Recall our main theorem we are interested in.

Theorem (Ergodic theorem)

Let $(X, \mathcal{B}(X), T, \mu)$ be MPT and assume that T is invertible. Given a measurable function $f \in L^p(X)$, $p \geq 1$, polynomial $P \in \mathbb{Z}[n]$ and an integer $N \in \mathbb{N}$ we define the polynomial ergodic averages

$$A_N^P f(x) = \frac{1}{N} \sum_{n=0}^{N-1} f(T^{P(n)} x) \quad , \quad x \in X.$$

Then (i) (Mean ergodic theorem) If $1 < p < \infty$, then the averages $A_N^P f$ converge in $L^p(X)$ norm as $N \to \infty$

(ii) (Pointwise ergodic theorem) If $1 < p < \infty$, then the averages $A_N^P f$ converge pointwise almost everywhere as $N \to \infty$

(iii) (Maximal ergodic theorem) If $1 < p < \infty$, then one has

$$\| \sup_{N \in \mathbb{N}} |A_N^P f| \|_{L^p(X)} \lesssim_p \|f\|_{L^p(X)}.$$

Comments: 1) To prove this theorem we cannot proceed as in the case of Birkhoff's theorem $(P(n) = n)$. The problem is that it is impossible to find a natural dense class since for instance in the case of squares $P(n)=n^2$ the gaps $(n+1)^2-n^2=2n+1 \xrightarrow[n\to\infty]{} \infty$.

2) The above theorem is true in the multidimensional situation:

Theorem (multi-dimensional ergodic theorem)

Let $d,k\in\mathbb{Z}_+$ be fixed and let $\mathcal{P}=(P_1,\dots,P_d):\mathbb{Z}^k\to\mathbb{Z}^d$, be a polynomial mapping where $P_j:\mathbb{Z}^k\to\mathbb{Z}$ are polynomials with integer coefficients and such that $P_j(0)=0$, $j=1,\dots,d$. Further, let $(X,\mathcal{B}(X),\mu)$ be a σ-finite measure space endowed with a family of commuting invertible measure preserving transformations $T_1,\dots,T_d: X\to X$. Then for $N\geq 1$ we consider ergodic means

$$A_N^{\mathcal{P}}f(x)=\frac{1}{N^k}\sum_{m\in[1,N]^k} f\left(T_1^{P_1(m)}\cdots T_d^{P_d(m)}x\right),\ x\in X.$$

even more general $m\in\Omega_N\dots$

Then (i) (Mean ergodic theorem) If $1<p<\infty$, then the averages $A_N^P f$ converge in $L^p(X)$ norm as $N\to\infty$

(ii) (Pointwise ergodic theorem) If $1<p<\infty$, then the averages $A_N^P f$ converge pointwise almost everywhere as $N\to\infty$

(iii) (Maximal ergodic theorem) If $1<p<\infty$, then one has

$$\|\sup_{N\in\mathbb{N}} |A_N^P f|\|_{L^p(X)} \lesssim_p \|f\|_{L^p(X)}.$$

Comments: 1) In this version it was proved in a series of papers by:

- variational estimate : Mirek, Trojan, Stein 2017
- jump estimates : Mirek, Stein, Zorin-Kranich 2020
- oscillation estimates : Mirek, Słomian, TZS.

2) Commutation assumption imposed on $(T_1, \dots, T_d)$ is essential in order to get the ergodic theorem if $\deg P \geq 2$. In 2002 Bergelson and Leibman have shown that this theorem is not true when $(T_1, \dots, T_d)$ generates solvable group. On the other hand, Walsh in 2012 proved that (i) is true if $(T_1, \dots, T_d)$ generates nilpotent group.

However, nothing is known in the case of pointwise ergodic theorem. Only recently it was shown that (ii) is true for nilpotent groups of step 2. [Heisenberg type groups etc.]

3) If $\deg P \geq 2 \Rightarrow$ (ii) is not true for $p=1$. in general [Buczolich, Mauldin (2010), Le Victoire (2011)]
$P(n)=n^2$ $P(n)=n^k$, $k\geq 2$.

4) If $\deg P \geq 2 \Rightarrow$ there is no natural dense class.

Therefore we show stronger result than (iii) which will immediately imply (iii) and (ii).

Notation: • $\mathbb{D}=\{2^n : n\in\mathbb{N}\}$, $\mathbb{N}=\{0,1,\dots\}$, $\mathbb{R}_+=(1,\dots)$

Def. For any $(a_t(x) : t\in\mathbb{I})\subseteq\mathbb{C}$ and $1\leq r<\infty$ we define r-Variation seminorm by

$$V_r(a_t(x) : t\in\mathbb{I}) := \sup_{J\in\mathbb{R}_+}\ \sup_{\substack{n_0<n_1<\dots<n_J\\ n_j\in\mathbb{I}}} \left(\sum_{j=0}^{J-1} |a_{n_{j+1}}(x)-a_{n_j}(x)|^r\right)^{\frac{1}{r}}.$$

Remarks: 1) $V^r(a_t : t \in \mathbb{I})$ defines a seminorm.

[triangle inequality + scaling + no uniqueness]

$a_t \equiv \text{const} \Rightarrow V^r(a_t : t \in \mathbb{I}) = 0.$

2) For every $a < b < c$ we have

$$V^r(a_t : t \in [a,c]) \leq V^r(a_t : t \in [a,b]) + V^r(a_t : t \in [b,c])$$

3) $V^r(a_t : t \in \mathbb{I}) \lesssim \left(\sum_{t \in \mathbb{I}} |a_t|^r \right)^{\frac{1}{r}}.$

4) $V^r(a_t : t \in \mathbb{I}) \searrow$ when $r \nearrow$.

Instead of proving maximal ergodic theorem we focus on variation ergodic theorem: For any $r > 2$ and $1 < p < \infty$

$$\| V^r(A_N^P f : N \in \mathbb{Z}_+) \|_p \lesssim_r \|f\|_p.$$

Lemma Assume that $(a_N(x) : N \in \mathbb{Z}_+) \subseteq \mathbb{C}$. Then for every $1 \leq p \leq \infty$ and $1 \leq r < \infty$ we have

$$\| \sup_{N \in \mathbb{Z}_+} |a_N| \|_{L^p} \leq \sup_{N \in \mathbb{Z}_+} \|a_N\|_{L^p} + \| V^r(a_N : N \in \mathbb{Z}_+) \|_p$$

Proof: Observe that

$$\| \sup_{N \in \mathbb{Z}_+} |e_N| \|_{L^p} = \lim_{M \to \infty} \| \sup_{N \in [1,M]} |e_N| \|_{L^p}$$

$$\leq \|e_1\|_p + \lim_{M \to \infty} \underbrace{\| \sup_{N \in [1,M]} |e_N - e_1| \|_p}_{\leq V^r(e_N : N \in \mathbb{Z}_+)}$$

Lemma If $1 \leq p, r < \infty$, then the estimate

$$\| V^r(e_N : N \in \mathbb{Z}_+) \|_p < \infty$$

implies that e_N converges as $N \to \infty$ for a.e. $x \in X$.

Proof: We know that for a.e. $x \in X$ we have $V^r(e_N(x) : N \in \mathbb{Z}) < \infty$.

Suppose that $\lim_{N \to \infty} e_N(x)$ does not exists.

There is small $\delta > 0$ s.t

$$\lim_{N \to \infty} \sup_{m,n \geq N} |e_m(x) - e_n(x)| > 2\delta$$

This means that there is N_0 s.t. for $N \geq N_0$ we have

$$\sup_{n \geq N} |a_n(x) - a_N(x)| > \delta$$

This allows us to find a sequence $N_0 < N_1 < \dots$ s.t.

$$|a_{N_{k+1}}(x) - a_{N_k}(x)| > \frac{\delta}{2}, \quad k \in \mathbb{N}.$$

This means that $V^r(a_N(x) : N \in \mathbb{Z}_+) = \infty \Rightarrow$ ↯. ☒

Therefore using the above two lemmas we see that in order to prove the whole theorem it suffices to show variation ergodic theorem i.e.:

$$\| V^r(A_N^P f : N \in \mathbb{Z}_+) \|_p \lesssim_{p,r} \|f\|_p.$$

for every $1 < p < \infty$ and $2 < r < \infty$ fixed.

Step 1: Using Calderón transference principle we reduce the above to the special case of MPT on $\mathbb{Z}^d$ and P being canonical polynomials.

Recall that $\mathcal{P}=(P_1,\dots,P_d)$, $P_j:\mathbb{Z}^k\to\mathbb{Z}$ are polynomials with $P_j(0)=0$ and integer coefficients.

__Notation:__ Fix a finite set $\Gamma\subseteq\mathbb{N}^k\setminus\{0\}$. We define the canonical polynomial mapping by

$$\mathbb{R}^k\ni x=(x_1,\dots,x_k)\longmapsto Q(x)=\underbrace{(x^\gamma:\gamma\in\Gamma)}_{(Q_\gamma(x)\,:\,\gamma\in\Gamma)}\in\mathbb{R}^\Gamma.$$

where $x^\gamma=x_1^{\gamma_1}\dots x_k^{\gamma_k}$ for any $\gamma\in\Gamma$

Here $\mathbb{R}^\Gamma\simeq\mathbb{R}^{|\Gamma|}$ denotes the space of tuples of real numbers labeled by multi-indices $\gamma=(\gamma_1,\dots,\gamma_k)\in\Gamma$.

__Example:__ $\Gamma=\{(0,1,2),(1,2,3),(0,0,1),(1,1,0)\}$

$$Q(x)=(x_2x_3^2,\ x_1x_2^2x_3^3,\ x_3,\ x_1x_2)$$

$x=(x_1,x_2,x_3)\in\mathbb{R}^\Gamma\simeq\mathbb{R}^3$.

For this special polynomial mapping we have

$$A_N^Q f(x)=\frac{1}{N^k}\sum_{m\in[1,N]^k}f\Big(\prod_{\gamma\in\Gamma}T_\gamma^{Q_\gamma(m)}x\Big),\ x\in X,$$

where $(T_\gamma)_{\gamma\in\Gamma}$ is a family of commuting MPT.

$T_\gamma: X\to X$

Considering a special case $X = \mathbb{Z}^\rho$ and $T_\rho : \mathbb{Z}^\rho \to \mathbb{Z}^\rho$ being a shift operator

$$T_\rho x = x - e_\rho .$$

where e_ρ is the ρth coordinate vector in $\mathbb{Z}^\rho$ we see that A_N^Q becomes

$$M_N f(x) = \frac{1}{N^k} \sum_{m \in [1,N]^k} f(x - Q(m)), \quad x \in \mathbb{Z}^\rho.$$

M_N are called discrete averaging Radon operators

We now reduce showing

$$\| V^r (A_N^P f : N \in \mathbb{Z}_+) \|_p \lesssim_{p,r} \|f\|_p .$$

for every $1 < p < \infty$ and $2 < r < \infty$ fixed

to proving analogous estimate for $M_N f$ instead of $A_N^P f$.

Reduction: Observe that we have a homomorphism

$$T : \mathbb{Z}^d \to \mathcal{S} := \{ S : X \to X \text{ which are measure preserving and invertible} \}$$

given by $T(n) = T_1^{n_1} T_2^{n_2} \dots T_d^{n_d}$, $n = (n_1, \dots, n_d)$

$[\, T(n+n') = T(n) T(n') \,]$

Further, $T(\mathcal{P}(m)) = T_1^{P_1(m)} T_2^{P_2(m)} \dots T_d^{P_d(m)}$, $m \in \mathbb{Z}^k$.

Next, we show that there is $\Gamma \subseteq \mathbb{N}^k \setminus \{0\}$ and a homomorphism $\Phi : \mathbb{Z}^\Gamma \to \mathbb{Z}^d$ $[\Phi(x+y) = \Phi(x) + \Phi(y)]$ such that

$$\Phi(Q(m)) = \mathcal{P}(m), \quad m \in \mathbb{Z}^k.$$

We can choose Γ which depends only on d, k and $\deg \mathcal{P}$ but does not depend on the coefficients of $\mathcal{P}$.

Indeed, there is $\Gamma \subseteq \mathbb{N}^k \setminus \{0\}$ and coefficients $e_{j,\gamma} \in \mathbb{Z}$ s.t

$$\mathcal{P}_j(m) = \sum_{\gamma \in \Gamma} e_{j,\gamma}\, m^\gamma \qquad \text{for all } j = 1, \dots, d.$$

Then setting

$$\Phi(x) = \sum_{\gamma \in \Gamma} e_{j,\gamma}\, x_\gamma, \quad x = (x_\gamma)_{\gamma \in \Gamma} \in \mathbb{Z}^\Gamma,$$

we see that Φ is linear and

$$\Phi(Q(m)) = \mathcal{P}(m) \quad \text{for all } m \in \mathbb{Z}^k.$$

Since T is a homomorphism and

$$T(\mathcal{P}(m)) = T_1^{\mathcal{P}_1(m)}\, T_2^{\mathcal{P}_2(m)} \cdots T_d^{\mathcal{P}_d(m)}, \quad m \in \mathbb{Z}^k,$$

we put

$$T \circ \Phi(-Q(m)) = T(\varphi(m)) = T_1^{p_1(m)} T_2^{p_2(m)} \cdots T_d^{p_d(m)}, \quad m \in \mathbb{Z}^k.$$

$\tilde{T} = T \circ \Phi : \mathbb{Z}^\Gamma \to \mathcal{A}$ homomorphism.

Let us introduce the dilation structure on $\mathbb{R}^\Gamma$. For $A > 0$ and $x \in \mathbb{R}^\Gamma$ we denote

$$(A \circ x)_\gamma = A^{|\gamma|} x_\gamma$$

Then we have

$$Q(Ax) = A \circ Q(x), \quad x \in \mathbb{R}^\Gamma, \quad A > 0.$$

For $f \in L^p(X)$, $x \in X$ and $M > 0$ we define

$$\varphi_x^M(n) = f(\tilde{T}(n)x) \mathbb{1}_{|M^{-1} \circ n|_\infty \leq 2}.$$

Then for $1 \leq N \leq M$ and $|M^{-1} \circ n|_\infty \leq 1$ we have

$$A_N^P f(\tilde{T}(n)x) = \frac{1}{N^k} \sum_{m \in [1,N]^k} f(\underbrace{T_1^{p_1(m)} \cdots T_d^{p_d(m)}}_{= \tilde{T}(-Q(m))} \tilde{T}(n)x)$$

$$= \frac{1}{N^k} \sum_{m \in [1,N]^k} f(\tilde{T}(n - Q(m))x)$$

$$= \frac{1}{N^k} \sum_{m \in [1,N]^k} \varphi_x^M(n - Q(m))$$

$$= m_N(\varphi_x^m)(n), \qquad x \in X.$$

Consequently, for $x \in X$ and $|m^{-1} \circ n|_\infty \leq 1$ we have

$$V^r(A_N^P f(\tilde{T}(n)x) : 1 \leq N \leq m)$$

$$= V^r(m_N(\varphi_x^m)(n) : 1 \leq N \leq m)$$

Therefore '$\sum_n$' and '$\int_X$' we get

$$\sum_{|m^{-1} \circ n|_\infty \leq 1} \int_X |V^r(A_N^P f(\tilde{T}(n)x) : 1 \leq N \leq m)|^p d\mu(x)$$

$$= \int_X \sum_{|m^{-1} \circ n|_\infty \leq 1} |V^r(m_N(\varphi_x^m)(n) : 1 \leq N \leq m)|^p d\mu(x)$$

$$\lesssim \int_X \sum_{n \in \mathbb{Z}^P} |V^r(m_N(\varphi_x^m)(n) : N \in \mathbb{Z}_+)|^p d\mu(x)$$

$$\lesssim \int_X \| V^r(m_N(\varphi_x^m) : N \in \mathbb{Z}_+) \|_{\ell^p(\mathbb{Z}^P)}^p d\mu(x)$$

Assuming that we have a r-variation estimate for m_N we further get

$$\lesssim \int_X \|\varphi_x^m\|^p_{\ell^p(\mathbb{Z}^r)}\, d\mu(x)$$

$$= \int_X \sum_{n\in\mathbb{Z}^r} |f(\tilde{T}(n)x)|^p \, \mathbb{1}_{|m^{-1}\circ n|_\infty \le 2}\, d\mu(x)$$

$$\simeq \Big(\prod_{r\in\Gamma} M^{|r|}\Big)\, \|f\|^p_{L^p}$$

here measure preserving of $\tilde{T}(n)$ is used.

Thus

$$\int_X \big| V^r(A_N^P f(x) : 1\le N\le M)\big|^p d\mu(x)$$

$$\lesssim \|f\|^p_{L^p}.$$

Letting now $M\to\infty$ we get the desired conclusion. Thus we have reduced our problem to showing that

$$\| V^r(M_N f : N\in\mathbb{Z}_+)\|_{\ell^p(\mathbb{Z}^r)} \lesssim \|f\|_{\ell^p(\mathbb{Z}^r)}.$$

where

$$M_N f(x) = \frac{1}{N^k}\sum_{m\in[1,N]^k} f(x - Q(m)),\quad x\in\mathbb{Z}^\Gamma,\ N\in\mathbb{Z}_+.$$

From now on we can forget about the ergodic theory

The techniques we will be using are the mixture of harmonic / Fourier analysis and analytic number theory.
The advantage of working with M_N instead of A_N^P is that on $\mathbb{Z}^P$ we have Fourier transform techniques and we have 'dilation structure' inherited from $\mathbb{R}^P$.
The problem we need to deal with is that in contrast to $\mathbb{R}^P$ in $\mathbb{Z}^P$ we do not have 'a change of variable technique'. Observe that in one-dimensional situation, when $Q(n) = n^2$ the continuous counterpart of M_N would be

$$\mathcal{M}_N f(x) = \frac{1}{N}\int_0^N f(x-y^2)\,dy \quad , \quad N \in \mathbb{Z}_+$$

Then the estimate

$$\|\sup_{N\in\mathbb{Z}_+} |\mathcal{M}_N f|\,\|_{L^p(\mathbb{R})} \lesssim_p \|f\|_{L^p(\mathbb{R})} \quad , \quad 1<p<\infty$$

is a straightforward consequence of the L^p-boundedness of the Hardy–Littlewood maximal operator.
Indeed, changing the variable $y = \sqrt{z}$ we get

$$\mathcal{M}_N f(x) = \frac{1}{N}\int_0^{N^2} f(x-z)\frac{1}{2\sqrt{z}}\,dz$$

$$= \phi_{r^2} * f(x),$$

where for $t>0$ we denote $\phi_t(x) = \frac{1}{t}\phi\left(\frac{x}{t}\right)$

and
$$\phi(z) = 1_{[0,1]}(z)\frac{1}{2\sqrt{z}}.$$

Then
$$\sup_{r \in \mathbb{R}_+} |\mathcal{M}_r f(x)| \leq \sup_{t>0} |\phi_t * f(x)|$$
$$\lesssim \mathcal{M}_{H-L} f(x)$$

Since $\phi \in L^1$ and 'radially decreasing'

Fact:

If $\phi \in L^1(\mathbb{R}^d)$, ϕ is radial and its profile is a decreasing function, then
$$\sup_{t>0} |\phi_t * f(x)| \leq \left(\int \phi\right) M_{H-L} f(x),$$

where
$$M_{H-L} f(x) = \sup_{R>0} \frac{1}{|B(x,R)|} \int_{B(x,R)} |f(y)| dy.$$

Proof: Use a layer cake formula, i.e.
$$\|f\|_p^p = \int_0^\infty p \lambda^{p-1} \mu(x : |f(x)| > \lambda) \, d\lambda$$

We focus only on $p=2$, i.e. on

$$\| V^r(M_N f : N \in \mathbb{Z}_+) \|_{\ell^2(\mathbb{Z}^n)} \lesssim_r \|f\|_{\ell^2(\mathbb{Z}^n)}$$

for $2<r<\infty$. Other values of $p \neq 2$ would require the introduction of the Ionescu – Wainger projections (modern approach) or tedious techniques of Bourgain from 1980's which strongly use the positivity of the averages M_N.
$[f \geq 0 \Rightarrow M_N f \geq 0]$

Step 2: Reduction to long and short variations

Lemma We have

$$\| V^r(M_N f : N \in \mathbb{Z}_+) \|_{\ell^2(\mathbb{Z}^n)}$$

$$\lesssim \| V^r(M_N f : N \in \mathbb{D}) \|_{\ell^2(\mathbb{Z}^n)}$$

$$+ \Big\| \Big[\sum_{n=0}^{\infty} V^2(M_N f : N \in [2^n, 2^{n+1}))^2 \Big]^{\frac{1}{2}} \Big\|_{\ell^2(\mathbb{Z}^n)}$$

where $\mathbb{D} = \{2^n : n \in \mathbb{N}\}$.