Contents of volume LX, number 2

The journal prints original papers in English, French, German, and Russian, mainly on functional analysis, abstract methods of mathematical analysis and on the theory of probabilities. Usually 3 issues constitute a volume.

The papers submitted should be typed on one side only and they should be accompanied by abstracts, normally not exceeding 200 words. The authors are requested to send two copies, one of them being the typed, not Xerox copy. Authors are advised to retain a copy of the paper submitted for publication.

Manuscripts and the correspondence concerning editorial work should be addressed to

STUDIA MATHEMATICA
ul. Śniadeckich 8
00–950 Warszawa, Poland

Correspondence concerning exchange should be addressed to:
INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
ul. Śniadeckich 8
00–950 Warszawa, Poland

The journal is available at your bookseller or at
ARS POLONA
Krakowskie Przedmieście 7
00–068 Warszawa, Poland

PRINTED IN POLAND

WROCŁAWSKA DRUKARNIA NAUKOWA

Inequalities for the ergodic maximal function*

by

ROGER L. JONES (Chicago)

Abstract. A decomposition introduced by Kakutani is used to give an ergodic theory analog to the classical Calderón–Zygmund decomposition. This decomposition is first used to prove certain known results. These proofs show the strong relationship between classical results on the real line, and ergodic theory results. The decomposition is also used to study the ergodic square function. This square function is an analogue to the martingale square function, and a generalization of it in certain cases. Many properties of the square function are studied.

Introduction. On $\boldsymbol{R}^n$, the Hardy–Littlewood maximal function, Mf, is defined by

$$Mf(x) = \sup_{Q\ni x} \frac{1}{|Q|} \int_Q f(t)\,dt, \qquad f \geqslant 0.$$

This operator has played an important role in the study of other operators on $\boldsymbol{R}^n$.

We consider an ergodic measure preserving transformation T, acting on a probability space (X, Σ, m). In this setting, the ergodic maximal function, defined by

$$f^*(x) = \sup_n \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x), \qquad f \geqslant 0,$$

may be expected to play the role of Mf on $\boldsymbol{R}^n$. Using a decomposition of the space X, introduced by Kakutani, we obtain a decomposition of functions, analogous to the Calderón–Zygmund decomposition on $\boldsymbol{R}^n$. With this decomposition, the methods used by Calderón and Zygmund can be adapted to problems arising in ergodic theory. In particular, proofs of the maximal ergodic theorem, and Ornstein's $L\log^+L$ theorem on the integrability of f^*, become especially clear when viewed in this way.

* The work presented here is contained in the author's Ph. D. thesis written under Professor Richard F. Gundy at Rutgers University.

We also introduce the ergodic square function, defined by

$$S(f)(x) = \Big[\sum_{n=0}^{\infty}\big(f_{n+1}(x)-f_n(x)\big)^2\Big]^{1/2},$$

where $f_n(x) = \frac{1}{n}\sum_{k=0}^{n-1} f(T^n x)$, $(f_0 \equiv 0)$. This operator arose from the study of certain martingales, where it has played an important role. We first note, by example, that no pointwise inequality can occur between $S(f)$ and f^*. We then prove that $S(f)$ and f^* are related by a Φ inequality. That is,

$$\int \Phi\big(S(f)(x)\big)\,dm(x) \leqslant C_\Phi \int \Phi\big(f^*(x)\big)\,dm(x).$$

We can take $\Phi(\lambda) = \lambda^p$, showing $\|S(f)\|_p \leqslant C_p\|f^*\|_p$, $0 < p < \infty$. The proof uses the Kakutani construction as an analogue to the Calderón–Zygmund decomposition. It is related to the proof by Coifman [3] that the Hilbert transform is related to the Hardy–Littlewood maximal function by such a Φ inequality.

These methods can also be used to study the ergodic Hilbert transform, defined as

$$\lim_{n\to\infty}\sum_{k=1}^{n}\frac{f(T^k x)-f(T^{-k}x)}{k},$$

and probably other operators.

Preliminaries. Let (X, Σ, m) denote a nonatomic, complete probability space, and T a point transformation mapping X onto itself. In this connection, we use the following notation. The transformation T is *measurable* if $A \in \Sigma$ implies $T^{-1}(A) \in \Sigma$, where $T^{-1}(A) = \{x \mid Tx \in A\}$. Here and below, sets *A and B are equal* if they agree up to a set of measure zero. The transformation is *invertible* if there is a measurable transformation S, such that $TS(A) = ST(A) = A$. In this case, S is unique and is denoted by T^{-1}. A set A is called *invariant* if $T(A) = A$; if the only invariant sets are X and φ, then T is said to be *ergodic*. Finally, T is *measure preserving* if $m(A) = m(T^{-1}A)$. Our transformations are measurable, invertible, ergodic, and measure preserving.

Some of the theorems remain true under fewer hypotheses, but the modifications are usually clear. For example, let $X = X_1 \cup X_2$, with X_1 and X_2 disjoint invariant sets of positive measure. Assume T is ergodic on X_1 and X_2, separately. Then T is not ergodic on X, but the results can be applied to X_1 and X_2, separately.

The following two decompositions play a major role in what follows.

THEOREM 1.1 (The Kakutani decomposition). *If T is ergodic, invertible, and measure preserving, and X is a probability space, then given B, with $m(B) > 0$, there exist disjoint sets*

$$B_i^j, \quad 0 \leqslant i \leqslant j-1,\ 1 \leqslant j < \infty,$$

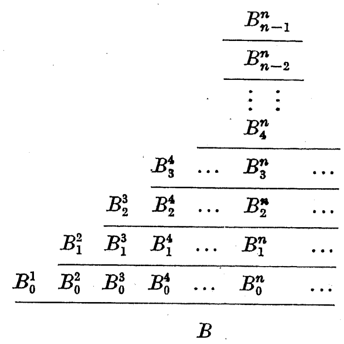
such that

$$B = \bigcup_{j=1}^{\infty} B_0^j,$$

$$X = \bigcup_{j=1}^{\infty}\bigcup_{i=0}^{j-1} B_i^j,$$

and

$$T(B_i^j) = B_{i+1}^j \quad \text{unless } i = j-1.$$



THEOREM 1.2 (The Calderón–Zygmund decomposition). *Given f in $L^1(\mathbf{R}^n)$, $f \geqslant 0$ and $\lambda > \|f\|_1$, there exists a collection of non-overlapping cubes $\{Q_N\}_{N=1}^{\infty}$ such that*

(1) *for a.e. $x \notin \bigcup_{N=1}^{\infty} Q_N$, $f(x) < \lambda$;*

(2) *there exist c_1, c_2, independent of f, λ, and Q_N, such that*

$$c_1\lambda \leqslant \frac{1}{|Q_N|}\int_{Q_N} f(x)\,dx \leqslant c_2\lambda.$$

The maximal function. We define the ergodic maximal function by

$$f^*(x) = \sup_{n>0}\frac{1}{n}\sum_{K=0}^{n-1}|f(T^K x)|.$$

This maximal function should be compared to the Hardy–Littlewood maximal function, Mf defined by

$$Mf(x) = \sup_{Q\ni x} \frac{1}{|Q|}\int_Q f(x)\,dx$$

for $f \geqslant 0$. We establish a decomposition, using f^*, similar to the one given by Calderón–Zygmund using Mf.

THEOREM 2.1. *Given f in $L^1(X)$, $f \geqslant 0$, and $\lambda > \|f\|_1$, there exists a collection of disjoint sets $\{C_N\}_{N=2}^{\infty}$ such that*

(1) *for a.e. $x \notin \bigcup_{N=2}^{\infty} C_N$, $f(x) \leqslant \lambda$;*

(2) *for each N, $\lambda \leqslant \frac{1}{m(C_N)}\int_{C_N} f(x)\,dm(x) \leqslant 2\lambda$.*

Proof. Let

$$B = \{x \in X \mid f^*(x) \leqslant \lambda\}.$$

With this B, we form the Kakutani decomposition (Theorem 1.1). Let

$$C_N = \bigcup_{j=1}^{N-1} B_j^N = \bigcup_{j=1}^{N-1} T^j B_0^N.$$

These C_N correspond to the cubes Q_N of the Calderón–Zygmund decomposition (Theorem 1.2).

If $x \notin \bigcup_{N=2}^{\infty} C_N$, then $x \in B$ so that $f^*(x) \leqslant \lambda$, and, in particular, $f(x) \leqslant \lambda$. The first assertion is proved.

LEMMA 2.2. *For each C_N,*

$$\frac{1}{m(C_N)}\int_{C_N} f(x)\,dm(x) \leqslant 2\lambda.$$

Proof. If $x \in B_0^N$, then $f^*(x) \leqslant \lambda$. Consequently,

$$\sum_{K=0}^{N-1} f(T^K x) < N\lambda.$$

Since $f \geqslant 0$, this tells us that

$$\sum_{K=1}^{N-1} f(T^K x) \leqslant N\lambda.$$

We now use this estimate to prove the lemma. Recall that

$$C_N = \bigcup_{K=1}^{N-1} B_K^N.$$

Consequently,

$$\frac{1}{m(C_N)}\int_{C_N} f(x)\,dm(x) = \frac{1}{(N-1)\,m(B_0^N)}\sum_{K=1}^{N-1}\int_{B_K^N} f(x)\,dm(x)$$

$$= \frac{1}{(N-1)\,m(B_0^N)}\sum_{K=1}^{N-1}\int_{T^K B_0^N} f(x)\,dm(x)$$

$$= \frac{1}{(N-1)\,m(B_0^N)}\int_{B_0^N}\sum_{K=1}^{N-1} f(T^K x)\,dm(x)$$

$$\leqslant \frac{1}{(N-1)\,m(B_0^N)}\int_{B_0^N}\lambda N\,dm(x) \leqslant 2\lambda.$$

This proves the right-hand inequality in the second assertion of the theorem.

LEMMA 2.3. *For each N,*

$$\frac{1}{m(C_N)}\int_{C_N} f(x)\,dm(x) \geqslant \lambda.$$

Proof. The notation is neater if we replace f by $g = f-\lambda$. Then

$$B = \{x \in X \mid \sup_{n>0}\frac{1}{n}\sum_{K=0}^{n-1} g(T^K x) \leqslant 0\}.$$

We need to show

$$\frac{1}{m(C_N)}\int_{C_N} g(x)\,dm(x) \geqslant 0,$$

or equivalently,

$$\int_{C_N} g(x)\,dm(x) \geqslant 0.$$

It is sufficient to show

$$\sum_{K=1}^{N-1} g(T^K x) \geqslant 0 \quad \text{for a.e. } x \in B_0^N,$$

since then

$$\int_{C_N} g(x)\,dm(x) = \int_{B_0^N}\sum_{K=1}^{N-1} g(T^K x)\,dm(x) \geqslant 0.$$

By construction, we have the following two facts:

(1) For $x \in B$, we have $\sum_{K=0}^{n} g(T^K x) \leqslant 0$ for all n.

(2) For $x \notin B$, there exists an integer $I = I(x)$ such that $\sum_{K=0}^{I} g(T^K x) > 0$, and I is the smallest such integer.

We first show that if $x \in B_{N-1}^N$ (the top of column C_N), then $g(x) > 0$. This follows from the fact that $T(x) \in B$ and hence

$$\sum_{K=0}^{n} g\big(T^K(Tx)\big) \leqslant 0 \quad \text{for all } n \geqslant 0.$$

If $g(x) \leqslant 0$, then we have

$$\sum_{K=0}^{I} g(T^K x) = g(x) + \sum_{K=1}^{I} g(T^K x) = g(x) + \sum_{K=0}^{I-1} g\big(T^K(Tx)\big) \leqslant 0.$$

This is a contradiction to the definition of I. Consequently, we must have $g(x) > 0$.

More generally, for any x in the column, we see, by a similar argument, that the sum

$$g(x) + g(Tx) + \ldots + g(T^I x)$$

is positive, and $T^I(x)$ is in the column.

Let $x \in B_1^N$, then

$$\sum_{K=0}^{I} g(T^K x) > 0.$$

Since $T^I(x)$ is in the column, it is either in the top or $T^{I+1}(x)$ is also in the column. If $T^I(x)$ is in the top, then we are done. Otherwise, sum from $T^{I+1}(x)$. Continue until we reach the top. We than have the desired conclusion.

THEOREM 2.4 (The maximal ergodic theorem). *If $f \in L^1(X)$, $f \geqslant 0$, and $\lambda > \|f\|_1$, then*

$$m\{f^* > \lambda\} \leqslant \frac{1}{\lambda} \int_{\{f^* > \lambda\}} f(x)\, dm(x).$$

Proof. By Lemma 2.3, we have

$$\frac{1}{m(C_N)} \int_{C_N} f(x)\, dm(x) \geqslant \lambda$$

or

$$\frac{1}{\lambda} \int_{C_N} f(x)\, dm(x) \geqslant m(C_N).$$

Recalling that $B = \{f^* \leqslant \lambda\}$, we have

$$B^C = \bigcup_{N=2}^{\infty} C_N = \{f^* > \lambda\}.$$

Consequently, by summing over N, we get

$$m\{f^* > \lambda\} = \sum_{N=2}^{\infty} m(C_N) \leqslant \sum_{N=2}^{\infty} \frac{1}{\lambda} \int_{C_N} f(x)\, dm(x) \leqslant \frac{1}{\lambda} \int_{\{f^* > \lambda\}} f(x)\, dm(x).$$

THEOREM 2.5 (Ornstein [4]). *For $f \in L^1(X)$, $f \geqslant 0$, the maximal function f^* is integrable if and only if $f \log^+ f$ is integrable, where*

$$f(x) \log^+ f(x) = f(x) \log \max\big(f(x), 1\big).$$

This theorem was first proved in this setting by D. S. Ornstein [4]. The proof given here is related to his, but follows an argument by Stein [5], who proves an analogous result for Mf. (Recall Mf is the Hardy–Littlewood maximal function.)

Proof. We first prove a distribution inequality and integrate to obtain the result.

LEMMA 2.6. *For $\lambda > \|f\|_1$, we have*

$$\frac{1}{\lambda} \int_{\{f > \lambda\}} f(x)\, dm(x) \leqslant 2m\{f^* > \lambda\}.$$

Proof. Recalling that

$$\bigcup_{N=2}^{\infty} C_N = \{f^* > \lambda\}$$

and that the C_N are disjoint, we have

$$\sum_{N=2}^{\infty} \int_{C_N} f(x)\, dm(x) = \int_{\bigcup_{N=2}^{\infty} C_N} f(x)\, dm(x) = \int_{\{f^* > \lambda\}} f(x)\, dm(x).$$

By Lemma 2.2 and the above, we have

$$\int_{\{f^* > \lambda\}} f(x)\, dm(x) = \sum_{N=2}^{\infty} \int_{C_N} f(x)\, dm(x) \leqslant \sum_{N=2}^{\infty} 2\lambda m(C_N) = 2\lambda m\{f^* > \lambda\}.$$

Dividing by λ, we have

$$\frac{1}{\lambda} \int_{\{f^* > \lambda\}} f(x)\, dm(x) \leqslant 2m\{f^* > \lambda\}.$$

The result now follows by replacing $\{f^*>\lambda\}$, on the left side, by the smaller set, $\{f>\lambda\}$.

We now integrate the distribution inequality of Lemma 2.6, obtaining half of Theorem 2.5:

$$\begin{aligned}\int_X f^*(x)\,dm(x) &= \int_0^\infty m\{f^*>\lambda\}\,d\lambda\\ &= \int_0^{\|f\|_1} m\{f^*>\lambda\}\,d\lambda + \int_{\|f\|_1}^\infty m\{f^*>\lambda\}\,d\lambda\\ &\geqslant \|f\|_1 + \int_{\|f\|_1} \frac{1}{2\lambda}\int_{\{f>\lambda\}} f(x)\,dm(x)\,d\lambda\\ &= C + \int_1^\infty \frac{1}{2\lambda}\int_{\{f>\lambda\}} f(x)\,dm(x)\,d\lambda\\ &= C + \frac{1}{2}\int_X f(x)\int_1^{f(x)} \frac{1}{\lambda}\,d\lambda\,dm(x)\\ &\geqslant C + \frac{1}{2}\int_X f(x)\log^+ f(x)\,dm(x).\end{aligned}$$

To prove the other half of Theorem 2.5, we use the maximal ergodic theorem and a truncation argument. Write

$$f \leqslant f1_{\{f>\lambda\}} + \lambda 1_{\{f\leqslant\lambda\}},$$

and note

$$f^* \leqslant (f1_{\{f>\lambda\}})^* + \lambda.$$

Consequently,

$$\begin{aligned}\lambda m\{f^*>2\lambda\} &\leqslant \lambda m\{(f1_{\{f>\lambda\}})^*>\lambda\}\\ &\leqslant \int_{\{(f1_{\{f>\lambda\}})^*>\lambda\}} f1_{\{f>\lambda\}}\,dm(x) \leqslant \int_{\{f>\lambda\}} f(x)\,dm(x).\end{aligned}$$

We have obtained a distribution inequality,

$$\lambda m\{f^*>2\lambda\} \leqslant \frac{1}{\lambda}\int_{\{f>\lambda\}} f(x)\,dm(x).$$

If we integrate this inequality, we find that $f^*\in L^1$ if f belongs to $L\log^+L$.

Remark. This theorem requires the assumption that $m(X)<\infty$. If $M(X)=\infty$, then Ornstein [4] has shown that $f\geqslant 0$ and $\int f(x)\,dm(x)\neq 0$ imply f^* is not integrable. In Stein's paper [5] he considers the space to be a compact subset of $\boldsymbol{R}^n$.

The square function. Recall that a *martingale* is a sequence of functions $f=(f_0,f_1,\ldots)$, where $E(f_n\|f_0,f_1,\ldots,f_{n-1})=f_{n-1}$. The *martingale square function*, S_Mf, is defined by

$$S_Mf(x) = \Big[\sum_{n=0}^\infty (f_{n+1}^{(x)}-f_n^{(x)})^2\Big]^{1/2}.$$

The *martingale maximal function*, f_M^*, is defined by

$$f_M^*(x) = \sup_n |f_n(x)|.$$

Burkholder and Gundy [2] have shown that these two martingale operators are closely related. They have shown that for a large class of martingales, there exist constants c_p and C_p such that

$$c_p\|f_M^*\|_p \leqslant \|S_Mf\|_p \leqslant C_p\|f_M^*\|_p, \qquad 0<p<\infty.$$

This result has had important implications in the theory of H_p spaces of classical function theory.

In ergodic theory we are interested in the averages

$$f_n(x) = \frac{1}{n}\sum_{K=0}^{n-1} f(T^kx).$$

These averages form a martingale in the special case where T is an independent shift. The obvious generalization of the martingale square function is the ergodic square function defined by

$$S(f)(x) = \Big[\sum_{n=0}^\infty (f_{n+1}(x)-f_n(x))^2\Big]^{1/2},$$

where the f_n are the Cesàro averages defined above.

We would like to extend the martingale results to this setting. Unfortunatly, this can only be a partial success. In particular, the inequality $\|f^*\|_1\leqslant C\|S(f)\|_1$, which is true in the martingale setting, is easily shown to be false in this setting. However, many things are true about this square function.

If we let $d_n(x)=f_{n+1}(x)-f_n(x)$, then it is easy to see that $dn = \frac{1}{n+1}f_n(x)+\frac{f(T^nx)}{n+1}$.

Consequently,

$$S(f)(x) = \Big[\sum_{n=0}^{\infty} \big(d_n(x)\big)^2\Big]^{1/2}$$

$$\leqslant \Big(\sum_{n=0}^{\infty}\Big(\frac{f_n(x)}{n+1}\Big)^2 + 2\sum_{n=0}^{\infty}\Big(\frac{1}{n+1}\Big)^2 f_n(x) f(T^n x)\Big)^{1/2} + \Big[\sum_{n=0}^{\infty}\Big(\frac{(f(T^n x)}{n+1}\Big)^2\Big]^{1/2}$$

$$\leqslant Cf^*(x) + S^*f(x),$$

where

$$S^*f(x) = \Big[\sum_{n=0}^{\infty}\Big(\frac{f(T^n x)}{n+1}\Big)^2\Big]^{1/2}.$$

Since the operator f^* is well understood, it is usually enough to study $S^*f(x)$.

THEOREM 3.1. *The square function is weak type* (1, 1). *That is, for* $\lambda > \|f\|_1$, *we have*

$$m\{S(f) > \lambda\} \leqslant \frac{c}{\lambda}\|f\|_1.$$

Proof. First note that

$$m\{S(f) > \lambda\} \leqslant m\{Cf^* > \lambda/2\} + m\{S^*f > \lambda/2\}.$$

Since f^* is weak type (1, 1) it is enough to show that S^*f is weak type (1, 1).

Proof that S^*f is weak type (1, 1). Assume $f \geqslant 0$, and let $f_k(x) = f(T^k x)$. Define $g_k(x)$ by

$$g_k(x) = \begin{cases} f_k(x) & \text{if} \quad f_k(x) \leqslant \lambda(k+1), \\ 0 & \text{if} \quad f_k(x) > \lambda(k+1). \end{cases}$$

Define $b_k(x)$ by

$$b_k(x) = f_k(x) - g_k(x).$$

Then

$$S^*(f)(x) = \Big[\sum_{k=0}^{\infty}\Big(\frac{f_k(x)}{k+1}\Big)^2\Big]^{1/2} = \Big[\sum_{k=0}^{\infty}\Big(\frac{g_k(x)}{k+1} + \frac{b_k(x)}{k+1}\Big)^2\Big]^{1/2}$$

$$\leqslant \Big[\sum_{k=0}^{\infty}\Big(\frac{g_k(x)}{k+1}\Big)^2\Big]^{1/2} + \Big[\sum_{k=0}^{\infty}\Big(\frac{b_k(x)}{k+1}\Big)^2\Big]^{1/2}.$$

Consequently,

$$m\{S^*(f) > \lambda\} \leqslant m\Big\{\Big[\sum_{k=0}^{\infty}\Big(\frac{g_k(x)}{k+1}\Big)^2\Big]^{1/2} > \lambda/2\Big\} + m\Big\{\Big[\sum_{k=0}^{\infty}\Big(\frac{b_k(x)}{k+1}\Big)^2\Big]^{1/2} > \lambda/2\Big\}.$$

We study these two pieces separately. For the first piece we have

$$m\Big\{\Big[\sum_{k=0}^{\infty}\Big(\frac{(g_k(x)}{k+1}\Big)^2\Big]^{1/2} > \lambda/2\Big\} \leqslant \frac{c}{\lambda^2}\int_X \sum_{k=0}^{\infty}\Big(\frac{g_k(x)}{k+1}\Big)^2 dm(x)$$

$$= \frac{c}{\lambda^2}\sum_{k=0}^{\infty}\Big(\frac{1}{k+1}\Big)^2 \int_X g_k(x)^2 dm(x)$$

$$= \frac{c}{\lambda^2}\sum_{k=0}^{\infty}\Big(\frac{1}{k+1}\Big)^2 \int_0^{\infty} am\{g_k > a\}\,da$$

$$\leqslant \frac{c}{\lambda^2}\sum_{k=0}^{\infty}\Big(\frac{1}{k+1}\Big)^2 \int_0^{\lambda(k+1)} am\{f > a\}\,da$$

$$\leqslant \frac{c}{\lambda^2}\int_0^{\infty} a \sum_{k=[a/\lambda]}^{\infty}\Big(\frac{1}{k+1}\Big)^2 m\{f > a\}\,da$$

$$\leqslant \frac{c}{\lambda^2}\int_0^{\infty} a\Big(\frac{\lambda}{a}\Big) m\{f > a\}\,da$$

$$= \frac{c}{\lambda}\int_0^{\infty} m\{f > a\}\,da = \frac{c}{\lambda}\|f\|_1.$$

For the second term we have

$$m\Big\{\Big[\sum_{k=0}^{\infty}\Big(\frac{b_k(x)}{k+1}\Big)^2\Big]^{1/2} > \lambda/2\Big\} \leqslant m\Big\{\sum_{k=0}^{\infty}\Big(\frac{b_k(x)}{k+1}\Big)^2 > 0\Big\}$$

$$\leqslant \sum_{k=0}^{\infty} m\{b_k > 0\} = \sum_{k=0}^{\infty} m\{f > \lambda(k+1)\}$$

$$\leqslant \int_0^{\infty} m\{f > \lambda a\}\,da = \frac{1}{\lambda}\int_0^{\infty} m\{f > a\}\,da = \frac{1}{\lambda}\|f\|_1.$$

Combining the two pieces, we obtain the desired resu't. This weak type inequality implies that $S(f)$ is finite a.e. for $f \in L^1(X)$.

Theorem 3.2. *For $1 < p \leqslant \infty$ there exists a constant c_p, depending only on p, such that*

$$\|S(f)\|_p \leqslant c_p \|f\|_p.$$

Proof. Except for a set of measure zero, we have

$$S^*(f)^2(x) = \sum_{k=0}^{\infty}\left(\frac{f(T^k x)}{k+1}\right)^2 \leqslant \sum_{k=0}^{\infty}\left(\frac{\|f\|_\infty}{k+1}\right)^2$$

$$\leqslant \|f\|_\infty^2 \sum_{k=0}^{\infty}\left(\frac{1}{k+1}\right)^2 = c_p \|f\|_\infty^2.$$

Consequently, $\|S(f)\|_\infty \leqslant c_p\|f\|_\infty$. Combining this with 3.1 and applying the Marcinkiewicz interpolation theorem, we have the result.

Remark. There is no constant c such that

$$S(f)(x) \leqslant cf^*(x)$$

for all f in $L^1(x)$. This can be seen by the following example. Choose N very large. Find a set A such that $A, TA, \ldots, T^{N-1}A$ are pairwise disjoint. We can always do this by Rokhlin's theorem.

We define $f_N(x)$ as follows:

$$f_N(x) = \begin{cases} 2^n & \text{if } \quad T^{-(2^n-1)}(x) \in A,\ 1 \leqslant 2^n - 1 < N, \\ 0 & \text{otherwise.} \end{cases}$$

For $x \in A$, we estimate $f^*(x)$ by

$$f^*(x) = \sup_n \frac{1}{n}\sum_{k=0}^{n-1} f(T^k x) = \sup_{\substack{n\\ 2^n < N+1}} \frac{1}{2^n}\sum_{k=1}^{n} f(T^{2^k-1}x)$$

$$= \sup_{\substack{n\\ 2^n < N+1}} \frac{1}{2^n}\sum_{k=1}^{n} 2^k = \frac{2^{n+1}}{2^n} = 2.$$

However, if $2^m \leqslant N < 2^{m+1}$, then we can estimate $S^*(f)(x)$, for x in A, by

$$S(f)(x) = \left[\sum_{n=0}^{\infty}\big(f_{n+1}(x) - f_n(x)\big)^2\right]^{1/2} \geqslant \left[\sum_{n=0}^{2m}\big(f_{n+1}(x) - f_n(x)\big)^2\right]^{1/2}$$

$$\geqslant \left[\sum_{n=0}^{m}\big(f_{2^n}(x) - f_{2^{n-1}}(x)\big)^2\right]^{1/2} \geqslant \sum_{n=0}^{m}\left(\frac{2^{n+1}}{2^n} - \frac{2^n}{2^n}\right)^2\Bigg]^{1/2} = \sqrt{m}.$$

Therefore, the ratio $S(f)/f^*$ is unbounded.

The same example also shows that no pointwise inequality can exist between $S^*(f)(x)$ and $f^*(x)$.

The non-existence of a pointwise inequality still does not mean that $S^*(f)$ is much larger than f^* in some sense. Specifically, the set where $S^*(f)$ is much larger than f^* may have very small measure. For example, consider the case of the martingale square function and maximal function.

Let Φ be any continuous nondecreasing real valued function on $[0, \infty)$ satisfying $\Phi(0) = 0$ and the growth condition $\Phi(2\lambda) \leqslant c\Phi(\lambda)$. Burkholder and Gundy [2] have shown

$$\int \Phi\big(S_M f(x)\big)\,dm(x) \leqslant c_\Phi \int \Phi\big(f_M^*(x)\big)\,dm(x).$$

However, again, it is easy to construct examples where for any given c, $S_M f(x) > cf_M^*(x)$.

In their proof of the above fact, Burkholder and Gundy use martingale stopping times as an analogue of the Calderón–Zygmund decomposition. They prove a joint distribution function inequality for $S_M f$ and f_M^*. This distribution inequality is then integrated to obtain the integral inequalities.

We now show an analogous result for $S(f)$ and f^*.

Theorem 3.3. *For $f \geqslant 0$, we have*

$$\int \Phi\big(S(f(x))\big)\,dm(x) \leqslant c_\Phi \int \Phi\big(f^*(x)\big)\,dm(x)$$

for all continuous nondecreasing Φ with $\Phi(0) = 0$ and $\Phi(2\lambda) \leqslant c\Phi(\lambda)$. In particular, $\Phi(\lambda) = \lambda^p$ gives $\|S(f)\|_p \leqslant c_p\|f^*\|_p$.

Proof. To prove this result, we introduce a two sided maximal function, A^*f, defined by

$$A^*f(x) = \sup_{\substack{n\neq 0\\ -\infty<n<\infty}} \frac{1}{|n|+1}\sum_{k=0}^{n} |f(T^k x)|.$$

Actually, we prove the above result with f^* replaced by A^*f, and then show that

$$\int \Phi\big(A^*f(x)\big)\,dm(x) \leqslant c_\Phi \int \Phi\big(f^*(x)\big)\,dm(x).$$

Lemma 3.4. *There exists a constant c such that for all $\lambda > 0$, all $\beta > 2$ and all $\delta < 1$, we have*

$$m\{S^*(f) > \beta\lambda,\ A^*f < \delta\lambda\} \leqslant \frac{c\delta}{\beta}\, m\{S^*(f) > \lambda\}.$$

We note that this lemma is the basic step that leads to the theorem. Once the lemma is proved, we can obtain the Φ inequalities by using the following extension of a method developed by Burkholder and Gundy [2].

Lemma 3.5 (Burkholder [1]). *Let Φ be a continuous, non-decreasing function on $[0,\infty)$, with $\Phi(0)=0$ and $\Phi(2\lambda)\leqslant c\Phi(\lambda)$. Let f and g be non-negative measurable functions on a probability space (X,Σ,m). Assume $\beta>1$, $\delta>0$ and $\varepsilon>0$ are real numbers such that*

$$m\{g>\beta\lambda,\ f>\delta\lambda\}\leqslant m\{g>\lambda\},\qquad \lambda>0.$$

Let γ and η be real numbers satisfying

$$\Phi(\beta\lambda)\leqslant\gamma\Phi(\lambda),\qquad \Phi(\delta^{-1}\lambda)\leqslant\eta\Phi(\lambda),\qquad \lambda>0.$$

Finally, suppose that $\gamma\varepsilon<1$. Then

$$\int\Phi(g)\leqslant\frac{\gamma\eta}{1-\gamma\varepsilon}\int\Phi(f).$$

Note that the existence of γ and η satisfying

$$\Phi(\beta\lambda)\leqslant\gamma\Phi(\lambda)\quad\text{and}\quad\Phi(\delta^{-1}\lambda)\leqslant\eta\Phi(\lambda)$$

is assured by

$$\Phi(2\lambda)\leqslant c\Phi(\lambda).$$

For example, a possible choice for γ is c^k, where k is a positive integer atisfying

$$2^{k-1}<\beta<2^k$$

since then

$$\Phi(\beta\lambda)\leqslant\Phi(2^k\lambda)\leqslant c^k\Phi(\lambda).$$

Proof of Lemma 3.4. Consider the set

$$\{S^*(f)\leqslant\lambda\}.$$

Use this set as the base for the Kakutani decomposition (Theorem 1.1). Thus

$$\{S^*(f)\leqslant\lambda\}=\bigcup_{j=1}^{\infty}B_0^j$$

and

$$X=\bigcup_{j=1}^{\infty}\bigcup_{i=0}^{j-1}B_i^j.$$

Let

$$f=f_1+f_2$$

where f_1 is supported on the parts of the columns where A^*f is small. That is,

$$f_1(x)=\begin{cases}f(x) & \text{if}\quad x\in B_i^j,\ j>1,\ \text{and}\\ & \quad A^*f(T^kx)<\delta\lambda\ \text{for some}\ k,\ -i\leqslant k\leqslant j-i-1,\\ 0 & \text{elsewhere.}\end{cases}$$

For x in {Support of f_1, $A^*f<\delta\lambda$} we have

$$S^*(f_2)(x)\leqslant(1+c\delta)\lambda.$$

In fact, let $x\in${Support of f_1, $A^*f<\delta\lambda$}. Assume $x\in B_S^n$, and let $a\in B_0^n$ so that $T^S(a)=x$. Then $S^*(f)(a)\leqslant\lambda$ and $A^*f(x)<\delta\lambda$. Furthermore,

$$|S^*(f_2)^2(x)-S^*(f_2)^2(a)|=\left|\sum_{k=0}^{\infty}\left(\frac{f_2(T^kx)}{k+1}\right)^2-\sum_{k=0}^{\infty}\left(\frac{f_2(T^ka)}{k+1}\right)^2\right|$$
$$=\left|\sum_{k=0}^{\infty}[f_2(T^kx)]^2\left[\left(\frac{1}{k+1}\right)^2-\left(\frac{1}{k+S+1}\right)^2\right]\right|.$$

If we set

$$b_k=f_2(T^kx)\left[\left(\frac{1}{k+1}\right)^2-\left(\frac{1}{k+S+1}\right)^2\right]$$

the above becomes

$$\left|\sum_{k=0}^{\infty}f_2(T^kx)b_k\right|=\left|\sum_{k=0}^{\infty}f_2(T^kx)\sum_{j=k}^{\infty}(b_j-b_{j+1})\right|$$
$$=\left|\sum_{j=0}^{\infty}(b_j-b_{j+1})(j+1)\frac{1}{j+1}\sum_{k=0}^{i}f_2(T^kx)\right|$$
$$\leqslant A^*f(x)\left|\sum_{j=0}^{\infty}(b_j-b_{j+1})(j+1)\right|$$
$$\leqslant\delta\lambda\left|\sum_{j=0}^{\infty}(b_j-b_{j+1})(j+1)\right|.$$

Now we show that

$$\left|\sum_{j=0}^{\infty}(b_j-b_{j+1})(j+1)\right|\leqslant c\delta\lambda.$$

We have

$$\left|\sum_{j=0}^{\infty}(b_j-b_{j+1})(j+1)\right|=\left|\sum_{j=0}^{\infty}b_j\right|=\left|\sum_{j=0}^{\infty}f_2(T^jx)\left[\left(\frac{1}{j+1}\right)^2-\left(\frac{1}{j+S+1}\right)^2\right]\right|.$$

If

$$a_j=\left[\left(\frac{1}{j+1}\right)^2-\left(\frac{1}{j+S+1}\right)^2\right],$$

the last expression may be written as

$$\Big|\sum_{j=0}^{\infty} f_2(T^j x)\, a_j\Big| = \Big|\sum_{j=0}^{\infty} f_2(T^j x) \sum_{k=j}^{\infty} (a_k - a_{k+1})\Big|$$

$$= \Big|\sum_{k=0}^{\infty} (a_k - a_{k+1})(k+1)\frac{1}{k+1}\sum_{j=0}^{k} f_2(T^j x)\Big|$$

$$\leqslant A^* f(x)\Big|\sum_{k=0}^{\infty} (a_k - a_{k+1})(k+1)\Big| \leqslant \delta\lambda \Big|\sum_{k=0}^{\infty} a_k\Big|$$

$$\leqslant \delta\lambda \Big|\sum_{k=0}^{\infty}\Big(\frac{1}{k+1}\Big)^2 - \Big(\frac{1}{k+S+1}\Big)^2\Big| \leqslant c\delta\lambda.$$

Thus, we have established the fact that

$$|S^*(f_2)^2(x) - S^*(f_2)^2(a)| \leqslant c\delta^2\lambda^2.$$

Hence we have

$$S^*(f_2)^2(x) \leqslant c\delta^2\lambda^2 + S^*(f_2)^2(a) \leqslant c\delta^2\lambda^2 + \lambda^2.$$

Consequently,

$$S^*(f_2)(x) \leqslant (c\delta^2\lambda^2 + \lambda^2)^{1/2} \leqslant (c\delta + 1).$$

We now complete the proof of the lemma. If $S^*(f)(x) > \beta\lambda$, then $S^*(f_1)(x) + S^*(f_2)(x) > \beta\lambda$. If $x \in \{\text{Support of } f_1,\ A^* f < \delta\lambda\}$, then we have shown $S^*(f_2)(x) < (c\delta+1)\lambda$. Consequently, if $S^*(f)(x) > \beta\lambda$, then

$$S^*(f_1)(x) > \beta\lambda - (c\delta+1)\lambda \geqslant [\beta - (c\delta+1)]\lambda.$$

Choose β and δ so that $c\delta + 1 < \frac{1}{2}\beta$. We then have

$$S^*(f_1)(x) > \tfrac{1}{2}\beta\lambda.$$

Thus,

$$m\{S^*(f) > \beta\lambda,\ A^* f < \delta\lambda\} \leqslant m\{S^*(f_1) > \tfrac{1}{2}\beta\lambda\} \leqslant \frac{c}{\beta\lambda}\int_X f_1(x)\, dm(x).$$

The last inequality is obtained by an application of the weak type (1, 1) estimate (3.1) for $S^*(f_1)$.

We now need to show that

$$\int_X f_1(x)\, dm(x) \leqslant c\delta\lambda m\{S^*(f) > \lambda\}.$$

Recall that f_1 is supported on columns where $A^* f$ is small. That is,

$$\{\text{Support } f_1\} = \bigcup_{j=2}^{\infty}\bigcup_{i=0}^{j-1} A_i^j \quad \text{with} \quad A_i^j \subseteq B_i^j,$$

and $x \in A_i^j$ implies there is a k, $-i \leqslant k \leqslant j-1$ such that $T^k(x) \in \{A^* f < \delta\lambda\}$. We now have

$$\int_X f_1(x)\, dm(x) = \int_{\bigcup_{j=2}^{\infty}\bigcup_{i=0}^{j-1} A_i^j} f(x)\, dm(x) = \sum_{j=2}^{\infty}\int_{A_0^j}\sum_{i=0}^{j-1} f(T^i x)\, dm(x).$$

On each column, we can estimate the sum by $\delta\lambda$, then obtain the final estimate by summing up the columns. In detail, we proceed as follows:

Let

$$A_0^j = \bigcup_{k=0}^{j-1} D_k^j$$

where D_k^j are disjoint sets such that $x \in D_k^j$ implies $T^k(x) \in \{A^* f < \delta\lambda\}$, and, if $S < k$, then $T^S(x) \notin \{A^* f < \delta\lambda\}$. Hence,

$$\sum_{j=2}^{\infty}\int_{A_0^j}\sum_{i=0}^{j-1} f(T^i x)\, dm(x)$$

$$= \sum_{j=2}^{\infty}\int_{\bigcup_{k=0}^{j-1} D_k^j}\sum_{i=0}^{j-1} f(T^i x)\, dm(x)$$

$$= \sum_{j=2}^{\infty}\sum_{k=2}^{j-1}\int_{D_k^j}\sum_{i=0}^{j=1} f(T^i x)\, dm(x)$$

$$\leqslant \sum_{j=2}^{\infty}\sum_{k=0}^{j-1}\int_{D_k^j}\sum_{i=0}^{k} f(T^i x) + \sum_{i=k}^{j-1} f(T^i x)\, dm(x)$$

$$= \sum_{j=2}^{\infty}\sum_{k=0}^{j-1}\int_{D_k^j}(k+1)\frac{1}{k+1}\sum_{i=0}^{k} f(T^i x) + (j-k)\frac{1}{j-k}\sum_{i=k}^{j-1} f(T^i x)\, dm(x)$$

$$\leqslant \sum_{j=2}^{\infty}\sum_{k=0}^{j-1}(k+1)A^* f(x)(D_k^j) + (j-k)A^* f(x)\, m(D_k^j)$$

$$\leqslant \sum_{j=2}^{\infty}\sum_{k=0}^{\infty}(j+1)\,\delta\lambda m(D_k^j)$$

$$\leqslant \delta\lambda\sum_{j=2}^{\infty}(j+1)\, m(A_0^j) \leqslant 2\delta\lambda m\{S^*(f) > \lambda\}.$$

That is, we have shown

$$\int_X f_1(x)\,dm(x) \leqslant c\delta\lambda m\{S(f) > \lambda\},$$

and thereby, proved the lemma.

To obtain the inequality for f^*, we use the following lemmas.

LEMMA 3.6. *For all* $\lambda > 0$,

$$m\{A^*f > 2\lambda\} \leqslant 2m\{f^* > \lambda\}.$$

Proof. Using the maximal inequality (Theorem 2.4), we see that

$$m\{A^*f > 2\lambda\} \leqslant \frac{1}{\lambda} \int_{\{f>\lambda\}} f(x)\,dm(x).$$

Whenever $m\{f^* > \lambda\} < 1$, we also have 2.6,

$$\frac{1}{\lambda} \int_{\{f>\lambda\}} f(x)\,dm(x) \leqslant 2m\{f^* > \lambda\}.$$

Hence the lemma is true if $m\{f^* > \lambda\} < 1$. But, if $m\{f^* > \lambda\} = 1$, the result is obvious. Thus we have the inequality for all λ.

LEMMA 3.7. *For Φ as in 3.5, we have*

$$\int_X \Phi\big(A^*f(x)\big)\,dm(x) \leqslant c_\Phi \int_X \Phi\big(f^*(x)\big)\,dm(x).$$

Proof. Recall that Φ determines a Lebesgue–Stieltjes measure. Consequently,

$$\int_X \Phi\big(A^*f(x)\big)\,dm(x) = \int_0^\infty m\{A^*f > 2\lambda\}\,d\Phi(2\lambda)$$

$$\leqslant \int_0^\infty 2m\{f^* > \lambda\}\,d\Phi(2\lambda) \leqslant c \int_X \Phi\big(f^*(x)\big)\,dm(x).$$

Thus the proof of Theorem 3.3 is complete.

Remarks. At this time a number of questions remain unanswered.

(I) As noted in the paper, the double inequality $C\|f^*\|_1 \leqslant \|S(f)\|_1 \leqslant C_2\|f^*\|_1$ does not hold in the ergodic theory setting, but does hold in the martingale setting. Is there another "square function" which extends this martingale result, at least for some large class of transformations?

(II) If we replace $f^*(x)$ by $Mf(x) = \sup_n \frac{1}{n}\Big|\sum_{k=0}^{n} f(T^k x)\Big|$, then does it follow that $\|S(f)\|_1 \leqslant C\|Mf\|_1$. This would be a further generalization of the martingale result. Since Mf can be much smaller than f^*, this would be a much sharper inequality.

(III) The fact that $S^*(f)(x) < \infty$ a.e. also implies for a.e. x that $F(x,t) = \sum_{k=0}^{\infty} \frac{f(T^k x) r_k(t)}{k+1}$ exists for a.e. t, where $r_k(t)$ are the Rademacher functions defined on $[0,1]$. We can also look at

$$\overline{F}(x,t) = \sum_{k=1}^{\infty} \frac{f(T^k x) - f(T^{-k}x)}{k} r_k(t).$$

The square function also shows that this operator exists for a.e. (x, t). The ergodic Hilbert transform is defined by

$$Hf(x) = \lim_{n\to\infty} \sum_{k=1}^{\infty} \frac{f(T^k x) - f(T^{-k}x)}{k}.$$

It is easy to see that

$$Hf(x) = \lim_{n\to\infty} 2^n \int_0^{\frac{1}{2^n}} \overline{F}(x,t)\,dt.$$

Thus, the Hilbert transform is the derivative, at 0, of $\overline{F}(x,t)$. Can this information be used to study the Hilbert transform? What is the relationship between the square functions and the Hilbert transform?

References

[1] D. L. Burkholder, *Distribution function inequalities for martingales*, Ann. of Prob. 1 (1973), pp. 19–42.

[2] D. L. Burkholder and R. F. Gundy, *Extrapolation and interpolation of quasi-linear operators on martingales*, Acta Math. 124 (1970), pp. 249–304.

[3] R. R. Coifman, *Distribution function inequalities for singular integrals*, Proc. Nat. Acad. Sci. U.S.A. Vol. 1, No 10 (1972), pp. 2838–2839.

[4] D. S. Ornstein, *A remark on the Birkhoff ergodic theorem*, Illinois J. Math. 15 (1971), pp. 77–79.

[5] E. M. Stein, *Note on the class $L\log^+L$*, Studia Math. 32 (1969), pp. 305–310.

DEPARTMENT OF MATHEMATICS
DE PAUL UNIVERSITY, CHICAGO, U.S.A.

Received June 30, 1975 (1038)