Math 639: Lecture 15

Multiple ergodic averages

Bob Hough

April 4, 2017

Multiple ergodic averages

The goal of this lecture is to prove the following mean ergodic theorem.

Theorem (Walsh, 2012)

Let G be a nilpotent group of measure preserving transformations of a probability space $(X, \mathscr{X}, \mu)$. Then, for every $T_1, ..., T_l \in G$, for every $f_1, ..., f_d \in L^\infty(X, \mathscr{X}, \mu)$, for every collection of integer valued polynomials $\{p_{i,j}, 1 \leqslant i \leqslant l, 1 \leqslant j \leqslant d\}$, the averages

$$\frac{1}{N}\sum_{n=1}^{N}\prod_{j=1}^{d}\left(T_1^{p_{1,j}(n)}\cdots T_l^{p_{l,j}(n)}\right) f_j$$

converge in $L^2(X, \mathscr{X}, \mu)$.

Background

- The proof is combinatorial in nature, and is based on a structure vs. randomness dichotomy.
- The following structural statements are based on the paper "Decompositions, approximate structure, transference, and the Hahn-Banach theorem," by Tim Gowers (Bull. London Math Soc., 42 (2010) pp. 573–606).

Structure theorems

Theorem (Hahn-Banach)

Let K be a convex body in $\mathbb{R}^n$ and let f be an element of $\mathbb{R}^n$ that is not contained in K. Then there is a constant β and a non-zero linear functional ϕ such that $\langle f, \phi \rangle \geqslant \beta$ and $\langle g, \phi \rangle \leqslant \beta$ for every $g \in K$.

Structure theorems

Theorem

Let $K_1, \ldots, K_r$ be closed convex subsets of $\mathbb{R}^n$, each containing 0, let $c_1, \ldots, c_r$ be positive real numbers, and suppose that f is an element of $\mathbb{R}^n$ that cannot be written as a sum

$$f_1 + \cdots + f_r, \qquad f_i \in c_i K_i.$$

Then there is a linear functional ϕ such that $\langle f, \phi \rangle > 1$ and $\langle g, \phi \rangle \leqslant c_i^{-1}$ for every $i \leqslant r$ and every $g \in K_i$.

Structure theorems

Proof.

- Define $K = \sum_i c_i K_i$.
- Since K is closed, there exists $\epsilon > 0$ and a small Euclidean ball B such that $(1+\epsilon)^{-1} f \notin B + K$.
- Apply Hahn-Banach to find ϕ and β such that $(1+\epsilon)^{-1}\langle f, \phi\rangle \geqslant \beta$ and $\langle g, \phi \rangle \leqslant \beta$ for every $g \in B + K$.
- Since $0 \in K$ we may take $\beta = 1$.

□

Structure theorems

Theorem

Let $K_1, ..., K_r$ be closed convex subsets of $\mathbb{R}^n$, each containing 0 and suppose that f is an element of $\mathbb{R}^n$ that cannot be written as a convex combination

$$c_1 f_1 + \cdots + c_r f_r, \; f_i \in K_i.$$

Then there is a linear functional ϕ such that $\langle f, \phi \rangle > 1$ and $\langle g, \phi \rangle \leqslant 1$ for every $i \leqslant r$ and every $g \in K_i$.

Structure theorems

Proof.

- Let K be the set of all convex combinations $c_1 f_1 + \cdots + c_r f_r$ with $f_i \in K_i$.
- Since K is closed and convex, there is an $\epsilon > 0$ such that $(1+\epsilon)^{-1} f \notin K$.
- By Hahn-Banach, there is a functional ϕ and a constant β such that $(1+\epsilon)^{-1}\langle f, \phi\rangle \geqslant \beta$ and $\langle g, \phi\rangle \leqslant \beta$ for all $g \in K$.
- As before, β may be taken equal to 1, since K is closed.

□

Definition

If $\|\cdot\|$ is a norm on $\mathbb{R}^n$, the dual norm $\|\cdot\|^*$ is defined by the formula

$$\|\phi\|^* = \max\{\langle f, \phi\rangle : \|f\| \leqslant 1\}.$$

The dual of a norm $\|\cdot\|$ defined on a subspace V of $\mathbb{R}^n$ is the seminorm

$$\|f\|^* = \max\{\langle f, g\rangle : g \in V, \|g\| \leqslant 1\}.$$

If $f \in \mathbb{R}^n$ a *support functional* for f is a linear functional $\phi \neq 0$ such that

$$\langle f, \phi\rangle = \|f\|\|\phi\|^*.$$

Structure theorems

Theorem

Let Σ be a set and, for each $\sigma \in \Sigma$, let $\|\cdot\|_\sigma$ be a norm defined on a subspace V_σ of $\mathbb{R}^n$. Suppose that $\sum_{\sigma\in\Sigma} V_\sigma = \mathbb{R}^n$, and define a norm $\|\cdot\|$ on $\mathbb{R}^n$ by the formula

$$\|x\| = \inf\{\|x_1\|_{\sigma_1} + \cdots + \|x_k\|_{\sigma_k} : x_1 + \cdots + x_k = x, \sigma_1, ..., \sigma_k \in \Sigma\}.$$

The dual norm $\|\cdot\|^*$ is given by the formula

$$\|z\|^* = \sup\{\|z\|_\sigma^* : \sigma \in \Sigma\}.$$

Structure theorems

Proof.

- First suppose that $\|z\|_\sigma^* \geqslant 1$ for some $\sigma \in \Sigma$. Then there exists $x \in V_\sigma$ such that $\|x\|_\sigma \leqslant 1$ and $|\langle x, z\rangle| \geqslant 1$. Since $\|x\| \leqslant 1$, $\|z\|^* \geqslant \|z\|_\sigma^*$.
- Now suppose $\|z\|^* > 1$. Then there is x with $\|x\| \leqslant 1$ and $|\langle x, z\rangle| \geqslant 1 + \epsilon$ for some $\epsilon > 0$. Choose $x_1, ..., x_k$ such that $x_i \in V_{\sigma_i}$ for each i, $x_1 + \cdots + x_k = x$ and $\|x_1\|_{\sigma_1} + \cdots + \|x_k\|_{\sigma_k} < 1 + \epsilon$. Then

$$\sum_i |\langle x_i, z\rangle| > \|x_1\|_{\sigma_1} + \cdots + \|x_k\|_{\sigma_k},$$

so there is i with $|\langle x_i, z\rangle| > \|x_i\|_{\sigma_i}$, and $\|z\|_i^* > 1$.

□

Corollary

Let $\Sigma \subset \mathbb{R}^n$ *be a set that spans* $\mathbb{R}^n$ *and define a norm* $\|\cdot\|$ *on* $\mathbb{R}^n$ *by the formula*

$$\|f\| = \inf\left\{\sum_{i=1}^k |\lambda_i| : f = \sum_{i=1}^k \lambda_i\sigma_i,\ \sigma_1, ..., \sigma_k \in \Sigma\right\}.$$

Then this formula does indeed define a norm, and its dual norm $\|\cdot\|^*$ *is defined by the formula*

$$\|f\|^* = \sup\{|\langle f, \sigma\rangle| : \sigma \in \Sigma\}.$$

This is the special case in which V_σ is the span of σ and $\|\lambda\sigma\|_\sigma = |\lambda|$.

Structure theorems

Theorem

Let $\| \cdot \|$ *be any norm on* $\mathbb{R}^n$ *and let* $f \in \mathbb{R}^n$. *Then* f *can be written as* $g + h$ *in such a way that* $\|g\| + \|h\|^* \leqslant \|f\|_2$.

Structure theorems

Proof.

- Let K_1 and K_2 be the unit balls in the $\|\cdot\|$ and $\|\cdot\|^*$ norms.
- Suppose for contradiction that the claim is false. Then $f/\|f\|_2$ is not a convex combination $c_1 g_1 + c_2 g_2$ with $g_i \in K_i$.
- By Hahn-Banach, we obtain ϕ with $\langle f, \phi \rangle > \|f\|_2$ and $\|\phi\|^*$ and $\|\phi\|$ both at most 1.
- The first claim implies $\|\phi\|_2 > 1$, while the second implies $\|\phi\|_2^2 = \langle \phi, \phi \rangle \leqslant \|\phi\| \|\phi\|^* \leqslant 1$, a contradiction.

□

Structure theorems

- Note that by scaling the norm, for any $\epsilon > 0$ it is possible to find g, h, $f = g + h$ with $\epsilon\|g\| + \epsilon^{-1}\|h\|^* \leqslant \|f\|_2$.
- By admitting a small L^2 error, the following decomposition theorem does better by replacing the inverse relationship ϵ, ϵ^{-1} in the two norms, with an arbitrary growth function.

Structure theorems

Theorem

Let $f \in \mathbb{R}^n$ with $\|f\|_2 \leqslant 1$, and let $\|\cdot\|$ be any norm on $\mathbb{R}^n$. Let $\epsilon > 0$ and let $\eta : \mathbb{R}_+ \to \mathbb{R}_+$ be any decreasing positive function. Let $r = \lceil 2\epsilon^{-1} \rceil$ and define a sequence $C_1, ..., C_r$ by setting $C_1 = 1$ and

$$C_i = 2\eta(C_{i-1})^{-1}, \; i > 1.$$

Then there exists $i \leqslant r$ such that f can be decomposed as $f_1 + f_2 + f_3$ with

$$C_i^{-1}\|f_1\|^* + \eta(C_i)^{-1}\|f_2\| + \epsilon^{-1}\|f_3\|_2 \leqslant 1.$$

In particular, $\|f_1\|^* \leqslant C_i$, $\|f_2\| \leqslant \eta(C_i)$ and $\|f_3\|_2 \leqslant \epsilon$.

Proof.

- Suppose, for contradiction, that no such decomposition exists. Applying Hahn-Banach for each i with the convex set

$$K_i = \left\{g = g_1 + g_2 + g_3 : C_i^{-1}\|g_1\|^* + \eta(C_i)^{-1}\|g_2\| + \epsilon^{-1}\|g_3\|_2 \leqslant 1\right\},$$

  there exists ϕ_i satisfying $\langle \phi_i, f\rangle > 1$ and such that

$$\|\phi_i\| \leqslant C_i^{-1},\ \|\phi_i\|^* \leqslant \eta(C_i)^{-1},\ \|\phi_i\|_2 \leqslant \epsilon^{-1}.$$

- Notice

$$\|\phi_1 + \cdots + \phi_r\|_2 \geqslant \langle \phi_1 + \cdots + \phi_r, f\rangle \geqslant r.$$

□

Structure theorems

Proof.

- If $i < j$ then

$$\langle \phi_i, \phi_j \rangle \leqslant \|\phi_i\| \|\phi_j\|^* \leqslant \eta (C_i)^{-1} C_j^{-1} \leqslant \frac{1}{2}$$

so that

$$\|\phi_1 + \cdots + \phi_r\|_2^2 \leqslant \epsilon^{-1} r + \frac{r(r-1)}{2}.$$

This is a contradiction, since $r \geqslant \frac{2}{\epsilon}$.

□

Structure theorems

Walsh uses the following variant of the last structure theorem, in which $\mathbb{R}^n$ is replaced by a Hilbert space $\mathscr{H}$, and on which there are a family of equivalent norms.

Structure theorems

Theorem (Hilbert space decomposition theorem)

Let $\mathscr{H}$ be a Hilbert space with norm $\|\cdot\|$.

- *Let $(\|\cdot\|_N)_{N\in\mathbb{N}}$ be a family of norms on $\mathscr{H}$ equivalent to $\|\cdot\|$, and satisfying $\|\cdot\|_{N+1}^* \leqslant \|\cdot\|_N^*$ for every N.*
- *Let $0 < \delta < c < 1$ be positive real numbers, $\eta : \mathbb{R}_+ \to \mathbb{R}_+$ a decreasing function, and $\psi : \mathbb{N} \to \mathbb{N}$ a function satisfying $\psi(N) \geqslant N$.*
- *Define constants $C_{\lceil 2\delta^{-2}\rceil} := 1$, $C_{n-1} = \max\{C_n, 2\eta(C_n)^{-1}\}$ for $n \geqslant 2$.*
- *For every integer $M_- > 0$ there exists a sequence*

$$M_- \leqslant M_1 \leqslant \cdots \leqslant M_{\lceil 2\delta^{-2}\rceil} \leqslant M_+ = O_{M,\delta,c,\psi}(1), \qquad s.t.$$

for any $f \in \mathscr{H}$ with $\|f\| \leqslant 1$ there is $1 \leqslant i \leqslant \lceil 2\delta^{-2}\rceil$ and integers A, B, $M_- \leqslant A < cM_i < \psi(M_i) \leqslant B$, so that $f = f_1 + f_2 + f_3$,

$$\|f_1\|_B < C_i, \; \|f_2\|_A^* < \eta(C_i), \; \|f_3\| < \delta.$$

Structure theorems

- When $\|f\|_A^*$ is small, we say that f is 'random', while when $\|f\|_B$ is small we say that f is 'structured.' This terminology comes from thinking of $\|f\|_A^*$ as $\|\hat{f}\|_\infty$, the ∞-norm on the Fourier transform, so that $\|f\|_B = \|\hat{f}\|_1$ is the 1-norm on the F.T.
- The win in Walsh's version of the structure theorem is that the structured part in the decomposition is at a higher level than the random part.

Structure theorems

Proof of Walsh's structure theorem.

- Set $A_1 = M_-$, $M_1 := \lceil c^{-1}A_1 + 1 \rceil$ and $B_1 = \psi(M_1)$. If no decomposition with $i = 1$ exists then obtain $\phi_1 \in \mathscr{H}$ satisfying

$$\langle \phi_1, f \rangle \geqslant 1,\ \|\phi_1\|_{B_1}^* \leqslant C_1^{-1},\ \|\phi_1\|_{A_1}^{**} \leqslant \eta(C_1)^{-1},\ \|\phi_1\| \leqslant \delta^{-1}.$$

- Recursively define parameters $A_j := B_{j-1}$, $M_j := \lceil c^{-1}A_j + 1 \rceil$, $B_j := \psi(M_j)$, and, if no decomposition exists with these parameters, find ϕ_j satisfying the corresponding estimates.
- For $i < j$ bound

$$|\langle \phi_j, \phi_i \rangle| \leqslant \|\phi_j\|_{A_j}^{**}\|\phi_i\|_{A_j}^{*} \leqslant \|\phi_j\|_{A_j}^{**}\|\phi_i\|_{B_i}^{*} \leqslant \eta(C_j)^{-1}C_i^{-1} \leqslant \frac{1}{2},$$

and hence $\|\phi_1 + \cdots + \phi_r\|_2^2 \leqslant \delta^{-2}r + \frac{r^2-r}{2}$, which forces the process to terminate as before.

□

Systems of finite complexity

Definition

Fix a probability space X and a nilpotent group G of measure preserving transformations on X.

- A *G-sequence* is a sequence $\{g(n)\}_{n\in\mathbb{Z}}$ taking values in G.
- A tuple $g = (g_1, ..., g_j)$ of G-sequences is a *G-system*.
- Two systems are *equivalent* if they contain the same set of G-sequences, so, for instance, if g and h are G-sequences then (h, g), (g, h) and (g, h, h) are equivalent.

Systems of finite complexity

Definition

- To a pair of G-sequences g, h and positive integer m, associate the G-sequence

$$\langle g|h\rangle_m(n) := g(n)g(n+m)^{-1}h(n+m).$$

- The m-reduction of a system $g = (g_1, ..., g_j)$ is the system

$$g_m^* = (g_1, ..., g_{j-1}, \langle g_j|\mathbf{1}_G\rangle_m, \langle g_j|g_1\rangle_m, \cdots, \langle g_j|g_{j-1}\rangle_m).$$

Systems of finite complexity

Definition (Complexity of a system)

- We say a system g has *complexity* 0 if it is equivalent to the trivial system (1_G).
- Recursively, a system g has complexity d for some positive integer $d \geqslant 1$ if it is not of complexity d' for some $0 \leqslant d' < d$, and it is equivalent to some system h for which every reduction h_m^* has complexity at most $d-1$.
- A system has finite complexity if it has complexity d for some $d \geqslant 0$.

Systems of finite complexity

Definition

- For integer $N \geqslant 1$, and $f_1, \ldots, f_j \in L^\infty(X)$, define *ergodic average*

$$\mathscr{A}_N^g[f_1, \ldots, f_j] = \mathrm{E}_{n \in [N]}\left[\prod_{i=1}^{j} g_i(n) f_i\right].$$

- Convergence of the ergodic averages of a system for all test functions implies convergence of the ergodic averages of an equivalent system for all test functions, since $T(f_1)T(f_2) = T(f_1 f_2)$.
- Given a pair of positive integers N, N', define

$$\mathscr{A}_{N,N'}^g[f_1, \ldots, f_j] = \mathscr{A}_N^g[f_1, \ldots, f_j] - \mathscr{A}_{N'}^g[f_1, \ldots, f_j].$$

Finite complexity theorem

Theorem (Finite complexity theorem)

- *Let G and X as above, and let $d \geqslant 0$.*
- *Let $F : \mathbb{N} \to \mathbb{N}$ be some nondecreasing function $F(N) \geqslant N$ for all N, and let $\epsilon > 0$.*
- *For every integer $M > 0$ there exists a sequence of integers, depending on F, ϵ and d,*

$$M \leqslant M_1 \leqslant \cdots \leqslant M_{K(\epsilon,d)} \leqslant M(\epsilon, F, d)$$

such that

  - *for every system $g = (g_1, ..., g_j)$ of complexity at most d*
  - *for every choice of functions $f_1, ..., f_j \in L^\infty(X)$ with $\|f_i\|_\infty \leqslant 1$*

there exists some $1 \leqslant i \leqslant K_{\epsilon,d}$ such that, for every $M_i \leqslant N, N' \leqslant F(M_i)$, $\left\|\mathscr{A}^g_{N,N'}[f_1, ..., f_j]\right\|_{L^2(X)} \leqslant \epsilon$.

Finite complexity theorem

The finite complexity theorem implies the L^2-convergence of all finite complexity ergodic averages since if $\mathscr{A}_N^g[f_1, ..., f_j]$ fails to converge, then there exists $\epsilon > 0$ and increasing function $F(N)$ so that

$$\left\|\mathscr{A}_{N,F(N)}^g[f_1, ..., f_j]\right\|_{L^2(X)} > \epsilon$$

for every positive integer N.

Reducible functions

From now on we work with the specific choices in the structure theorem

$$\delta := \frac{\epsilon}{96}, \ \eta(x) = \frac{\epsilon^2}{216x}, \ C^* = C_1.$$

Definition (Reducible functions)

Given a positive integer L, we say $\sigma \in L^\infty(X)$, $\|\sigma\|_\infty \leqslant 1$, is an *$L$-reducible function with respect to g* if there exists some integer $M > 0$ and a family $b_0, b_1, ..., b_{j-1} \in L^\infty(X)$ with $\|b_i\|_\infty \leqslant 1$, such that for every positive integer $l \leqslant L$,

$$\left\| g_j(l)\sigma - \mathbf{E}_{m\in[M]} \left[(\langle g_j | \mathbf{1}_G \rangle_m (l)) b_0 \prod_{i=1}^{j-1} (\langle g_j | g_i \rangle_m (l)) b_i \right] \right\|_{L^\infty(X)} < \frac{\epsilon}{16 C^*}.$$

Weak inverse result

Theorem (Weak inverse result for ergodic averages)

Assume the inequality

$$\left\| \mathscr{A}_N^g[f_1, \ldots, f_{j-1}, u] \right\|_2 > \frac{\epsilon}{6}$$

holds for some u, $\|u\|_\infty \leqslant 3C$, *some* $1 \leqslant C \leqslant C^*$ *and some* $f_1, \ldots, f_{j-1} \in L^\infty(X)$ *with* $\|f_i\|_\infty \leqslant 1$. *Then there exists a constant* $0 < c_1 < 1$, *depending only on* ϵ, *such that for every positive integer* $L < c_1 N$ *there is an* L*-reducible function* σ *with*

$$\langle u, \sigma \rangle > 2\eta(C).$$

Weak inverse result

Proof.

Expand the square in the L^2 norm to find

$$\left\|\mathscr{A}_N^g\right\|_2^2 = \left\langle \mathscr{A}_N^g[f_1, \ldots, f_{j-1}, u], \mathrm{E}_{n\in[N]}\left[\left(\prod_{i=1}^{j-1} g_i(n) f_i\right) g_j(n) u\right]\right\rangle$$
$$= \left\langle \mathrm{E}_{n\in[N]}\left[g_j(n)^{-1}\mathscr{A}_N^g[f_1, \ldots, f_{j-1}, u]\prod_{i=1}^{j-1} g_j(n)^{-1} g_i(n) f_i\right], u\right\rangle.$$

Define

$$h := \mathrm{E}_{n\in[N]}\left[g_j(n)^{-1}\mathscr{A}_N^g[f_1, \ldots, f_{j-1}, u]\prod_{i=1}^{j-1} g_j(n)^{-1} g_i(n) f_i\right].$$

Set $\sigma = \frac{h}{3C}$. We claim that σ is L-reducible for every $L > c_1 N$, some $0 < c_1 < 1$. This suffices since $\langle u, \sigma\rangle > 2\eta(C)$. □

Weak inverse result

Proof.

Let $c_1 := \frac{\epsilon}{96(C^*)^2}$ and let $0 < l < c_1 N$. Use $\left\|\mathscr{A}_N^g[f_1, \ldots, f_{j-1}, u]\right\|_\infty \leqslant 3C \leqslant 3C^*$. Since the average is short,

$$\left\| h - \mathrm{E}_{n\in[N]}\left[g_j(l+n)^{-1}\mathscr{A}_N^g[\cdot]\prod_{i=1}^{j-1} g_j(l+n)^{-1}g_i(l+n)f_i\right]\right\|_{L^\infty(X)} < \frac{\epsilon}{16C^*}.$$

Shifting by $g_j(l)$,

$$\left\| g_j(l)h - \mathrm{E}_{n\in[N]}\left[\langle g_j|\mathbf{1}_G\rangle_n(l))\mathscr{A}_N^g[\cdot]\prod_{i=1}^{j-1}(\langle g_j|g_i\rangle_n(l))f_i\right]\right\|_{L^\infty(X)} < \frac{\epsilon}{16C^*}.$$

Choose $M := N$, $b_0 = \frac{1}{3C}\mathscr{A}_N^g[\cdot]$ and $b_i = f_i$. □

Bounds for structured functions

Theorem (Stability of averages for structured functions)

For every positive integer M_* there exists $\tilde{K} = \tilde{K}(\epsilon, d)$, and a sequence

$$M_* \leqslant M_1 \leqslant \cdots \leqslant M_{\tilde{K}} \leqslant M^*$$

depending on M_*, ϵ, d, F such that if

- $f_1, ..., f_{j-1} \in L^\infty(X)$, $\|f_i\|_\infty \leqslant 1$
- $f = \sum_{t=0}^{k-1} \lambda_t \sigma_t$, $\sum_{t=0}^{k-1} |\lambda_t| \leqslant C^*$ *and each σ_t is an L-reducible function for some $L \geqslant F(M^*)$*

then there exists some $1 \leqslant i \leqslant \tilde{K}$ such that

$$\left\| \mathscr{A}^g_{N,N'}[f_1, ..., f_{j-1}, f] \right\|_{L^2(X)} \leqslant \frac{\epsilon}{4}$$

for every pair $M_i \leqslant N, N' \leqslant F(M_i)$.

Bounds for structured functions

Proof.

Since σ_t is L-reducible, choose corresponding integer $M^{(t)}$ and functions $b_i^{(t)} \in L^\infty(X)$. Using the reducibility, replace $\mathscr{A}_N^g[f_1, \ldots, f_{j-1}, \sigma_t]$ with

$$\mathrm{E}_{[M^{(t)}]}\left[\mathrm{E}_{[N]}\left(\prod_{i=1}^{j-1} g_i(n) f_i\right)\left((\langle g_j|\mathbf{1}_G\rangle_m(n))b_0^{(t)}\right)\left(\prod_{i=1}^{j-1}(\langle g_j|g_i\rangle_m(n))b_i^{(t)}\right)\right]$$

making error at most $\frac{\epsilon}{16C*}$. Thus, for $N, N' \leqslant L$, $\left\|\mathscr{A}_{N,N'}^g[f_1, \ldots, f_{j-1}, f]\right\|_2$ is bounded by

$$\frac{\epsilon}{8} + \sum_{t=0}^{k-1} |\lambda_t| \, \mathrm{E}_{m\in[M_t]} \left\|\mathscr{A}_{N,N'}^{g_m^*}\left[f_1, \ldots, f_{j-1}, b_0^{(t)}, b_1^{(t)}, \ldots, b_{j-1}^{(t)}\right]\right\|_{L^2(X)}.$$

□

Bounds for structured functions

Proof.

- Let $\gamma = \frac{\epsilon}{16C*}$.
- Since g_m^* is lower complexity than g, we invoke the bounded complexity theorem inductively. Recall that this theorem provides for some $1 \leqslant i \leqslant K_{\gamma,d-1}$ a range $M_i^{\gamma,F,d} \leqslant N \leqslant F\left(M_i^{\gamma,F,d}\right)$, such that the average at length N varies by at most γ over the interval.
- Our goal now is to find an interval $[M, M']$ over which this is valid for many t.

□

Bounds for structured functions

Proof.

- Let $r = O_{\epsilon,d}(1)$ and define functions $F_1, F_2, \ldots, F_r : \mathbb{N} \to \mathbb{N}$ given by

$$F_r = F, \qquad F_{i-1}(N) := \max_{1 \leqslant M \leqslant N} F_i(M^{\gamma, F_i, d-1}).$$

- For each tuple $1 \leqslant i_1, \ldots, i_s \leqslant K$, $s \leqslant r$ and integer M, define

$$M^{(i_1,\ldots,i_s)} := \left(\cdots \left(\left(M_{i_1}^{\gamma, F_1, d-1} \right)_{i_2}^{\gamma, F_2, d-1} \right) \ldots \right)_{i_s}^{\gamma, F_s, d-1}.$$

Thus $M^{(i_1)}$ is the integer $M_{i_1}^{\gamma, F_1, d-1}$ found by starting the sequence at M using F_1, $M^{(i_1,i_2)}$ the result of starting at $M^{(i_1)}$ using F_2, etc. Thus

$$\left[M^{(i_1)}, F_1\left(M^{(i_1)}\right) \right] \supset \left[M^{(i_1,i_2)}, F_2\left(M^{(i_1,i_2)}\right) \right] \supset \ldots$$

□

Bounds for structured functions

Proof.

- Note $\left\| \mathscr{A}_{N,N'}^{g_m^*}\left[f_1, ..., f_{j-1}, b_0^{(t)}, ..., b_{j-1}^{(t)}\right]\right\|_{L^\infty(X)} \leqslant 2$. Hence

$$\sum_{t=0}^{k-1} |\lambda_t| \, \mathbf{E}_{m\in[M_t]} \left\| \mathscr{A}_{N,N'}^{g_m^*}\left[f_1, ..., f_{j-1}, b_0^{(t)}, b_1^{(t)}, ..., b_{j-1}^{(t)}\right]\right\|_{L^2(X)} \leqslant 2C^*.$$

- Applying the finite complexity theorem inductively, the reduced average at t is bounded by γ for all pairs $N, N' \in \left[M_*^{(i)}, F_1\left(M_*^{(i)}\right)\right]$ for some $1 \leqslant i \leqslant K$ which depends on t.
- By the pigeonhole principle we can pick i_1 so that the sum of $|\lambda_t|$ for which $i \neq i_1$ is at most $\left(1 - \frac{1}{K}\right) C^*$.

□

Bounds for structured functions

Proof.

- Iterate the argument using $M^{(i_1)}$, F_2, etc. r times to find $M^{(i_1,\ldots,i_r)}$ such that the contribution of $|\lambda_t|$ for which $\mathscr{A}^{g_m^*}_{N,N'}[\cdot] > \gamma$ for some

$$M^{(i_1,\ldots,i_r)} \leqslant N, N' \leqslant F\left(M^{(i_1,\ldots,i_r)}\right)$$

  is at most $\left(\frac{K-1}{K}\right)^r C^* < \frac{\epsilon}{32}$.
- The contribution of the remaining part is at most $\sum_t |\lambda_t|\gamma < \frac{\epsilon}{16}$.
- Putting together the estimates gives, for all N, N' in the interval,

$$\left\|\mathscr{A}^{g}_{N,N'}[f_1, \ldots, f_{j-1}, f]\right\|_2 < \frac{\epsilon}{4}.$$

□

Finite complexity theorem

- The weak inverse theorem bounds ergodic averages for functions which do not correlate strongly with a reducible function, while the previous theorem shows that the averages for reducible functions are slowly varying.
- We now combine these estimates using the structure decomposition theorem to prove the theorem on finite complexity.

Finite complexity theorem

Proof of finite complexity theorem.

- Fix X, G, F, ϵ, d and g as in the theorem, and assume that all reductions g_m^* of g have complexity at most $d-1$.
- The proof is by induction. We assume the statement for all $d' < d$.
- Let M_0 be the starting point of the sequence in the theorem.
- Let $\delta := \frac{\epsilon}{2^5 3}$ and $\eta(x) := \frac{\epsilon^2}{2^3 3^3 x}$ as previously. This determines the constants $C_1, C_2, ...$ and C^* which appear in the structure decomposition theorem.

$\square$

Finite complexity theorem

Proof of finite complexity theorem.

- Given a positive integer L, write Σ_L for the set of L-reducible functions, and

$$\Sigma_L^+ := \Sigma_L \cup B_2\left(\frac{\delta}{C*}\right).$$

- Define the norm $\|\cdot\|_L = \|\cdot\|_{\Sigma_L^+}$ by

$$\|f\|_{\Sigma_L^+} := \inf\left\{\sum_{j=0}^{k-1} |\lambda_j| : f = \sum_{j=0}^{k-1} \lambda_j \sigma_j, \sigma_j \in \Sigma_L^+\right\}.$$

□

Finite complexity theorem

Proof of finite complexity theorem.

- Define $\psi(M) = F(M^*)$ where M^* is the upper bound on the sequence started from $M = M_*$ in the theorem on structured functions.
- Given $f_1, f_2, ..., f_j \in L^\infty(X)$, $\|f_i\|_\infty \leqslant 1$.
- Since $\Sigma_{L+1}^+ \subset \Sigma_L^+$, $\|\cdot\|_{L+1}^* \leqslant \|\cdot\|_L^*$, perform decomposition of f_j according to $(\|\cdot\|_L)_{L\in\mathbb{N}}$, ψ, δ, η and with $c_1 = c$ the constant from the weak inverse theorem.
- We thus find a constant $1 \leqslant C_i \leqslant C^*$, an M with $M_0 \leqslant M = O(1)$ and

$$f_j = \sum_{t=0}^{k-1} \lambda_t \sigma_t + u + v$$

where $\sum_{t=0}^{k-1} |\lambda_t| \leqslant C_i$, each $\sigma_t \in \Sigma_B^+$ for some $B \geqslant \psi(M)$, $\|u\|_A^* \leqslant \eta(C_i)$ for some $A < c_1 M$ and $\|v\|_2 \leqslant \delta$.

□

Finite complexity theorem

Proof of finite complexity theorem.

- By absorbing any $\sigma_t \in B_2(\delta/C^*)$ into v, so that $\|v\|_2 \leqslant 2\delta$, we may assume that all $\sigma_t \in \Sigma_{\psi(M)}$.
- Applying the bound for structured theorems, we obtain that

$$\left\| \mathscr{A}^g_{N,N'}\left[f_1, ..., f_{j-1}, \sum_{t=0}^{k-1} \lambda_t \sigma_t\right]\right\|_{L^2(X)} < \frac{\epsilon}{3}$$

  for all $M_i \leqslant N, N' \leqslant F(M_i)$, for some index i.
- The contribution of the L^2 error is controlled by using that $\|f_i\|_\infty \leqslant 1$.

□

Finite complexity theorem

Proof of finite complexity theorem.

- To handle u, we first control it's large values. Let S be the set of points where $|v(s)| \leqslant C_i$.
- Note $\mu(S^c) \leqslant \left(\frac{2\delta}{C_i}\right)^2$
- Since $\|\sigma_t\|_{L^\infty(X)} \leqslant 1$, one has $|u\mathbf{1}_{S^c}(x)| \leqslant 3|v(x)|$, so $\|u\mathbf{1}_{S^c}\|_2 \leqslant 3\|v\|_2$
- Similarly, $\|u\mathbf{1}_S\|_\infty \leqslant 3C_i$. Also, for every $\sigma \in \Sigma_A$,

$$\begin{aligned} |\langle u\mathbf{1}_S, \sigma\rangle| &\leqslant |\langle u, \sigma\rangle| + |\langle u1_{S^c}, \sigma 1_{S^c}\rangle| \\ &\leqslant \|u\|_A^* + \|u1_{S^c}\|_2 \|\sigma 1_{S^c}\|_2 \leqslant 2\eta(C_i). \end{aligned}$$

- By the weak inverse theorem, $\left\|\mathscr{A}_{N,N'}[f_1, \ldots, f_{j-1}, u1_S]\right\|_2 \leqslant \frac{\epsilon}{3}$.

□

Polynomial systems

Definition

- Given a G-sequence $\{g(n)\}_{n\in\mathbb{Z}}$ taking values in a nilpotent group G and an integer m, define operator D_m by $(D_m g)(n) := g(n)g(n+m)^{-1}$. Thus $\langle g|h\rangle_m(n) = (D_m g)(n)h(n+m)$.
- A G-sequence g is *polynomial* if there exists some positive integer d such that, for every choice of integers $m_1, ..., m_d$,

$$D_{m_1} D_{m_2} \cdots D_{m_d} g = \mathbf{1}_G.$$

Polynomial systems

Definition

Let $\mathbb{Z}_* = \{0, 1, 2, ...\} \cup \{-\infty\}$. A vector $\overline{d} = (d_1, ..., d_c) \in \mathbb{Z}_*^c$ is *superadditive* if $d_i \leqslant d_j$ for all $i < j$ and $d_i + d_j \leqslant d_{i+j}$ for all i, j with $i + j \leqslant c$.
For $d \in \mathbb{Z}_*$ and $t \in \mathbb{Z}_+$, let

$$d -_* t = \begin{cases} d - t & t \leqslant d \\ -\infty & t > d \end{cases}.$$

If $\overline{d} = (d_1, ..., d_c) \in \mathbb{Z}_*^c$, let $\overline{d} -_* t = (d_1 -_* t, ..., d_c -_* t)$.

In what follows we write just $-$ for $-_*$. Notice that $(\overline{d} - t_1) - t_2 = \overline{d} - (t_1 + t_2)$. Also, subtraction preserves the property of being superadditive.

Polynomial systems

Definition

Let G be nilpotent of class c, and let

$$G = G_{(1)} \supset G_{(2)} \supset \cdots \supset G_{(c)} \supset G_{(c+1)} = \{1_G\}$$

be the lower central series of F, $G_{(i+1)} = [G_{(i)}, G], i = 1, 2, ..., c$. Let $\phi : \mathbb{Z} \to G$ be a polynomial mapping, and let $\overline{d} = (d_1, ..., d_c) \in \mathbb{Z}_*^c$ be a superadditive vector. We say ϕ has *lc-degree* $\leqslant \overline{d}$ if for each $i = 1, ..., c$,

- If $d_i = -\infty$, then $\phi(\mathbb{Z}) \in G_{(i+1)}$
- If $d_i \geqslant 0$ then for any $h_1, ..., h_{d_i+1}$, $D_{h_1} \cdots D_{h_{d_i+1}} \phi(\mathbb{Z}) \subset G_{(i+1)}$.

Notice that if ϕ has lc-degree $\overline{d}$ then $D_h \phi$ has lc-degree $\overline{d} - 1$.

Polynomial systems

Leibman proved the following theorem regarding polynomial sequences.

Theorem (Leibman's theorem on polynomial sequences)

Let $\overline{d} = (d_1, ..., d_s)$ be a superadditive vector, and let $t, t_1, t_2 \geqslant 0$ be non-negative integers. Then we have the following properties:

1. *If g is a polynomial sequence of degree $\leqslant \overline{d} - t$, then $D_m g$ is a polynomial sequence of degree $\leqslant \overline{d} - (t+1)$ for every $m \in \mathbb{Z}$.*
2. *The set of polynomial sequences of degree $\leqslant \overline{d} - t$ forms a group.*
3. *If g is a polynomial sequence of degree $\leqslant \overline{d} - t_1$ and h is a polynomial sequence of degree $\leqslant \overline{d} - t_2$, then $[g, h]$ is a polynomial sequence of degree $\leqslant \overline{d} - (t_1 + t_2)$, where $[g, h](n) := g^{-1}(n)h^{-1}(n)g(n)h(n)$.*

Polynomial systems

Proof of Leibman's theorem on polynomial sequences.

- The first claim is immediate.
- The proof of the remaining claims is a joint downward induction on t and $t_1 + t_2$.
- Note that the second claim is trivial if $t \geqslant d_c$, since in that case, $h \equiv 1_G$. Similarly, the third claim is trivial if $t_1 + t_2 \geqslant 2d_c$.
- Thus we assume both claims hold for $t \geqslant s+1$, $t_1 + t_2 \geqslant s+1$ and prove that they hold for $t = t_1 + t_2 = s$.

□

Polynomial systems

Proof of Leibman's theorem on polynomial sequences.

- We first check the multiplication law.

$$\begin{aligned} D_m(g_1g_2)(n) &= g_1(n)g_2(n)g_2(n+m)^{-1}g_1(n+m)^{-1} \\ &= g_1(n)D_mg_2(n)g_1(n)^{-1}D_mg_1(n) \\ &= D_mg_2(n)[D_mg_2(n), g_1^{-1}(n)]D_mg_1(n). \end{aligned}$$

  This has lc-degree $\leqslant \overline{d} - t - 1$ by applying the inductive assumption.

- To check the inverse property, use induction in

$$\begin{aligned} D_m(g^{-1})(n) &= g^{-1}(n)g(n+m) \\ &= g^{-1}(n)D_{-m}g(n+m)g(n) \\ &= [g(n), D_{-m}g(n+m)^{-1}](D_{-m}g(n+m))^{-1}. \end{aligned}$$

□

Polynomial systems

Proof of Leibman's theorem on polynomial sequences.

- To prove the claim regarding commutators, we use the identity

$$\begin{aligned}[xy, uv] =& [x,u][x,v]\,[v,[u,x]]\,[[x,v]\,[v,[u,x]]\,,[x,u]] \\ & \cdot [[x,v]\,[v,[u,x]]\,[x,u],y]\,[y,v]\,[v,[u,y]]\,[y,u]\end{aligned}$$

  in the expression

$$\begin{aligned}D_m[g_1,g_2](n) &= [g_1(n),g_2(n)][g_1(n+m),g_2(n+m)]^{-1} \\ &= [g_1(n),g_2(n)][D_{-m}g_1(n+m)g_1(n),g_2(n)(D_{-m}g_2(n+m))^{-1}]^{-1}.\end{aligned}$$

  In making the expansion, $[y,u] = [g_1(n),g_2(n)]$, and this cancels the leading term. All remaining commutators are lower degree, so that the claim follows by induction.

□

Polynomial systems

Definition

Let $g = (g_1, g_2, ..., g_j)$ be a polynomial system in a nilpotent group G. A *step* consists of replacing g with an equivalent sytem, then reducing by an integer m. We write the reduction of g as

$$g^* = (g_1, ..., g_{j-1}, \langle g_j | 1_G \rangle, \langle g_j | g_1 \rangle, ..., \langle g_j | g_{j-1} \rangle),$$
$$\langle g | h \rangle(n) = Dg(n)(Dh(n))^{-1} h(n),$$

omitting the dependence on m.

The *complete reduction* of a system g is the system

$$g^{**} = (g_1, ..., g_{j-1}, \langle g_j | g_1 \rangle, ..., \langle g_j | g_{j-1} \rangle).$$

A *complete step* consists of replacing g with an equivalent system, then performing a complete reduction.

Polynomial systems

Walsh proves the following reduction theorem which reduces the main theorem on multiple ergodic averages to his theorem on systems of bounded complexity.

Theorem (Reduction theorem)

Let g be a polynomial system of size $|g| \leqslant C_1$ and degree $\leqslant \overline{d}$ for some superadditive vector $\overline{d} = (d_1, ..., d_s)$. Then

- *One can go from g to the trivial system (1_G) in $O_{C_1,\overline{d}}(1)$ steps.*
- *One can go from g to a system consisting of a single sequence of degree $\leqslant \overline{d}$ in $O_{C_1,\overline{d}}(1)$ complete steps.*

Lemma

Suppose s_1, s_2 *are sequences of degree* $\leqslant \overline{d}$ *and* h_i, h_j *are sequences of degree* $\leqslant \overline{d} - 1$. *Then*

$$\langle s_1 h_1 | s_2 h_2 \rangle = s_2 h$$

where h *has degree* $\leqslant \overline{d} - 1$. *Also,* $\langle s_1 h_1 | s_1 h_2 \rangle = s_1 \langle h_1 | h_2 \rangle$.

Polynomial systems

Proof.

Calculate

$$\begin{aligned}\langle s_1 h_1 | s_2 h_2 \rangle &= D(s_1 h_1) D(s_2 h_2)^{-1} s_2 h_2 \\ &= s_2 D(s_1 h_1) D(s_2 h_2)^{-1} [D(s_1 h_1) D(s_2 h_2)^{-1}, s_2] h_2 \\ &=: s_2 h.\end{aligned}$$

Also,

$$\begin{aligned}\langle s_1 h_1 | s_1 h_2 \rangle_m(n) &= s_1(n) h_1(n) h_1(n+m)^{-1} s_1(n+m)^{-1} s_1(n+m) h_2(n+m) \\ &= s_1(n) \langle h_1 | h_2 \rangle_m(n).\end{aligned}$$

□

Polynomial systems

Proof of reduction theorem.

- Write
$$g = \underline{h}_0 \oplus \bigoplus_{i=1}^{l} s_i \underline{h}_i$$
where each s_i is a polynomial sequence of degree $\leqslant \overline{d}$ and each $\underline{h}_i$ is a polynomial system of degree $\leqslant \overline{d} - 1$, and where $s(h_1, h_2, ..., h_j) = (sh_1, sh_2, ..., sh_j)$.
- We argue that in $O(C_1, \overline{d})$ steps we can produce a system $\tilde{g} = \underline{\tilde{h}}_0 \oplus \bigoplus_{i=1}^{l-1} s_i \underline{\tilde{h}}_i$ with $|\tilde{g}| \leqslant O(C_1, \overline{d})|g|$.
- Notice $\langle s_l h_{l,j_l}, 1_G \rangle$ has degree $\leqslant \overline{d} - 1$. Thus, when a single step is performed, $\underline{h}_0$ is replaced with a system of size $\leqslant 2|\underline{h}_0| + 1$, while $\underline{h}_i$ is replaced by a system of size $\leqslant 2|\underline{h}_i|$ for $i \leqslant l - 1$, and $s_l \underline{h}_l$ is replaced with $s_l \underline{h}_l^{**}$.

□

Polynomial systems

Proof of reduction theorem.

- By the inductive assumption on complete steps, $\underline{h}_l$ may be reduced to (1_G) in $O(C_1, \overline{d})$ steps, and eliminated in the following step.
- We need to prove the corresponding inductive statement for reducing complete steps, but the proof is the same.

□