Pointwise ergodic theorems and discrete harmonic analysis

Tomasz Z. Szarek
BCAM - Basque Center for Applied Mathematics

17th Discussion Meeting in Harmonic Analysis, School of Math. Sciences, NISER, India

January 2022

Around 1910 Bohl (1909), Sierpiński (1910) and Weyl (1910) proved independently the following (now famous) result

Theorem (equidistribution)

If α is an irrational number then the sequence $n\alpha$, $n \in \mathbb{N}$ is uniformly distributed mod 1, i.e. for every subinterval $I \subseteq [0,1)$ we have

$$\lim_{N\to\infty} \frac{\#\{1\le n\le N : \{n\alpha\}\in I\}}{N} = |I|, \qquad (1)$$

where $\{x\} = x - \lfloor x \rfloor$ is the fractional part of $x \in \mathbb{R}$
($\lfloor x \rfloor = \max\{n \in \mathbb{Z} : n \le x\}$ is the floor function)

and $|I|$ is the length of I.

Some (historical) comments:

- Then in 1916 Weyl proved that the sequence $n^2\alpha$ is equidistributed (still $\alpha \notin \mathbb{Q}$)
- Later on in 1930's Vinogradov proved that $p_n\alpha$ is equidistributed (here p_n is the nth prime number).
- On the other hand it is not hard to see that for some irrational α the sequence $2^n\alpha$ is not equidistributed mod 1.

Indeed, take for instance $\alpha = 0{,}01001000100001\ldots$ written in the binary number system.

Natural question: What happens if we replace the interval I in (1) by any arbitrary Lebesgue measurable subset of $[0,1]$.

We cannot expect 'word for word' generalization of (1) since the (Lebesgue) measure of any sequence is 0 and so I may be even disjoint from the sequence!!!

In the beginning of 1930's Birkhoff (1931) and Khintchine (1933) proved the appropriate generalization of (1):

Theorem For any fixed Lebesgue measurable $I \subseteq [0,1)$ we have

$$\lim_{N\to\infty} \frac{\#\{1 \le n \le N : \{x + n\alpha\} \in I\}}{N} = |I| \qquad (2)$$

for almost every $x \in [0,1)$.

Remark: • It is not clear at first sight that $(2) \Rightarrow (1)$.

• It took ~ 50 years to obtain the analogues of (2) for the corresponding results by Weyl and Vinogradov. In 1980's Bourgain in a series of papers proved that for any fixed measurable $I \subseteq [0,1)$

$$\lim_{N\to\infty} \frac{\#\{1 \le n \le N : \{x + p(n)\alpha\} \in I\}}{N} = |I| \qquad (3)$$

for almost every $x \in [0,1)$. Here p is a polynomial with integer coefficients.

Bourgain's method is a combination of analytic number theory, Fourier analysis and ergodic theory. During these lectures we give a sketch of the proof of (3).

Before we focus on Bourgain's result we want to recall quite standard condition which allows us to check if the sequence is equidistributed.

Recall first:

Def $(a_n)_{n=0}^{\infty}$ is equidistributed mod 1 (equivalently in $\mathbb{T}=[0,1)$)

if $$\lim_{N\to\infty} \frac{\#\{0\le n\le N-1 : a_n\in I\}}{N} = |I|$$

for every interval $I\subseteq[0,1)$.

Theorem The following statements are equivalent

(a) $(a_n)_{n=0}^{\infty}$ is equidistributed mod 1

(b) For every smooth function on $\mathbb{T}$ we have

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} f(a_n) = \int_{\mathbb{T}} f(x)\,dx \qquad (*)$$

(c) For every $m\in\mathbb{Z}\setminus\{0\}$ we have

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} e^{2\pi i m a_n} = 0.$$

Proof

We first show that (a) $\Leftrightarrow$ (b).

Observe that (a) implies that $(*)$ holds for $f=\mathbb{1}_I$, i.e.

$(*)$ holds for step functions.

Given $f\in C^\infty(\mathbb{T})$ for every $\varepsilon>0$ we can find a step function g s.t. $\|f-g\|_\infty<\frac{\varepsilon}{3}$. Since $(*)$ is true for g we can find N_0

s.t for $N \geq N_0$ we have

$$\left| \frac{1}{N} \sum_{n=0}^{N-1} g(a_n) - \int_{\mathbb{T}} g(x)dx \right| < \frac{\varepsilon}{3}.$$

Since

$$\left| \frac{1}{N} \sum_{n=0}^{N-1} g(a_n) - \frac{1}{N} \sum_{n=0}^{N-1} f(a_n) \right| \leq \|g-f\|_\infty < \frac{\varepsilon}{3}$$

and

$$\left| \int_{\mathbb{T}} g(x)dx - \int_{\mathbb{T}} f(x)dx \right| \leq \|g-f\|_\infty < \frac{\varepsilon}{3}$$

we obtain for $N \geq N_0$

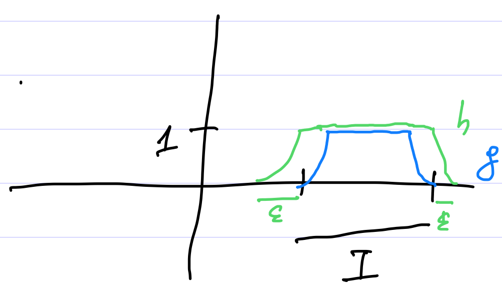
$$\left| \frac{1}{N} \sum_{n=0}^{N-1} f(a_n) - \int_{\mathbb{T}} f(x)dx \right| < \varepsilon$$

and (b) is proved.

Now we prove that (b) $\Rightarrow$ (a). We fix an interval $I \subseteq \mathbb{T}$.

We can find two smooth functions g and h s.t.

$$\mathbb{1}_{(1-\varepsilon)I} \leq g \leq \mathbb{1}_I \leq h \leq \mathbb{1}_{(1+\varepsilon)I}.$$



Thanks to that we have

$$(1-\varepsilon)|I| \leq \int g(x)dx \leq |I| \leq \int h(x)dx \leq (1+\varepsilon)|I|$$

and

$$\frac{1}{N} \sum_{n=0}^{N-1} g(a_n) \leq \frac{1}{N} \sum_{n=0}^{N-1} \mathbb{1}_I(a_n) \leq \frac{1}{N} \sum_{n=0}^{N-1} h(a_n)$$

Therefore by the sandwich theorem we get

$$(1-\varepsilon)|I| \leq \liminf_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mathbb{1}_I(a_n) \leq \limsup_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mathbb{1}_I(a_n)$$
$$\leq (1+\varepsilon)|I|$$

Since $\varepsilon > 0$ is arbitrary we get (a) and consequently (a) $\Leftrightarrow$ (b).

Further observe that the implication (b) => (c) is trivial so it suffices to show that (c) => (b).

Using the inverse Fourier transform theorem we may write for any $f \in C^\infty(\mathbb{T})$

$$\frac{1}{N}\sum_{n=0}^{N-1} f(a_n) = \frac{1}{N}\sum_{n=0}^{N-1}\left(\sum_{m\in\mathbb{Z}} \hat{f}(m)\, e^{2\pi i m a_n}\right) \qquad \left\{\hat{f}(m) = \int_{\mathbb{T}} e^{-2\pi i m s} f(s)\,ds\right\}$$

$$= \hat{f}(0) + \sum_{m\in\mathbb{Z}\setminus\{0\}} \hat{f}(m) \underbrace{\left[\frac{1}{N}\sum_{n=0}^{N-1} e^{2\pi i m a_n}\right]}_{\xrightarrow{N\to\infty} 0 \text{ by (c)}}$$

Thanks to the rapid decay of $\hat{f}(m)$ as $|m| \to \infty$ we see that

$$\frac{1}{N}\sum_{n=0}^{N-1} f(a_n) \xrightarrow[N\to\infty]{} \hat{f}(0) = \int_{\mathbb{T}} f(x)\,dx$$

and (c) is proved.

$\blacksquare$

Remark: Using the above one can easily check that the sequence $\{n\alpha\}_{n\in\mathbb{N}}$ is equidistributed mod 1 iff $\alpha \notin \mathbb{Q}$.

We now come back to the question of existence of the limit (3), i.e. for any characteristic function of the Lebesgue measurable set $I \subseteq [0,1)$ we ask if the limit

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} f(\{x + p(n)\alpha\}) \qquad (4)$$

exists for almost every $x \in [0,1)$. Here p is a polynomial with integer coefficients.

In fact we are interested in a more general situation when the function $f \in L^p[0,1)$ for some $p \in [1, \infty]$.
We post this question in a more abstract situation

Ergodic theory

Def. A measure preserving transformation (MPT) is a quartet $(X, \mathcal{B}(X), \mu, T)$, where
1) $(X, \mathcal{B}(X), \mu)$ is a measure space and μ is σ-finite.
2) $T: X \to X$ is measurable, i.e. $E \in \mathcal{B}(X) \Rightarrow T^{-1}(E) \in \mathcal{B}(X)$
3) μ is T invariant, i.e. $\mu(T^{-1}(E)) = \mu(E)$ for all $E \in \mathcal{B}(X)$

Further, if μ is probability measure, then we call the above quartet a probability preserving transformation (PPT).

Let $(X, \mathcal{B}(X), \mu)$ be a MPT. We are interested in the following question: Does the limit

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^{P(n)} x) \qquad (5)$$

exist in the pointwise (a.e) sense (and also in norm) when $f \in L^p(X)$, $1 \le p \le \infty$.

Def Suppose $(X, \mathcal{B}(X), \mu, T)$ is MPT. A measurable set $E \in \mathcal{B}(X)$ is called invariant if $T^{-1}(E) = E$. (or T invariant)
A MPT $(X, \mathcal{B}(X), \mu, T)$ is ergodic if every invariant set E is trivial, i.e. $\mu(E) = 0$ or $\mu(X \setminus E) = 0$

Examples of MPT/PPT

1) Shift on $\mathbb{Z}$

The most important example from our point of view is
$(\mathbb{Z}, P(\mathbb{Z}), \#_{\mathbb{Z}}, S)$,
where $\#_{\mathbb{Z}}$ - counting measure on $\mathbb{Z}$
$S: \mathbb{Z} \to \mathbb{Z}$ is given by $Sx = x+1$ (shift operator)
$(\mathbb{Z}, P(\mathbb{Z}), \#_{\mathbb{Z}}, S)$ is ergodic

2) Circle rotations:

$(\mathbb{T}, \mathcal{B}(\mathbb{T}), dx, T_\alpha)$, where for any $\alpha \in \mathbb{R}$ we define
$T_\alpha : \mathbb{T} \to \mathbb{T}$ by $T_\alpha x = x + \alpha \pmod 1$
$(\mathbb{T}, \mathcal{B}(\mathbb{T}), dx, T_\alpha)$ is ergodic iff $\alpha \notin \mathbb{Q}$

Comment: Observe that taking $T = T_\alpha$ in (5) gives the answer to (4)!

3) The angle doubling map

$(\mathbb{T}, \mathcal{B}(\mathbb{T}), dx, T)$, where

$$T: \mathbb{T} \to \mathbb{T}, \quad T(x) = 2x \pmod 1.$$

$(\mathbb{T}, \mathcal{B}(\mathbb{T}), dx, T)$ is ergodic.

4) Bernoulli Schemes

Let S be a finite set (alphabet) and let $X = S^{\mathbb{N}}$.

We introduce a metric

$$d((x_n)_n, (y_n)_n) = 2^{-\min\{k:\, x_k \neq y_k\}}.$$

This generates the product topology on X generated by cylinders

$$[e_0, \dots, e_{n-1}] := \{x \in X : x_i = e_i \text{ for } 0 \leq i \leq n-1\}$$

Let $T: X \to X$ be given by

$$T(x_0, x_1, x_2, \dots) = (x_1, x_2, x_3, \dots)$$

To introduce μ on X we fix a vector $(p_s)_{s \in S}$ s.t $0 \leq p_s \leq 1$ and $\sum_{s \in S} p_s = 1$. Then the Bernoulli measure μ_p corresponding to $(p_s)_{s \in S}$ is the unique measure on the Borel σ-algebra on X s.t

$$\mu_p[e_0, \ldots, e_{n-1}] = p_{e_0} p_{e_1} \cdots p_{e_{n-1}}$$

for any cylinder $[e_0, \ldots, e_{n-1}]$.

The existence of μ_p is guaranteed by the famous Kolmogorov theorem.

Every Bernoulli scheme is ergodic.

There is a difference between PPT and MPT, which can easily be seen at the level of Poincare recurrence theorem.

Def. Let $(X, \mathcal{B}(X), \mu, T)$ be a MPT. Given a set $A \in \mathcal{B}(X)$ and $x \in A$ we say that x returns to A if $T^n x \in A$ for some $n > 0$ $[\iff x \in A \cap T^{-n}[A]]$

Theorem (Poincare recurrence theorem)

Let $(X, \mathcal{B}(X), \mu, T)$ be a PPT. If $\mu(A) > 0$, then μ a.e. $x \in A$ return to A.

Comment: Observe that this is not true for MPT.

Take shift on $\mathbb{Z}$! $Tx = x + 1$.

Proof Step 1: We first show that if $\mu(A)>0$, then there is $n>0$ s.t. $\mu(A\cap T^{-n}[A])>0$.

Indeed, consider the sequence

$$(**)\qquad T^{-1}[A],\ T^{-2}[A],\ T^{-3}[A],\ldots T^{-k}[A]$$

Since T is measure preserving for every i we have

$$\mu(T^{-i}[A])=\mu(A).$$

Therefore if $k>\frac{1}{\mu(A)}$, then the sets $(**)$ cannot be pairwise disjoint mod μ. [If they were we would have

$$1\geq\sum_{i=1}^{k}\mu(T^{-i}[A])=k\mu(A)>1\implies ↯\ ]$$

Therefore there are $0<i<j\leq k$ such that

$$\mu(T^{-i}[A]\cap T^{-j}[A])>0.$$

Since

$$T^{-i}[A]\cap T^{-j}[A]=T^{-i}\left[A\cap T^{-(j-i)}[A]\right]$$

and T is measure preserving we get the desired property with $n=j-i>0$.

Step 2: Let us consider the set

$$E = \{x \in A : \forall n > 0 \quad x \notin T^{-n}[A]\}$$
$$= A \setminus \bigcup_{n=1}^{\infty} T^{-n}[A]$$

Observe that $E \subseteq A$ and $T^{-n}[E] \cap E \subseteq T^{-n}[A] \cap E = \emptyset$ for any $n \geq 1$. By Step 1 it means that $\mu(E) = 0$

$\boxtimes$

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

$$A_N^P f(x) = \frac{1}{N} \sum_{n=0}^{N-1} f(T^{P(n)} x), \quad x \in X. \quad (6)$$

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

Further if we also assume that μ is a probability measure and that T is totally ergodic, then the limit in (i) and (ii) is equal to $\int_X f(x)\,d\mu(x)$.

Def: T is totally ergodic iff $\forall n \in \mathbb{N}\setminus\{0\}$ T^n is ergodic.

Remarks: 1) The above theorem was proved by Bourgain in the series of papers in 1980's. During these lectures we will focus on (the most interesting case) $p=2$.

2) Notice that (i) is a simple consequence of (ii) and (iii) and the dominated convergence theorem.

3) As we will see in a moment the standard procedure in harmonic analysis to problems of (a.e) pointwise convergence is based on two step procedure:

(a) First we show the corresponding maximal estimate, see (iii) which implies that the class of functions for which almost everywhere convergence holds is closed in $L^p(X)$.

(6) The second step is based on finding the dense class of functions in L^p for which the convergence holds.

4) Finding a dense class of functions for which (6) converges in the abstract setup of MPT seems to be impossible if $\deg P \geq 2$, therefore Bourgain introduce several other tools like oscillation, variation seminorms as well as jumps functions, which dominate the maximal function and which immediately show pointwise convergence of the underlying sequence without finding a dense class.

5) pointwise convergence of ergodic means was a question asked by Bellow and Furstenberg independently in the early 1980's. Furstenberg was motivated by some results about finding patterns.

We now show the convergence in mean for $P(n) = n$, which was proved in 1930's by von Neumann.

Theorem (von Neumann's Mean Ergodic Theorem)

Let $(X, \mathcal{B}(X), T, \mu)$ be MPT. Then for any $f \in L^2(X)$ the means

$$A_N f(x) = \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x)$$

converge in $L^2(X)$. If T is ergodic, then $A_N f \to \int f d\mu$ in $L^2(X)$.

and μ is a probability measure

<u>Proof</u> Let $\mathcal{E} = \{ g - g \circ T : g \in L^2(X) \}$. Then by the telescoping we have

$$A_N (g - g\circ T)(x) = \frac{1}{N} \left[g(x) - g(T^N x) \right]$$

Therefore

$$\| A_N (g - g \circ T) \|_2 \leq \frac{2 \|g\|_2}{N} \xrightarrow[N \to \infty]{} 0.$$

The last estimate is true because T is measure preserving and consequently for any $f \in L^2(X)$ and $n \in \mathbb{N}$ we have

$$\int |f(T^n x)|^2 d\mu(x) = \int |f(x)|^2 d\mu(x) \qquad (*)$$

Therefore for $f \in \mathcal{E}$ we have convergence. Further using $(*)$ we see that A_N are contractions, i.e

$\| A_N f \|_2 \leq \|f\|_2$, $N \in \mathbb{N}$.

This implies that $A_N f \xrightarrow[N \to \infty]{} 0$ in $L^2(X)$ for $f \in \overline{\mathcal{E}}$.

Further, $\overline{\mathcal{E}}^{\perp} = \{ f \in L^2 : f \circ T = f \} \leftarrow T$ invariant functions. $(**)$

Indeed, $f \in \overline{\mathcal{E}}^{\perp} \iff f \perp g - g\circ T \quad \forall g \in L^2 \iff$

$$\iff \langle f, g - g\circ T \rangle = 0 \quad \forall g \in L^2$$

$$\iff \langle f, g \rangle = \langle f, g \circ T \rangle = \langle T^* f, g \rangle \quad \forall g \in L^2.$$

Here T^* is the dual operator to $Tf = f \circ T$.

The above shows that
$$f = T^* f.$$

It suffices to show the following claim in general Hilbert spaces.

Claim If H is a Hilbert space and $T : H \to H$ is linear and contraction then for every $x \in H$ we have $Tx = x$ iff $T^* x = x$.

Observe that the claim above gives the desired conclusion. Indeed, since T is a contraction we see that $f = Tf = f \circ T$. Consequently $(**)$ holds true. Therefore we have

$$L^2(X) = \overline{\mathcal{E}} \oplus \underbrace{\{ f \in L^2 : f \circ T = f \}}_{\mathcal{I}}.$$

Since for $f \in \mathcal{I}$ we have $A_n f = f$, the proof is finished.

Observe that if T is ergodic, then $\mathcal{I} = \{ C \mathbb{1}_X : C \in \mathbb{C} \}$.
Consequently
$$A_N f \to \int_X f \, d\mu.$$

$\boxtimes$

Finally, we prove the claim.

Proof of the Claim:

For symmetry reasons $[(T^*)^* = T]$ it suffices to show that
$$T^* x = x \implies Tx = x.$$

Observe that
$$\|x - Tx\|^2 = \langle x - Tx, x - Tx \rangle$$
$$= \|x\|^2 + \|Tx\|^2 - \underbrace{\langle Tx, x \rangle}_{\substack{\| \\ \langle x, T^* x \rangle = \langle x, x \rangle \\ = \|x\|^2}} - \langle x, Tx \rangle$$

$$= \|Tx\|^2 - \|x\|^2 \leq 0$$

contraction.

This shows that $Tx = x$.

$\boxtimes$

Now we show pointwise convergence for $P(n) = n$. In the proof we will need the following result.

For $N \in \mathbb{N}$ and $f \in \ell^1(\mathbb{Z})$ let

$$M_N f(m) = \frac{1}{N} \sum_{n=1}^{N} f(n+m).$$

This is ergodic mean for the shift operator and the (discrete) Hardy–Littlewood maximal operator is given by

$$M_* f(m) = \sup_{N \in \mathbb{N}} |M_N f(m)|, \quad m \in \mathbb{Z}.$$

It is well known that M_* is bounded on $\ell^p(\mathbb{Z})$ for $p > 1$ and of weak type $(1,1)$.

Weak type $(1,1)$ $\Leftrightarrow$ $\sup_{\lambda > 0} \lambda \, \#\{m \in \mathbb{Z} : M_* f(m) > \lambda\} \lesssim \|f\|_{\ell^1(\mathbb{Z})}$,

for every $f \in \ell^1(\mathbb{Z})$.

This goes from the Vitali type argument as in the continuous setup.

Theorem (Birkhoff Pointwise Ergodic Theorem)

Let $(X, \mathcal{B}(X), T, \mu)$ be MPT. Then for any $f \in L^p(X)$, $1 \leq p < \infty$, the means
$$A_N f(x) = \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x)$$
converge pointwise almost everywhere. Further, if T is ergodic and μ is a probability measure then $A_N f \to \int f d\mu$ a.e.

Proof We first show the result for the dense class in $L^2(X)$. Similarly as in the mean ergodic theorem, let
$$\mathcal{C} = \{g - g \circ T : g \in L^2(X) \cap L^\infty(X)\}.$$
Then for $f = g - g \circ T$ we have
$$A_N f(x) = \frac{g(x) - g(T^N x)}{N} \xrightarrow[N \to \infty]{} 0$$
since $\|g\|_\infty < \infty$.

On the other hand for $f \in \mathcal{I} = \{f \in L^2 : f = f \circ T\}$ we have
$$A_N f(x) = f(x)$$
and the convergence is trivial. This shows that we have the convergence for $f \in \mathcal{C} \oplus \mathcal{I}$ which is dense in $L^2(X)$.

Now we will use the Calderón transference principle to show that the maximal operator

$$A_* f(x) = \sup_n |A_n f(x)|$$

is bounded on $L^p(X)$ for $p > 1$ and is of weak type $(1,1)$. $\Big\}$ $(\square)$

Assume for a moment that this holds. Then by the well known argument we see that the convergence holds for all $f \in L^2(X)$. Since $L^p(X) \cap L^2(X)$ is dense in $L^p(X)$ for every $1 \le p < \infty$ we get the result for all $1 \le p < \infty$.

Finally we justify $(\square)$. To do that we will use the so-called Calderón transference principle developed in 1960's which will reduce $(\square)$ to the special case of the discrete Hardy-Littlewood max. operator. M_*.

Let us consider $p > 1$. Let us fix $M \in \mathbb{N}$ $f \in L^p(X)$, $x \in X$ and define $\phi_x^M : \mathbb{Z} \to \mathbb{C}$ by

$$\phi_x^M(n) = f(T^n x) \mathbb{1}_{1 \le n \le 2M}, \quad n \in \mathbb{Z}$$

Then for $N \le M$ and $1 \le m \le M$ we have

$$A_N f(T^m x) = \frac{1}{N}\sum_{n=0}^{N-1} \underbrace{f(T^{n+m} x)}_{\phi_x^M(n+m)}$$

$$= M_N(\phi_x^M)(m)$$

Therefore for $1 \leq m \leq N$, $N \leq M$ we have

$$M \int \sup_{1 \leq N \leq m} |A_N f(x)|^p \, d\mu(x)$$

$$= \sum_{m=1}^{M} \int \sup_{1 \leq N \leq m} |A_N f(T^m x)|^p \, d\mu(x)$$

$$= \sum_{m=1}^{M} \int \sup_{1 \leq N \leq m} |M_N(\phi_x^M)(m)|^p \, d\mu(x)$$

$$\leq \int \sum_{m \in \mathbb{Z}} |M_*(\phi_x^M)(m)|^p \, d\mu(x)$$

$$\lesssim \sum_{m \in \mathbb{Z}} \int |\phi_x^M(m)|^p \, d\mu(x)$$

$$= 2M \int |f|^p \, d\mu = 2M \|f\|_p^p .$$

Dividing by M and taking $M \to \infty$ we see that

$$\|A_* f\|_p^p \lesssim \|f\|_p^p$$ and ($\square$) is proved.

Weak type $(1,1)$ in ($\square$) can be proved in a similar way $\square$