SOME TOPICS IN PROBABILITY & ANALYSIS

Conference Board of the Mathematical Sciences
REGIONAL CONFERENCE SERIES IN MATHEMATICS

supported by the
National Science Foundation

Number 70

SOME TOPICS IN PROBABILITY AND ANALYSIS

Richard F. Gundy

Published for the
Conference Board of the Mathematical Sciences
by the
American Mathematical Society
Providence, Rhode Island

Lectures given at DePaul University,
Chicago, Illinois, July 14–18, 1986
Supported by the Conference Board
of the Mathematical Sciences

Research supported by National Science Foundation Grant DMS-8602950.

1980 *Mathematics Subject Classifications* (1985 *Revision*). 60G44, 26D15, 26B15, 44A15, 47D05, 60J05, 60J25.

Library of Congress Cataloging-in-Publication Data

Some topics in probability and analysis / Richard F. Gundy.
p. cm. —(Conference Board of the Mathematical Sciences
Regional conference series in mathematics; no. 70.)
Bibliography: p.
ISBN 0-8218-0721-8 (alk. paper)
1. Inequalities (Mathematics) 2. Harmonic functions. I. Series: Regional conference series in mathematics; no. 70.
QA1.R33 no. 70 [QA 295] 510 s—dc19 [512.9'7] 89-303 CIP

The paper used in this book is acid-free and falls within the guidelines
established to ensure permanence and durability. ∞

Contents

Introduction

In these lectures I concentrated on three topics: (1) Local time theory for Brownian motion and some geometrical inequalities for harmonic functions in the upper half-plane $\mathbf{R}_+^{n+1}$. (2) A probabilistic treatment of Riesz transforms in $\mathbf{R}_+^{n+1}$ and semimartingale inequalities. (3) A discussion of the Ornstein-Uhlenbeck semigroup and P. A. Meyer's extension of the Riesz inequalities for the infinite-dimensional version of this semigroup, introduced by Malliavin (see [26, 27]).

Regarding topic (1), we sketch a proof of the inequalities obtained by Barlow and Yor in [1] for the maximal local time functional. These inequalities led the author to some new inequalities for a geometric functional, defined on harmonic functions in $\mathbf{R}_+^{n+1}$, called the *density of the area integral*.

Topic (2) is a probabilistic approach to the Riesz transform inequalities in $\mathbf{R}^n$. This method of proof was first introduced by the author and Varopoulos [24]. The method was elaborated further by the author with Silverstein [23]. We show that the same ideas are effective in proving semimartingale inequalities of the type usually obtained for martingales.

The final topic in the series is a discussion of the Ornstein-Uhlenbeck semigroup. We give a proof of Nelson's hypercontractivity inequality, following the ideas of Neveu [29]. Then, we present a proof of P. A. Meyer's inequalities, the analogues of the classical Riesz transform inequalities in which the role of the Laplacian Δ is replaced by the Ornstein-Uhlenbeck generator, $\Delta - X \cdot \nabla$. This approach, found in [21], is a direct extension of the methods discussed in the previous section (topic (2)).

The author would like to express his warmest gratitude to the organizers of the Conference at DePaul University, the participants, and especially, to Roger Jones for his enthusiastic hospitality.

The Barlow-Yor Inequalities

For H^p-theory of harmonic functions in $\mathbf{R}_+^{n+1}$, two functionals have been studied in detail: the nontangential maximal function and the Lusin area function. These two geometric objects are analogs of two probabilistic functionals associated with continuous martingales. If $X = \{X_t, t \geq 0\}$ is a continuous martingale, then

$$X^* = \sup_t |X_t| \quad \text{and} \quad S(X) = \langle X \rangle^{1/2}$$

(where $\langle X \rangle$ is the quadratic variation of X). The basic relation between $S(X)$ and X^* is well known now:

$$\|S(X)\|_p \simeq \|X^*\|_p$$

for all $0 < p < \infty$. Until the appearance of Barlow and Yor's paper [1], no other functional "of significance" had been found. They discovered another functional that is, in some sense, in between the maximal function and the quadratic variation: the maximal local time. Suppose for simplicity that X is Brownian motion run up to a stopping time τ. The local time of the Brownian motion may be defined as follows. For each trajectory ω consider the mapping $X(\omega) : [0, \tau] \to \mathbf{R}^1$, and the "push-forward" map X_*. That is, $X_*(dt)$ is the image measure of Lebesgue measure on $[0, \tau]$ on $\mathbf{R}^1$, under the mapping X. The measure $X_*(dt)$ on $\mathbf{R}^1$ is absolutely continuous, by a result due to P. Lévy, and its density $L(r) = L(r, \omega, \tau)$ is called the *local time at r* (up to time τ for the trajectory ω). Consider $L(r)$ as a process in the real parameter r. Of course, the filtration, based on the space parameter r, is completely different from the time filtration. Now, take the maximal function $L^* = \sup_r L(r)$.

We seek to prove a good-λ inequality of the form

$$P(L^* > \beta\lambda, S(X) \leq \delta\lambda) \leq \varepsilon(\beta, \delta) P(L^* > \lambda) \tag{1}$$

with $\beta > 1$, $\delta < 1$, valid for all $\lambda > 0$. From this we can obtain an inequality for norms of the form

$$\|L^*\|_p \leq C_p \|S(X)\|_p,$$

$p > 1$, by now standard arguments [7]. The converse inequality is obtained by another good-λ inequality with the roles of L^* and $S(X)$ interchanged.

Because L^* is a maximal function across a *different* filtration than the one giving time evolution, Barlow and Yor appealed to the Ray-Knight theorem which specifies the structure of $L(r)$ as a process. However, during these lectures, Burgess Davis observed that this was not necessary. The scaling properties of these functionals already foretell the good-λ inequalities and their consequences. As it turns out, Richard Bass had observed the same thing some months before. Their articles appear in **[3, 11]**. (That the scaling and Markov property implies the good-λ inequality between $S(X)$ and X^* already was observed by Burkholder **[8]**.)

Suppose we wish to prove (1). Let X^t be the process X stopped at a stopping time t. The key observation is that, for *constant* times t, the random variable $L^*(X^t)$ scales like Brownian motion:

$$L^*(X^t) \simeq \lambda L^*(X^{t/\lambda^2})$$

and, like $(X^t)^*$ and $S(X^t)$, it is an increasing, subadditive functional in t (since $L(X^t, r)$ is additive for each fixed r). Now the proof of the good-λ inequality is standard: Let

$$\mu = \inf\{t: L^*(X^t) = \lambda\}, \quad \nu = \inf\{t: L^*(X^t) = \beta\lambda\}.$$

If X^μ, X^ν and θ_μ are the stopped Brownian motions, and shift operators, respectively, then $L^*((X \circ \theta_\mu)^{\nu-\mu}) = (\beta - 1)\lambda$ on the set $\{\mu < \infty\}$. Now suppose in addition, that $\tau^{1/2} \le \delta\lambda$. Then the lifetime of $(X \cdot \theta_\mu)$ is bounded by $(\delta\lambda)^2$, so that

$$\begin{aligned} P(L^* > \beta\lambda, \tau^{1/2} \le \delta\lambda) &= P(L^*((X \circ \theta_\mu)^{\nu-\mu}) = (\beta - 1)\lambda, \tau^{1/2} \le \delta\lambda, \mu < \infty) \\ &\le P(L^*(X \circ \theta_\mu)^{(\delta\lambda)^2} \ge (\beta - 1)\lambda, \mu < \infty). \end{aligned}$$

Now $(\delta\lambda)^2$ is a constant stopping time, and we may rescale:

$$L^*(X \circ \theta_\mu)^{(\delta\lambda)^2} \simeq (\delta\lambda)L(X \circ \theta_\mu)^1$$

so that

$$\begin{aligned} P(L^*(X \circ \theta_\mu)^{(\delta\lambda)^2} &\ge (\beta - 1)\lambda, \mu < \infty) \\ &= P((\delta\lambda)L^*(X \circ \theta_\mu)^1 \ge (\beta - 1)\lambda, \mu < \infty) \\ &= \int P\left(L^*(X \circ \theta_\mu)^1 \ge \frac{(\beta - 1)}{\delta} \| X_\mu\right) P(\mu < \infty). \end{aligned}$$

Now by the joint continuity (t, r) of $L(X^t, r)$, and the Markov character of local time, $P(L^*(X \circ \theta_\mu)^1 \ge C \| X_\mu)$ is uniformly small in X_μ as C tends to infinity. Thus,

$$P(L^* > \beta\lambda, \tau^{1/2} \le \delta\lambda) \le O\left(\frac{\delta}{\beta - 1}\right) P(L^* > \lambda)$$

as desired. The converse inequality is obtained in much the same way. For the details, see **[11]**.

REMARKS. What is L^* good for? The answer lies in the theorems one can prove with the new device.

(1) *Factorizations, and ratios.* The most straightforward application is the following observation. Suppose X is a continuous martingale. Following what we did for Brownian motion, we have a measure $d\langle X\rangle$ on $[0, \langle X\rangle)$ and a push-forward $X_*(d\langle X\rangle) = L(r)dr$, where $-X^* \leq r \leq X^*$.

By definition,

$$\langle X\rangle = \int_{-X^*}^{X^*} L(r)\,dr \leq 2X^* \cdot L^*.$$

This intriguing "factorization" of $\langle X\rangle$ is perhaps the principal feature of L^*. It prompted this writer to wonder whether such a factorization could be obtained when $\langle X\rangle$ and X^* were replaced by the geometrically defined analogs, A and N, the Lusin area function and the nontangential maximal function, respectively. This will be the subject of the next section.

(2) $L \log L$ characterizations for *nonnegative* martingales X: X^* belongs to L^1 if and only if X_∞ belongs to $L\log^+ L$ [17]. This result was recently extended in an interesting way by Brossard and Chevalier [6]. Their theorem concerns a martingale X of *arbitrary* sign: *The martingale X is in $L\,\mathrm{Log}^+ L$ if and only if X^* belongs to L^1 and, $L(0)$ belongs to $L\log^+ L$.*

Notice that for a positive martingale, $L(0) = 0$, and that zero is *not* distinguished here: $L(0)$ is in $L\log^+ L$ if and only if $L(r)$, $-\infty < r < \infty$, is in $L\log^+ L$.

Here is a proof, drawn from [6]. Assume for simplicity that X is a Brownian motion stopped at a stopping time τ. By Tanaka's formula, the terminal variable

$$|X_\infty| = Y_\infty + L(X,0)$$

where Y_∞ is the random variable

$$Y_\infty = \int_0^\tau \operatorname{sgn} X_S\, dX_S.$$

Now recall the $L\log L$ results: For any martingale M, we have

$$\|M^*\|_1 \leq C\left\{E(M_\infty \log^+ M_\infty) + 1\right\}$$

and for nonnegative martingales (see [17])

$$E(M_\infty \log^+ M_\infty) \leq E(M^*) + \|M_\infty\|_1 \log^+ \|M_\infty\|_1.$$

We propose to show that for X, the stopped Brownian motion,

$$\sup_t E(|X_t| \log^+ |X_t|) \leq C\left\{E(L(X,0)\log^+ L(X,0)) + E(X^*)\right\}.$$

If we take conditional expectations, then

$$E(|X|\,\|F_t) = |X_\infty|_t \text{ and } E(L(X,0)\|F_t) = L_t(X,0)$$

are nonnegative martingales. By the above inequalities, and Jensen's inequality,

$$\begin{aligned} CE(|X_t|\log^+|X_t|) &\le E(|X|^*) + \|X\|_1 \log^+ \|X\|_1 \\ &\le E(Y^*) + E\left(\sup_t L_t(X,0)\right) + \|X\|_1 \log^+ \|X_1\|_1 \\ &\le E(Y^*) + CE(L(X,0)\log^+ L(X,0)) + C + \|X\|_1 \log^+ \|X\|_1 \\ &\le CE(X^*) + CE(L(X,0)\log^+ L(X,0)) \\ &\quad + C + \|X\|_1 \log^+ \|X\|_1 \end{aligned}$$

since $E(Y^*) \le CE(X^*)$ by the standard H^1-inequality. The other direction,

$$E(L(X,0)\log^+ L(X,0)) \le C\left\{\sup_t E(|X_t|\log^+|X_t|) + E(X^*_\infty)\right\}$$

is proved in the same manner.

The Brossard-Chevalier theorem, too, has a geometric version, based on what we call the density of the area integral, the subject of the next section. Our goal will be to take from the probability theory the idea of local time and replace it by a geometric functional. With some regret, we shall also deprive ourselves of the probability ideas in order to find a purely analytic proof of Barlow-Yor type inequalities in this setting.

The Density of the Area Integral

The references for this section are [18, 20, and 23]. The geometric analogue of the quadratic variation of a martingale is, formally, the area integral

$$A^2(u)(x_0) = \iint_{\Gamma(x_0)} y^{1-n}|\nabla u|^2(x, y)\, dx\, dy$$

where $\Gamma(x_0)$ is the cone with base x_0 and axis $y > 0$. The aperture is sometimes indicated with a subscript $a > 0$. Thus,

$$\Gamma_a(x_0) = \{(x, y): |x - x_0| \leq ay\};$$

if $a = 1$, we omit the subscript.

If $\langle X \rangle$, the quadratic variation of a martingale, has an associated local time, one is led to inquire whether $A^2(u)$ can be treated in the same manner. Local time for X is the "push-forward" $X_*(d\langle X \rangle)$. In the same spirit, replace X by the harmonic function u: consider $u(x, y)$ as a mapping $u: \Gamma(x_0) \to \mathbf{R}^1$. Then $u_*(y^{1-n}|\nabla u|^2(x, y)\, dx\, dy)$ is the 'push-forward" of the indicated measure to another measure on $\mathbf{R}^1$. The specification of this measure may be accomplished in two ways:

(1) The coarea formula (elementary version). Let $u(x, y)$ be a C^∞-function on $\mathbf{R}^{n+1}_+$ to $\mathbf{R}^1$ and ψ, f be a pair of functions with compact support in $\mathbf{R}^{n+1}_+$ and $\mathbf{R}$, respectively. Suppose that on the support of ψ, $|\nabla u| \neq 0$. Then by the standard change of variables theorem from advanced calculus, we can find functions v_i, $i = 1, 2, \ldots, n$, defined on $\mathbf{R}^{n+1}_+$ such that the $n + 1$ vector $(v(x, y), u(x, y))$ forms a coordinate system with the property that the Jacobian

$$|\det(\partial(v, u)/\partial(x, y))| = |\nabla u|.$$

This translates integrals in the following way:

$$\iint f(u(x, y))\psi(x, y)|\nabla u(x, y)|^2\, dx\, dy$$
$$= \iint f(u)(\psi \cdot |\nabla u|)(x(v, u), y(v, u))\, dv\, du.$$

The loss of a power on $|\nabla u|^2$ going from left to right comes from the interpretation of $|\nabla u|$ as the Jacobian of a transformation. In this context, we do not have to worry about the complications of the general coarea formula of Federer [12] since the presence of the "extra" $|\nabla u|$ on the right-hand side ($|\nabla u|^2$ as opposed to $|\nabla u|$) means that we can avoid discussion of critical values of the transformation $(v(x,y), u(x,y))$. Finally, notice that for ψ with compact support in $\mathbf{R}_+^{n+1}$, $\int(\psi \cdot |\nabla u|)(x(v,r), y(v,r))\,dv$ is a *continuous* function of the parameter r, indexing the level surfaces $\{(x,y): u(x,y) = r\}$.

Now we apply these considerations to the area integral

$$\int y|\nabla u|^2 \psi_y(x_0 - x)\,dx\,dy = A^2(u)(x_0)$$

where ψ_y is the dilated "bump" function given above. We have

$$\begin{aligned} A^2(u)(x_0) &= \int_{-N(x_0)}^{N(x_0)} \int y|\nabla u|(x,y)\psi_y(x_0 - x)\sigma_r(dx\,dy)\,dr \\ &= \int_{-N(x_0)}^{N(x_0)} D(u;r)(x_0)\,dr \end{aligned}$$

where $\sigma_r(dx\,dy)$ is the n-dimensional Hausdorff surface measure of the level set $u = r$. Notice that $D(u;r)$ is lower semicontinuous as a function of r, so that the supremum and the essential supremum (over r) coincide. We define

$$D(u)(x_0) = \sup_r D(u;r)(x_0).$$

(2) A second change of variables formula. In the previous change of variables formula, we have assumed that u was smooth, but not necessarily harmonic. Now we suppose u to be harmonic. We wish to prove

$$D(u;r)(x_0) = \iint y\psi_y(x_0 - x)\Delta(u - r)^+(dx, dy).$$

Write

$$F(s) = \int_{-\infty}^{\infty} (s - r)^+ f(r)\,dr$$

so that $F''(s) = f(s)$, and if $u(x,y)$ is harmonic, $\Delta F[u] = f(u)|\nabla u|^2$. Again for any $\psi(x,y)$ with compact support strictly in the interior of $\mathbf{R}_+^{n+1}$, we have

$$\begin{aligned} \iint \psi(x,y)\Delta F[u](x,y)\,dx\,dy &= \iint \Delta\psi F[u]\,dx\,dy \\ &= \iiint (\Delta\psi) f(r)(u - r)^+\,dr\,dx\,dy \\ &= \iint \psi(x,y) f(r)\Delta(u - r)^+(dx\,dy)\,dr. \end{aligned}$$

Now form an integral using the dilated function $\psi_y(x)$ as before, and $F(r) = r^2$.

We have

$$\int y\psi_y(x_0 - x)\Delta(u - r)^+(dx\,dy) = D(u;r)(x_0) \quad \text{a.e.}$$

as defined above since the two change of variables formulas give the same result when integrated in r. We would like to establish that the two procedures give the same result for *all* r. We conjectured that $\Delta(u-r)^+$ never charges the set $|\nabla u| = 0$, but we could not prove this. Nevertheless, by Sard's theorem, the set of r such that $u^{-1}(r)$ contains points $\{(x,y): \nabla u(x,y) = 0\}$ has Lebesgue measure zero. Furthermore, the integral

$$\iint y\psi_y(x_0 - x)\Delta(u-r)^+(dx\,dy)$$

is lower semicontinuous as a function of r, so that again, the supremum over r is equal to the essential supremum over r. Therefore, the supremum functional $D(u)(x_0)$ may be defined by procedure (1) or procedure (2).

The question that we could not answer, that is, whether $\Delta(u-r)^+$ charges the set where $|\nabla u| = 0$ has recently been answered by Jean Brossard **[5]**. In fact, Brossard shows, using the Weierstrass preparation theorem, that locally, $\{|\nabla u| = 0\}$ is a surface of dimension at most $n-1$ in $\mathbf{R}^{n+1}$. This means that this surface is of capacity zero (see Carleson **[10**, p. 28**]**) and therefore, cannot support any measure with finite potential, such as $\Delta(u-r)^+$. Conclusion: $\Delta(u-r)^+$ does not charge $\{|\nabla u| = 0\}$. Here is an alternative proof of the fact that, at most, the set $\{\nabla u = 0\}$ is of dimension $n-1$ in $\mathbf{R}^{n+1}_+$. First of all, since u is assumed to be nonconstant on $\mathbf{R}^{n+1}_+$, there is a partial derivative v of some finite order such that the set $\{\nabla u = 0\}$ is contained in $\{\nabla v = 0; D^2 v \neq 0\}$. Observe that the latter surface has dimension at most $n-1$: The matrix $D^2 v$ is symmetric and of trace zero. Certainly $\{\nabla v = 0\}$ has dimension *at most n*, the exact dimension given by the number of independent rows in the matrix $D^2 v$. However, since $D^2 v$ is not trivial, and of trace zero, there must be at least two independent rows. (To see this, simply diagonalize the matrix.)

Norm Inequalities for D

With the maximal density D defined in the most natural way, it is tempting to ask whether the Barlow-Yor inequalities hold for D. That is what we wish to show.

THEOREM. $\|D\|_p \simeq \|A\|_p$ *for* $0 < p < \infty$.

There are two proofs known for this set of inequalities. The first, given in [18], relied on the Barlow-Yor inequalities together with a probabilistic argument relating the local time of the martingale $u(B_t)$ to the area integral. The second proof [23] relies on a technique of Barlow and Yor and the machinery of the Calderón-Zygmund theory. It is this method that we present here. The proof is not quite perfected. For example, the good-λ inequalities proved by Barlow and Yor for L^* have escaped us, so that the most general integral inequalities of the form

$$\int \Phi(D)(x_0)\, dx_0 \simeq \int \Phi(A)(x_0)\, dx_0$$

are still open.

The original selling point for the D function (see [18]) was a ratio theorem of the form

$$\|A^2(u)/N(u)\|_p \simeq \|A\|_p, \qquad 0 < p < \infty.$$

One direction is easy: Write

$$(A/N^{1/2})(N^{1/2}) = A.$$

and use Schwarz' inequality to obtain

$$\|A\|_p^p \leq \|A^2/N\|_p^{p/2} \times \|N\|_p^{p/2}.$$

Then use $\|N\|_p \leq C_p\|A\|_p$ to arrive at the inequality

$$\|A\|_p \leq C_p\|A^2/N\|_p.$$

The other direction uses the pointwise relation

$$A^2(u) = \int_{-N(u)}^{N(u)} D(u;r)\, dr$$

and its consequence $A^2 \leq 2N \cdot D$. This means that $\|A^2/N\|_p \leq 2\|D\|_p$; the problem now is to prove that $\|D\|_p \leq C_p\|A\|_p$.

The study of ratios of the functionals A and N was continued in **[13]**, using the technique of the good-λ inequality. However, it was always necessary in **[13]** to require a larger cone aperture for N than for A, so that the statement of the norm inequalities is not as succinct as above. The D functional, like local time, seems to measure another aspect of the behavior of u. Recall that, in the previous section we gave the Brossard-Chevalier characterization of the Zygmund class $L \log L$ using the local time. The same authors show that $D(u;0)$ *belonging to* $L \log L$ *characterizes* $L \log L$ *as a subspace of* H^1 **[6]**.

So where are we now? We wish to prove the main theorem of this section. We have seen, in the examination of ratios, that for $0 < p < \infty$,

$$\|A\|_p \leq C_p\|A^2/N\|_p \leq 2C_p\|D\|_p.$$

Therefore, our main task is to prove that $\|D\|_p \leq C_p\|A\|_p$.

When confronted with the problem of proving the boundedness of D, the first problem is to choose the appropriate definition of D. The two equivalent definitions are not symmetrical in all respects: viewing $D(u;r)$ via the coarea formula seems to lead to a dead end. The Laplacian definition, however, opens the possibility of using Green's theorem. The drawback of this approach is clear, in that we are limited to harmonic functions. It would be interesting, in the light of everything we know about area integrals, **[14, 9]** to develop a proof of the boundedness of D that does not require $u(x,y)$ to be the harmonic extension of a distribution f defined on $\mathbf{R}^n$.

Let us pass to the details of the proof. It is convenient to work with a smoothed version of the area integral. Let $\psi(x)$ be a smooth, nonnegative, radial function on $\mathbf{R}^n$. We assume that $\psi(x)$ decreases monotonically as $|x|$ increases $\psi(x) \geq \frac{1}{2}\psi(0)$ when $|x| = \frac{3}{4}$ and that $\psi(x)$ has integral one. For $y > 0$, let $\psi_y(x) = y^{-n}\psi(x/y)$. Our version of the Lusin area integral is then

$$A^2(u)(x_0) = \int\int \psi_y(x_0 - x)y|\nabla u|^2(x,y)\,dx\,dy$$

with associated densities

$$D(u;r)(x_0) = \int\int \psi_y(x_0 - x)y\Delta(u - r)^+(dx\,dy).$$

Here we have an abuse of notation with D defined with respect to cones and with respect to the smooth approximation of the identity ψ. As will become clear in the proof, the precise shape or spread of ψ is not important, in so far as the support of ψ is compact. Thus D(conical) can be majorized by D(smooth) and conversely. Therefore, the boundedness of D is independent of the exact shape or spread of the approximate identity that defines it.

Local Estimates

The exposition is based on [23]; however, since that paper was written, Murai and Uchiyama [28] have improved the technique for doing such inequalities.

We introduce W_0, a subset of $\mathbf{R}^n$, and associated with it the union of cones $W = \bigcup_{x \in W_0} \Gamma(x)$ and also the union of larger cones $W_\alpha = \bigcup_{x \in W_0} \Gamma_\alpha(x)$ with $\alpha > 1$ fixed once and for all. In the rest of this section we work with the "cut down" functions $N_\alpha(W_\alpha)(x_0) = \sup_{\Gamma_\alpha(x_0)} |u| I_{W_\alpha}$,

$$D(W,t)(x_0) = \int\int \psi_y(x_0 - x) I_W(x, y) y \Delta(u-t)^+(dx, dy)$$

with I_{W_α}, I_W denoting the indicators (characteristic functions to nonprobabilists) of W, W_α.

Lemma 1. *For each $t \in \mathbf{R}$ and for $1 < p < \infty$*

$$\int (D(W,t)(x))^p \, dx \leq C \int (N_\alpha(W_\alpha)(x))^p \, dx$$

with $C > 0$ depending only on p and the dimension n.

Proof. We can obtain the inequality as a consequence of the method of Murai and Uchiyama [28]. They introduced a new technique for obtaining the good-λ inequality that significantly improved the estimates given in [13]. (The method allowed Uchiyama to complete McConnell's subharmonic function inequality. See [34].) We must establish a good-λ inequality of the form

(1) $\quad m(D(W,t) > \beta\lambda, N_\alpha(W_\alpha) \leq \delta\lambda) \leq c \exp(-c\beta/\delta) m(D(W,t) > \lambda)$

for $\delta > 0$ and $\beta > 1$ sufficiently large. For a fixed $p > 1$, this inequality implies

$$
\begin{aligned}
(2)\quad \int |D(W,t)|^p &= \beta^p \int \left|\frac{D(W,t)}{\beta}\right|^p \\
&\le \beta^p p \int_0^\infty \lambda^{p-1} m(D(W,t) > \beta\lambda, N_\alpha(W_\alpha) \le \delta\lambda)\, d\lambda \\
&\quad + \beta^p p \int_0^\infty \lambda^{p-1} m(N_\alpha(W_\alpha) > \delta\lambda)\, d\lambda \\
&\le C \exp\left(-\left(c\frac{\beta}{\delta}\right)\right) p \int_0^\infty \lambda^{p-1} m(D(W_t) > \lambda)\, d\lambda \\
&\quad + \left(\frac{\beta}{\delta}\right)^p \int |N_\alpha(W_\alpha)|^p \\
&= c \exp\left(-\left(c\frac{\beta}{\delta}\right)\right) \beta^p \int |D(W,t)|^p + C\left(\frac{\beta}{\delta}\right)^p \int |N_\alpha(W_\alpha)|^p
\end{aligned}
$$

If β is large, $c\exp(-c\beta/\delta)\beta^p < 1/2$, so that, if the integrals are all finite, we may subtract the right-hand side from the left to finish the proof. In general we can truncate W and replace u to be a harmonic function with boundary function in $C^\infty_{\text{com}}(\mathbf{R}^n)$, so that the right side of (2) is finite, argue as above, and then pass to the limit to establish (5) for the given u and W. Thus the proposition will follow if we prove (1).

Let Q be an arbitrary cube in $\mathbf{R}^n$ and

$$V_0 = \{x\colon D(W,t) > \beta\lambda, N_\alpha(W_\alpha) \le \delta\lambda\} \cap Q; \quad V = \left\{\bigcup_{x \in V_0} \Gamma(x) \cap W\right\}.$$

We shall prove that $\Delta(u-t)^+$, restricted to $W \cap V$, is a Carleson measure: If $B(Q)$ is the box over Q of height $m(Q)^{1/n}$ (that is, $\{(x,y)\colon x \in Q, 0 < y < m(Q)^{1/n}\}$) then

$$(3)\qquad \Delta(u-t)^+(B(Q) \cap W \cap V) \le c\delta\lambda m(Q).$$

Once we have this result, then it follows that $D(W \cap V, t)$, being the balayage of a Carleson measure, is a BMO function with BMO norm less than $c\delta\lambda$. (See Garnett [15, pp. 229–330].) By an extension of the John-Nirenberg theorem (see Murai-Uchiyama, [28]), we have

$$m(D(W \cap V, t) > \beta\lambda) \le c\exp(-c\beta/\delta) m(D(W \cap V, t) > \lambda).$$

Since $D(W \cap V, t) \le D(W,t)$ in any case, and $D(W \cap V, t) = D(W,t)$ where $N_\alpha(W_\alpha) \le \delta\lambda$ we have obtained the good-λ estimate from the BMO inequality. It remains to show that $\Delta(u-t)^+$, restricted to $W \cap V$, is a Carleson measure. For this we use Green's theorem. Note, first of all, that we may assume that $|t| \le \delta\lambda$: if not, the measure $\Delta(u-t)^+ \equiv 0$ on $W \cap V$. The integration is taken over $B = B(Q)$ intersected with $W \cap V$, so that

$$\int_{B\cap W\cap V} y\Delta(u-t)^{+} = \int_{\partial(B\cap W\cap V)} y\frac{\partial}{\partial\eta}(u-t)^{+} - \frac{\partial y}{\partial\eta}(u-t)^{+}.$$

Here we are applying Green's theorem although some justification is required. The function $(u-t)^{+}$ as well as the boundary of $B\cap W\cap V$ may be smoothed, and the integral computed before passing to the limit. This gives an estimate

$$\begin{aligned} &m(D(W;t) > \beta\lambda, N_\alpha(W_\alpha) \le \delta\lambda, Q) \\ &\qquad \le \frac{2}{\beta\lambda}\int_{\partial\{W\cap V\}} y\left|\frac{\partial}{\partial\eta}(u-t)^{+}\right| + (u-t)^{+}\left|\frac{\partial y}{\partial\eta}\right| d\sigma. \end{aligned}$$

Since $W \subset W_\alpha$, we conclude that $y|\nabla u| \le C\delta\lambda$ on $\partial(W\cap V)$ (This is because $y|\nabla u| \le CN_\alpha(W_\alpha)$ on W; see [30, p. 207].) The smooth function that approximates $(u-a)^{+}$ can be chosen with uniformly bounded derivatives, so that $y|\frac{\partial}{\partial\eta}(u-t)^{+}| \le C\delta\lambda$ on $\partial(W\cap V)$. Also $(u-t)^{+} \le 2\delta\lambda$ since we assume $|t| \le N(W)$ and $|\frac{\partial y}{\partial\eta}| \le C$; we conclude that $(u-t)^{+}|\frac{\partial y}{\partial\eta}| \le C\delta\lambda$. Finally, we see that

$$\int_{\partial(B\cap W\cap V)} d\sigma \le Cm(Q)$$

by the examination of the region. Thus we have shown that $\Delta(u-t)^{+}$ is a Carleson measure. This concludes the proof of the lemma.

The passage from $D(u;r)$ to the supremum functional $D(u)$ is made via an inequality of Garsia, Rodemich, and Rumsey [16]. The idea for this passage, due to Barlow and Yor [2], goes as follows: We first establish a smoothness result of the form

$$|D(u;r)(x) - D(u;s)(x)| \le CB(x)|r-s|^{1/2-2/p}$$

for some $p > 4$, provided r,s are contained in an interval $I = \{r\colon |r| < \lambda\}$. Here all the points x belong to the set $\{x\colon N_\alpha(W_\alpha)(x) \le \lambda\}$ so that $D(u;r)(x) \equiv 0$ outside of I. Now the function $B(x)$ is defined as follows:

$$B^p(x) = \int\!\!\int_{I\times I}\left|\frac{D(u;r)(x) - D(u;s)(x)}{|r-s|^{1/2}}\right|^p dr\,ds.$$

It turns out that $B^p(x)$ is integrable over the set $\{N_\alpha(W_\alpha) \le \lambda\}$. The GRR inequality should be understood as follows: If $D(u;r)(x)$ were Hölder continuous of order $\frac{1}{2}$, then clearly $B^p(x) = C(x)\cdot|I|^2 < \infty$. If $B^p(x)$ satisfies this estimate, however, we have no guarantee that $D(u;r)(x)$ is Hölder continuous as a function of r. The GRR inequality tells us that it just misses by the amount $\varepsilon = 2/p$. We estimate the supremum functional $D(u)(x)$ by its maximum oscillation, since it vanishes off I. Thus

$$D(u)(x) \le CB(x)|I|^{1/2-2/p} = CB(x)\lambda^{1/2-2/p}$$