[R2] W. Rudin, *Real and Complex Analysis*, McGraw-Hill, 1974.
[S] K. Seip, *Beurling type density theorems in the unit disk*, Invent. Math. 113 (1993), 21–39.
[S1] —, *On Korenblum's density condition for the zero sequences of $A^{-\alpha}$*, J. Anal. Math. 67 (1995), 307–322.
[S2] —, *Developments from nonharmonic Fourier series*, in: Proc. ICM 1998, Vol. II, 713–722.
[Sh] J. O. Shapiro, *Composition Operators and Classical Function Theory*, Springer, 1993.
[SW] A. L. Shields and D. L. Williams, *Bounded projections and the growth of harmonic conjugates in the disk*, Michigan Math. J. 29 (1982), 3–25.
[SZ] Z. Słodkowski and W. Żelazko, *On joint spectra of commuting families of operators*, Studia Math. 50 (1974), 127–148.
[V] D. Vukotić, *Pointwise multiplication operators between Bergman spaces on simply connected domains*, Indiana Univ. Math. J., to appear.
[Z] W. Żelazko, *An axiomatic approach to joint spectra I*, Studia Math. 64 (1979), 249–261.

Departamento de Matemática Aplicada
Universidad Politécnica de Valencia
E-46071 Valencia, Spain
E-mail: jbonet@pleione.cc.upv.es

Department of Mathematics
Åbo Akademi University
FIN-20500 Åbo, Finland
E-mail: mlindstr@abo.fi

Institute of Mathematics (Poznań branch)
Polish Academy of Sciences
Matejki 48/49
60-769 Poznań, Poland
E-mail: domanski@amu.edu.pl

Received January 7, 1999
Revised version August 23, 1999 (4240)

STUDIA MATHEMATICA 137 (2) (1999)

An exponential estimate for convolution powers

by

ROGER L. JONES (Chicago, IL)

Abstract. We establish an exponential estimate for the relationship between the ergodic maximal function and the maximal operator associated with convolution powers of a probability measure.

1. Introduction. Let $\tau : X \to X$ denote a measurable, invertible, ergodic point transformation from a probability space (X, Σ, m) to itself. For $f \in L^1(X)$, define

$$f^\star(x) = \sup_{m,n\ge 0} \frac{1}{m+n+1} \sum_{k=-m}^{n} |f(\tau^k x)|.$$

Let μ denote a probability measure on $\mathbb{Z}$ and define

$$\mu f(x) = \sum_{j=-\infty}^{\infty} \mu(j) f(\tau^j x).$$

For $n > 1$ define

$$\mu^n f(x) = \mu(\mu^{n-1} f)(x).$$

(See [2] for a discussion of these averaging operators, and conditions associated with a.e. convergence for $f \in L^p$, $p > 1$. Also see [1] where for a large class of measures, μ, Bellow and Calderón establish a.e. convergence for all $f \in L^1$.)

In [2] the following condition was introduced.

Definition 1.1. A probability measure μ on $\mathbb{Z}$ has *bounded angular ratio* if $|\widehat{\mu}(\gamma)| = 1$ only for $\gamma = 1$, and

$$\sup_{|\gamma|=1} \frac{|\widehat{\mu}(\gamma) - 1|}{1 - |\widehat{\mu}(\gamma)|} < \infty.$$

1991 *Mathematics Subject Classification*: Primary 42B25; Secondary 28D05.
Key words and phrases: maximal functions, exponential estimates, convolution powers.
R. Jones is partially supported by NSF Grant DMS-9531526.

The reason for this condition was the following theorem.

THEOREM 1.2 ([2]). *Let μ have bounded angular ratio.*

1. *For $f \in L^p$, $1 < p \le \infty$, $\mu^n f(x)$ converges a.e.*
2. *For $1 < p \le \infty$ we have*

$$\|\sup_n |\mu^n f|\|_p \le c(p)\|f\|_p.$$

Further, we establish in [2] that if the bounded angular ratio condition fails, then there are bounded functions f such that the averages $\mu^n f$ diverge a.e. Hence, the bounded angular ratio condition is essential to have a convergence result.

We also established the following theorem, which shows there is a large class of measures with the bounded angular ratio property. In particular, the theorem implies any symmetric measure with finite second moment will satisfy the required property.

THEOREM 1.3. *If*

$$\sum_{k=-\infty}^{\infty} k\mu(k) = 0 \quad \text{and} \quad \sum_{k=-\infty}^{\infty} k^2\mu(k) < \infty$$

then μ has bounded angular ratio, and $\mu^n f(x)$ converges a.e. for $f \in L^p$, $1 < p < \infty$.

Let $\mu^\star f(x) = \sup_n |\mu^n f(x)|$.

2. The exponential estimate. In this paper we establish the following result, which shows that for a large class of measures, μ, the set where $f^\star$ is large, and the set where $\mu^\star f$ is large, have substantial intersection. This is somewhat surprising, since at least for small values of n, $\mu^n f$ can be very different from the Cesàro averages of the iterates of τ applied to f.

THEOREM 2.1. *Assume μ has finite second moment and $\sum k\mu(k) = 0$. Then there are constants α and β (depending only on μ) such that for all $y > 0$ and all $\lambda > 0$ we have*

$$m\{x : f^\star(x) < y,\ \mu^\star f(x) > \lambda y\} \le \alpha e^{-\beta\lambda}$$

for all $f \in L^1(X)$.

REMARK 2.2. Let $\widetilde{f}$ denote the conjugate function, and let Mf denote the Hardy–Littlewood maximal function. In [3] Hunt showed there exist positive constants c and C such that the following relationship holds:

$$m\{x \in (-\pi,\pi) : Mf(x) \le y,\ |\widetilde{f}(x)| > \lambda y\} \le Ce^{-c\lambda},$$

for all $y > 0$, $\lambda > 0$ and $f \in L^1(-\pi,\pi)$. Theorem 2.1 shows that there is a similar relationship between two maximal functions that arise in ergodic theory. The argument below is similar to the argument given by Hunt.

REMARK 2.3. During the course of the proof we will need the assumption of finite second moment and that $E(\mu) = 0$, since we will need to apply a result of Bellow and Calderón which has been established only in that setting.

Proof of Theorem 2.1. In the following, α and β will denote positive constants, but α and β may not always denote the same constants from one occurrence to the next.

We begin by using the Calderón–Zygmund decomposition to write f as a sum of two functions with special properties.

LEMMA 2.4 (Calderón–Zygmund decomposition). *Fix $\lambda > 0$ and $f \ge 0$. Let $B = \{x : f^\star(x) \le \lambda\}$. We can now write $f = g + b$, where g and b have the following properties.*

1. *The "good" function g satisfies $\|g\|_\infty \le 2\lambda$.*
2. *The set B can be decomposed into disjoint sets $B_n, n = 0, 1, \ldots$, such that for $n > 0$, the "bad" function b satisfies*
   - (a) *$b = \sum_{n=1}^{\infty} b_n$ where each b_n is supported on a set C_n of the form $\bigcup_{k=1}^{n} \tau^k B_n$.*
   - (b) *For each n, if $x \in B_n$ we have $\sum_{k=1}^{n} b_n(\tau^k x) = 0$.*
   - (c) *For each n, if $x \in B_n$ we have $\sum_{k=1}^{n} |b_n(\tau^k x)| \le 2\lambda$.*
   - (d) *$\sum_{n=1}^{\infty} m(C_n) \le \frac{2}{\lambda}\|f\|_1$.*

Proof. Let $B = \{x : f^\star(x) \le \lambda\}$ be the base for the standard Kakutani sky-scraper construction. Define $B_n = \{x \in B : \tau^k x \notin B,\ k = 1, \ldots, n$, but $\tau^{n+1} \in B\}$. For $n \ge 1$, let $C_n = \bigcup_{k=1}^{n} \tau^k B_n$.

We first show that for $n \ge 1$ and $x \in B_n$, we have

$$\frac{\lambda}{2} \le \frac{1}{n}\sum_{k=1}^{n} f(\tau^k x) \le 2\lambda.$$

To see this, note that $x \in B_n$ implies $\frac{1}{n+1}\sum_{k=0}^{n} f(\tau^k x) \le \lambda$. Hence

$$\frac{1}{n}\sum_{k=0}^{n} f(\tau^k x) \le \frac{n+1}{n}\lambda \le 2\lambda,$$

and since $f(x) \ge 0$, the right hand side of the statement follows.

We next show that for $x \in B_n$, we have

$$\frac{1}{n}\sum_{k=1}^{n} f(\tau^k x) \ge \frac{\lambda}{2}.$$

To see this we use a covering argument. Let $x \in B_n$ and define $x_0 = \tau x$. Since $x_0 \notin B$, we know there is an integer r_0 such that $\frac{1}{r_0+1}\sum_{k=0}^{r_0} f(\tau^k x_0) > \lambda$. (Note that here we only need to look in the positive direction, since $\tau^{-1}x_0 \in B$, and $f^\star(\tau^{-1}x_0) \le \lambda$. Also note that all the points $x_0, \tau x_0, \ldots, \tau^{r_0}x_0$ are in C_n since if not, the set would include a point of B, and we know this cannot happen.) Let $l_0 = 0$.

If l_{j-1} and r_{j-1} have been chosen, let $x_j = \tau^{r_{j-1}+1}x_{j-1}$. If this point is not in the column, we are done. Otherwise, since $x_j \in C_n$, we know there exist non-negative integers l and r such that

$$\frac{1}{r+l+1}\sum_{k=-l}^{r} f(\tau^k x_j) > \lambda.$$

We first note that all the points $\tau^k x_j$ for $-l \le k \le r$ are contained in the column C_n, since otherwise there would be a point in B with maximal function greater than λ. There may be more than one possible choice for l and r. From the set of possible pairs (l,r), select the pair such that r is as large as possible, and denote this pair by (l_j, r_j). Since the x_j's selected by this process are distinct, and there are only n possible points in the set, the process terminates.

We now have a finite collection of sets,

$$S_j = \{\tau^{-l_j}x_j, \tau^{-l_j+1}x_j, \ldots, \tau^{r_j}x_j\},$$

such that the average of f over each of these sets is greater than λ, and no point is in more than two of the sets. (It is to make sure that this bounded overlap condition holds that we selected r_j to be the largest possible r in the above construction. This ensures that if some point is in three or more of the sets, then the selection procedure must have been violated.) Consequently, we have

$$2\sum_{k=1}^{n} f(\tau^k x) \ge \sum_j \sum_{y\in S_j} f(y) > \sum_j \lambda(r_j + l_j + 1) \ge n\lambda.$$

We now define $g(x) = f(x)$ for $x \in B$. If $x \in C_n$ for some n, let $x_0 \in B_n$ be such that $\tau^k x_0 = x$ for some $0 < k \le n$. Define

$$g(x) = \frac{1}{n}\sum_{k=1}^{n} f(\tau^k x_0).$$

Hence we have $\|g\|_\infty \le 2\lambda$. Define $b(x) = f(x) - g(x)$. Then b is supported in the union of the columns C_n. Let $b_n(x) = b(x)\chi_{C_n}(x)$. By the construction, and the above observations, all the desired properties of b follow, with the final property holding because $\bigcup_n C_n \subset \{x : f^\star > \lambda\}$, the C_n are disjoint, and the maximal ergodic theorem gives us $m\{x : f^\star(x) > \lambda\} \le \frac{2}{\lambda}\|f\|_1$. ■

REMARK 2.5. In [5] a version of this decomposition is introduced in the ergodic theory setting. However, there the maximal function considered is the usual "forward looking" maximal function, hence some of the necessary estimates are easier. Also see Stein [9] for a discussion in the $\mathbb{R}^n$ setting. In Jones, Kaufman, Rosenblatt and Wierdl [6] and in Jones, Ostrovskii and Rosenblatt [7] a version of this decomposition on $\mathbb{Z}$ is used to prove square function inequalities.

We now apply the Calderón–Zygmund decomposition to the function f, but at height $2y$.

Let $\widetilde{C}_i = C_i \cup \tau^i C_i \cup \tau^{-i}C_i$, and let $\widetilde{C} = \bigcup_{i=1}^\infty \widetilde{C}_i$. For $x \in \widetilde{C}^c$ and $k > 0$, define

$$\begin{aligned}
\nu^\star(1,i,x) &= \inf\{l > 0 : \tau^l(x) \in C_i\},\\
\nu^\star(k,i,x) &= \inf\{l > \nu^\star(k-1,i,x) + i : \tau^l(x) \in C_i\},\\
\nu_\star(1,i,x) &= \inf\{l > 0 : \tau^{-l}(x) \in C_i\},\\
\nu_\star(k,i,x) &= \inf\{l > \nu^\star(k-1,i,x) + i : \tau^{-l}(x) \in C_i\}.
\end{aligned}$$

We have

$$\begin{aligned}
m\{x : f^\star(x) < y,\ \mu^\star f(x) > \lambda y\} \le{}& m\{x : f^\star(x) < y,\ \mu^\star g(x) > \lambda y/2\}\\
&+ m\{x : f^\star(x) < y,\ \mu^\star b(x) > \lambda y/2\}.
\end{aligned}$$

We need to show that both of these terms can be dominated by $\alpha e^{-\beta\lambda}$.

The first term is easy. Since $\|g\|_\infty \le 4y$, and $\mu^\star$ is a contraction on L^∞, we clearly have

$$m\{x : f^\star(x) < y,\ \mu^\star g(x) > \lambda y/2\} = 0$$

for $\lambda > 8$. Hence all we need to do is select α and β so that $\alpha e^{-8\beta} \ge 1$. Then for $\lambda \le 8$ the result is obvious since we are on a probability space. For $\lambda > 8$, clearly $0 \le \alpha e^{-\beta\lambda}$.

Thus it remains to work with the second term. Note that for $x \in \widetilde{C}$, we have $f^\star(x) \ge y$, so it will be enough to consider only $x \in \widetilde{C}^c$.

We first need an estimate of $\mu^\star b(x)$. For $x \in \widetilde{C}^c$, we have

$$\begin{aligned}
\mu^\star b(x) &= \sup_n \Big|\mu^n\Big(\sum_i b_i(x)\Big)\Big| \le \sum_{i=1}^\infty \sup_n \Big|\sum_{j=-\infty}^{\infty} \mu^n(j) b_i(\tau^j x)\Big|\\
&\le \sum_{i=1}^\infty \sup_n \Big(\sum_{k=1}^\infty \Big|\sum_{l=\nu^\star(k,i,x)}^{\nu^\star(k,i,x)+i} \mu^n(l) b_i(\tau^l x)\Big| + \Big|\sum_{l=\nu_\star(k,i,x)}^{\nu_\star(k,i,x)+i} \mu^n(l) b_i(\tau^{-l}x)\Big|\Big)\\
&\le \sum_{i=1}^\infty \sup_n \sum_{k=1}^\infty \Big|\sum_{l=\nu^\star(k,i,x)}^{\nu^\star(k,i,x)+i} \mu^n(l) b_i(\tau^l x)\Big|
\end{aligned}$$

$$+\sum_{i=1}^{\infty}\sup_n\sum_{k=1}^{\infty}\Big|\sum_{l=\nu_*(k,i,x)}^{\nu_*(k,i,x)+i}\mu^n(l)b_i(\tau^{-l}x)\Big|$$

$$= A_+(x)+A_-(x).$$

The estimates for $A_+(x)$ and $A_-(x)$ are the same. To see this, we just replace τ by $\sigma=\tau^{-1}$. Hence we only need to estimate $A_+(x)$.

For the next step, we will need the following lemma, due to Bellow and Calderón [1].

LEMMA 2.6 (Bellow–Calderón). *If μ has bounded angular ratio and $\sum_{k=-\infty}^{\infty}|k|^2\mu(k)<\infty$ then there exists a constant c_μ, which depends only on μ, such that*

$$\sup_n|\mu^n(x-y)-\mu^n(x)|\le c_\mu|y|/|x|^2.$$

REMARK 2.7. This lemma was the key to the Bellow–Calderón proof that $\mu^\star$ is a weak type (1,1) operator. The lemma gave them exactly the same control of the "smoothness" of the convolution powers which is used in the standard proof that the Hilbert transform is weak type (1,1). Here we use it because it gives us the same type of "smoothness" that was used by Hunt in his argument involving the conjugate function.

We deduce, using the fact that the average of b on the column C_i is zero, that

$$A_+(x)\le\sum_{i=1}^{\infty}\sup_n\sum_{k=1}^{\infty}\sum_{l=\nu^\star(k,i,x)}^{\nu^\star(k,i,x)+i}|\mu^n(l)-\mu^n(\nu_+(k,i,x)|\cdot|b_i(\tau^l x)|$$

$$\le c_\mu\sum_{i=1}^{\infty}\sum_{k=1}^{\infty}\frac{i}{\nu^\star(k,i,x)^2}\sum_{l=\nu^\star(k,i,x)}^{\nu^\star(k,i,x)+i}|b_i(\tau^l x)|$$

$$\le 4yc_\mu\sum_{i=1}^{\infty}\sum_{k=1}^{\infty}\frac{i^2}{\nu^\star(k,i,x)^2}\le 8yc_\mu\sum_{i=1}^{\infty}\sum_{k=1}^{\infty}i\sum_{j=\nu^\star(k,i,x)}^{\nu^\star(k,i,x)+i}\frac{\chi_{C_i}(\tau^j x)}{j^2}$$

$$\le 8yc_\mu\sum_{i=1}^{\infty}\sum_{j=\nu^\star(1,i,x)}^{\infty}\frac{i}{j^2}\chi_{C_i}(\tau^j x).$$

Define the function $d(x)$ by $d(x)=\inf\{l:\tau^l(x)\in\widetilde{C}^c$ or $\tau^{-l}x\in\widetilde{C}^c\}$. Note that for $x\in C_i$, we have $d(x)\ge i$. We can now write

$$A_+(x)\le 8yc_\mu\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\frac{d(\tau^j x)\chi_{C_i}(\tau^j x)}{j^2}\le 8yc_\mu\sum_{j=1}^{\infty}\frac{d(\tau^j c)\chi_C(\tau^j x)}{j^2}$$

$$=8yc_\mu\Phi_+(\chi_C)(x),$$

where $\Phi_+(\chi_C)(x)=\sum_{j=1}^{\infty}d(\tau^j x)\chi_C(\tau^j x)/j^2$. Using the same estimate for $A_-(x)$, we see that for $x\in\widetilde{C}^c$ we have $\mu^\star b(x)\le 8c_\mu y\Phi(\chi_C)(x)$ where

$$\Phi(\chi_C)(x)=\Phi_+(\chi_C)(x)+\Phi_-(\chi_C)(x)=\sum_{j\ne0}\frac{d(\tau^j x)\chi_C(\tau^j x)}{j^2}.$$

Write $E=\{x:\Phi(\chi_C)(x)>\lambda/(8c_\mu)\}$ and define

$$\Psi(x)=\frac{\chi_E(x)}{m(E)\log(2/m(E))}.$$

We then have

$$\begin{aligned}
\frac{\frac{\lambda}{8c_\mu}}{\log(2/m(E))}&\le\int\frac{\chi_E(x)}{m(E)\log(2/m(E))}\Phi(\chi_C)(x)\,dx\\
&\le\int\Psi(x)\sum_{j\ne0}\frac{d(\tau^j x)\chi_C(\tau^j x)}{j^2}\,dx\\
&\le\int d(x)\chi_C(x)\sum_{j\ne0}\frac{\Psi(\tau^{-j}x)}{j^2}\,dx\\
&\le\sum_{i=1}^{\infty}\int_{C_i}d(x)\sum_{|j|\ge d(x)}\frac{\Psi(\tau^{-j}x)}{j^2}\,dx\\
&\le\sum_{i=1}^{\infty}\int_{C_i}d(x)\sum_{|j|\ge d(x)}\Psi(\tau^{-j}x)\sum_{k=|j|}^{\infty}\Big(\frac{1}{k^2}-\frac{1}{(k+1)^2}\Big)\,dx\\
&\le\sum_{i=1}^{\infty}\int_{C_i}d(x)\sum_{k=d(x)}^{\infty}\frac{2}{k^2}\frac{1}{k}\sum_{|j|\le k}\Psi(\tau^{-j}x)\,dx\\
&\le 4\sum_{i=1}^{\infty}\int_{C_i}d(x)\sum_{k=d(x)}^{\infty}\frac{1}{k^2}\Psi^\star(x)\,dx\le 4\sum_{i=1}^{\infty}\int_{C_i}\Psi^\star(x)\,dx\\
&\le 4\int_X\Psi^\star(x)\,dx\le 4\int_X\Psi(x)\log^+(\Psi(x))\,dx+C\le\beta,
\end{aligned}$$

where in the next to the last step we used the fact that the L^1 norm of the maximal function is controlled by the $L\log^+L$ norm of the function; see [5].

Consequently, we see that $\lambda\le\beta\alpha\log(2/m(E))$, or (with a different choice of α and β) that $m(E)\le\alpha e^{-\beta\lambda}$, as required. ■

REMARK 2.8. Letting $y=\lambda$, we see that

$$m\{f^\star<\lambda,\ \mu^\star f>\lambda^2\}\le\alpha e^{-\beta\lambda}.$$

Now letting $\lambda \to \infty$, we see that $m\{f^* < \infty,\ \mu^* f = \infty\} = 0$. Hence $\mu^* f < \infty$ a.e. Applying the Stein–Sawyer principle, we have a proof that $\mu^* f$ is weak $(1,1)$. However the proof by Bellow and Calderón [1] is easier.

Remark 2.9. If μ has finite support and mean value zero, Reinhold [8] has shown that $\mu^* f(x) \le c_\mu f^*(x)$, giving the much stronger relationship that for large enough λ, the sets $\{f^* < y\}$ and $\{\mu^* f > \lambda y\}$ are actually disjoint. However, in general, μ may not have finite support, and we cannot apply her result.

References

[1] A. Bellow and A. P. Calderón, *A weak type inequality for convolution products*, to appear.

[2] A. Bellow, R. L. Jones and J. Rosenblatt, *Almost everywhere convergence of convolution powers*, Ergodic Theory Dynam. Systems 14 (1994) 415–432.

[3] R. A. Hunt, *An estimate of the conjugate function*, Studia Math. 44 (1972), 371–377.

[4] R. L. Jones, *Ergodic theory and connections with analysis and probability*, New York J. Math. 3A (1997), 31–67.

[5] —, *Inequalities for the ergodic maximal function*, Studia Math. 60 (1977), 111–129.

[6] R. L. Jones, R. Kaufman, J. Rosenblatt and M. Wierdl, *Oscillation in ergodic theory*, Ergodic Theory Dynam. Systems 18 (1998), 889–935.

[7] R. L. Jones, I. Ostrovskii and J. Rosenblatt, *Square functions in ergodic theory*, ibid. 16 (1996), 267–305.

[8] K. Reinhold, *Convolution powers in L^1*, Illinois J. Math. 37 (1993), 666–679.

[9] E. M. Stein, *Singular Integrals and Differentiability Properties of Functions*, Princeton Univ. Press, Princeton, N.J., 1970.

Department of Mathematics
DePaul University
2219 N. Kenmore
Chicago, IL 60614, U.S.A.
E-mail: rjones@condor.depaul.edu

Received January 18, 1999
Revised version September 24, 1999 (4245)