalities which can be seen as different versions of so-called "concentration of (Gaussian) measure phenomenon." A question of

* Partially supported by the National Science Foundation and the Air Force Office of Scientific Research Grant 91-0030, the Army Research Office Grant DAAL03-92-G-0008, Grant 93-011-1454 of Russian Foundation for Scientific Research and Grant NXZ000 from the International Science Foundation.

0022-1236/96 $12.00

interest is whether or not analytic inequalities can contain the isoperimetric inequality (1) (or, its equivalent) as a partial case; the answer is positive.

THEOREM. *For any smooth function g on $\mathbb{R}^n$ with values in $[0, 1]$,*

$$\varphi(\Phi^{-1}(\mathbf{E}g)) - \mathbf{E}\varphi(\Phi^{-1}(g)) \leqslant \mathbf{E}\,|\nabla g|. \tag{2}$$

Conversely, (2) *implies the isoperimetric property of the Gaussian measure.*

As usual, ∇g denotes gradient of g, and mathematical expectations in (2) are understood with respect to measure γ_n.

Remark 1. The function $I(p) = \phi(\Phi^{-1}(p))$ is called the isoperimetric function of γ_n, in the sense that the minimal value of "surface Gaussian measure"

$$\gamma_n^+(A) = \liminf_{\varepsilon \to 0} \frac{\gamma_n(A^\varepsilon) - \gamma_n(A)}{\varepsilon},$$

while $\gamma_n(A) = p$ is fixed, is equal to $I(p)$. This property, i.e., the inequality

$$\gamma_n^+(A) \geqslant I(p), \tag{3}$$

represents a differential analog of (1): it follows from (1) by taking lim inf as $h \to 0$. Conversely, integrating (3) over h gives (1). Let us note that the inequality (2) being applied to indicator (characteristic) functions $g = \chi_A$ becomes (3); it is in this sense we say that (2) contains (1).

Remark 2. The function I is concave and continuous on $[0, 1]$, so the left-hand side of (2) is non-negative by Jensen's inequality. Asymptotically, as $p \to 0$, $I(p)$ is equivalent to the function $p\sqrt{2\log(1/p)}$, and (2) has thus resemblance with the Gross' logarithmic inequality [7]

$$\mathbf{E}g^2 \log g^2 - \mathbf{E}g^2 \log \mathbf{E}g^2 \leqslant 2\mathbf{E}\,|\nabla g|^2 \tag{4}$$

which holds true for all smooth g without boundary conditions. It is shown in Ledoux [11, pp. 98–100] how (4) can be derived from (1), although the original proof of (4) was independent on isoperimetry. Apparently, one can not deduce the isoperimetric inequality from (4) (as well as from the Gaussian Poincaré inequality $\mathbf{E}g^2 - (\mathbf{E}g)^2 \leqslant \mathbf{E}\,|\nabla g|^2$ which is weaker than (4)) since extremal functions in (4) are exponential (respectively, linear) but not indicator. However, one can deduce from (4) a concentration inequality (isoperimetric in nature) which is very close to (1). First, (4) implies a deviation inequality

$$\mathbf{E}\exp(\lambda(g - \mathbf{E}g)) \leqslant \exp(\lambda^2/2), \tag{5}$$

valid for all Lipschitz g with Lipschitz seminorm $\leqslant 1$ and for all real λ, by applying (4) to functions $\exp(\lambda g/2)$ and further integrating a differential inequality ([11, p. 101], see also [4] and [1] for a similar application of (4)). A different proof of (5) which is of interest itself was found by Ledoux [10]; with a worse constant in the exponent, (5) is a partial case of a more general inequality due to Pisier [14]. Now, it follows from (5) by Chebyshev's inequality that, for all $h > 0$,

$$\gamma_n\{g - \mathbf{E}g \geqslant h\} \leqslant \exp(-h^2/2). \tag{6}$$

Take $g(x) = dist(x, A)$ the shortest distance from the point x to a closed set A of γ_n-measure $p \in (0, 1)$. Since $g \geqslant 0$, $\gamma_n\{g = 0\} = p$ and $\mathbf{E}g^2 - (\mathbf{E}g)^2 \leqslant 1$ (the last is due to the above-mentioned Poincaré-type inequality), we easily have $\mathbf{E}g \leqslant c_p = \sqrt{(1-p)/p}$. At last, noting that $\{g - \mathbf{E}g < h\} = A^{h + \mathbf{E}g}$, we obtain from (6)

$$\gamma_n(A^h) \geqslant 1 - \exp\left\{-\frac{(h - c_p)^2}{2}\right\}, \qquad h \geqslant c_p. \tag{7}$$

The same argument of getting (7) from (5) with other estimation of $\mathbf{E}g$ but similar right-hand side in (7) was used in [11, pp. 20–21]. (7) can replace (1) in many applications where large values of h are important but for h small it becomes useless and, even, does not imply a weaker form of (3), for example, the inequality

$$\gamma_n^+(A) \geqslant cI(p) \tag{8}$$

with some numerical constant $c \in (0, 1)$. Regardless isoperimetry, Ledoux [11] proved (8) for some universal $c \in (0, 1)$ with the help of a semigroup technique; he also pointed out there that it does not seem likely to reach in this way the value $c = 1$. It might be worthwile noting that (8) can also be derived from a Talagrand's logarithmic inequality [17] on the discrete cube–discrete analog of (8) for multidimensional Bernoulli distribution, simply applying the central limit theorem as it was done in the proof of (4) in [8]. However, it is still an open problem to know how to prove (8) with $c = 1$ analytically. Note also, that (8) in the integral form can be written equivalently as

$$\gamma_n(A^h) \geqslant \Phi(\Phi^{-1}(p) + ch), \qquad h \geqslant 0,$$

hence, for h large, this inequality becomes worse than (7) in case $c < 1$.

The inequality (8) and even (3) can be contained in some analytic inequalities that involve the first power of the modulus of gradient and where indicator functions are extremal. The inequality (2) gives an example but there are other ones. Consider the inequality of the form

$$c \, \|g - \mathbf{E}g\|_N \leqslant \mathbf{E} \, |\nabla g|, \tag{9}$$

where g is arbitrary smooth function from the Orlicz space $L^N = L^N(\mathbb{R}^n, \gamma_n(dx))$ equipped with the norm

$$\|g\|_N = \inf\{t > 0 : \mathbf{E}N(g/t) \leqslant 1\}.$$

If the Young function N behaves at infinity like $N(x) = |x| \sqrt{\log |x|}$, then (9) may be viewed as L^1-analog of the logarithmic inequality (4); existence of $c > 0$ in (9) for such N was proved in Ledoux [9]. Clearly, (9), being applied to indicator functions, implies (8) (maybe with another constant c). This result was sharpened by Pelliccia and Talenti [13]. They defined a function N_0 on the interval $[\sqrt{2\pi}, +\infty)$ by equality

$$2p(1-p) \, N_0\left(\frac{1}{I(p)}\right) = 1, \qquad 0 < p < 1,$$

and linearly on the interval $[0, \sqrt{2\pi}]$ so that $N_0(0) = 0$, $N_0(\sqrt{2\pi}) = 2$, and found the optimal value $c = 1$. At the same time, their proof, which was based on Federer's coarea formula with application of the isoperimetric inequality (3), showed (as well as a theorem by Rothaus [15] who considered more general inequalities on Rimannian manifolds) that the inequality (9) holds true for all smooth g if and only if it is fulfilled (in asymptotic sense) for all indicator functions $g = \chi_A$. That is, (9), for simplicity with $c = 1$, is equivalent to

$$\gamma_n^+(A) \geqslant I_N(p), \qquad p = \gamma_n(A), \tag{10}$$

where $I_N(p) = \|\chi_A - p\|_N$. Consequently, (10) becomes the isoperimetric inequality (3) if and only if $I_N(p) = I(p)$, for all $p \in (0, 1)$, i.e., by definition of the Orlicz space norm, if and only if

$$pN\left(\frac{1-p}{I(p)}\right) + (1-p) \, N\left(\frac{p}{I(p)}\right) = 1, \qquad 0 < p < 1. \tag{11}$$

Since $pN((1-p)\,x) + (1-p)\,N(px) \leqslant 2p(1-p)\,N(x)$, whenever $p \in [0, 1]$, $x \geqslant 0$, we have $I_{N_0} \leqslant I(p)$. Since N_0 is linear on $[0, \sqrt{2\pi}]$, we also have $pN_0((1-p)\,x) + (1-p)\,N_0(px) = 2p(1-p)\,N_0(x)$ for all $x \in [0, \sqrt{2\pi}]$ with strong inequality for $x > \sqrt{2\pi}$. Therefore, $I_{N_0}(p) = I(p)$ if and only if $1/I(p) \leqslant \sqrt{2\pi}$. The last is possible in case $p = 1/2$, only. Thus, at indicator functions $g = \chi_A$, Pelliccia's–Talenti's inequality

$$\|g - \mathbf{E}g\|_{N_0} \leqslant \mathbf{E} \, |\nabla g|$$

coincides with the isoperimetric inequality (3) for the sets A of measure $p = 1/2$, and it is weaker than (3) for $p \neq 1/2$. One may wonder if there exist Young functions N which satisfy (11) so that (11) and (3) coincide (and thus (9) involves (3)) for all p. This turns out to be true ([2]).

Remark 3. By Ehrhard ([5], p. 325), if the function $g^*(x) = g^*(x_1)$ which depends on the first variable, only, is non-decreasing, and g^* and g are γ_n-equimeasurable, then

$$\mathbf{E}\, |\nabla g^*| \leqslant \mathbf{E}\, |\nabla g|. \tag{12}$$

It implies that the inequality (2) is needed to prove in case $n = 1$, and for non-decreasing g, only. In brief, the last can be done as follows. One can write (2) as

$$I(\mathbf{E}g) - \mathbf{E}I(g) \leqslant \int_0^1 I(1 - F_g(t))\, dt, \tag{13}$$

where $F_g(t) = \gamma_1\{x \in \mathbb{R}: g(x) \leqslant t\}$ is distribution function of g with respect to γ_1. Fix $m \in [0, 1]$ and consider the family $M(m)$ of all Borel probability measures on $[0, 1]$ with mean $\int_0^1 x\, dF(x) = m$. $M(m)$ is a convex, compact set equipped with the topology of weak convergence. Replacing in (13) F_g with arbitrary Borel measure F on $[0, 1]$, we observe that the left-hand side and the right-hand side of (13) represent a linear functional and a concave functional on $M(m)$, respectively. Therefore, to prove (13) for all F from $M(m)$, it suffices to state it for extremal "points" of $M(m)$, only. The extremal measures F of $M(m)$ are discrete and have at most two atoms: $F = p\delta_x + (1-p)\,\delta_y$, for some $0 \leqslant p \leqslant 1$, $0 \leqslant y \leqslant x \leqslant 1$, where δ_x denotes the unit mass at x. For such measures F, (13) takes the form

$$I(px + (1-p)\, y) - (pI(x) + (1-p)\, I(y)) \leqslant I(p)(x - y),$$

that is fulfilled for all p, x, y as above and can be proved in an elementary way. However, we will deduce (2) directly from (1), not using the Ehrhard's theorem (property (12)).

Proof of Theorem. First we show that (1) implies the inequality

$$\mathbf{E} g_h \geqslant R_h(\mathbf{E}R_{-h}(g)) \tag{14}$$

which holds for any Borel measurable $g: \mathbb{R}^n \to [0, 1]$ and $h > 0$. Here

$$g_h(x) = \sup_{|e| < 1} g(x + he), \qquad x \in \mathbb{R}^n, \quad h > 0,$$

$$R_h(p) = \Phi(\Phi^{-1}(p) + h), \qquad 0 \leqslant p \leqslant 1, \quad h \in \mathbb{R}.$$

Note that (1) is a partial case of (14) when $g=\chi_A$ is the indicator function of the set A. The following idea is that (14) turns into (2) for small h, i.e., for smooth g, letting $h\to 0$, we obtain (2). To obtain (1) from (2), one can observe that, if h_1 and h_2 satisfy (14), then h_1+h_2 satisfies (14), so only small h are important; we need to verify this property for indicator functions. Below this plan is performed in detail according to the chain $(1)\Rightarrow(14)\Rightarrow(2)\Rightarrow(1)$.

Step 1: $(1)\Rightarrow(14)$. Obviously, the function g_h is lower semicontinuous, therefore, Borel measurable. We prove that (1) implies (14) in a more general situation. Let μ be a probability measure on $\mathbb{R}^n$. Given $h>0$, suppose that, for all Borel measurable sets $A\subset\mathbb{R}^n$,

$$\mu(A^h)\geqslant R(\mu(A)),$$

where R is a concave increasing function from $[0,1]$ onto itself. In particular, $R(0)=0$, $R(1)=1$. Denote $S(p)=1-R(1-p)$. Let R^{-1} be the inverse of R.

LEMMA. *If for all $p, q\in[0,1]$,*

$$R(pq)\leqslant R(p)\,R(q), \tag{15}$$

$$S(pq)\geqslant S(p)\,S(q), \tag{16}$$

then, for any Borel measurable g: $\mathbb{R}^n\to[0,1]$,

$$\mathbf{E}g_h\geqslant R(\mathbf{E}R^{-1}(g)). \tag{17}$$

Mathematical expectations in (17) are with respect to μ.

Proof of Lemma. Put $A(t)=\{x\in\mathbb{R}^n\colon g(x)>t\}$. Since $\{x\in\mathbb{R}^n\colon g_h(x)>t\}=A(t)^h$, for all t real, and since $0\leqslant g_h\leqslant 1$, we have

$$\begin{aligned}\mathbf{E}g_h&=\int_0^1\mu\{g_h>t\}\,dt=\int_0^1\mu(A(t)^h)\,dt\\&\geqslant\int_0^1 R(\mu(A(t)))\,dt=\int_0^1 R(1-F(t))\,dt,\end{aligned}$$

where $F(t)=1-\mu(A(t))$ is distribution function of g with respect to μ. Hence, (17) will follow from

$$\int_0^1 R(1-F(t))\,dt\geqslant R\left(\int_0^1 R^{-1}(x)\,dF(x)\right), \tag{18}$$

where F is arbitrary probability Borel measure on $[0,1]$. Let $M=C[0,1]^*$ denote the family of all signed finite Borel measures on

$[0, 1]$. Let $M = C[0, 1]^*$ denote the family of all signed finite Borel measures on $[0, 1]$ equipped with the topology of weak convergence. Let $M_1 \subset M$ consist of all probability distributions. M_1 is considered as a compact, convex subset of M, and the extremal points in M_1 are just unit masses δ_x, $x \in [0, 1]$. Again, as it was suggested in Remark 3, we fix $m \in [0, 1]$ and define $M(m)$ as the family of all probability distributions F on $[0, 1]$ such that $\int_0^1 R^{-1}(x)\, dF(x) = m$. We observe that the left-hand side and the right-hand side of (18) represent a concave functional and a constant $(= R(m))$ on $M(m)$, respectively. Therefore, to prove (18) for all F from $M(m)$, it suffices to state it for extremal points of $M(m)$, only. Since $M(m)$ is the intersection of the simplex M_1 with a hyperplane in M, the extremal points of $M(m)$ lie on one-dimensional edges of M_1, hence, they are of the form $F = p\delta_x + (1-p)\,\delta_y$, for some $0 \leqslant p \leqslant 1$, $0 \leqslant y \leqslant x \leqslant 1$. For such F, since $R(0) = 0$, $R(1) = 1$, (18) turns into $y + R(p)(x-y) \geqslant R(pR^{-1}(x) + (1-p)\, R^{-1}(y))$. Changing the variables x and y with $R(x)$ and $R(y)$, respectively, we come to the inequality

$$R(px + (1-p)\, y) \leqslant R(p)\, R(x) + (1 - R(p))\, R(y) \tag{19}$$

under the same assumptions on x, y and p. Given p, $c \in [0, 1]$, we observe that the left-hand side of (19) is constant and the right-hand side of (19) is a concave function on the segment

$$\Delta(p, c) = \{(x, y) : 1 \geqslant x \geqslant y \geqslant 0, px + (1-p)\, y = c\}.$$

Therefore, one needs to check (19) at the end points of $\Delta(p, c)$, only. One of these points lies on the diagonal $x = y$, where (19) becomes identity. The other lies either on the line $y = 0$, or on the line $x = 1$. In the first case, (19) becomes $R(px) \leqslant R(p)\, R(x)$ that is fulfilled by the assumption (15) on R. In the second case, we come to the inequality $R(p + (1-p)\, y) \leqslant R(p) + (1 - R(p))\, R(y)$ that may be expressed via the function S as $S((1-p)(1-y)) \geqslant S(1-p)(1-y)$. The last holds by (16). Lemma is thus proved.

Now we explain why the above proposition may be applied to $\mu = \gamma_n$. The family R_h, $h \in \mathbb{R}$, forms one-parametric group of increasing bijections in $[0, 1]$, i.e., R_h is superposition of R_{h_1} and R_{h_2} if $h = h_1 + h_2$, and R_{-h} is the inverse of R_h. Therefore, for $R = R_h$, we have $S = R^{-1}$. Hence, (16) follows from (15) which can be proved directly using log-concavity of ϕ/Φ. But (15) is also a partial case of (1) when $n = 2$, and A is the cube $(-\infty, \Phi^{-1}(p)) \times (-\infty, \Phi^{-1}(q))$. Concavity of R_h, $h > 0$, is evident. Thus, the isoperimetric property of Gaussian measure implies (14).

Step 2: (14) $\Rightarrow$ (2). We prove (2) a little in more general situation when g is arbitrary locally Lipschitz function with values in $[0, 1]$. By

Rademacher's theorem (see, e.g., [6], p. 216), the function g is differentiable at almost all $x \in \mathbb{R}^n$, and for such x, it follows that $(g_h(x) - g(x))/h \to |\nabla g(x)|$ as $h \to 0$.

Step 2.1: *g is arbitrary Lipschitz function with values in* $[0, 1]$. Let, for some c, $|g(x) - g(y)| \leqslant c\,|x - y|$, for all $x, y \in \mathbb{R}^n$. One may assume that g is not constant so that $0 < \mathbf{E}g < 1$ (otherwise, there is nothing to prove). Since, for all $x \in \mathbb{R}^n$, $h > 0$, $g_h(x) - g(x) \leqslant ch$, we get

$$\lim_{h \to 0} \mathbf{E}\, \frac{g_h(x) - g(x)}{h} = \mathbf{E}\, |\nabla g|, \tag{20}$$

by Lebesgue's dominated convergence theorem. Using Taylor's expansion, we have $R_h(p) = p + \phi(\Phi^{-1}(p))\, h + c_p(h)\, h^2$ for all $p \in [0, 1]$, $h \in \mathbb{R}$, where $c_p(h)$ is bounded by the constant $\sup_x \phi'(x)/2$. So,

$$\mathbf{E}R_{-h}(g) = \mathbf{E}g - \mathbf{E}\varphi(\Phi^{-1}(g))\, h + c(h)\, h^2,$$

where $c(h)$ is a bounded function. Hence,

$$\begin{aligned} R_h(\mathbf{E}R_{-h}(g)) = {} & (\mathbf{E}g - \mathbf{E}\varphi(\Phi^{-1}(g))\, h + c(h)\, h^2) \\ & + \varphi(\Phi^{-1}(\mathbf{E}R_{-h}(g)))\, h + O(h^2). \end{aligned} \tag{21}$$

Since $0 < \mathbf{E}g < 1$, the value $\mathbf{E}R_{-h}(g)$ is separated from 0 and 1 for h small enough. Therefore, $\phi(\Phi^{-1}(\mathbf{E}R_{-h}(g)) = \phi(\Phi^{-1}(\mathbf{E}g)) + O(h)$ as $h \to 0$, and we finally obtain from (21)

$$R_h(\mathbf{E}R_{-h}(g)) = \mathbf{E}g + (\varphi(\Phi^{-1}(\mathbf{E}g)) - \mathbf{E}\varphi(\Phi^{-1}(g)))\, h + O(h^2),$$

that is,

$$\lim_{h \to 0} \frac{R_h(\mathbf{E}R_{-h}(g)) - \mathbf{E}g}{h} = \varphi(\Phi^{-1}(\mathbf{E}g)) - \mathbf{E}\varphi(\Phi^{-1}(g)). \tag{22}$$

It remains to rewrite (14) as

$$\mathbf{E}\, \frac{g_h(x) - g(x)}{h} \geqslant \frac{R_h(\mathbf{E}R_{-h}(g)) - \mathbf{E}g}{h}, \qquad h > 0,$$

and take the limit as $h \to 0$, $h > 0$, using (20) and (22).

Step 2.2: *g is arbitrary locally Lipschitz function.* Consider a truncation of g putting $g_N(x) = g(x)\, T_N(x)$ where N is positive integer and

$$T_N(x) = \begin{cases} 1, & \text{if } \ |x| \leqslant N \\ N + 1 - |x|, & \text{if } \ N \leqslant |x| \leqslant N + 1 \\ 0, & \text{if } \ N + 1 \leqslant |x|. \end{cases}$$

Then $0 \leqslant g_N \leqslant 1$, g_N is Lipschitz and, furthermore, $|\nabla g_N(x)| = |\nabla g(x)|$ for $|x| < N$, $|\nabla g_N(x)| = 0$ for $|x| > N+1$. If g is differentiable at a point x with $N < |x| < N+1$, then $\partial g_N(x)/\partial x_i = \partial g(x)/\partial x_i (N+1-|x|) - g(x)\, x_i/|x|$ $(1 \leqslant i \leqslant n)$, hence $|\nabla g_N(x)| |\nabla g(x)| + |g(x)| \leqslant |\nabla g(x)| + 1$ almost everywhere. Consequently, $\mathbf{E}\, |\nabla g_N| \leqslant \mathbf{E}\, |\nabla g| + \gamma_n\{x \in \mathbb{R}^n\colon N < |x| < N+1\} \to \mathbf{E}\, |\nabla g|$ as $N \to \infty$. On the other hand, according to Step 2.1,

$$\varphi(\Phi^{-1}(\mathbf{E} g_N)) - \mathbf{E}\varphi(\Phi^{-1}(g_N)) \leqslant \mathbf{E}\, |\nabla g_N|. \tag{23}$$

Since g_N converges to g pointwise, we obtain (2) from (23) taking limit as $N \to \infty$.

Step 3: (2) $\Rightarrow$ (1). Clearly, using approximation argument, one may assume that the set A in (1) is a finite union of non-empty, open balls in $\mathbb{R}^n$. In particular, $0 < \gamma_n(A) < 1$. The family of such sets A is closed under operations $A \to A^h$, $h > 0$.

Assume that (2) holds for all smooth g. Then (2) holds for all Lipschitz g. Indeed, for any Lipschitz function g with values in $[0, 1]$, of Lipschitzian constant c, there exists a sequence g_n of smooth functions g_n with values in $[0, 1]$ such that $|\nabla g| \leqslant c$, g_n tends to g everywhere and ∇g_n tends to ∇g almost everywhere. Taking the limit in (2) for g_n as $n \to \infty$, we obtain (2) for g.

Now we approximate the indicator function $g = \chi_A$ by Lipschitz functions: for any $\varepsilon > 0$, there exists a Lipschitz function $g^{(\varepsilon)}$ of Lipschitz constant $\leqslant 1/\varepsilon$ such that $g^{(\varepsilon)} = 1$ on A, $g^{(\varepsilon)} = 0$ on $\mathbb{R}^n \backslash A^\varepsilon$ and $0 \leqslant g^{(\varepsilon)} \leqslant 1$. Then $|\nabla g^{(\varepsilon)}| = 0$ on A and $\mathbb{R}^n \backslash clos(A^\varepsilon)$, and $|\nabla g^{(\varepsilon)}| \leqslant 1/\varepsilon$ everywhere. Note that, $\partial(A^\varepsilon)$ is of measure 0 for all $\varepsilon > 0$. Therefore, according to (2),

$$\varphi(\Phi^{-1}(\mathbf{E} g^{(\varepsilon)})) - \mathbf{E}\varphi(\Phi^{-1}(g^{(\varepsilon)})) \leqslant \mathbf{E}\, |\nabla g^{(\varepsilon)}| \leqslant (\gamma_n(A^\varepsilon) - \gamma_n(A))/\varepsilon. \tag{24}$$

Since $g^{(\varepsilon)}$ converges pointwise to $g = \chi_{clos(A)} = \chi_A$ a.e. as $\varepsilon \to 0$, $\varepsilon > 0$, it follows from (24) that

$$\varphi(\Phi^{-1}(\gamma_n(A))) \leqslant \gamma_n^+(A). \tag{25}$$

Now (1) easily follows from (25). Consider the family of functions $R_h^\sigma(p) = \Phi_\sigma(\Phi_\sigma^{-1}(p) + h)$ with the extra parameter $\sigma > 1$, where $\Phi_\sigma(x) = \Phi(x/\sigma)$, Φ_σ^{-1} is the inverse of Φ_σ. Like the case $\sigma = 1$, the family R_h^σ, $h \in \mathbb{R}$, forms one-parameter group of increasing bijections in $[0, 1]$. If we show that, for all $\sigma > 1$ and $h > 0$, $R_h^\sigma(\gamma_n(A)) \leqslant \gamma_n(A^h)$, then, letting $\sigma \to 1$, we get (1). For this purpose, fix $\sigma > 1$, and put

$$J = \{h > 0 : R_t^\sigma(\gamma_n(A)) \leqslant \gamma_n(A^t) \text{ for all } t \in (0, h]\}.$$

The functions $h \to R_h^\sigma(\gamma_n(A))$ and $h \to \gamma_n(A^h)$ are continuous. Hence, to show that $J = (0, +\infty)$, it suffices to see that

(a) $\varepsilon \in J$, for all $\varepsilon > 0$ small enough;

(b) if $h \in J$, then $h + \varepsilon \in J$, for all ε small enough.

For small h, $R_h^\sigma(p) = p + \phi_\varepsilon(\Phi_\sigma^{-1}(p))\, h + O(h^2)$, where $\phi_\sigma = \Phi'_\sigma$, therefore,

$$R_h^\sigma(\gamma_n(A)) = \gamma_n(A) + \varphi_\sigma(\Phi_\sigma^{-1}(\gamma_n(A)))\, h + O(h^2), \qquad \text{as} \quad h \to 0. \tag{26}$$

On the other hand,

$$\gamma_n(A^h) = \gamma_n(A) + \gamma_n^+(A)\, h + O(h^2), \qquad \text{as} \quad h \to 0. \tag{27}$$

Taking into account (25) and comparing (26) and (27), we prove (a), because $0 < \gamma_n(A) < 1$ and, therefore, $\phi_\sigma(\Phi_\sigma^{-1}(\gamma_n(A))) = \phi(\Phi^{-1}(\gamma_n(A)))/\sigma < \phi(\Phi^{-1}(\gamma_n(A)))$. Let $h \in J$. Then, $R_h^\sigma(\gamma_n(A)) \leqslant \gamma_n(A^h)$. If this inequality is strong, then it remains true for all $h + \varepsilon$ with $\varepsilon > 0$ small enough. Let $R_h^\sigma(\gamma_n(A)) = \gamma_n(A^h)$. Set $B = A^h$. According to (25), $\phi(\Phi^{-1}(\gamma_n(B))) \leqslant \gamma_n^+(B)$ and, therefore, again by (26) and (27), $R_\varepsilon^\sigma(\gamma_n(B)) < \gamma_n(B^\varepsilon)$ for all ε small enough. It remains to note that

$$\begin{aligned} R_{h+\varepsilon}^\sigma(\gamma_n(A)) &= R_\varepsilon^\sigma(R_h^\sigma(\gamma_n(A))) = R_\varepsilon^\sigma(\gamma_n(A^h)) = R_\varepsilon^\sigma(\gamma_n(B)) \\ &\leqslant \gamma_n(B^\varepsilon) = \gamma_n(A^{h+\varepsilon}), \end{aligned}$$

because $B^\varepsilon = A^{A+\varepsilon}$. This proves (b) and, therefore, theorem.

Acknowledgment

The author is grateful to the Department of Statistics of the University of North Carolina at Chapel Hill for the invitation to visit the Center for Stochastic processes. He thanks the referee for careful reading and useful comments (especially for the idea of Remark 3).

References

1. S. Aida, T. Masuda, and I. Shigekawa, Logarithmic Sobolev inequalities and exponential integrability, preprint (1993).
2. S. G. Bobkov and Ch. Houdré, On some connections between Sobolev-type inequalities and isoperimetry, preprint (1994).
3. C. Borell, The Brunn–Minkowski inequality in Gauss space, *Invent. Math.* **30**, Fasc. 2 (1975), 207–216.
4. J.-D. Davies and B. Simon, Ultracontractivity and the heat kernel for Schrödinger operators and Dirichlet Laplacians, *J. Funct. Anal.* **59** (1984), 335–395.
5. A. Ehrhard, Inégalités isopérimétriques et intégrales de Dirichlet gaussiennes, *Ann. Sci. Ecole Norm. Sup.* **17** (1984), 317–332.
6. H. Federer, "Geometric Measure Theory," Springer-Verlag, Berlin, 1969.
7. L. Gross, Logarithmic Sobolev inequalities, *Amer. J. Math.* **97** (1975), 1061–1083.

8. L. Gross, Logarithmic Sobolev inequalities and contractivity properties of semigroups, *in* "Varenna, 1992," Lecture Notes in Mathematics, Vol. 1563, pp. 54–88, 1993.
9. M. Ledoux, Isopérimétrie et inégalités de Sobolev logarithmiques gaussiennes, *C.R. Acad. Sci. Paris* **306** (1988), 79–82.
10. M. Ledoux, Semigroup proofs of the isoperimetric inequality in Euclidean and Gauss space, *Bull. Sci. Math*, to appear.
11. M. Ledoux, Isoperimetry and Gaussian analysis, *in* "Ecole d'Eté de Probabilités de Saint-Flour," Lecture Notes in Mathematics Ser., Springer-Verlag, Berlin, 1994.
12. R. Osserman, The isoperimetric inequality, *Bull. Amer. Math. Soc.* **84** (1978), 1182–1238.
13. E. Pelliccia and G. Talenti, A proof of a logarithmic Sobolev inequality, *Calc. Var.* **1** (1993), 237–242.
14. G. Pisier, Probabilistic methods in the geometry of Banach spaces, *in* "Probability and Analysis, Varenna 1985," pp. 167–214, Lecture Notes in Mathematics, Vol. 1206, Springer-Verlag, Berlin, 1986.
15. O. S. Rothaus, Analytic inequalities, isoperimetric inequalities and logarithmic Sobolev inequalities, *J. Funct. Anal.* **64** (1985), 296–313.
16. V. N. Sudakov and B. S. Tsirel'son, Extremal properties of half-spaces for spherically invariant measures, *J. Soviet Math.* **9** (1978), 9–18 [Engl. transl.]; *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. inst. Steklov.* (*LOMI*) **41** (1974), 14–24.
17. M. Talagrand, Isoperimetry, logarithmic Sobolev inequalities on the discrete cube, and Margulis' graph connectivity theorem, *Geom. Funct. Anal.* **3**, No. 3 (1993), 295–314.