Acta Math., 173 (1994), 135–154

A generalization of Birkhoff's pointwise ergodic theorem

by

AMOS NEVO(1) and ELIAS M. STEIN(2)

Princeton University
Princeton, NJ, U.S.A.

Princeton University
Princeton, NJ, U.S.A.

0. Introduction, definitions and statements of results

0.1. Introduction

Given an arbitrary invertible measure preserving transformation T on a probability space X, Birkhoff's pointwise ergodic theorem asserts that for any $f\in L^1(X)$, the averages of f along an orbit of T, namely the expressions $(f(T^{-n}x)+...+f(T^nx))/(2n+1)$ converge, for almost all $x\in X$, to the limit $\tilde{f}(x)$, where $\tilde{f}$ is the conditional expectation of f with respect to the σ-algebra of T-invariant sets.

It is natural to wonder whether, given *two* arbitrary invertible measure preserving transformations T and S, there is a natural way to average a function f along the orbits of the group generated by T and S, so as to obtain the same conclusion.

If T and S happen to commute, then, as is well known, e.g. [OW], the expressions $(2n+1)^{-2}\sum_{-n\leqslant n_1,n_2\leqslant n} f(T^{n_1}S^{n_2}x)$ converge for almost all $x\in X$, for any $f\in L^1(X)$, and again the limit is the conditional expectation of f with respect to the σ-algebra of sets invariant under T and S. In other words, the pointwise ergodic theorem holds for finite-measure-preserving actions of the free Abelian group on two generators, namely $\mathbf{Z}^2$.

To answer the question posed above, we need to prove a pointwise ergodic theorem for finite-measure-preserving actions of the free non-Abelian group on two generators. Note that such a result implies a corresponding one for factor groups, when the weights used are those induced by the canonical factor map.

The problem is thus naturally part of the following general framework:

---

(1) The first author is supported by the Israel National Academy of Science and Humanities–Wolfson Grant.

(2) The second author is supported by an NSF grant.

0.2. Definition of pointwise ergodic sequences

Let Γ be a countable group, and let $l^1(\Gamma)=\{\mu=\sum_{\gamma\in\Gamma}\mu(\gamma)\gamma:\sum_{\gamma\in\Gamma}|\mu(\gamma)|<\infty\}$ denote the group algebra. Let $(X,\mathcal{B},m)$ be a standard Lebesgue probability space, and assume Γ acts on X by measurable automorphisms preserving the probability measure m. The action $(\gamma,x)\mapsto\gamma x$ induces a representation of Γ by isometries on the $L^p(X)$ spaces, $1\leqslant p\leqslant\infty$, and this representation can be extended to the group algebra by $(\mu f)(x)=\sum_{\gamma\in\Gamma}\mu(\gamma)f(\gamma^{-1}x)$.

Let $\mathcal{B}_1=\{A\in\mathcal{B}:m(\gamma A\triangle A)=0\ \forall\gamma\in\Gamma\}$ denote the sub-σ-algebra of invariant sets, and denote by E_1 the conditional expectation operator on $L^1(X)$ which is associated with $\mathcal{B}_1$. Now consider:

Definition 1. A sequence $\nu_n\in l^1(\Gamma)$ is called a *pointwise ergodic sequence in* L^p if, for any action of Γ on a Lebesgue space X which preserves a probability measure, and for every $f\in L^p(X)$, $\nu_n f(x)\to E_1 f(x)$ for almost all $x\in X$, and in the norm of $L^p(X)$.

It is natural to consider sequences in $l^1(\Gamma)$ which are given in an explicit geometric form. To that end, assume Γ is finitely generated, and let S be a finite generating set which is symmetric: $S=S^{-1}$ (we will assume $e\notin S$). S induces a length function on Γ, given by $|\gamma|=|\gamma|_S=\min\{n:\gamma=s_1\dots s_n, s_i\in S\}$, $|e|\overset{d}{=}0$. Consider the following sequences:

Definition 2. (i) $\sigma_n=(\#S_n)^{-1}\sum_{w\in S_n}w$, where $S_n=\{w:|w|=n\}$ is the sphere of radius n, with center e. Define also $\sigma'_n=\frac{1}{2}(\sigma_n+\sigma_{n+1})$.

(ii) $\mu_n=(n+1)^{-1}\sum_{k=0}^{n}\sigma_k$, the average of the first $n+1$ spheres.

(iii) $\beta_n=(\#B_n)^{-1}\sum_{w\in B_n}w$, where $B_n=\{w:|w|\leqslant n\}$ denotes the ball of radius n with center e.

When we consider the free group $\mathbf{F}_r$, the set of generators S will always be taken to be a set of free generators (and their inverses).

0.3. Statement of the ergodic theorems

We can now formulate the following result, the first part of which is a direct analog of Birkhoff's pointwise ergodic theorem:

THEOREM 1. *Consider the free group* $\mathbf{F}_r$, $r\geqslant 2$. *Then:*

(1) *The sequence* μ_n *is a pointwise ergodic sequence in* L^p, *for all* $1\leqslant p<\infty$.

(2) *The sequence* σ'_n *is a pointwise ergodic sequence in* L^p, *for* $1<p<\infty$.

(3) σ_{2n} *converges to an operator of conditional expectation with respect to an* $\mathbf{F}_r$-*invariant sub-*σ-*algebra.* β_{2n} *converges to the operator* $E_1+((r-1)/r)E$, *where* E *is a*

projection disjoint from E_1. *Given* $f \in L^p(X)$, $1<p<\infty$, *the convergence is pointwise almost everywhere, and in the* L^p*-norm.*

The proof of Theorem 1 utilizes a strong L^p maximal inequality, of the following form: Given a sequence $\nu_n \in l^1(\Gamma)$, define the associated maximal function $f^*_\nu(x) = \sup_{n\geqslant 0} |\nu_n f(x)|$. Let $(X, \mathcal{B}, m)$ be an $\mathbf{F}_r$-space with an invariant σ-finite measure, not necessarily finite. Then:

THEOREM 2. *For each* $\mathbf{F}_r$, $r\geqslant 2$, *there exist positive constants* $C_p(r)$ *such that for any* $f\in L^p(X)$ *the following inequalities hold*:

(1) $\|f^*_\mu\|_p$, $\|f^*_\sigma\|_p$ *and* $\|f^*_\beta\|_p$ *are all bounded by* $C_p(r)\|f\|_p$, *for* $1<p<\infty$.

(2) f^*_μ *satisfies the maximal inequality of weak type* $(1,1)$, *namely*:

$$\lambda\{x : |f^*_\mu(x)| \geqslant \delta\} \leqslant C_1(r)\delta^{-1}\|f\|_1, \quad \textit{for every } \delta>0, \textit{ and } f\in L^1(X).$$

1. The method of proof and some historical remarks

1.1. Some historical remarks

The search for pointwise ergodic theorems has been a central theme in ergodic theory ever since the publication of G. D. Birkhoff's theorem [B]. The basic problem is to establish, for a general sequence of Markov operators T_k acting in $L^p(X, \mathcal{B}, m)$, the existence of the limit $\lim_{k\to\infty} T_k f(x) = \tilde{f}(x)$, for m-almost all points $x\in X$, and in the L^p norm. A particularly interesting case is when the operators in the sequence arise from a measure preserving action of a countable group, namely they are (finite) convex combinations of the unitary operators associated with the group elements. In this case it is natural to consider uniformly distributed weights, so that one seeks in effect a sequence of averaging (finite) sets in the group. The standard example arises when one fixes a natural translation invariant distance $|\cdot|$ on the group, for example the length with respect to a finite symmetric generating set defined above. Then a natural choice to consider is the sequence of operators β_k of averaging a function on balls of radius k.

If the group is for example a finitely generated Abelian group Γ, consider the following three main ingredients which figure in the proof of ergodic theorems for the sequence β_k:

(1) To establish the mean ergodic theorem, one notes that the sequence of balls in the group has the Følner property of being asymptotically invariant under translation, i.e., $\lim_{n\to\infty} \#(B_n g \triangle B_n)/\#B_n = 0$. Using this fact it is easy to see that $\beta_k f$ has a limit in norm, which is invariant under the group action. The assumption of ergodicity of the measure preserving action then guarantees that $\tilde{f} = \int_X f\, dm$.

To establish pointwise convergence rather than just norm convergence one usually takes the route of proving an L^p maximal inequality. The following two arguments are the essential ones:

(2) The transfer principle, due to N. Wiener [W] and formulated generally by A. P. Calderón [Cal1]: Suppose that the maximal inequality has been established for the action of the averaging sequence β_k in $l^p(\Gamma)$. Then, given an action of Γ on a probability space X, fix $x \in X$ and consider the restriction f_N of a function $f \in L^p(X)$ to a subset of a Γ-orbit in X of the form $B_N x = \{gx : |g| \leqslant N\}$. Regard the function obtained as a function on Γ, and use the assumption that (for any $n \geqslant 0$) $\| \sup_{0 \leqslant k \leqslant n} |\beta_k * f_N| \|_{l^p(\Gamma)} \leqslant C \|f_N\|_{l^p(\Gamma)}$. It follows that $\sum_{|g| \leqslant N-n} |\sup_{0 \leqslant k \leqslant n} \beta_k f(gx)|^p \leqslant C^p \sum_{|g| \leqslant N} |f(gx)|^p$. Since the action is measure preserving, integrating over X we obtain: $\#B_{N-n} \| \sup_{0 \leqslant k \leqslant n} |\beta_k f(x)| \|^p_{L^p(X)} \leqslant C^p \#B_N \|f\|^p_{L^p(X)}$. Since $\lim_{N \to \infty} \#B_{N-n}/\#B_N = 1$, the maximal inequality follows.

(3) The covering argument due to N. Wiener [W]: In order to prove the maximal inequality for the sequence β_k in $l^p(\Gamma)$, one uses geometric covering properties of balls in the group. The basic covering argument needed is a Vitali type "disjointification lemma" which asserts that given a cover of a finite set F in the group using a finite family of balls, there exists a subfamily of disjoint balls which covers at least $\varepsilon(\Gamma)\#F$ of the points of F, where $\varepsilon(\Gamma)$ is a fixed positive constant. For an elegant proof of the disjointification lemma and the resulting L^1 pointwise ergodic theorems for a class of (amenable) groups we refer to [OW]. The crucial assumption used, originating in [Cal2], is that the balls in the group (or more generally some nested asymptotically invariant sequence) satisfy $\#(B_n \cdot B_n^{-1}) \leqslant C \#B_n$ for some fixed C. As pointed out in [OW], the only examples known to satisfy this condition are groups with polynomial growth, and the condition certainly fails for B_n when the group has exponential growth.

1.2. The method of proof

Evidently, all three ingredients are unavailable when the group under consideration is a finitely generated free non-Abelian group $\mathbf{F}_r$. The sequence of balls (which we take with respect to a free generating set) is not asymptotically invariant and the mean ergodic theorem cannot be obtained as above. The sequence has exponential growth and

$$\lim_{N \to \infty} \frac{\#B_N}{\#B_{N-n}} = (2r-1)^n,$$

so the transfer principle does not apply. The disjointification lemma also fails, as one sees by considering any covering of B_n by translates of B_1.

The proof of the pointwise ergodic theorems in this case rests on the following two observations:

(1) The convolution identity

$$\sigma_1 * \sigma_n = \frac{1}{2r}\sigma_{n-1} + \left(1 - \frac{1}{2r}\right)\sigma_{n+1}$$

holds in the group algebra $l^1(\mathbf{F}_r)$, as is easily established. It implies, by induction, that the elements σ_n are linear combinations of the convolution powers σ^k, $0 \leqslant k \leqslant n$. Therefore the spheres σ_n generate a commutative convolution $*$-algebra, which we denote $A(\mathbf{F}_r)$. We note that the algebra $A(\mathbf{F}_r)$ has been introduced into ergodic theory in [AK]. It has a simple and explicit spectral theory, discussed e.g. in [Sa], [Car], [Mac], [Mat], [FP]. This fact makes it possible to use spectral methods to establish ergodic theorems. Such methods were utilized to prove the mean ergodic theorem for the sequence σ_n' [G], and the pointwise ergodic theorem for functions in L^2 [N1].

(2) The Markov operators β_k and σ_k are comparable, in the sense that a maximal inequality for one sequence implies the same inequality for the other (up to changing the constant). This fact follows since $\sigma_n \leqslant C_r \beta_n$, for some fixed positive constant C_r, in sharp contrast to the situation in the Abelian case. Hence one might as well establish the ergodic theorems for spheres rather than for balls. The utility of this observation is in showing that the right approach is to apply methods developed to handle convergence of singular means [St1], [St2]. Originally the methods were devised [St1] to prove pointwise convergence for the (even) powers of a self-adjoint Markov operator, improving the Hopf–Dunford–Schwartz theorem for uniform averages of powers. These methods will serve as a replacement for the covering arguments and asymptotic invariance of balls, used in the Abelian case.

The method of singular means proceeds in our case as follows [St1]:

(1) The first step is to prove an L^p maximal inequality for uniform, or Cesaro averages of the singular means σ_n. The proof of the first inequality of Theorem 2, namely $\|f_\mu^*\|_p \leqslant C_p(r)\|f\|_p$, is based on the observation that $\sum_{k=0}^{N} \sigma_k \leqslant C_r \sum_{k=0}^{3N} \sigma_1^k$, which enables one to use the Hopf–Dunford–Schwartz maximal inequality [DS]. For the free group this estimate will be proved in Lemma 1 below, using the fact that the word-length distribution of the convolution powers σ_1^n is well approximated by the binomial distribution $b_n(k,p)$ with positive expectation p. Therefore, expanding σ_1^n, the weight attached to an element σ_k for k in the interval $[np-\sqrt{n}, np+\sqrt{n}]$ is at least $a_1/\sqrt{n}$. Since σ_k appears as an element of such an interval at least $a_2\sqrt{n}$ times, the estimate follows.

(2) The second step is to embed the Cesaro averages of the singular means in an analytic family of complex Cesaro averages S_n^λ, defined for any $\lambda \in \mathbf{C}$. The embedding is implemented using the complex binomial coefficients, so that the ordinary Cesaro averages correspond to the sequence S_n^0, and the singular means to S_n^{-1}. There are associated maximal operators S_*^λ, and these satisfy maximal inequalities in L^p, $p>1$,

when $\operatorname{Re}\lambda>0$, since the ordinary Cesaro averages do (Lemma 4). When $\operatorname{Re}\lambda\leqslant 0$, it is possible to show (Lemma 5), that S_*^λ satisfy a strong maximal inequality in L^2. For $\lambda\in-\mathbf{N}$ (the set of poles of the Γ function), the method is to consider the operators on the Fourier transform side and use the Littlewood–Paley square-function estimates. Here the essential property used is that the $*$-characters of the algebra $A(\mathbf{F}_r)$ decay exponentially when evaluated at σ_n'. For $\lambda\notin-\mathbf{N}$, the maximal inequality is established using standard estimates on the complex binomial coefficients and the Γ function, and repeated summation by parts.

(3) The third step is to use the L^p maximal inequalities for S_*^λ for $\operatorname{Re}\lambda>0$, and the L^2 maximal inequality for S_*^λ for $\operatorname{Re}\lambda\leqslant 0$, together with bounds established on the constants, and apply the analytic interpolation theorem [St3]. The result is a maximal inequality for the singular averages S_*^{-1}, in every L^p, $p>1$.

(4) The fourth and final step is to construct a dense set of functions $f\in L^p$ for which the sequence $\sigma_n' f(x)$ converges almost everywhere. It is enough to construct such a set in L^2, and that is easily done using again the spectral theory of $A(\mathbf{F}_r)$, or more specifically the decay estimates on the characters.

1.3. The scope of the method

In essence, the method described above shows that one can embed a given sequence of self adjoint Markov operators P_n in an analytic family of operators, and that if the sequence spans a commutative algebra in $\operatorname{End} L^2(X)$ whose characters decay exponentially (in n) it is possible to obtain maximal inequalities in L^p for the sequence. As noted above, originally the method was devised [St1] to handle the case of the algebra $l^1(\mathbf{N})$, or in other words, to prove the maximal inequalities for the (even) powers of a single self-adjoint Markov operator, but its scope is quite general.

We note that in the group theoretic set-up there is a great abundance of important examples where the method can be expected to apply. In fact, given any simple algebraic group G over a local field, there is a natural commutative convolution algebra associated to it, namely the algebra of bi-K-invariant L^1 functions, where K is a suitable maximal compact subgroup. As is well known, the characters of such an algebra can be identified with K-spherical functions on the group. Spherical functions generically decay exponentially as a function of the distance $d(K,KgK)$ in G/K [Mac], [Mat], [Sa]. Consequently, for suitable self-adjoint singular averages on G the same arguments should apply. Moreover, it is sometimes possible to embed the algebra in question as a subalgebra of the group algebra of a lattice subgroup $\Gamma\subset G$, and obtain ergodic theorems for actions of Γ. The case of the free group considered here properly belongs in this context, and we refer to [N1] for more on the group theoretic point of view.

Moreover, it is of course interesting to consider spherical singular averages on simple real Lie groups as well, and here it is possible to use ideas from the theory of spherical singular averages acting on $L^p(\mathbf{R}^n)$ [St4], [SW] to obtain pointwise ergodic theorems for group actions in a similar fashion. We refer to [J] for the case of $\mathbf{R}^n$ actions, and to [N2], [N3] for the case of actions of the real hyperbolic groups $\mathrm{SO}^0(n,1)$.

It would seem, then, that the natural context for the applications of the theory of singular spherical averages to pointwise ergodic theorems for group actions is the context of spherical functions on Gelfand pairs, see e.g. [Fa].

2. Cesaro sums and maximal inequalities

2.1. The maximal inequality for μ_n

We begin by establishing the maximal inequalities for the sequence μ_n of ordinary Cesaro averages:

LEMMA 1.
$$\mu_N = \frac{1}{N+1}\sum_{n=0}^{N}\sigma_n \leqslant \frac{C(r)}{3N+1}\sum_{n=0}^{3N}\sigma_1^n.$$

Hence the strong maximal inequality $\|f_\mu^*\|_p \leqslant \|f\|_p$ *holds for all* $1<p<\infty$, *as well as the weak type* $(1,1)$ *maximal inequality in* L^1.

Proof. As noted above, the following holds in $l^1(\Gamma)$, (where we denote $q=2r-1$):

$$\sigma_1 * \sigma_n = \frac{1}{q+1}\sigma_{n-1} + \left(1-\frac{1}{q+1}\right)\sigma_{n+1}.$$

Since the convolution powers σ_1^n are convex combinations of the radial measures σ_k, $0\leqslant k\leqslant n$, we can write $\sigma_1^n = \sum_{k=0}^n a_n(k)\sigma_k$. To estimate the coefficients $a_n(k)$ consider the Markov chain on the non-negative integers, with nearest neighbour transition probabilities of $1/(q+1)$ to the left and $1-1/(q+1)$ to the right, except at the point 0, which is a reflecting barrier. Clearly, $a_n(k)$ is the probability that having started at 0, after n steps the chain is at the point k. Now consider the number of paths of length n of the chain that start at 0 and end at k. It is at least the number of paths of length n that start at 0 and end at k without having visited 0 again along the way. The latter is given by the ballot problem: It is the number of ways in which, counting a ballot containing a total of $n>0$ votes, the candidate that ended up receiving $\frac{1}{2}(n+k)$ votes was always in the lead (note that k and n have the same parity). This number, e.g. [K, Chapter 9, §2], is given by

$$\frac{k}{n}\binom{n}{\frac{1}{2}(n-k)}.$$

Hence

$$a_n(k) \geqslant \frac{k}{n}\binom{n}{\frac{1}{2}(n-k)}\left(\frac{1}{q+1}\right)^{\frac{1}{2}(n-k)}\left(\frac{q}{q+1}\right)^{\frac{1}{2}(n+k)-1}.$$

Now recall that for k satisfying $\left|\frac{1}{2}(n-k)-n/(q+1)\right| \leqslant \sqrt{n}$ we have, using the standard approximation of the binomial coefficient, e.g. [Fe, Chapter 7, §2], for all $n>0$:

$$\binom{n}{\frac{1}{2}(n-k)}\left(\frac{1}{q+1}\right)^{\frac{1}{2}(n-k)}\left(\frac{q}{q+1}\right)^{\frac{1}{2}(n+k)-1} \geqslant \frac{a_q}{\sqrt{n}}.$$

Clearly, for $k>0$ in the prescribed range, also $k/n \geqslant b_q > 0$. Moreover, denoting $(q+1)/(q-1)$ by τ, it is easily seen that given k ($\geqslant k_0(q)$, say), if n satisfies $|n-\tau k| \leqslant \frac{1}{2}\sqrt{k}$, then k falls in the range prescribed above, so that the estimates hold. Now compute:

$$\begin{aligned}
\sum_{n=0}^{3N} \sigma_1^n &= \sum_{k=0}^{k_0(q)-1} \sigma_1^k + \sum_{n=k_0(q)}^{3N} \sum_{k=0}^{n} a_n(k)\sigma_k \geqslant c_q \sum_{k=0}^{k_0(q)} \sigma_k + \sum_{k=k_0(q)}^{3N}\left(\sum_{n=k}^{3N} a_n(k)\right)\sigma_k \\
&\geqslant c_q \sum_{k=0}^{k_0(q)} \sigma_k + \sum_{k=k_0(q)}^{N}\left(\sum_{n=\tau k-\sqrt{k}/2}^{\tau k+\sqrt{k}/2} a_n(k)\right)\sigma_k \\
&\geqslant c_q \sum_{k=0}^{k_0(q)} \sigma_k + \sum_{k=k_0(q)}^{N} \sqrt{k}\cdot\frac{a_q b_q}{\sqrt{\tau k/2}}\sigma_k \geqslant C\sum_{k=0}^{N}\sigma_k
\end{aligned}$$

The second part of the lemma, namely the strong maximal inequality for f_μ^* in every L^p, $1<p<\infty$, and the weak type $(1,1)$ maximal inequality in L^1, now follows from the Hopf–Dunford–Schwartz theorem [DS] for Markov operators, applied to σ_1. □

2.2. Complex Cesaro sums

We recall the following facts and definitions [Z, III, §1]: Let $\lambda=\alpha+i\beta$ be an arbitrary complex number, and consider the sequence of complex binomial coefficients

$$A_n^\lambda = \frac{(\lambda+1)(\lambda+2)\cdot\ldots\cdot(\lambda+n)}{n!} = \prod_{k=1}^{n}\left(1+\frac{\lambda}{k}\right), \quad A_0^\lambda \stackrel{d}{=} 1,\ A_{-n}^\lambda \stackrel{d}{=} 0.$$

Given any sequence u_n, $n\geqslant 0$, consider the sequence defined by $S_n^\lambda = \sum_{k=0}^{n} A_{n-k}^\lambda u_k$. S_n^λ is called the Cesaro sum of order λ of the sequence $u_0, \ldots, u_n$.

Note that for a negative integer $-m$, the binomial coefficients corresponding to A_n^{-m-1} are given by $A_n^{-m-1}=(-1)^n\binom{m+n}{n}$ when $0\leqslant n\leqslant m$ and $A_n^{-m-1}=0$ otherwise. Therefore, the operator S_n^{-m-1} is the operator Δ^m of discrete differentiation of order m, namely: $S_n^{-2}=u_n-u_{n-1}\overset{d}{=}\Delta u_n$, $S_0^{-2}=u_0$, and similarly $S_n^{-3}=u_n-2u_{n-1}+u_{n-2}=\Delta(\Delta u_n)$. In general $S_n^{-m-1}=\Delta^m u_n=\sum_{k=0}^m(-1)^k\binom{m}{k}u_{n-k}$, $n\geqslant m$.

We collect the properties of A_n^λ that will be used later, in the following

LEMMA 2. *The complex binomial coefficients satisfy*:

(1) *The convolution formula*: $S_n^{\lambda+\delta}=\sum_{k=0}^n A_{n-k}^{\delta-1}S_k^\lambda$.

(2) $\Delta A_n^\lambda=A_n^\lambda-A_{n-1}^\lambda=A_n^{\lambda-1}$.

(3) $\Delta S_n^\lambda=S_n^\lambda-S_{n-1}^\lambda=S_n^{\lambda-1}$.

Proof. As is well known, the following identity holds, for $|y|<1$:

$$\frac{1}{(1-y)^{1+\lambda}}=\sum_{k=0}^\infty A_k^\lambda y^k.$$

If $g(y)$ denotes the formal power series $\sum_{k=0}^\infty u_k y^k$, then clearly multiplication of formal power series gives

$$\frac{g(y)}{(1-y)^{1+\lambda}}=\sum_{k=0}^\infty S_k^\lambda y^k.$$

Since

$$\frac{g(y)}{(1-y)^{1+\lambda+\delta}}=\frac{g(y)}{(1-y)^{1+\lambda}}\cdot\frac{1}{(1-y)^\delta},$$

the convolution formula follows: $S_n^{\lambda+\delta}=\sum_{k=0}^n A_{n-k}^{\delta-1}S_k^\lambda$. Substituting $\delta=1$, and $\lambda-1$ for λ, we obtain the identity $S_n^\lambda=\sum_{k=0}^n S_k^{\lambda-1}$. Taking the sequence $u_0=1$ and $u_n=0$ for $n>0$, which amounts to setting $g(y)=1$, we obtain for $\delta=1$ and $\lambda-1$: $A_n^\lambda=\sum_{k=0}^n A_k^{\lambda-1}$. Parts (2) and (3) follow immediately. □

We now collect the estimates needed later on the coefficients A_n^λ, $n\geqslant 0$, in the following

LEMMA 3. *Let* $\lambda=\alpha+i\beta$. *Then*:

(1) *For* $\alpha>-1$, $0<b_\alpha^{-1}\leqslant A_n^\alpha/(n+1)^\alpha\leqslant b_\alpha<\infty$.

(2) *For* $\alpha>-1$, $1\leqslant|A_n^{\alpha+i\beta}/A_n^\alpha|\leqslant a_\alpha\exp 2\beta^2$.

(3) *For* $m\in\mathbf{N}$, $|(n+1)^m A_n^{-m+i\beta}|\leqslant B_m\exp 3\beta^2$.

Proof. Part (1) is an immediate consequence of the Euler product formula for the Γ-function, which states that for $\alpha>-1$, $n^{-\alpha}A_n^\alpha\to\Gamma(\alpha+1)$.

For part (2) write, using the explicit formula for $A_n^{\alpha+i\beta}$,

$$\left|\frac{A_n^{\alpha+i\beta}}{A_n^\alpha}\right|^2=\prod_{k=1}^n\left(1+\frac{\beta^2}{(k+\alpha)^2}\right)\leqslant\left(1+\frac{\beta^2}{(1+\alpha)^2}\right)\prod_{k=2}^n\left(1+\frac{\beta^2}{(k-1)^2}\right).$$

Since $1+x\leqslant\exp x$ for a non-negative x, we have

$$\leqslant\left(1+\frac{\beta^2}{(1+\alpha)^2}\right)\exp\left(\beta^2\sum_{k=2}^{n}\frac{1}{(k-1)^2}\right)\leqslant(1+(1+\alpha)^{-2})\exp\beta^2\exp\tfrac{1}{6}\pi^2\beta^2\leqslant a_\alpha^2\exp 4\beta^2,$$

where we have used $1+(1+\alpha)^{-2}\beta^2\leqslant(1+(1+\alpha)^{-2})\exp\beta^2$, and $\sum_{k=1}^{\infty}k^{-2}=\frac{1}{6}\pi^2\leqslant 3$.

For part (3), note first that the case $m=0$ is covered by parts (1) and (2). By definition of the binomial coefficients, for $n>m\geqslant 1$,

$$A_n^{-m+i\beta}=\frac{(-m+1+i\beta)\cdot\ldots\cdot i\beta}{(n-m+1)\cdot\ldots\cdot n}\cdot\frac{(1+i\beta)\cdot\ldots\cdot(n-m+i\beta)}{(n-m)!}.$$

Therefore, by definition of $A_{n-m}^{i\beta}$,

$$|n^m A_n^{-m+i\beta}|=\left|\prod_{k=1}^{m}(-m+k+i\beta)\right|\left(\frac{n^m}{(n-m+1)\cdot\ldots\cdot n}\right)|A_{n-m}^{i\beta}|.$$

Estimate the first factor by $(m+|\beta|)^m\leqslant m^m\exp|\beta|$. The second factor is given by $\prod_{k=0}^{m-1}(1-k/n)^{-1}$ and converges to 1 as $n\to\infty$. Finally estimate $|A_{n-m}^{i\beta}|$ using parts (1) and (2) of the lemma. Putting the estimates together, part (3) follows:

$$|(n+1)^m A_n^{-m+i\beta}|\leqslant m^m\exp|\beta|\cdot B_m'\cdot a_0\exp 2\beta^2\leqslant B_m\exp 3\beta^2. \qquad \square$$

2.3. Maximal operators

Consider now a sequence P_k, $k\geqslant 0$, of bounded linear operators defined on $L^p(X)$, and a function $f\in L^p(X)$. There is an associated sequence of Cesaro sums given by $S_n^\lambda f(x)=\sum_{k=0}^{n}A_{n-k}^{\lambda}P_kf(x)$. Consider the maximal functions given by

$$S_*^\lambda f(x)\stackrel{d}{=}\sup_{n\geqslant 0}\left|\frac{S_n^\lambda f(x)}{(n+1)^{\lambda+1}}\right|.$$

Note that, in an $\mathbf{F}_r$-space X, taking $P_k\stackrel{d}{=}\sigma_k$, we obtain $S_*^0 f(x)=f_\mu^*(x)$ and $S_*^{-1}f(x)=f_\sigma^*(x)$. To interpolate between the L^p maximal inequality for $S_*^0 f$ $(1\leqslant p<\infty)$ proved in Lemma 1, and the desired one for $S_*^{-1}f$, we begin with the following

LEMMA 4. *Let P_k be a sequence of Markov operators of norm 1 in $L^1(X)$ and $L^\infty(X)$. Then, for $\alpha>0$, there exist positive constants C_α such that for $f\in L^p(X)$, $\|S_*^{\alpha+i\beta}f\|_p\leqslant C_\alpha\exp(2\beta^2)\|S_*^0|f|\|_p$, $1\leqslant p<\infty$.*

Proof. Using the estimate in Lemma 3 (1),

$$|S_n^\alpha f(x)|=\left|\sum_{k=0}^{n}A_{n-k}^{\alpha}P_kf(x)\right|\leqslant b_\alpha\sum_{k=0}^{n}(n-k+1)^\alpha P_k|f|(x)\leqslant b_\alpha(n+1)^{\alpha+1}S_*^0|f|(x).$$

By the convolution formula of Lemma 2 (1) and the foregoing estimate,

$$\begin{aligned}|S_n^{\alpha+i\beta}f(x)| &= \left|\sum_{k=0}^{n} A_{n-k}^{\alpha/2-1+i\beta} S_k^{\alpha/2} f(x)\right| \\ &\leqslant b_{\alpha/2} \sum_{k=0}^{n} |A_{n-k}^{\alpha/2-1+i\beta}|(k+1)^{\alpha/2+1} S_*^0|f|(x).\end{aligned}$$

Using Lemma 3 (2), the last expression is bounded by

$$b_{\alpha/2}a_{\alpha/2-1}\exp(2\beta^2)\left(\sum_{k=0}^{n}(n-k+1)^{\alpha/2-1}(k+1)^{\alpha/2+1}\right)S_*^0|f|(x).$$

But since

$$\frac{1}{(n+1)^{\alpha+1}}\sum_{k=0}^{n}(n-k+1)^{\alpha/2-1}(k+1)^{\alpha/2+1} \longrightarrow \int_0^1 (1-t)^{\alpha/2-1}t^{\alpha/2+1}\,dt < \infty,$$

we obtain, for $\alpha>0$, $\|S_*^{\alpha+i\beta}f\|_p \leqslant \exp(2\beta^2)C_\alpha\|S_*^0|f|\|_p$. □

3. The Littlewood–Paley square-function method

We now turn to a discussion of the operators $S_*^{\alpha+i\beta}$, $\alpha\leqslant 0$. For reasons that will be explained in §4.1 below, we now use the sequence of operators $P_k=\sigma_{2k}$, the associated operators S_n^λ, S_*^λ, and prove:

LEMMA 5. *The L^2 maximal inequality $\|S_*^{-m+i\beta}f\|_2 \leqslant C_m\exp(3\beta^2)\|f\|_2$ holds, for every non-positive integer $-m\leqslant 0$ and $\beta\in\mathbf{R}$.*

Proof. (1) First note that if the desired maximal inequality holds for each non-positive integer $-m\leqslant 0$, then it also holds for all complex values of the form $-m+i\beta$, $\beta\in\mathbf{R}$. Indeed, fixing $\beta\neq 0$, by the convolution formula of Lemma 2 (1),

$$|S_n^{-m+i\beta}f(x)| = \left|\sum_{k=0}^{n_2} + \sum_{k=n_2}^{n} A_{n-k}^{i\beta} S_k^{-m-1} f(x)\right|$$

where we put $n_2=\left[\frac{1}{2}n\right]$. Consider the second sum first. Using Lemma 3 (2) we obtain

$$(n+1)^{m-1}\left|\sum_{k=n_2}^{n} A_{n-k}^{i\beta} S_k^{-m-1}f(x)\right| \leqslant a_0\exp(2\beta^2)(n-n_2)\max_{n_2\leqslant k\leqslant n}(n+1)^{m-1}|S_k^{-m-1}f(x)|.$$

The last expression is bounded by $C_m\exp(2\beta^2)S_*^{-m-1}f(x)$, by definition of S_*^{-m-1}.

Now consider the first sum, to which we apply Abel's summation formula, namely: if $A_k=\sum_{j=0}^k a_j$, then $\sum_{k=0}^n A_k(b_{k+1}-b_k)=-\sum_{k=0}^n a_k b_k+A_n b_{n+1}$. Making use of parts (2) and (3) of Lemma 2, the result obtained is

$$\left|-\sum_{k=0}^{n_2} A_{n-k}^{-1+i\beta} S_k^{-m} f(x)+A_{n-n_2}^{i\beta} S_{n_2+1}^{-m} f(x)\right|.$$

We can now appeal to Lemma 3 (3), which implies that the remainder term in the foregoing formula is estimated by $B_0 \exp(3\beta^2)(n_2+1)^{-m+1} S_*^{-m} f(x)$. Another summation by parts on the main term will yield $\left|\sum_{k=0}^{n_2} A_{n-k}^{-2+i\beta} S_k^{-m+1} f(x)-A_{n-n_2}^{-1+i\beta} S_{n_2+1}^{-m+1} f(x)\right|$. Appealing again to Lemma 3 (3), the reminder term is estimated by

$$B_1 \exp(3\beta^2)(n_2+1)^{-m+1} S_*^{-m+1} f(x).$$

Similarly, performing the summation by parts on the main term m times, the expression left to estimate is

$$\left|\sum_{k=0}^{n_2} A_{n-k}^{-m+i\beta} S_k^{-1} f(x)+A_{n-n_2}^{-(m-1)+i\beta} S_{n_2+1}^{-1} f(x)\right|.$$

Using Lemma 3 (3) once again, the bound obtained for the foregoing expression is

$$\begin{aligned}(n_2+1)B_m \exp(3\beta^2)(n_2+1)^{-m} S_*^{-1} f(x)+B_{m-1}\exp(3\beta^2)(n-n_2)^{-m+1} S_*^{-1} f(x)\\ \leqslant C_m(n+1)^{-m+1}\exp(3\beta^2) S_*^{-1} f(x).\end{aligned}$$

Recalling that the properly normalized maximal operator is

$$S_*^{-m+i\beta} f(x)=\sup_{n\geqslant 0}\left|(n+1)^{m-1} S_n^{-m+i\beta} f(x)\right|$$

we see that the maximal inequality holds for $S_*^{-m+i\beta}$ if it holds for S_*^{-m}, $m\in\mathbf{N}$.

(2) To prove the L^2 maximal inequality for S_*^{-m}, $m\in\mathbf{N}$, we use the Littlewood–Paley square-function method. Given an arbitrary complex number $\lambda=\alpha+i\beta$, we have by Abel's summation formula (using even indices for notational convenience)

$$\sum_{k=n}^{2n}(k+1)(S_k^\lambda-S_{k-1}^\lambda)=\sum_{k=n}^{2n}(k+1)\Delta S_k^\lambda=(2n+1)S_{2n}^\lambda-\sum_{k=n}^{2n} S_k^\lambda.$$

Rewrite and use the Cauchy–Schwarz inequality:

$$\begin{aligned}\left|\sum_{k=n}^{2n}(k+1)\Delta S_k^\lambda\right| &= \left|\sum_{k=n}^{2n}(k+1)^{\lambda+3/2}(k+1)^{-\lambda-1/2}\Delta S_k^\lambda\right|\\ &\leqslant \left(\sum_{k=n}^{2n}\left|(k+1)^{2\lambda+3}\right|\right)^{1/2}\left(\sum_{k=n}^{2n}\left|(k+1)^{-2\lambda-1}\right|\left|\Delta S_k^\lambda\right|^2\right)^{1/2}.\end{aligned}$$

Note that the first factor in the last equation can be estimated as follows:

$$\left|\frac{1}{(2n+1)^{\lambda+2}}\right|\left(\sum_{k=n}^{2n}|(k+1)^{2\lambda+3}|\right)^{1/2}=\left(\sum_{k=n}^{2n}\frac{(k+1)^{2\alpha+3}}{(2n+1)^{2\alpha+3}}\cdot\frac{1}{2n+1}\right)^{1/2}\to\left(\int_{1/2}^{1}t^{2\alpha+3}\,dt\right)^{1/2}.$$

Denote the integral by c_α, and now divide the first formula in this subsection by $(2n+1)^{\lambda+2}$, and use the foregoing to obtain the following estimate:

$$\left|\frac{S_{2n}^{\lambda}}{(2n+1)^{\lambda+1}}\right|\leqslant c_\alpha\left(\sum_{k=n}^{2n}(k+1)^{-2\alpha-1}|\Delta S_k^{\lambda}|^2\right)^{1/2}+(2n+1)^{-\alpha-2}\left|\sum_{k=0}^{2n}S_k^{\lambda}-\sum_{k=0}^{n-1}S_k^{\lambda}\right|.$$

Recall that the following identity holds: $\sum_{k=0}^{n}S_k^{\lambda}=S_n^{\lambda+1}$. Let us now fix a negative integer $-m$ for the parameter λ. Using the foregoing, and noting that the last formula also holds for odd indices by the same argument, we can write

$$\left|\frac{S_n^{-m}}{(n+1)^{-m+1}}\right|\leqslant c_{-m}\left(\sum_{k=1}^{\infty}(k+1)^{2m-1}|\Delta S_k^{-m}|^2\right)^{1/2}+2\sup_{n\geqslant 0}\left|\frac{S_n^{-m+1}}{(n+1)^{-m+2}}\right|.$$

The Littlewood–Paley square-function of order m is defined by

$$R_m(f,x)^2=\sum_{k=0}^{\infty}(k+1)^{2m-1}|S_k^{-m-1}f(x)|^2.$$

Recalling that $\Delta S_k^{-m}=S_k^{-m-1}$, and taking the supremum over $n\geqslant 0$, we obtain

$$S_*^{-m}f(x)\leqslant c_{-m}R_m(f,x)+2S_*^{-(m-1)}f(x).$$

Consequently, if $\|R_m(f,\cdot)\|_2\leqslant K_m\|f\|_2$ for all $f\in L^2(X)$ and $m\in\mathbf{N}$, then the maximal inequality $\|S_*^{-m}f\|_2\leqslant C_{-m}\|f\|_2$, follows from the foregoing estimate by induction on m, using Lemma 1 for the case $m=0$.

Note that the function $R_m(f,x)^2$ is a sum of squares, and therefore, in the special case that the sequence of operators P_k span a commutative $*$-subalgebra of $\operatorname{End}L^2(X)$, one can use the spectral theory of the algebra to estimate the square norm of R_m on the Fourier transform side, as we proceed to do.

4. Spectral theory

4.1. The algebra $A(\mathbf{F}_r)$ and its characters

As already noted, the spherical measures σ_n satisfy the convolution identity $\sigma_1*\sigma_n=(1/(q+1))\sigma_{n-1}+(1-1/(q+1))\sigma_{n+1}$, and so σ_n is a linear combination of the convolution powers σ_1^k, $0 \leqslant k \leqslant n$. Therefore, σ_n, μ_n and β_n belong to the cyclic algebra generated by σ_1 in $l^1(\Gamma)$, the closure of which we denote by $A(\mathbf{F}_r)$. By the convolution identity, a character $\varphi: A(\mathbf{F}_r) \to \mathbf{C}$, satisfies the recurrence relations

$$\varphi(\sigma_1)\varphi(\sigma_n) = \frac{1}{q+1}\varphi(\sigma_{n-1}) + \left(1 - \frac{1}{q+1}\right)\varphi(\sigma_{n+1}).$$

A continuous character is determined completely by its value $\varphi(\sigma_1)$, since $A(\mathbf{F}_r)$ is cyclic. The two linearly independent solutions of the foregoing second order difference equation are q^{-nz} and $q^{-n(1-z)}$, when $z \neq \frac{1}{2} + ij\pi/\log q$, and q^{-nz}, nq^{-nz} otherwise. These solutions correspond to the eigenvalue $\varphi_z(\sigma_1) = \gamma(z) = (q^z + q^{1-z})/(q+1)$. Any character φ_z is a linear combination of the two solutions, with coefficients obtained by solving the linear equations $\varphi_z(\sigma_1) = \gamma(z)$, $\varphi_z(\sigma_0) = 1$. The results are [Car], [Mac], [Mat], [FP]:

$$\varphi_z(\sigma_n) = \mathbf{c}(z)q^{-nz} + \mathbf{c}(1-z)q^{-n(1-z)}, \quad z \neq \frac{1}{2} + \frac{ij\pi}{\log q}, \quad \mathbf{c}(z) = \frac{q^{1-z} - q^{z-1}}{(q+1)(q^{-z} - q^{z-1})}$$

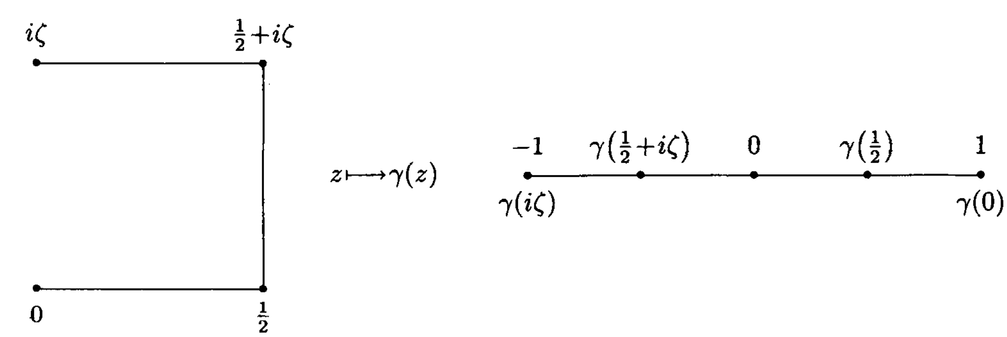
and

$$\varphi_z(\sigma_n) = \left(1 + n\frac{q-1}{q+1}\right)(-1)^{jn} q^{-n/2}, \quad z = \frac{1}{2} + \frac{ij\pi}{\log q}.$$

A necessary and sufficient condition for φ_z to be continuous is that it be bounded, and this condition is equivalent to $0 \leqslant \operatorname{Re} z \leqslant 1$. The unitary representation of $\mathbf{F}_r$ in $L^2(X)$, extended to $l^1(\mathbf{F}_r)$, assigns to σ_1 a self-adjoint operator. Consequently, the values $\varphi_z(\sigma_1) = \gamma(z)$ are real, for those φ_z that occur in the spectrum of σ_1 in $L^2(X)$. Note that $\gamma(z)$ is real if and only if $\operatorname{Re} z = \frac{1}{2}$, or $\operatorname{Im} z = ij\pi/\log q$. This set, the image of which under γ is the real spectrum of $A(\mathbf{F}_r)$, will be denoted by spA_r. Note that for z and $1-z$ the same character obtains, so we can assume that $0 \leqslant \operatorname{Re} z \leqslant \frac{1}{2}$.

Note also that the characters corresponding to $z = s$ and to $z = s + ij\pi/\log q$ differ by sign only: $\varphi_{s+ij\pi/\log q}(\sigma_n) = (-1)^{jn}\varphi_s(\sigma_n)$. In particular, the sign character ε, given by $\varepsilon(\sigma_n) = (-1)^n$, is obtained at the points $z = i(2j+1)\pi/\log q$.

The spectrum is depicted in the following figure, where $\zeta = i\pi/\log q$:



4.2. Spectral estimates: conclusion of the proof of Lemma 5

To conclude the proof of Lemma 5, we estimate the L^2 norm of the square-function

$$R_m(f,x)^2=\sum_{k=0}^{\infty}(k+1)^{2m-1}|S_k^{-m-1}f(x)|^2.$$

Recall that the operators $S_k^{\lambda}f(x)$ are defined using the sequence $P_k=\sigma_{2k}$. Now since S_k^{-m-1} is a linear combination of the operators σ_{2j}, $0\leqslant j\leqslant k$, and σ_{2j} are linear combinations of powers of σ_1, we have, by the spectral theorem,

$$\|S_k^{-m-1}f\|_2^2=\int_{spA_r}|\varphi_z(S_k^{-m-1})|^2\,d\nu_f(z),$$

where ν_f is the spectral measure determined by $f\in L^2(X)$. Consequently,

$$\|R_m(f,\cdot\,)\|_2^2=\int_{spA_r}\sum_{k=0}^{\infty}(k+1)^{2m-1}|\varphi_z(S_k^{-m-1})|^2\,d\nu_f(z).$$

An estimate of $\|R_m(f,\cdot\,)\|_2$ is possible for functions $f\in L^2(X)$ which are orthogonal to the space $\mathcal{H}_{\pm1}=\{f\in L^2(X):\sigma_1 f=\pm f\}$. The spectral measure ν_f of such a function assignes zero mass to the trivial character and the sign character ε. Therefore, to obtain the desired L^2 norm bound on $R_m(f,\cdot\,)$, it is enough to show that the integrand appearing in the last formula is a bounded function of z, as z ranges over the set $spA_r\setminus\{ij\pi/\log q\}$, i.e., omitting the trivial character 1 and the sign character ε. We proceed to estimate the integrand:

(1) First note that the sequence q^{-nz}, $n\geqslant m$, transforms as follows under the discrete differentiation operators Δ^m: $\Delta^m q^{-nz}=(q^{-z}-1)^m q^{-(n-m)z}$. Consider the characters of the form $\varphi_z(\sigma_{2n})=\mathbf{c}(z)q^{-2nz}+\mathbf{c}(1-z)q^{-2n(1-z)}$, where $z=s+ij\pi/\log q$, $0<s<\frac14$. Note that the choice of the even index operators σ_{2k} implies that it is enough to consider $z=s$, $j=0$. Now

$$\sum_{k=0}^{\infty}(k+1)^{2m-1}|\Delta^m q^{-2ks}|^2\leqslant L_m+|q^{-2s}-1|^{2m}\sum_{k=m}^{\infty}(k+1)^{2m-1}q^{-4(k-m)s}.$$

The last infinite series is a sum of derivatives of $1/(q^{-4s}-1)$, up to the $(2m-1)$st order. Therefore the denominator $(q^{-2s}-1)(q^{-2s}+1)$ appears to at most the $2m$th power, and cancels out with the factor preceding it. Taking into account that the $\mathbf{c}$-function is bounded in the strip $0\leqslant\operatorname{Re}z\leqslant\frac14$ (see the previous section), we obtain a bound K_m independent of z.

(2) Now consider the characters corresponding to $\frac14\leqslant\operatorname{Re}z\leqslant\frac12$. A glance at the explicit expression for $\varphi_z(\sigma_n)$ for $z=\frac12+it$ and $z=\frac12-\delta$ (where $0\leqslant\delta\leqslant\frac14$) shows that

$|\varphi_z(\sigma_n)| \leqslant Cnq^{-n/4}$. (Writing out the expression explicitly this estimate is a consequence of $|\sin nt/\sin t| \leqslant n$ for $0<t<\pi$, and $\sinh n\delta/\sinh\delta \leqslant n\cosh n\delta$ for $\delta>0$.) Therefore, if $n \geqslant m$,

$$|\Delta^m \varphi_z(\sigma_{2n})| = \left|\sum_{k=0}^{m}(-1)^k \binom{m}{k} \varphi_z(\sigma_{2n-2k})\right| \leqslant L'_m n q^{-(n-m)/2}.$$

The square norm of the square-function itself is estimated by

$$L_m + L'_m \sum_{k=0}^{\infty}(m+k+1)^{2m-1} q^{-k} \leqslant K_m.$$

The bound $\|R_m(f,\cdot)\|_2 \leqslant K_m\|f\|_2$ has now been established for functions f orthogonal to the spaces $\ker(\sigma_1 \pm I)$. For functions in each of these spaces the maximal inequality is obvious, by definition of S_*^{-m}, and our choice of the operators σ_{2k}. Therefore $\|S_*^{-m}f\|_2 \leqslant C_{-m}\|f\|_2$, for any $f \in L^2(X)$. The proof of Lemma 5 is now complete. □

5. Analytic interpolation

To conclude the proof of Theorem 2 we will make use of the following result [St2], [St3]:

ANALYTIC INTERPOLATION THEOREM. *Let $(X,\mathcal{B},m)$ be a σ-finite measure space, and let $T(\lambda)\colon L^p(X) \to L^p(X)$, $0 \leqslant \operatorname{Re}\lambda \leqslant 1$, be a family of bounded linear operators defined simultaneously for all $1 \leqslant p < \infty$. Assume the following:*

(1) *For each $f \in L^p(X)$ and $g \in L^{p'}(X)$ $(1/p+1/p'=1)$ the function*

$$\lambda \mapsto \int_X T(\lambda) f g \, dm$$

is analytic in the interior of the strip and continuous in its closure.

(2) $\|T(i\beta)f\|_{p_0} \leqslant C_0\|f\|_{p_0}$ *for* $f \in L^{p_0}$.

(3) $\|T(1+i\beta)f\|_{p_1} \leqslant C_1\|f\|_{p_1}$ *for* $f \in L^{p_1}$.

If p_t is defined by

$$\frac{1}{p_t} = \frac{1-t}{p_0} + \frac{t}{p_1},$$

where $0 \leqslant t \leqslant 1$, then $\|T(t)f\|_{p_t} \leqslant C_t\|f\|_{p_t}$.

We can now turn to

Conclusion of the proof of Theorem 2. (1) We use interpolation in order to complete the proof of the L^p maximal inequality, $1<p<\infty$, for the maximal operator $\sup_{n\geqslant 0}|\sigma_{2n}f|$. Recall the maximal inequalities proved in Lemmas 4 and 5:

$$\|S_*^{-m+i\beta}f\|_2 \leqslant C_{-m}\exp(3\beta^2)\|f\|_2, \quad \|S_*^{1+i\beta}f\|_p \leqslant C_1\exp(2\beta^2)\|f\|_p,$$

where $-m$ is a negative integer and $1 \leqslant p < \infty$.

Now take a linear approximation of the maximal operator, as follows: for an arbitrary measurable simple function $N: X \to \mathbf{N}$ let us define the operator $S_N^\lambda f(x) \stackrel{d}{=} (N(x)+1)^{-\lambda-1} S_{N(x)}^\lambda f(x)$. Every operator S_N^λ is linear and $|S_N^\lambda f(x)|$ is bounded pointwise by the corresponding maximal operator $S_*^\lambda f(x)$. Fix a negative integer $-m$, a measurable simple function N, and define the analytic family of linear operators $T(w) = \exp(3w^2) S_N^{(1-w)(-m)+w}$, in the strip $0 \leqslant \operatorname{Re} w \leqslant 1$. Using the norm bounds on $S_*^\lambda f$ above, the analytic interpolation theorem implies that given $1 < p < \infty$ and $0 \leqslant t \leqslant 1$, the operator $\exp(-3t^2) T(t) = S_N^{(1-t)(-m)+t}$ has a maximal inequality in $L^{p_t}(X)$, where $1/p_t = \frac{1}{2}(1-t) + t/p$.

(2) Now solve for the desired operator S_N^{-1}, namely set $-1 = (1-t)(-m) + t$, or $t = (m-1)/(m+1)$. Note that the condition $1/p_t = \frac{1}{2}(1-t) + t/p$ is equivalent to

$$p = \frac{p_t(m-1)}{m+1-p_t} > p_t \frac{m-1}{m+1}.$$

Therefore, given any value $1 < p_t < \infty$, it is possible to choose m large enough so that the foregoing expression is larger than 1. Then, interpolating with these values of m, p and t, the maximal inequality for S_N^{-1} in L^{p_t} follows. Since the constants do not depend on the function N, taking the supremum over all such functions concludes the proof of the maximal inequality for $\sup_{n \geqslant 0} |\sigma_{2n} f(x)|$.

(3) The proof of the maximal inequality for the operator $\sup_{n \geqslant 0} |\sigma_n f(x)|$ is a straightforward consequence of the previous maximal inequality, since, for a non-negative function $f \in L^p(X)$,

$$\frac{1}{q+1} \sigma_{2n-1} f(x) \leqslant \left(\frac{1}{q+1} \sigma_{2n-1} + \frac{q}{q+1} \sigma_{2n+1} \right) f(x) = \sigma_{2n} * \sigma_1 f(x)$$

where we have used the convolution identity governing the radial averages.

(4) The maximal inequality for f_μ^* and f_β^* is an immediate consequence of the maximal inequality for f_σ^*, since μ_n and β_n are convex averages of σ_k, $0 \leqslant k \leqslant n$.

This concludes the proof of Theorem 2. □

Remark. It is interesting to note that the differentiation operators $(n+1)^k \Delta^k \sigma_n f(x)$ satisfy a maximal inequality in L^p, for each $k \geqslant 0$ and each $1 < p < \infty$. This follows by the same argument as above, solving for the operator S^{-k-1} in (2) instead of the operator S^{-1}.

Conclusion of the proof of Theorem 1. As is well known, it is enough to prove Theorem 1 under the additional assumption that the action of $\mathbf{F}_r$ on X is ergodic.

(I) Start with the case of $L^2(X)$:

(1) Fix $0<\delta<\frac{1}{2}$, and let $U_\delta=\{z\in spA_r : \delta\leqslant \operatorname{Re} z\leqslant 1-\delta\}$. Denote by $\mathcal{H}_\delta$ the subspace of vectors in $L^2(X)$ whose spectral measure has its support in U_δ. By the spectral estimates of §4.2, for $f\in\mathcal{H}_\delta$,

$$\|\sigma_n f\|_2^2 = \int_{spA_r} |\varphi_z(\sigma_n)|^2\, d\nu_f \leqslant C^2 n^2 q^{-2n\delta}\|f\|_2^2.$$

Hence $\sum_{k=0}^{\infty} |\sigma_n f(x)|^2$ is a function in $L^1(X)$, and in particular, $\lim_{n\to\infty}\sigma_n f(x)=0=\int_X f\,dm$, for almost all $x\in X$.

(2) The orthogonal complement of $\bigcup_{\delta>0}\mathcal{H}_\delta$ is clearly $\ker(\sigma_1-I)+\ker(\sigma_1+I)$. On the latter space, σ_n' acts as the projection E_1 onto $\ker(\sigma_1-I)$, σ_{2n} acts as the identity, μ_n converges to E_1, and β_{2n} converges to $E_1+((r-1)/r)E_{-1}$, where E_{-1} is the projection on $\ker(\sigma_1+I)$. Therefore, given any $f\in L^2(X)$, each of the four sequences $\nu_n f$, $\nu_n=\sigma_{2n}$, σ_n', μ_n or β_{2n}, has the limit $E_\nu f$ stated in Theorem 1, pointwise almost everywhere. In particular, the pointwise limit of $\nu_n f$ exists and is given by the above for every *bounded* function.

(II) Given $f\in L^p(X)$, $1\leqslant p<\infty$, and a bounded function h satisfying $\|f-h\|_p\leqslant\delta$, write

$$\|\nu_n f-E_\nu f\|_p\leqslant \|\nu_n f-\nu_n h\|_p+\|\nu_n h-E_\nu h\|_p+\|E_\nu h-E_\nu f\|_p.$$

The middle term converges to zero by Lebesgue's dominated convergence theorem. The other two terms are bounded by δ. Therefore the stated limit of $\nu_n f$ exists in L^p norm. As noted above, it is also the limit pointwise almost everywhere for the dense set of bounded functions. The same conclusion holds for any function in L^p, where $1<p<\infty$ if $\nu_n=\sigma_{2n}$, σ_n', β_{2n} (and $1\leqslant p<\infty$ if $\nu_n=\mu_n$). This fact follows from the strong maximal inequality using the following standard argument: Choose $f\in L^p$ and suppose $f_k\to f$ in norm, where $\nu_t f_k(x)$ converges almost everywhere for each k. For fixed $\varepsilon>0$ and k, consider

$$\begin{aligned}|\nu_t f(x)-\nu_s f(x)| &\leqslant |\nu_t(f-f_k)(x)|+|\nu_s(f-f_k)(x)|+|(\nu_t-\nu_s)f_k(x)|\\ &\leqslant 2|f-f_k|_\nu^*(x)+|(\nu_t-\nu_s)f_k(x)|.\end{aligned}$$

By assumption $\limsup_{t,s\to\infty}|(\nu_t-\nu_s)f_k(x)|=0$ a.e. Taking $k\to\infty$ and using the maximal inequalities of Theorem 2, it follows that the set $\{x:\limsup_{t,s\to\infty}|\nu_t f(x)-\nu_s f(x)|>2\varepsilon\}$ is a null set. Hence $\nu_t f(x)$ is a Cauchy sequence for almost all x. This concludes the proof of Theorem 1. □

Remark: *The L^1 problem.* The question of whether the sequence σ_n' satisfies a weak type (1,1) maximal inequality (or the pointwise ergodic theorem in L^1) is still unresolved.

We note however, that for the sequence of even powers of a self adjoint Markov operator [St1], which is our basic motivating case, pointwise convergence generally fails for L^1-functions. A counter example was constructed by D. Ornstein in [O].

References

[AK] ARNOLD, V. I. & KRYLOV, A. L., Uniform distribution of points on a sphere and some ergodic properties of solutions of linear ordinary differential equations in the complex plane. *Soviet Math. Dokl.*, 4 (1962), 1–5.

[B] BIRKHOFF, G. D., Proof of the ergodic theorem. *Proc. Nat. Acad. Sci. U.S.A.*, 17 (1931), 656–660.

[Cal1] CALDERÓN, A. P., Ergodic theory and translation-invariant operators. *Proc. Nat. Acad. Sci. U.S.A.*, 59 (1968), 349–353.

[Cal2] — A general ergodic theorem. *Ann. of Math.*, 58 (1953), 182–191.

[Car] CARTIER, P., Harmonic analysis on trees. *Proc. Sympos. Pure Math.*, 26 (1972), 419–424.

[DS] DUNFORD, N. & SCHWARTZ, J. T., *Linear Operators,* Vol. I and II. Interscience, New York, 1963.

[Fa] FARAUT, J., Analyse harmonique sur les paires de Guelfand et les espaces hyperboliques, in *Analyse harmonique* (P. Eymard, ed.). Le Cours du C.I.M.P.A., Nancy, 1980.

[Fe] FELLER, W., *An Introduction to Probability Theory and Its Applications,* Vol. I. Wiley Publications in Statistics, New York, 1950.

[FP] FIGA-TALAMANCA, A. & PICARDELLO, M. A., *Harmonic Analysis on Free Groups.* Lecture Notes in Pure and Appl. Math., 87. Dekker, New York, 1983.

[G] GUIVARC'H, Y., Généralisation d'un théorème de von Neumann. *C. R. Acad. Sci. Paris Sér. I Math.*, 268 (1969), 1020–1023.

[HM] HOWE, R. & MOORE, C. C., Asymptotic properties of unitary representations. *J. Funct. Anal.*, 32 (1979), 72–96.

[J] JONES, R. L., Ergodic averages on spheres. *J. Analyse Math.*, 61 (1993), 29–47.

[K] KARLIN, S., *A First Course in Stochastic Processes.* Academic Press, New York–London, 1966.

[Mac] MACDONALD, I. G., *Spherical Functions on a Group of p-adic Type.* Publications of the Ramanujan Institute, 2. University of Madras, 1971.

[Mat] MATSUMOTO, H., *Analyse harmonique dans les systemes de Tits bornologiques de type affine.* Lecture Notes in Math., 590. Springer-Verlag, Berlin–New York, 1977.

[N1] NEVO, A., Harmonic analysis and pointwise ergodic theorems for non-commuting transformations. To appear in *J. Amer. Math. Soc.*

[N2] — Pointwise ergodic theorems for radial averages on simple Lie groups I. To appear in *Duke Math. J.*

[N3] — Pointwise ergodic theorems for radial averages on simple Lie groups II. Preprint.

[O] ORNSTEIN, D., On the pointwise behavior of iterates of a self-adjoint operator. *J. Math. Mech.*, 18 (1968), 473–477.

[OW] ORNSTEIN, D. & WEISS, B., Entropy and isomorphism theorems for actions of amenable groups. *J. Analyse Math.*, 48 (1987), 1–141.

[Sa] SATAKE, I., Theory of spherical functions on reductive algebraic groups over p-adic fields. *Inst. Hautes Études Sci. Publ. Math.*, 18 (1962), 5–69.

[St1] STEIN, E. M., On the maximal ergodic theorem. *Proc. Nat. Acad. Sci. U.S.A.*, 47 (1961), 1894–1897.

[St2] — *Topics in Harmonic Analysis Related to the Littlewood-Paley Theory.* Ann. of Math. Stud., 63. Princeton Univ. Press, Princeton, NJ, 1970.

[St3] — Interpolation of linear operators. *Trans. Amer. Math. Soc.*, 83 (1956), 482–492.

[St4] — Maximal functions: Spherical means. *Proc. Nat. Acad. Sci. U.S.A.*, 73 (1976), 2174–2175.

[SW] STEIN, E. M. & WAINGER, S., Problems in harmonic analysis related to curvature. *Bull. Amer. Math. Soc.*, 84 (1978), 1239–1295.

[W] WIENER, N., The ergodic theorem. *Duke Math. J.*, 5 (1939), 1–18.

[Z] ZYGMUND, A., *Trigonometric Series*, Vol. I and II, 2nd edition. Cambridge Univ. Press, New York, 1959.

AMOS NEVO
Department of Mathematics
Princeton University
Fine Hall—Washington Road
Princeton, NJ 08544-1000
U.S.A.
nevo@math.princeton.edu

ELIAS M. STEIN
Department of Mathematics
Princeton University
Fine Hall—Washington Road
Princeton, NJ 08544-1000
U.S.A.
stein@math.princeton.edu

Received November 19, 1993