Discrete maximal functions and ergodic theorems related to polynomials

Akos Magyar[1]

Abstract

We describe below a series of results on the boundary of harmonic analysis, ergodic and analytic number theory. The central objects of study are maximal averages taken over integer points of varieties defined by integral polynomials. The mapping properties of such operators are intimately connected with those of certain exponential sums, studied in analytic number theory. They can be used to prove pointwise ergodic results associated to singular averages defined in terms of polynomials both in the commutative and non-commutative settings.

0. Introduction

We start by reviewing the correspondence between standard maximal functions and pointwise ergodic theorems in the general settings of amenable groups, we'll refer to that as standard theory. Next we discuss the polynomial ergodic results of J. Bourgain [B1], and our results on the uniform distribution of solutions of certain diophantine equations when mapped to measure spaces [M]. Finally the main theme of our discussion is to show pointwise convergence for polynomial averages of non-commuting transformations which generate discrete nilpotent groups. These are pointwise analogues of the general L^2- ergodic theorem of Bergelson and Liebman [BL]. We will discuss the example of the Heisenberg group to avoid the general theory of discrete subgroups of stratified nilpotent Lie groups [C].

In most cases we will not give full proofs, but rather will try to explain the main constructs an ideas behind these results. All our results (except Theorem 2.) are joint work with E.M.Stein and S.Wainger.

1. Standard Theory

A discrete group Γ is called amenable if there exists a sequence of sets: $B_N \subset B_{N+1} \subset \Gamma$ such that $\Gamma = \bigcup_{N=1}^{\infty} B_N$ and

$$(i)\quad \forall g \in \Gamma : \quad \frac{|g \cdot B_N \triangle B_N|}{|B_N|} \to 0 \quad \text{as} \quad N \to \infty$$

[1]Research supported in part by NSF Grant DMS-0202021

Amenable groups form a large class including both commutative and non-commutative groups such as discrete nilpotent groups. We say the sets $B_N(g) = g \cdot B_N$ form a family of balls, if $\exists C > 0$ such that

$$(ii) \quad |B_{2N}| \le C|B_N|$$

$$(i3) \quad B_N(g) \cap B_N(h) \neq 0 \;\Rightarrow\; B_N(h) \subseteq B_{CN}(g)$$

It is well-known that the above properties imply the standard

Maximal Theorem: Let $f \in l^2(\Gamma)$ and define:

$$(1.1) \quad B_N f(h) = \frac{1}{|B_N|} \sum_{g \in B_N} f(hg) \quad \text{and} \quad B^* f(h) = \sup_{N>0} |B_N f(h)|$$

Then

$$\|B^* f\|_{l^2(\Gamma)} \le C\|f\|_{l^2(\Gamma)}$$

Suppose Γ acts on a probability measure space (X, μ) via measure-preserving transformations: $x \to g \cdot x$. The one has corresponding

Maximal Ergodic Theorem: Let $F \in L^2(X)$ and define:

$$(1.2) \quad B_N F(x) = \frac{1}{|B_N|} \sum_{g \in B_N} F(g \cdot x) \quad \text{and} \quad B^* F(x) = \sup_{N>0} |B_N F(x)|$$

Then

$$\|B^* F\|_{L^2(X)} \le C\|F\|_{L^2(X)}$$

The passage between the above theorems is based on a general idea which implicitly contained in Riesz' proof of Birkhoff's ergodic theorem but explicitly described first by Calderon. We explain it below since we will need it later when considering more singular averages.

Transfer principle: For $K > 0$ define the truncated maximal function $B^*_K F = \max_{N \le K} |B_N F|$. It is enough to show that: $\|B^*_K F\|_{L^2(X)} \le C\|F\|_{L^2(X)}$ holds with constant C independent of K. For fixed $L > 0$ and $x \in X$ define:
$f_{L,x}(g) = F(g \cdot x)$ if $g \in B_L$ and $f_{L,x}(g) = 0$ otherwise. The one has

$$B_N F(h \cdot x) = \frac{1}{|B_N|} \sum_{g \in B_N} F(gh \cdot x) = \frac{1}{|B_N|} \sum_{g \in B_N} f_{L,x}(gh) = B_N f_{L,x}(h)$$

if $B_N \cdot h \subseteq B_L$, that is when $h \in \cap_{g \in B_N} B_L \cdot g^{-1} = B_{L,N}$. Taking the square of both sides, summing in h and integrating over X we get using the fact that action of Γ is measure preserving

$$(1.3)\quad |B_{L,K}|\|B_K^* F\|_{L^2}^2 \leq \int_X \|B_K^* f_{L,x}\|_{l^2}^2 \, d\mu \leq$$

$$\leq C \int_X \|f_{L,x}\|_{l^2}^2 \, d\mu = C|B_L|\|F\|_{L^2}^2$$

From property (i) it is easy to see that $\frac{|D_{L,K}|}{|B_L|} \to 1$ as $L \to \infty$ and K being fixed.

Now let $\pi(g)F(x) = F(g \cdot x)$, let $L_I(X) = \{F \in L^2(X) : \ \pi(g)F = F \ \forall g \in \Gamma\}$ be the space of Γ invariant functions, and finally let $P_I : \ L^2(X) \to L_I(X)$ be the corresponding orthogonal projection.

Pointwise Ergodic Theorem: For $F \in L^2(X)$ one has for a.e. $x \in X$

$$(1.4)\quad \lim_{N \to \infty} B_N F(x) = P_I F(x)$$

To see this first we remark that $L_0(X) = L_I(X)^\perp = \overline{\{\pi(g)F - F : \ F \in L^2(X) \cap L^\infty(X), \ g \in \Gamma\}}$. Indeed $\langle \pi(g)F - F, H\rangle = 0 \ \forall F$ is equivalent to $\pi(g)H = H$. If $F = \pi(g)H - H$ then one has

$$|B_N F(x)| \leq \frac{|g \cdot B_N \triangle B_N|}{|B_N|}\|H\|_\infty \to 0$$

as $N \to \infty$ by property (i). Thus the pointwise theorem holds for a dense subspace of $L_0(X)$ and extends to $L_0(X)$ by the maximal ergodic theorem. On the other hand clearly $B_N F = F \ \forall N$ if $F \in L_I(X)$.

The action of Γ on X is called ergodic if the only invariant functions are the constants. In this case the right side of inequality (1.4) is the constant: $c = \int_X F \, d\mu$.

2. Commuting transformations

Here we describe singular averages of commuting measure preserving transformations on a finite measure space defined in terms of polynomials. As in the Euclidean case these fall into two broadly defined category.

On the one hand averages are taken over finite parts of a fixed surface which is the graph of a polynomial mapping, this leads to via the transference principle, to subsequence ergodic theorems associated to arithmetic subsets of the integers. On the other hand averages are taken over the family level surfaces of a homogeneous integral polynomial, which implies equi-distribution type results for solution of diophantine equations.

2.1 Polynomial averages Let $\Gamma = \mathbf{Z}^n = \mathbf{Z}^k \times \mathbf{Z}^d$ and $P = (p_1, \ldots, p_d) : \mathbf{Z}^k \to \mathbf{Z}^d$ be an integral polynomial map, that is $p_j(m)$ are polynomials with integer coefficients. Assume the group Γ acts on a measure space (X, μ) via measure preserving transformations. Consider the family of surfaces:
$S_N = \{(n, p(n)) : \ |n_j| \leq N, \ \forall 1 \leq j \leq k\}$. Then one has

Theorem 1 *(J. Bourgain) For $F \in L^2(X)$ there exists a function $F_* \in L^2(X)$ such that for a.e. $x \in X$:*

$$\lim_{N \to \infty} \frac{1}{S_N} \sum_{g \in S_N} F(g \cdot x) = F_*(x) \tag{2.1}$$

This result was extended in a series of papers [B1]-[B3] to all L^p spaces when $p > 1$ and the case $p = 1$ remains one of the outstanding open problems in this subject.

To see explicitly the type of averages appearing in the theorem suppose the generators of the action of Γ are the commuting transformations $T_1, \ldots, T_k, S_1, \ldots, S_d$ then the right side of (2.1) is: $S_N F(x) = (2N + 1)^{-k} \sum_{|n_j| \leq N} F(T^n S^{p(n)} x)$. Note that (2.1) is equivalent to the (seemingly) special case when $T_1 = \ldots = T_k = I$, where I denotes the identity operator. The essential ideas already appear in the simplest settings, when $S_N F(x) = 1/N \sum_{n=1}^N F(S^{n^2} x)$. The ergodicity of S does not imply that $S_N F(x) \to \int_X F \, d\mu$ as $N \to \infty$ even in L^2 norm. This can be seen in the following simple example which also indicates the arithmetic nature of the problem. Suppose $SF = e^{2\pi i a/q} F$ where S is the shift $SF(x) = F(Sx)$. Then it is a moment to check that

$$S_N F \to \frac{1}{q} \sum_{n=1}^{q} e^{2\pi i \frac{an^2}{q}} F = G(a, q) F$$

the Gaussian sum $|G(a, q)| \approx q^{-1/2}$ thus depends on a and q. Indeed ergodicity is equivalent to the fact that 1 is not an eigenvalue of the shift S. It is called *fully ergodic* if S^q is ergodic for all q. In this case it cannot have rational eigenvalues and the averages converge to the mean value of F.

The key role is played again by the corresponding discrete maximal theorem, however the passage is much more difficult then in the standard case as there is no dense set of functions for which the convergence is immediate (telescopic sums cannot be used). Our approach will be slightly different than that in [B1] , so that it could be generalized more easily to the non-commutative settings.

The continuous maximal function:
We start with the continuous analogue, for $f \in L^2(\mathcal{R})$ let

$$S^* f(x) = \sup_N |S_N f(x)| \quad \text{where} \quad S_N f(x) = 1/N \int_y f(x-y^2)\phi(y/N)$$

where ϕ is a standard cut-off function. It is enough to take supremum over only dyadic values $N = 2^j$ and let us use the notation $S_j f$ for $S_{2^j} f$. Notice that $S_j f = k_j * f$ is a convolution operator where by stationary phase one has the estimate for the multiplier $\hat{k}_j(\theta) = \hat{k}(2^j\theta)$:

$$(2.2) \quad |\hat{k}_j(\theta)| = 2^{-j} |\int_x e^{2\pi i\theta x^2} \phi(2^{-j}x)| \le C(1 + 2^{2j}|\theta|)^{-1/2}$$

Let ψ be a bump function and decompose the multiplier:
$\hat{k}_j(\theta) = \hat{k}_j(\theta)\psi(2^{2j}\theta) + \hat{k}_j(\theta)(1-\psi(2^{2j}\theta))$ and correspondingly
$M_j = S_j + E_j$. The contribution of the "error" terms E_j can be handled by a square function. Indeed by Plancherels formula one has

$$(2.3) \quad \sum_j \|E_j f\|^2 = \int_\theta (\sum_j |\hat{k}_j(\theta)(1-\psi(2^{2j}\theta))|^2)\, |\hat{f}(\theta)|^2\, d\theta \le$$

$$\le C \int_\theta (\sum_{j:\ 2^{2j}|\theta| \ge 1} (2^{2j}|\theta|)^{-1}) \hat{f}(\theta)|^2\, d\theta \le C\|f\|^2$$

The main term $M_j f = m_j * f$ is a convolution operator with kernel $m_j(x) = (k_j * \psi_j)(x) = 2^{-2j} m(2^{-2j}x)$ as it is easy to see by scaling, thus the corresponding maximal function is majorized by the standard Hardy-Littlewood maximal function.

The discrete maximal function: Let us briefly highlight the proof of the $l^2 \to l^2$ boundedness of the discrete maximal function

$$A^* f(m) = \sup_N |A_N f(m)| \quad \text{where} \quad A_N f(m) = \sum_{n=1}^{N} f(m-n^2)$$

We can again consider dyadic averages smoothed by a cut-off

$$S_j f = k_j * f \quad \text{where} \quad \hat{k}_j(\theta) = 2^{-j} \sum_n e^{2\pi i\theta n^2} \phi(2^{-j}n)$$

However when $\theta = a/q$ for q being much smaller than 2^j the size of the multiplier $\hat{k}_j(\theta) \approx G(a,q) \approx q^{-1/2}$ Thus the main contribution is not coming from a neighborhood of 0 but from the neighborhood of rational numbers with small denominators. The role of the stationary phase estimate is played by the so-called Weyl summation.

Lemma 1 *Let $\theta \in [0,1]$ and assume that there is a rational a/q such that $|\theta - a/q| \leq q^{-2}$. Then one has*

$$(2.4) \quad |\hat{k}_j(\theta)|^2 \leq Cmax(q^{-1}, q2^{-2j})$$

The proof of this is standard, the left side of (2.4) is a double sum and after a change of variables $u = n - m$ the exponent becomes linear in each variable and one can estimate the sum, see for example [V].

Accordingly for $1 < \alpha < 2$ and $\varepsilon > 0$ we define a neighborhood of rational - the so-called major arcs - in terms of the cut-off function:

$$(2.5) \quad \omega_j^{\varepsilon}(\theta) = \sum_{q \leq j^{\alpha}, (a,q)=1} \omega(2^{(2-\varepsilon)j}(\theta - a/q))$$

and write $A_j = M_j + E_j$ where M_j and E_j) correspond to the multipliers $m_j(\theta) = \hat{k}_j(\theta)\omega_j^{\varepsilon}(\theta)$ and $e_j(\theta) = \hat{k}_j(\theta)(1 - \omega_j^{\varepsilon}(\theta)$. Here ω is a standard cut-off function and (a,q) denotes the greatest common divisor of a and q. The summation in (2.5) goes through all reduced rationals a/q whose denominator is at most j^{α}.

The reason for sucha splitting is that by Dirichlet's principle for every θ there is a/q with $q \leq 2^{(2-\varepsilon)j}$ such that: $|\theta - a/q)| \leq q^{-1}2^{-(2-\varepsilon)j} \leq q^{-2}$. However if θ is in the support of $e_j(\theta)$ then $j^{\alpha} < q$. Thus by Lemma 1. $\|E_j\|^2_{l^2 \to l^2} = \sup_{\theta} |e_j(\theta)|^2 \leq j^{-\alpha}$ which is summable in j and the error terms E_j are absorbed again into a square function.

Because of the presence of the parameter $\varepsilon > 0$ the support of the terms in (2.5) are not small enough in order be majorized by standard maximal operators as in the continuous case. However by a similar error estimate one can further reduce the size of the intervals to: $|\theta - a/q| \leq cq^{-1}2{-}2j$, that is to the case when $\varepsilon = 0$. According to (2.5) one can write: $M_j = M_{j,0} + \sum_{a,q} M_{j,a/q}$ and consider the corresponding maximal operators M_0^* and $M_{a/q}^*$ separately. Notice that the multiplier $m_{j,0}(\theta) = \hat{k}_j(\theta)\omega(2^{2j}\theta)$ is the same as the one appeared in the continuous case, and thus is majorized by the Hardy-Littlewood maximal operator on $l^2(\mathbf{Z})$. Similarly $M_{a/q}^*$ can be majorized by (a sum of) standard maximal operators which act on $l^2(q\mathbf{Z})$. To see this one decomposes the integers ($mod\ q$) and sum in each residue class separately first. This is the point when the arithmetic nature of the problem appears as the size of Gaussian sums become essential.

To be more precise let $\theta = \beta + a/q$ for some fixed rational a/q, where $|\beta| \leq q^{-1}2^{-2j}$. One writes $n := qn + \nu$, $\nu = 1, \ldots, q$ and

$$(2.6) \quad \hat{k}_j(\beta + a/q) = \sum_{n,\nu} e^{2\pi i (qn+\nu)^2(\beta + a/q)} \psi_j(qn + \nu)$$

$$= (\sum_n e^{2\pi i q^2 n^2 \beta} \psi_j(qn))(\sum_{\nu=1}^{q} e^{2\pi i \nu^2 a/q}) + e_j(\beta) = m_{j,q}(\beta) G(a,q) + e_j(\beta)$$

Here $m_{j,q}$ can be considered as a multiplier on $l^2(q\mathbf{Z})$ and the corresponding maximal operator is bounded by $1/q$ (there are just $2^{2j}/q$ elements of $q\mathbf{Z}$ in the support). However the size of the Gaussian sum is about $\sqrt{q}$ giving us a gain about $1/\sqrt{q}$ at each rational. For the error one has $|e_j(\beta)| \leq 2^{-j/2}$ say because of the small size of β and thus it can be handled again by a square function estimate.

The actual proof is technically more involved as one cannot simply add the norms of the maximal operators $M^*_{a/q}$ but one first has to group the denominators q into classes $\mathcal{C}$ and to each class define $Q = \prod_{q \in \mathcal{C}}$ and then decompose $\mathbf{Z}$ modulo Q, and assign to it a maximal operator M^*_Q.

This will be explained together with the passage to the ergodic theorem later in the non-commutative settings, see also [B1]

2.2 Diophantine equations.

If $P(m_1, \ldots, m_d)$ is a positive definite polynomial with integer coefficients, then a fundamental problem in number theory is to determine asymptotically the number of integer solutions of the corresponding diophantine equation: $P(m) = N$ as $N \to \infty$. A strikingly general result of Birch says that this is possible if P is also homogeneous of degree k and depends essentially on exponentially many variables w.r.t. its degree [Bi]. The precise condition is:

$$(2.7) \quad d - dim\, V_P > k 2^{k-1}$$

where $V_P = \{z \in \mathcal{C}^d : \; P'(z) = 0\}$ is the so-called complex singular variety of P. We will refer to polynomials satisfying all the above conditions as non-degenerate forms.

Suppose there is a commuting family of measure preserving transformations $T = (T_1, \ldots, T_d)$ acting on a finite measure space (X, μ). Then for each N and $x \in X$ the family T maps the solution set $S_N = \{m \in \mathbf{Z}^d :$

$P(m) = N\}$ into X. Indeed let:
$\Omega_{N,x} = \{T^m x = T_1^{m_1} \dots T_d^{m_d} x :\ m = (m_1, \dots m_d) \in S_N\}$. Our next theorem says that the sets $\Omega_{N,x}$ become equi-distributed on X as $N \to \infty$ for almost every $x \in X$ if the family T is fully ergodic, [M]. For a family this means that for each q and $F \in L^2(X)$: if $T_1^q F = \dots = T_d^q F = F$ then F is a constant. Notice that this means that the family $T^q = \{T_1^q, \dots, T_d^q\}$ is ergodic for each q.

Theorem 2 *Let T be a fully ergodic family and let P be an integral non-degenerate form. Then one has for $F \in L^2(X)$ and for a.e. $x \in X$*

$$(2.8) \qquad \lim_{N\to\infty} \frac{1}{|S_N|} \sum_{m\in S_N} F(T^m x) = \int_X F \, d\mu$$

and also the averages on the left side converges in L^2 norm:

$$(2.9) \qquad \| \frac{1}{|S_N|} \sum_{m\in S_N} F(T^m x) = \int_X F \, d\mu \|_{L^2} \to 0$$

as $N \to \infty$.

A special case worth to mention is when $X = \Pi^d$ the torus $\alpha_1, \dots, \alpha_d$ are irrational numbers and T_j is the shift by α_j in the j-th coordinate. Then $\Omega_N = (n_1\alpha_1, \dots n_j\alpha_j) : P(n) = N\}$.

Note that here the averages are taken over disjoint sets and assuming only the ergodicity of the family T the averages in (2.8) may not even converge in L^2-norm. As before the crucial point is to prove the l^2 boundedness of the corresponding discrete maximal operator. In the simplest case, when $P(n) = n_1^2 + \dots + n_d^2$ this is the discrete analogue of Stein's spherical maximal function: $S^* f = \sup_N |S_N f|$ where $S_N f(m) = 1/r_d(N) \sum_{|n|^2=N} f(m-n)$, where $r_d(N)$ is the number of ways of writing N as sum of d squares. It can be proved that for $d > 4$ the operator S^* is bounded in l^p exactly when $p > d/(d-2)$ and this is sharp. For $d = 4$ one might expect that it is bounded in l^p for $p > 2$ at least when the supremum is taken over odd values of N, however this remains an open question.

The asymptotic formula:

One of the key element of the proof is an asymptotic formula for the Fourier transform of the solution sets, i.e for the exponential sums:

$$(2.10) \qquad \hat{\sigma}_{P,N}(\xi) = \sum_{m\in\mathbf{Z}^n, P(m)=\lambda} e^{2\pi i m\cdot\xi} \ , \quad \xi \in \Pi^d$$

Here $\Pi^d = \mathbf{R}^d/\mathbf{Z}^d$ is the flat torus. Let us introduce the measure: $d\sigma_P(x) = \frac{dS_P(x)}{|P'(x)|}$, where $dS_Q(x)$ denotes the Euclidean surface area measure of the level surface $P(x) = 1$, and $|P'(x)|$ is the magnitude of the gradient of the form P, and its Fourier transform:

$$d\tilde{\sigma}_P(\xi) = \int_{\{x\in\mathbf{R}^d\,:\,P(x)=1\}} e^{2\pi i x\cdot\xi}\, d\sigma_P(x) \tag{2.11}$$

Now we can state

Lemma 2 *Let $P(m)$ be a positive integral non-degenerate form of degree k, then there exists $\delta > 0$, s.t.*

$$\hat{\sigma}_{P,N}(\xi) = C_P N^{\frac{d}{k}-1} \sum_{q=1}^{\infty} \sum_{l\in\mathbf{Z}^n} K(q,l,N)\psi(q\xi - l) d\tilde{\sigma}_P(N^{\frac{1}{k}}(\xi - s/q)) +$$

$$+\,\mathcal{E}_N(\xi) \quad , \quad \textit{where} \quad \sup_{\xi} |\mathcal{E}_N(\xi)| \le c_\delta N^{\frac{d}{k}-1-\delta} \tag{2.12}$$

Here $\psi(\xi)$ is a smooth cut-off function.

The approximation formula (2.12) means, that the Fourier transform (of the indicator function) of the solution set $P(m) = N$ is asymptotically a sum over all rational points, of pieces of the Fourier transform of a surface measure on the level set $P(x) = N$, multiplied by arithmetic factors and shifted by rationals.

We sketch below how to derive formula (2.12) and how to use it to prove the mean ergodic theorem. Let $M = N^{1/k}$, and let $\phi(x)$ be smooth cut-off function on $\mathbf{R}^d$ s.t. $\phi(x) = 1$ for $P(x) \le 2$. Then

$$\hat{\sigma}_{P,N}(\xi) = \sum_{m\in\mathbf{Z}^d} e^{2\pi i m\cdot\xi}\phi(m/M) \int_0^1 e^{2\pi\alpha i(P(m)-N)}\, d\alpha =$$

$$= \int_0^1 S(\alpha,\xi) e^{-2\pi i N\alpha}\, d\alpha$$

where $S(\alpha,\xi) = \sum_m e^{2\pi i(\alpha P(m) + m\cdot\xi)}\phi(m/M)$. As is usual in the Hardy-Littlewood method the main contribution to the integral comes from the major arcs, that is from the intervals : $|\alpha - a/q| \le q^{-1}M^{-d+\varepsilon}$ where $(a,q) = 1$ and $q \le M^\varepsilon$. Indeed by applying the Weyl-type estimates developed by Birch one estimates the integral over the complement of $|S(\alpha,\xi)|$ by $C\,N^{d/k-1-\delta}$ for some $\delta > 0$ uniformly in ξ.

Now let a/q be fixed with $(a,q)=1$ and $q \leq M^\varepsilon$ and write $\alpha = a/q+\beta$, $|\beta| \leq M^{-k+\varepsilon}$, $m = qm_1+s$. We have

$$S(\alpha,\xi) = \sum_{s\in\mathbf{Z}^d/q\mathbf{Z}^d} e^{2\pi i\frac{aP(s)}{q}} \sum_{m_1\in\mathbf{Z}^d} e^{2\pi i(\beta P(qm_1+s)+(qm_1+s)\cdot\xi)}\phi(\frac{qm_1+s}{P})$$

Let $H(x,\beta) = e^{2\pi i\beta P(x)}\phi(x/M)$, applying Poisson summation for the inner sum we get

$$\sum_{m_1} H(qm_1+s)e^{2\pi i(qm_1+s)\cdot\xi} = q^{-n}\sum_l e^{2\pi i\frac{l\cdot s}{q}}\tilde{H}(\xi-l/q,\beta)$$

where

$$(2.13)\quad \tilde{H}(\eta,\beta) = M^d\int_{\mathcal{R}^d} e^{-2\pi i(M^k\beta P(x)-Mx\cdot\eta}\phi(x)\,dx$$

stands for the Fourier transform of $H(x,\beta)$. This already gives an approximation resembling (2.12)

$$\hat{\sigma}_{P,N}(\xi) = \sum_{q\leq M^\varepsilon,\ (a,q)=1} e^{-2\pi iaN/q}\sum_{l\in\mathbf{Z}^n} G(a,l,q)J_N(\xi-l/q)\ +\ Error$$

where $J_N(\eta) = \int_{|\beta|\leq P^{-k+\varepsilon}}\tilde{H}(\eta,\beta)e^{-2\pi iN\beta}\,d\beta$, and $G(a,l,q)$ is the normalized exponential sum obtained by collecting the terms depending on s.

Notice that the gradient of phase $\Phi(x) = M^k\beta P(x) - Mx\cdot(\xi-l/q)$ in (2.13) is at least $M^{1-\varepsilon}|q\xi-l|$ for $|q\xi-l| \gg 1$ on the support of $\phi(x)$. Thus using partial integration one can insert the factors $\psi(q\xi-l)$ by making a small error.

Next one uses a uniform estimate for the integral

$$(2.14)\quad |\tilde{H}(\eta,\beta)| \leq CM^d(1+M^k|\beta|)^{-\frac{codim\ V_p}{(k-1)2^k}}$$

This estimate which is uniform in η is far from obvious even on the non-singular case when $V_P = 0$. It follows from a Weyl type estimate for exponential sums and a comparison of the integral to them, and it would be desirable to obtain such type of estimates directly. It allows one to extend the integration in β to the whole real line by making an error $\approx N^{d/k-1-\delta}$ which is smaller then the main term.

One can evaluate the integral in the sense of distributions, as follows

$$(2.15)\quad I_N(\eta) = \int_{\mathbf{R}}\tilde{H}(\eta,\beta)e^{-2\pi iN\beta}\,d\beta\ =$$

$$= \int_{\mathbf{R}^d} \int_{\mathbf{R}} e^{-2\pi i \beta (P(x)-N)} d\beta \, e^{2\pi i x\cdot\eta} \phi(x/M)\, dx =$$

$$= \int_{P(x)=N} e^{2\pi i x\cdot \eta}\, d\sigma_{P,N}(x) = d\tilde{\sigma}_{P,N}(\eta) = N^{d/k-1}\, d\tilde{\sigma}_P(N^{1/k}\eta)$$

Here the third inequality is an oscillatory integral expression for the measure $d\sigma_{P,N}$ supported on the surface $P(x) = N$ and the last equality follows by scaling. If we put together these transformations and extend the summation in q, which is possible by using standard estimates for the exponential sums $G(a,l,q)$, we get the asymptotic formula (2.9).

The mean ergodic theorem: We illustrate the use of this formula by proving the L^2 convergence of the averages in (2.9). First notice that by substituting $\xi = 0$ into (2.12) we get $r_P(N) = |S_N| \approx N^{d/k-1}$, and one can easily show

Proposition 1 *Let $\xi \notin \mathcal{Q}^d$, that is assume that ξ has at least one irrational coordinate, then one has*

$$\lim_{N\to\infty} \frac{1}{r_P(N)} |\hat{\sigma}_{P,N}(\xi)| = 0 \tag{2.16}$$

Indeed because of the factors $\psi(q\xi - l)$ there is at most one non-zero term in the l summation for each q. After normalizing with the factor $N^{d/k-1}$ each term is bounded by:

$$K(q,l,N) = q^{-d} \sum_{(a,q)=1} e^{-2\pi i a N/q} G(a,l,q) \ll q^{-1-\delta}$$

using standard estimates for exponential sums. Thus if we fix an $\epsilon > 0$ the sum in q for say $q \geq q_\epsilon$ is bounded by ϵ. However for fixed $q < q_\epsilon$ the non-zero term can be estimated by: $d\tilde{\sigma}(\frac{N^{1/k}}{q}\|q\xi\|) \ll (N^{1/2k}\|q\xi\|)^{-\delta}$ if N is large enough w.r.t. q_ϵ, where $\|q\xi\| = \min_{l\in\mathbf{Z}^d} |q\xi - l| > 0$ because $\xi \notin \mathcal{Q}^d$. Here we used the uniform decay of the Fourier transform of the measure $d\sigma_P$, which can be derived from (2.14). So as $N \to \infty$ each term in q individually tend to zero an this proves (2.16).

Now let (X,μ) be a probability measure space, $T = (T_1 \dots T_n)$ be a family of commuting, measure preserving and invertible transformations. By the Spectral theorem there exists a positive Borel measure ν_f on the torus Π^n, s.t.

$$\langle P(T_1,\dots,T_n)f, f\rangle = \int_{\Pi^n} p(\xi) d\nu_f(\xi) \tag{2.17}$$

for every polynomial $P(z_1, \ldots, z_n)$, where

$$p(\xi) = p(\xi_1, \ldots, \xi_n) = P(e^{2\pi i \xi_1}, \ldots, e^{2\pi i \xi_n})$$

and $\langle , \rangle$ denotes the inner product on $L^2(X, \mu)$. We recall two basic facts

i) For $r \in \Pi^n$, $\nu_f(r) > 0$ if and only if r is a joint eigenvalue of the shifts T_j, (i.e. there exists $g \in L^2(X)$ s.t. $T_j g = e^{2\pi i r_j} g$ for each j.

ii) If the family $T = (T_1, \ldots, T_n)$ is ergodic, then $\nu_f(0) = |\langle f, \mathbf{1}\rangle|^2 = |\int_X f d\mu|^2$. We first observe that the full ergodicity is in fact a condition on the joint spectrum of the shifts T_j.

Proposition 2 *Suppose the family $T = (T_1, \ldots, T_d)$ is ergodic. Then it is fully ergodic if and only if $\nu_f(r) = 0$ for every $r \in \mathbf{Q}^d$, $r \neq 0$.*

To see that suppose $\nu_f(l/q) > 0$ for some $l \neq 0$, then there exists $g \in L^2(X, \mu)$ s.t. $T_j g = e^{2\pi i l_j/q} g$ for all j. But then $T_j^q g = g$ for all j but $g \neq$ *constant* since $l \neq 0$. On the other hand suppose that $T_j^q g = g$, for all j for some $g \neq$ *constant*. Then the functions $g_{s_1 \ldots s_d}$ for $s \in \mathbf{Z}^d / q\mathbf{Z}^n$ defined by

$$g_{s_1 \ldots s_n} = \sum_{m \in \mathbf{Z}^n / q\mathbf{Z}^n} e^{-2\pi i \frac{m \cdot s}{q}} T_1^{m_1} \ldots T_n^{m_n} g$$

are joint eigenfunctions of with eigenvalues s_j/q. They cannot vanish for all $s \neq 0$ $(mod\ q)$, because then one would have $T_j g = g$ for every j, as can be seen easily by expressing $T_j g$ in terms of the functions $g_{s_1 \ldots s_n}$ this proves the proposition.

Proof of the Mean Ergodic Theorem. We start rewriting the statement in the form

$$\|S_N f - \langle f, \mathbf{1}\rangle \mathbf{1}\|_2^2 = \|S_N f\|_2^2 - |\langle f, \mathbf{1}\rangle|^2 = \int_{\Pi^n/\{0\}} \frac{|\hat{\sigma}_{P,N}(\xi)|^2}{r_P(N)^2}\, d\nu_f(\xi)$$

The point is that $\nu_f(\mathbf{Q}^d/\{0\}) = 0$ by the full ergodicity condition, moreover the integrand pointwise tends to zero on the irrationals by Lemma 2, and is majorized by $\mathbf{1}$. It follows from the Lebesgue dominant convergence theorem, that the integral also tends to 0 as $N \to \infty$. This proves the theorem.

The proof of the l^2 boundedness of the associated discrete maximal function and that of the pointwise ergodic theorem is much more involved. Both uses the approximation formula by looking the main term as a weighted sum of averaging operators over the level surfaces $P(x) = N$. The associated maximal operators are bounded on $\mathbf{R}^d$ and by using a general transference argument one can pass the estimates to $\mathbf{Z}^d$. Another difficulty is that one cannot seem to use square function arguments and the maximal operators associated with the error terms has to be bounded by using directly the structure of the operators. The interested reader can consult [MSW] for the case of spheres and [M] for the general case.

3 Polynomial averages on the discrete Heisenberg goup.

In this section we sketch the proof of the simplest non-commutative analogue of Bourgain's pointwise ergodic theorem. As before they key tool is to show the l^2 boundedness of the associated maximal function. In doing that one has to reformulate some of the basic estimates, such as the Weyl summation, into an operator valued settings. As the Fourier transform of the central variables is used, the method is somewhat limited, at the end we indicate the type of more general results which seem to be obtainable via these arguments.

3.1 The discrete maximal function on H^d

Let $\Gamma = H^d_{pol} = \{(m,l) \in \mathbf{Z}^{2d} \times \mathbf{Z} : \ (n,k)\cdot(m,l) = (n+m, k+l+n_2\cdot m_1\}$ where
$n = (n_1, \dots, n_d)$, $m = (m_1, m_2)$ and $n_2 \cdot m_1$ denotes the scalar product in $\mathbf{Z}^d$. We remark that H^d_{pol} is isomorphic to the standard Heisenberg group H^d with the product law $(n,k)\cdot(m,l) = (n+m, k+l+n\circ m\}$, where $n\circ m = n_2\cdot m_1 - n_1\cdot m_2$ thus the maximal theorem transfers from one group to the other.

Let $p(n) : \ \mathbf{Z}^{2d} \to \mathbf{Z}$ be an integral polynomial of degree at most 2. Consider the family of surfaces $S_N = \{(n,p(n)) : \ |n_i| \le N, \ \forall 1 \le i \le 2d\}$. Then one has

Theorem 3 *Let $f \in l^2(H^d)$ and define the averages and the corresponding maximal function by*

$$(3.1) \quad S_N f(h) = \frac{1}{|S_N|}\sum_{g\in S_N} f(g\cdot h) \quad , \quad S^*f(h) = \sup_{N>0} |S_N f(h)|$$

Then one has $\|S^*f\|_{l^2(H^d)} \le C_{p,d}\|f\|_{l^2(H^d)}$.

Note that if $h = (m,l)$ then
$S_N f(m,l) = (2N+1)^{-2d}\sum_{|n_i|\le N} f(n+m, l+n\circ m+p(n))$. As in the com-

mutative case it is enough to consider dyadic values $N = 2^j$ and smoothed averages, which after a change of variables $n \to n - m$ look like

$$(3.2)\quad S_j f(m,l) = 2^{-2jd} \sum_n \phi(\frac{n-m}{2^j}) f(n, l + n \circ m + p(n-m))$$

(using $m \circ m = 0$).

Suppose now that H^d_{pol} acts on a probability measure space (X, μ) via measure preserving invertible transformations $T_1, \ldots T_{2d}, S$ as its generators. This means that they satisfy the commutation relations:

$$(3.3)\quad [T_i, T_{i+d}] = S \quad \text{and} \quad [T_i, T_j] = [T_i, S] = I \quad \text{if} \quad |j - i| \neq d$$

where I denotes the identity on X. Indeed from (3.3) it is easy to see that $T^n S^k T^m S^l = T^{m+n} S^{k+l+n_2 \cdot m_1}$. We call the action of H^d_{pol} on X fully ergodic if for each q and $F \in L^2(X)$ one has that $T^q_1 F = \ldots = T^q_{2d} F = S^q F = F$ implies that F is constant. Now we can formulate

Theorem 4 *Let $F \in L^2(X)$, $S_N \subseteq H^d_{pol}$ defined as above. Assume that the action of H^d_{pol} on X is fully ergodic. The one has*

$$(3.4)\quad \lim_{N \to \infty} \frac{1}{|S_N|} \sum_{g \in S_N} f(g \cdot x) = \int_X F \, d\mu$$

for a.e. $x \in X$.

Let us remark that similarly as in Theorem 1., the averages on the left side of (3.4) converge a.e. to some function $F_* \in L^2(X)$ without assuming full ergodicity. However in our approach the essential part is to prove (3.4).

The discrete maximal function on H^d: By taking the Fourier transform in the central variable one has

$$(3.5)\quad S_j f(m,l) = \int_0^1 e^{-2\pi i l\theta} (T^\theta_j \hat{f}(\theta))(m) \, d\theta$$

where

$$\hat{f}(\theta)(n) = \hat{f}(n, \theta) = \sum_k e^{2\pi i k\theta} f(n,k)$$

and

$$(3.5)\quad T^\theta_j g(m) = \sum_n \phi_j(n-m) e^{-2\pi i\theta(n \circ m + p(n-m))} g(n)$$

here $\phi_j(n) = 2^{-2jd}\phi(2^{-j}n)$. It is immediate to see that:
$\|T^\theta_j\|_{l^2(\mathbf{Z}^{2d}) \to l^2(\mathbf{Z}^{2d})} \ll 1$. A crucial point is that $\|T^\theta_j\|_{l^2 \to l^2}$ is small unless *theta* is close to a rational a/q with small denominator q. The following is the analogue of Lemma 1. in section 2.1

Lemma 3 *Let $0 \leq \theta < 1$ and suppose there exists a rational a/q such that $|\theta - a/q| \leq 1/q^2$. Then*

$$(3.7)\quad \|T_j^\theta\|_{l^2 \to l^2} \leq C \max(q^{-1}, q2^{-2j})^{1/2}$$

The proof is based on estimating the kernel $K_j^\theta(m_1, m_2)$ of $T_j^\theta(T_j\theta)^*$ which is the following exponential sum,

$$(3.8)\quad \sum_n \phi_j(n-m_1)\phi_j(n-m_2)e^{-2\pi i\theta(n\cdot A(m_1-m_2)+R(m_1,m_2))} + O(2^{-j/2})$$

Indeed the phase is of the form:
$n \circ m_1 - n \circ m_2 + p(n-m_1) - p(n-m_2) = n \cdot A(m_1 - m_2) + R(m_1, m_2)$ where A is a $2d \times 2d$ matrix with rank at least 1, and "$\cdot$" denotes the dot product. This sum is similar to the one appeared in Lemma 1., and can be estimated analogously.

Major arcs decomposition: Let $1 < \alpha < 2$ and $\varepsilon > 0$ be fixed. We write $S_j f = M_j f + E_j F$ where

$$M_j f(m,l) = \sum_{a/q,\ q \leq j^\alpha} \int_0^1 e^{-2\pi i l\theta}\omega(2^{(2-\varepsilon)j}(\theta - a/q))\,(T_j^\theta \hat{f}(\theta))\,(m)\,d\theta$$

$$= M_{j,0}f(m,l) + \sum_{a/q,\ 1<q\leq j^\alpha} M_{j,a/q}f(m,l)$$

As before it follows from Dirichlet's principle and Lemma 3. that

$$(3.9)\quad \|E_j f\|_{l^2} \ll j^{-\alpha/2}\|f\|_{l^2}$$

and thus

$$\sup_j \|E_j f\|_{l^2} \leq (\sum_j \|E_j f\|_{l^2}^2)^{1/2} \ll 1$$

so it is enough to prove that

$$\|M^* f\|_{l^2} = \|\sup_j |M_j f|\,\|_{l^2} \leq C\|f\|_{l^2}$$

Comparison to standard averages: First we discuss the maximal operator $M_0^* f = \sup_j |M_{j,0}f|$ and write $M_{j,0}f = B_j f + E_{j,0}f$ where

$$B_j f(m,l) = \int_0^1 e^{-2\pi i l\theta}(T_j^\theta \hat{f}(\theta))\,(m)\omega(2^{2j}\theta)\,d\theta$$

The point is that $B_j f$ is a standard average. Indeed

$$B_j f(m,l) = \int_0^1 \sum_n 2^{-2jd} \phi(\frac{n-m}{2^j}) e^{-2\pi i\theta(l+n\circ m+p(n-m))} \hat{f}(n,\theta)\omega(2^{2j}\theta)\, d\theta$$

$$= 2^{-2jd} 2^{-2j} \sum_{n,k} \phi(\frac{n-m}{2^j}) \hat{\omega}(\frac{l+n\circ m+p(n-m)}{2^{2j}})\, f(n,k)$$

which is the smoothed average corresponding to the family of balls:

$$B_j(m,l) = \{(n,k):\ |n-m| < 2^j,\ |l+n\circ m+p(n-m)| < 2^{2j}\}$$

It is easy to see that they satisfy (i)-(i3) described in the introduction and hence $\|B^* f\|_{l^2} = \|\sup_j |B_j f|\,\|_{l^2} \ll \|f\|_{l^2}$. We still have to deal with the error term $E_{j,0}f$ which corresponds to the range $2^{-2j} \ll |\theta| \ll 2^{-(2-\varepsilon)j}$. However on such range a similar argument to that of Lemma 3. shows that:

$$(3.10)\quad \|T_j^\theta\|_{l^2\to l^2} \ll (1+2^{2j}|\theta|)^{-1/2} + 2^{-j/4}$$

Indeed the kernel of $T_j^\theta (T_j\theta)^*$ can be compared to the integral

$$\bar{K}_j^\theta(m_1,m_2) = \int \phi_j(x-m_1)\phi_j(x-m_2) e^{-2\pi i\theta(x\cdot A(m_1-m_2)+R(m_1,m_2))} dx$$

by writing $x = n+t$ with $0 \le t_i \le 1$ and making an error of $O(2^{-j/2}$ because θ is small. From this estimate (3.10) follows by a straightforward computation and the error terms $E_{j,0}f$ are handled by the corresponding square-function.

Reduction mod Q: Essentially the same arguments work on each major arc $I_{a/q} = \{\theta:\ |\theta - a/q| \le q^{-1}2^{-2j}\}$ after a decomposition mod q. An extra difficulty arising is that one has to avoid adding up all of these estimates, by handling a large number of major arcs simultaneously, this idea first appeared in [B1] in the commutative settings.

Let $1/2 < \rho < 1$ such that $\rho\alpha < 1$. Group the denominators $q \le j^\alpha$ into groups Λ of size $2^{\rho s}$ where for each $q \in \Lambda$ we have $2^s \le q < 2^{s+1}$. Further put $Q = \prod_{q\in\Lambda} q$. Notice that $Q \ll_\epsilon 2^{\epsilon j}\ \forall \epsilon > 0$. To each group of rationals Λ there correspond the maximal operator:

$$M_\Lambda^* f = \sup_j |M_{\Lambda,j} f| \quad \text{where} \quad M_{\Lambda,j} f(m,l) = \sum_{a/q,\ q\in\Lambda} M_{j,a/q} f(m,l)$$

It is now enough to prove the following estimate

Lemma 4 $\qquad (3.11)\quad \|M_\Lambda^* f\|_{l^2(H^d)} \le C 2^{-s/2} \|f\|_{l^2(H^d)}$

Indeed for each s there are at most $2^{(1-\rho)s}$ groups Λ and adding estimate (3.11) for each group gives the contribution $2^{(1/2-\rho)s} < 2^{-\epsilon s}$ since $\rho > 1/2$. Finally we add up these estimates for each s, using the subadditivity of the maximal operator. The proof of Lemma 4. is computationally involved, we restrict ourselves to highlight the main points. For fixed Λ and the corresponding Q, we decompose each element of the group H^d modulo Q. That is write $m := Qm+r,\ n := Qn+s,\ l := Ql+t$ where $0 \leq r_i, s_i, t_i < Q$. Next on a major arc write $\theta = \beta + a/q$ with $|\beta| \leq 2^{-2j}$. We have

$$\text{(3.12)}\quad M_{j,a/q}f(Qm+r,Ql+t) = \int_0^1 e^{-2\pi i\beta Q(l+m\circ r)}\omega(2^{2j}\beta)\cdot$$

$$\cdot(\sum_n \phi_j(Qm-Qn)e^{-2\pi i\beta(Qn\circ Qm+P_0(Qn-Qm))})\cdot$$

$$\cdot(\sum_s e^{-2\pi i(s\circ r+P(s-r)+t)a/q}\hat{f}(Qn+s,\beta+a/q)\omega(Q\beta))\,d\beta + E_{j,a/q}f$$

What we did here was to separate the terms involving multiples of Q from those involving only residue classes *mod* Q. One could do this by making a small error $E_{j,a/q}f$ which came from transforming "cross terms" like:
$\beta s\circ Qm \to \beta s\circ Qm$, $\phi_j(Qn-Qm+s-r) \to \phi_j(Qn-Qm)$, $\beta P(Qn-Qm+s-r) \to P_0(Qn-Qm)+\beta P(s-r)$. In each case one makes an error of $O(2^{-j/4})$. For fixed r,t this allows one to view the operator $M_{j,\Lambda} = \sum_{a/q\in\Lambda} M_{j,a/q}$ as a standard averaging operator corresponding to the balls $B_{j,Q} = B_j \cap H^d_Q = \{(Qm,Ql)\ :\ (m,l) \in H^d\}$ applied to some function $g_\Lambda \in l^2(H^d_Q)$. The boundedness of the maximal function M^*_Λ follows from the standard theory, while $\|g_\Lambda\|_{l^2}$ can be estimated by orthogonality arguments. To be more precise (3.12) can be rewritten in the form (for fixed r,t)

$$\text{(3.13)}\quad M_{j,\Lambda}f(Qm+r,Ql+t) = \int_0^1 e^{-2\pi i\beta Q(l+m\circ r)}\omega(2^{2j}\beta\ )\cdot$$

$$\cdot(T^\beta_{j,Q}\hat{g}_\Lambda(\beta))(Qm) + E_{\Lambda,j}f$$

where

$$T^\beta_{j,Q}h\,(Qm) = \sum_n\sum_n \phi_j(Qm-Qn)e^{-2\pi i\beta(Qn\circ Qm+P_0(Qn-Qm))}h(Qn)$$

and

$$\hat{g}_\Lambda(Qn,\beta) = \sum_{a/q\in\Lambda}\sum_s \omega(Q\beta)e^{-2\pi i(s\circ r+P(s-r)+t)a/q}\,\hat{f}(Qn+s,\beta+a/q)$$

If we perform the (scaled) inverse Fourier transform

$$\hat{g}_\Lambda(Qn,\beta) = Q\sum_k e^{2\pi i Qk\beta} g(Qk)$$

which is valid on $|\beta| < 1/2Q$ since $\hat{g}$ is supported on that interval we get the standard averages on H^d_Q. Finally we remark that the oscillatory phase and the disjointness of the supports of the the pieces of $\hat{f}$ in the expression for $\hat{g}$, makes it possible to get a good bound for $\|g_\Lambda(Qn,\beta)\|_{l^2(H^d_Q)}$ in terms $\|f\|_{l^2(H^d)}$.

3.2 The pointwise ergodic theorem

The proof of Theorem 4. consists of two parts, and was largely motivated by the arguments in in [B1]. First a refined version of the transfer principle is used to replace pointwise convergence by L^2 estimates for truncated maximal functions. Then the discrete maximal theorem as well as ideas from its proof are utilized to reduce matters to estimate a maximal function attached to a fixed rational a/q, which can be majorized by standard averages on the subgroup H^d_q.

Transfer principle for H^d. Let us start with some standard observations. The sets: $B_N = \{(n,k):\ |n_i| \le N,\ |k-p(n)| \le N^2\}$ and their shifts: $B_N(h) = \{(n,k):\ |n_i| \le N,\ |k-p(n)| \le N^2\}$ satisfy properties $(i)-(i3)$ and hence form a family of balls. Let $F \in L^2(X)$ and let $f_{L,x}$ be the associated function on H^d defined in the introduction. Let $N_k < N_{k+1} < \ldots$ be an increasing sequence of natural numbers and for each k define the truncated maximal functions: $B^*F(x) = max_{N_k \le N < N_{k+1}} B_N F(x)$ and $B^*f(h) = max_{N_k \le N < N_{k+1}} B_N f(h)$. Here $B_N F(x)$ and $B_N f(h)$ are the averages defined in (1.1)-1.2). Notice that $S_N \subset B_N$ hence $S_N F(h\cdot x) = S_N f_{L,x}(h)$ for all $h \in D_{L,N} = \cap_{g\in B_N} g^{-1}\cdot B_L$. The same is true for the maximal functions:

$$S^*_k F(h\cdot x) = S^*_k f_{L,x}(h) \quad \text{for all} \quad h \in D_{L,k} = g \in S_{k+1} g^{-1} \cdot B_L \tag{3.14}$$

Thus the maximal ergodic theorem readily follows from Theorem 3. and one can pass to dense subspace of $L^2(X)$, say to $L^2(X)\cap L^\infty(X)$. Since the surfaces $S_N \subseteq S_{N+1}$ are nested it is enough to consider lacunary values $N = a^j$ for some $a > 1$ let us assume for simplicity that $a = 2$ and use the notation S_j for S_{2^j}. Summarizing the above it is enough to prove that;

If $\|F\|_{L^2(X)} \le 1$, $\|F\|_{L^\infty(X)} \le 1$ and $\int_X F\,d\mu = 0$ and if the action of H^d_{pol} on X is fully ergodic, then one has

$$S_j F(x) = 2^{-2jd} \sum_{|n_i|\le 2^j} F(T^n S^{p(n)} x) \to 0 \tag{3.15}$$

for a.e. $x \in X$. Following [B1], the indirect assumption:

$limsup\,|S_jF(x)| > 0$ on a set of positive measure implies that $limsup\,|S_jF(x)| > \delta$ on a set of measure at least δ for some and $\delta > 0$, thus there exists an increasing sequence j_k such that $\forall\ k$: $\mu\{S_k^*F > \delta\} > \delta$. Then the L^2 norm of S_k^*F is at least $\delta^{3/2}$ and the same is true in Cesaro means:

$$(3.16)\qquad \frac{1}{K}\sum_{k\le K}\|S_k^*F\|^2_{L^2(X)} \ge \delta^3$$

for each $K > 0$, in fact at crucial point one needs to consider Cesaro means. By (3.14) this translates to

$$(3.17)\qquad \frac{1}{K}\sum_{k\le K}\frac{1}{|B_L|}\int_X \|S_k^*f_{L,x}\|^2_{L^2(X)}\,d\mu \ge \delta^3$$

Thus it is enough to show the opposite inequality to (3.17) for every $\delta > 0$ if $k > k_\delta$ and $L > L_{k,\delta}$. Let us remark at this point that if the averages S_jF are replaced by the ball averages B_jF then the above can be proved easily:

Lemma 5 *Let $\|F\|_{L^2(X)} \le 1$, $\|F\|_{L^\infty(X)} \le 1$ and $\int_X F\,d\mu = 0$ and assume that the action of H^d_{pol} on X ergodic. Then $\forall\ \delta > 0$ one has*

$$(3.18)\qquad \frac{1}{K}\sum_{k\le K}\frac{1}{|B_L|}\int_X \|S_k^*f_{L,x}\|^2_{l^2}\,d\mu \ge \delta^3$$

for $k > k_\delta$ and $L > L_{k,\delta}$.

Indeed using the transfer principle (3.18) can be reduced to the fact that: $\|B_k^*F\|^2_{L^2(X)} < \delta^3$ if $k > k_\delta$ which is easy to see from property (i).

Reduction to a fixed rational a/q. Fix $\delta > 0$ and let $f = f_{L,x}$. As before one has the decomposition: $S_jf = M_jf + E_jF$ where

$$M_jf = M_{j,0}f + \sum_{2^s\le j^\alpha} M_{j,s}f \quad\text{where}\quad M_{j,s}f = \sum_{a/q,\ 2^s\le q<2^{s+1}} M_{j,a/q}f$$

By (3.9) and (3.11) one has

$$\|E_jf\|_{l^2} \ll j^{-\alpha/2}\|f\|^2_{l^2} \quad\text{and}\quad \|M_{j,s}f\|^2_{l^2} \le C2^{-\epsilon s}\|f\|^2_{l^2}$$

thus for some $\epsilon > 0$ and by Schwarz' inequality

$$(3.19)\qquad \|M_k^*f\|^2_{l^2} \le \sum_{s\le s_\delta}|M_{s,k}^*f\|^2_{l^2} + c\delta^3\|f\|^2_{l^2}$$

if $k > k_\delta$ and k_δ, s_δ are chosen large enough. By subadditivity of the maximal operator the l.s. of (3.19) is majorized by a finite sum of the L^2 norm square of maximal operators: $M^*_{k,a/q}f = \max_{j_k \le j < j_{k+1}} |M_{j,a/q}f|$. Hence it is enough to show (3.18) for $M^*_{k,a/q}f$ in place of $B^*_k f$ for a fixed rational a/q. The idea is to replace the cut-off $\omega(2^{(2-\varepsilon)j}(\theta - a/q))$ by $\omega(2^{2j}(\theta - a/q))$ and use the discrete maximal theorem to bound the maximal function $M^*_{k,a/q}f$ by a standard type average $B_{j,a/q}f$.

To do this define the bump functions: $\omega_k(\beta) = \omega(2^{2j_{k-1}}\beta) - \omega(2^{2j_{k+1}}\beta)$. Note that the sequence j_k can be chosen quickly increasing, say $2j_k+2 < j_{k+1}$ and the the supports of $\omega_k(\beta)$ and $\omega_l(\beta)$ are disjoint for $|k-l| > 2$. Let $\beta = \theta - a/q$ and write

$$\omega(2^{(2-\varepsilon)j}\beta) = \omega(2^{2j}\beta) + \omega_k(\beta)\omega(2^{2j}\beta)$$

Accordingly one has the decomposition

$$M_{j,a/q}f = B_{j,a/q}f + M_{j,a/q}f_{k-1,a/q} - B_{j,a/q}f_{k-1,a/q} \tag{3.20}$$

where

$$\hat{f}_{k,a/q}(n,\beta) = \hat{f}(n,\beta)\omega_k(\beta)$$

and

$$B_{j,a/q}f(m,l) = \int_0^1 e^{-2\pi i l\theta}\omega(2^{2j}(\theta - a/q))\,(T^\theta_j \hat{f}(\theta))\,(m)\,d\theta \tag{3.21}$$

We remark that the l^2 boundedness of both maximal operators $B^*_{k,a/q}$ and $M^*_{k,a/q}$ has already been proved in section 3.1. By the essential disjointness of the supports of the functions $\hat{f}_{k,a/q}$ one has

$$\frac{1}{K}\sum_{k\le K} \|M^*_{k,a/q}f\|^2_{l^2} + \|B^*_{k,a/q}f\|^2_{l^2} \le \frac{C_q}{K}\sum_{k\le K}\|f_{k,a/q}\|^2_{l^2} \le \frac{5C_q}{K}\|f\|^2_{l^2}$$

and note that

$$\frac{1}{|B_L|}\int_X \|f_{L,x}\|^2_{l^2}\,d\mu = \|F\|^2_{L^2} \le 1$$

Now it is to enough to prove (3.18) for the averages: $B\ ^*_{k,a/q}f_{L,x}$. Notice that the only difference between this operator and B^*_k is only the shift by the rational: $\theta \to \theta - a/q$. If one writes out the averages $B_{j,a/q}f$ explicitly using the inverse Fourier transform, and takes the summation in fixed residue classes *mod* q first, then one gets a finite sum (over r, t *mod* q) of averages of the form $B_{j,q} f_{r,t}$, where $B_{j,q}$ is the standard average on the subgroup H_q^d corresponding to the balls $B_j \cap H_q^d$. Finally since $f = f_{L,x}$ corresponds to F, one gets that $f_{r,t}$ corresponds to $F_{r,t}(x) = F(T^r S^t x)$. For fixed r, t one can apply Lemma 5. on the subgroup H_q^d, and that proves Theorem 4.

3.3 General discrete nilpotent groups

As was shown by Bergelson and Liebman [BL], the mean ergodic theorem holds in an amazingly general setting. Let Γ be a finitely generated discrete nilpotent Lie group acting on a finite measure space via measure preserving transformations. In terms of generators this means a finite set of transformations $T_1, \ldots, T_d$ whose long enough commutators trivialize. A sequence $p : \mathbf{Z} \to \Gamma$ is called polynomial if the operation $Dp(n) = p(n)p(n+1)^{-1}$ trivialize after finitely many steps, that is there is a M such that $D^M p\,(n) = I$ for all n where I is the identity element. In terms of the generators this means that $p(n) = T_1^{p_1(n)} \ldots T_d^{p_d(n)}$ for some integral polynomials $p_i(n)$. Under these settings one has

Theorem 5 *(V. Bergelson, A. Liebman) Let $F \in L^2(X)$ the there exists a function $F_* \in L^2(X)$ such that*

$$S_N F = \frac{1}{N} \sum_{n \leq N} T_1^{p_1(n)} \ldots T_d^{p_d(n)} F \to F_* \tag{3.22}$$

in L^2 norm as $N \to \infty$. Moreover if the action of Γ on X is fully ergodic, then $F_* = \int_X F \, d\mu$.

Much of our effort has been devoted to see in what generality the corresponding pointwise ergodic theorem and the l^2 boundedness of the associated maximal operator holds. At this point the following case seems to be within reach;

Assume that Γ is a discrete co-compact subgroup of stratified nilpotent Lie group. Then by Malcev's theory one can introduce canonical coordinates in which Γ is represented by $\mathbf{Z}^d$, see [Co]. Consider maximal averages corresponding to hypersurface transversal to the center and can be parameterized by integral polynomials, we remark that this condition can be expressed in a coordinate free way. Then roughly one considers the averages:

$$S_N F = \frac{1}{(2N+1)^D} \sum_{|n_i| \leq N} F(\, T_1^{n_1)} \ldots T_d^{n_d} S_1^{p_1(n)} \ldots S_k^{p_k(n)} x \,)$$

where S_i are in the center and $p_i(n)$ are integral polynomials, moreover the transformations altogether generate a stratified nilpotent group. In this case it seems that both the maximal and the pointwise ergodic theorem can be proved by the methods described here. There are new essential difficulties arise; in the decomposition *mod* Q, and also if the polynomials have high degrees. The first problem can be resolved by a semi-direct type decomposition; the subgroup Γ_Q consisting of elements Qm whose each coordinate is divisible by Q is normal, and one can write every element $m' \in \Gamma$ in a unique way of the form: $m' = Qm \cdot r$ where $0 \leq r_i < Q$. The set of r's can be identified with the group: Γ/Γ_Q. This way one can avoid large cross terms which makes the reduction possible. A paper containing such results is under preparation [MSW2]. Of course there is still a huge gap, and finally let us mention the simplest settings when the above methods break down. Let T, U be measure preserving transformations on a finite measure space X, such that $S = [T, U] = TUT^{-1}U^{-1}$ commutes with both T and U.

It is plausible to expect that the averages:

$$S_N F = \frac{1}{N} \sum_{n \leq N} F(T^n U^{n^2} x) \to F_*$$

as $N \to \infty$ for a.e. x, and that the l^2 associated discrete maximal operator is bounded from $l^2(\Gamma)$ to itself. This remains however an open problem at present.

References

[BL] V.Bergelson, A.Liebman: *A nilpotent Roth theorem* Invent. Math. 147 (2002), no. 2, 429–470.

[Bi] J.Birch: *Forms in many variable* London Roy. Proc. Soc, Ser. A, (1961)

[B1] J.Bourgain: *On the maximal ergodic theorem for certain subsets of integers* Israeli J. Math., 61, 39-72 (1988)

[B2] J.Bourgain: *On the pointwise ergodic theorem on L^p for arithmetic sets* Israel J. Math. 61 (1988), no. 1, 73–84.

[B3] J.Bourgain: *Pointwise ergodic theorems for arithmetic sets* Inst. Hautes tudes Sci. Publ. Math. No. 69 (1989), 5–45.

[Co] L.Corwin: Irreducible represeantaions of nilpotent Lie groups Cambridge Univ. Press

[M] A.Magyar: *Diophantine equations and ergodic theorems* Amer J. Math. 124 (2002), pp. 921-953

[MSW] A.Magyar, E.M.Stein, S.Wainger: *Discrete analogues in Harmonic Analysis: Spherical Averages* Annals of Math. v. 155, 2002, 189-208

[MSW2] A.Magyar, E.M.Stein, S.Wainger: *Discerete analogues of integral operator: maximal functions on discrete stratified nilpotent groups with applications to ergodic theory*, in preparation

[V] R.C. Vaughan: *Hardy-Littlewood method* Cambridge tracts in mathematics 125 (1997) Cambridge Univ. Press