Cesàro convergence of spherical averages for measure-preserving actions of Markov semigroups and groups

Alexander I. Bufetov, Mikhail Khristoforov, Alexey Klimenko

Cesàro convergence of spherical averages for measure-preserving actions of Markov semigroups and groups

Alexander Bufetov*, Mikhail Khristoforov†, Alexey Klimenko‡

Abstract

Cesàro convergence of spherical averages is proven for measure-preserving actions of Markov semigroups and groups. Convergence in the mean is established for functions in L^p, $1 \leq p < \infty$, and pointwise convergence for functions in L^∞. In particular, for measure-preserving actions of word hyperbolic groups (in the sense of Gromov) we obtain Cesàro convergence of spherical averages with respect to any symmetric set of generators.

1 Introduction

1.1 Formulation of the main results

Let Γ be a finitely generated semigroup. Choice of a finite set of generators O endows Γ with a norm $|\cdot|_O$: for $g \in \Gamma$ the number $|g|_O$ is the length of the shortest word over the alphabet O representing g. Denote $S_O(n) = \{g : |g|_O = n\}$.

Assume that the semigroup Γ acts on a probability space (X, ν) by measure-preserving transformations, and for $g \in \Gamma$ let T_g be the corresponding map. Now take $\varphi \in L^1(X, \nu)$ and consider the sequence of its *spherical averages*

$$s_n(\varphi) = \frac{1}{\#S_O(n)} \sum_{g \in S_O(n)} \varphi \circ T_g$$

*Steklov Mathematical Institute, Moscow, Russia, and Rice University, Houston, Texas, USA.

†Chebyshev Laboratory, Department of Mathematics and Mechanics, Saint-Petersburg State University, Saint-Petersburg, Russia.

‡Steklov Mathematical Institute, Moscow, Russia.

(here and everywhere # stands for the cardinality of a finite set; if $S_O(n) = \varnothing$, then we set $s_n(\varphi) = 0$). Next, consider the Cesàro averages of the spherical averages:

$$c_N(\varphi) = \frac{1}{N} \sum_{n=0}^{N-1} s_n(\varphi).$$

The main result of this paper establishes mean convergence of the averages $c_N(\varphi)$ for $\varphi \in L^1(X, \nu)$ and pointwise convergence of $c_N(\varphi)$ for $\varphi \in L^\infty(X, \nu)$ in the case when Γ is a *Markov semigroup* with respect to the generating set O.

Recall the definition of Markov semigroups. As before, let Γ be a semigroup with a finite generating set O. For a finite directed graph $\mathbf{G}$ with the set of arcs $\mathcal{E}(\mathbf{G})$, a labelling on $\mathbf{G}$ is a map $\xi\colon \mathcal{E}(\mathbf{G}) \to O$. Let v_0 be a vertex of $\mathbf{G}$ and let $\mathcal{P}(\mathbf{G}, v_0)$ be the set of all finite paths in $\mathbf{G}$ starting at v_0. To each path $p = e_1 \dots e_n \in \mathcal{P}(\mathbf{G}, v_0)$ we assign an element $\xi(p) \in \Gamma$ by the formula

$$\xi(p) = \xi(e_1) \dots \xi(e_n).$$

The semigroup Γ is called *Markov* with respect to a finite generating set O if there exists a finite directed graph $\mathbf{G}$, a vertex v_0 of $\mathbf{G}$, and a labelling $\xi\colon \mathcal{E}(\mathbf{G}) \to O$ such that the lifted map $\xi\colon \mathcal{P}(\mathbf{G}, v_0) \to \Gamma$ is a bijection, and, furthermore, for a path $p \in \mathcal{P}(\mathbf{G}, v_0)$ of length n we have $|\xi(p)|_O = n$.

For example, a theorem by Gromov [15] states that a word hyperbolic group is Markov with respect to any symmetric set of generators (for cocompact groups of isometries of Lobachevsky spaces, the Markov property had been established earlier by Cannon [9]; a detailed exposition of the proof of Gromov's theorem can be found in the book of Ghys and de la Harpe [11]).

We are now ready to formulate the main result of the paper.

Theorem 1. *Let Γ be a Markov semigroup with respect to a finite generating set O. Assume that Γ acts by measure-preserving transformations on a probability space (X, ν). Then for any p, $1 \le p < \infty$, and any $\varphi \in L^p(X, \nu)$ the sequence of Cesàro averages of its spherical averages*

$$c_N(\varphi) = \frac{1}{N} \sum_{n=0}^{N-1} \frac{1}{\#S_O(n)} \sum_{g \in S_O(n)} \varphi \circ T_g$$

converges in $L^p(X, \nu)$ as $N \to \infty$. If, additionally, $\varphi \in L^\infty(X, \nu)$, then the sequence $c_N(\varphi)$ converges ν-almost everywhere as $N \to \infty$.

Corollary 1. *Let Γ be an infinite word hyperbolic group (in the sense of Gromov), and let O be a finite symmetric generating set for Γ. Assume that Γ acts by measure-preserving transformations on a probability space (X, ν). Then for any p, $1 \leq p < \infty$, and any $\varphi \in L^p(X, \nu)$ the sequence of Cesàro averages of its spherical averages*

$$c_N(\varphi) = \frac{1}{N} \sum_{n=0}^{N-1} \frac{1}{\#S_O(n)} \sum_{g \in S_O(n)} \varphi \circ T_g$$

converges in $L^p(X, \nu)$ as $N \to \infty$. If, additionally, $\varphi \in L^\infty(X, \nu)$, then the sequence $c_N(\varphi)$ converges ν-almost everywhere as $N \to \infty$.

Under additional assumption of exponential mixing of the action, pointwise Cesàro convergence for spherical averages of functions from L^2 for measure-preserving actions of word hyperbolic groups was obtained by Fujiwara and Nevo [10]. L. Bowen [2] proved convergence of spherical averages for actions of word hyperbolic groups on finite spaces. Both Fujiwara and Nevo [10] and L. Bowen [2] also proved that in their setting the limit is invariant under the action.

Our result applies to all measure-preserving actions of all finitely-generated infinite word hyperbolic groups. Our argument, however, does not give any information about the limit.

Question. In Theorem 1, when is it true that the limit is Γ-invariant?

We conjecture that it always is in Corollary 1.

1.2 History

First ergodic theorems for measure-preserving actions of arbitrary countable groups were obtained by Oseledets in 1965 [21]. Oseledets endows a countable group Γ with a probability distribution μ satisfying $\mu(g) = \mu(g^{-1})$, $g \in \Gamma$, and establishes pointwise convergence of the sequence of operators

$$S_{2n}^{(\mu)} = \sum_{g \in \Gamma} \mu^{(2n)}(g) T_g$$

as $n \to \infty$ (here $\mu^{(k)}$ stands for the k-th convolution of the measure μ). To prove pointwise convergence Oseledets uses the martingale theorem in the space of trajectories of the Markov chain corresponding to the self-adjoint Markov operator $S_1^{(\mu)}$; the argument of Oseledets is thus a precursor, in the self-adjoint case, of Rota's "Alternierende Verfahren" argument [22].

For uniform spherical averages corresponding to measure-preserving actions of free groups convergence in the mean was established by Y. Guivarc'h [16], who used earlier work of Arnold and Krylov [1] on equidistribution of two rotations of the sphere.

In 1986, R.I. Grigorchuk [12] (see also [13], [14]) obtained pointwise convergence of Cesàro averages of uniform spherical averages of L^1-functions for measure-preserving actions of free groups. The limit is invariant under the action of the group.

For functions in L^2, pointwise convergence of uniform spherical averages themselves was established in 1994 by Nevo [17], and for functions in L^p, $p > 1$, by Nevo and Stein [19]. The limit was proven to be invariant under the subgroup of elements of even length. Whether convergence of uniform spherical averages holds for functions in L^1 remains an open problem (recall that, as Ornstein showed [20], powers of a self-adjoint Markov operator applied to a function in L^1 need not converge almost surely).

In [7], pointwise convergence of uniform spherical averages is obtained by applying Rota's "Alternierende Verfahren" Theorem to a special Markov operator assigned to the action. This approach also yields pointwise convergence of *non-uniform* spherical averages corresponding to Markovian weights satisfying a symmetry condition [7].

Convergence of Cesàro averages on non-uniform spherical averages for actions of free groups and free semigroups holds for general Markovian (and, in fact, for general stationary) weights [4], [5], [6]. The motivation behind considering such Markovian weights is precisely to establish ergodic theorems for actions of Markov groups, in particular, of word hyperbolic groups.

The results of [6], however, can only be applied to groups that are coded by admissible words in an *irreducible* Markov chain; in fact, to prove invariance of the limit function, even a stronger condition is needed, which is called *strict irreducibility* in [6] and is equivalent to the triviality of the symmetric σ-algebra of the corresponding Markov chain with finitely many states.

For some groups, a Markov coding is known explicitly: for instance, for Fuchsian groups such a coding has been constructed by Series [23]. The Series coding does in fact have the strict irreducibility property, and pointwise convergence of Cesàro averages of uniform spherical averages for measure-preserving actions of Fuchsian groups and for functions in L^1 is established in [8], extending the earlier theorem of Fujiwara and Nevo [10] for functions in L^2.

For general word hyperbolic groups, however, it is not clear whether the Markov coding is irreducible. The main result of this paper is that convergence of Cesàro averages of spherical averages still holds without the irreducibility assumption.

1.3 Acknowledgements

We are deeply grateful to R.I. Grigorchuk, V.A. Kaimanovich, A. Nevo, and A.M. Vershik for useful discussions.

A. B. is an Alfred P. Sloan Research Fellow. He is supported in part by grant MK-4893.2010.1 of the President of the Russian Federation, by the Programme on Mathematical Control Theory of the Presidium of the Russian Academy of Sciences, by the Programme 2.1.1/5328 of the Russian Ministry of Education and Research, by the Edgar Odell Lovett Fund at Rice University, by the NSF under grant DMS 0604386, and by the RFBR-CNRS grant 10-01-93115.

M. K. is supported in part by the Chebyshev Laboratory (Department of Mathematics and Mechanics, Saint-Petersburg State University) under the grant 11.G34.31.2006 of the Government of the Russian Federation.

A. K. is partially supported by RFBR grant 08-01-00342a, by grants NSh-8508.2010.1 and MK-4893.2010.1 of the President of the Russian Federation, and by the Programme on Mathematical Control Theory of the Presidium of the Russian Academy of Sciences.

2 Paths and operators

Let us introduce some notation regarding a directed graph from the definition of Markov groups. Consider a finite directed graph $\mathbf{G}$ (loops and multiple edges are permitted). The sets of vertices and edges (arcs) of $\mathbf{G}$ are denoted as $\mathcal{V}(\mathbf{G})$ and $\mathcal{E}(\mathbf{G})$ respectively. For an edge e, $I(e)$ and $F(e)$ are its initial (tail) and terminal (head) vertices. Denote

$$\mathcal{E}(\mathbf{G}, u, v) = \{e \in \mathcal{E}(\mathbf{G}) \mid I(e) = u, F(e) = v\}.$$

Then, let $\mathcal{P}(\mathbf{G})$ be the set of finite paths in $\mathbf{G}$, that is,

$$\mathcal{P}(\mathbf{G}) = \{l = e_1 e_2 \dots e_k \mid I(e_j) = F(e_{j-1})\}.$$

Denote by $|l|$ the length of a path l.

Let (X, ν) be a probability space. Assume that to every arc $e \in \mathcal{E}(\mathbf{G})$ a measure-preserving transformation T_e of (X, ν) is assigned. In this case we say that $\mathbf{G}$ is *labelled* by measure-preserving transformations of (X, ν).

The map $e \mapsto T_e$ is naturally extended onto $\mathcal{P}(\mathbf{G})$ by formula

$$T_{e_1 \dots e_k} = T_{e_1} \dots T_{e_k}.$$

The action of T_l, $l \in \mathcal{P}(\mathbf{G})$, induces a standard action on the space $L^p(X, \nu)$: $T_l(\varphi) = \varphi \circ T_l$. For any finite subset $L \subset \mathcal{P}(\mathbf{G})$ introduce an operator $s(L)$

on $L^p(X,\nu)$ acting by the formula

$$s(L) = \frac{1}{\#L}\sum_{l\in L} T_l$$

if $L \neq \varnothing$; we set $s(\varnothing) = 0$.

In particular, denote

$$L^{\mathbf{G}}_{u,v,n} = \{l \in \mathcal{P}(\mathbf{G}) \mid I(l) = u, F(l) = v, |l| = n\},$$

that is, $L^{\mathbf{G}}_{u,v,n}$ is the set of all paths from u to v of length n. Define $s^{\mathbf{G}}_{u,v,n} = s(L^{\mathbf{G}}_{u,v,n})$ and let $c^{\mathbf{G}}_{u,v,N}$ be their Cesàro averages:

$$c^{\mathbf{G}}_{u,v,N} = \frac{1}{N}\sum_{n=0}^{N-1} s^{\mathbf{G}}_{u,v,n}.$$

Analogously, denote $L^{\mathbf{G}}_{u,*,n} = \bigcup_{v\in\mathcal{V}(\mathbf{G})} L^{\mathbf{G}}_{u,v,n}$ and define

$$s^{\mathbf{G}}_{u,*,n} = s(L^{\mathbf{G}}_{u,*,v}), \quad c^{\mathbf{G}}_{u,*,N} = \frac{1}{N}\sum_{n=0}^{N-1} s^{\mathbf{G}}_{u,*,n}.$$

Theorem 2. *Let* $\mathbf{G}$ *be a finite directed graph labelled by measure-preserving transformations of a probability space* (X,ν)*. Then for operators* $c^{\mathbf{G}}_{u,v,N}$ *and* $c^{\mathbf{G}}_{v_0,*,N}$ *defined above, the following statements hold.*
1a. *For any* $\varphi \in L^p(X,\nu)$*,* $p \in [1,\infty)$*, the sequence* $\{c^{\mathbf{G}}_{u,v,N}(\varphi)\}_{N=1}^{\infty}$ *converges in* $L^p(X,\nu)$*.*
1b. *For any* $\varphi \in L^\infty(X,\nu)$ *the sequence* $\{c^{\mathbf{G}}_{u,v,N}(\varphi)\}_{N=1}^{\infty}$ *converges* ν*-almost everywhere.*
2a. *For any* $\varphi \in L^p(X,\nu)$*,* $p \in [1,\infty)$*, the sequence* $\{c^{\mathbf{G}}_{v_0,*,N}(\varphi)\}_{N=1}^{\infty}$ *converges in* $L^p(X,\nu)$*.*
2b. *For any* $\varphi \in L^\infty(X,\nu)$ *the sequence* $\{c^{\mathbf{G}}_{v_0,*,N}(\varphi)\}_{N=1}^{\infty}$ *converges* ν*-almost everywhere.*

Statements 2a–b of Theorem 2 immediately imply Theorem 1. Indeed, if we assign the map $T_{\xi(e)}$ to an edge e, then

$$s_n(\varphi) = s^{\mathbf{G}}_{v_0,*,n}(\varphi).$$

Now we proceed to the proof of Theorem 2. Define a square matrix $M(\mathbf{G})$ of order $\#\mathcal{V}(\mathbf{G})$ with entries being operators on $L^1(X,\nu)$ by the formula

$$M(\mathbf{G})_{u,v} = \sum_{e\in\mathcal{E}(\mathbf{G},u,v)} T_e.$$

Denote also $M^\circ(\mathbf{G})_{u,v} = \#\mathcal{E}(\mathbf{G}, u, v)$. Note that if $\mathbf{1}$ is the function that equals 1 everywhere, then $T_e\mathbf{1} = \mathbf{1}$ for any $e \in \mathcal{E}(\mathbf{G})$. Define the following class of operators.

Definition 1. *A class* $\mathcal{B}^+$ of operators on $L^1(X, \nu)$ is a set of all operators $A\colon L^1(X, \nu) \to L^1(X, \nu)$ such that

1. there exists $\lambda(A) \in \mathbb{R}$ such that $A(\mathbf{1}) = \lambda(A) \cdot \mathbf{1}$,
2. if $f \geq 0$ (that is, $f(x) \geq 0$ for almost all $x \in X$) then $Af \geq 0$,
3. $A(L^p(X, \nu)) \subset L^p(X, \nu)$ for all $p \in [1, \infty]$,
4. $\|Af\|_p \leq \lambda(A)\|f\|_p$ for any $p \in [1, \infty]$, $f \in L^p(X, \nu)$.

It is clear that this class is a convex cone, that is, it is closed under linear combinations with nonnegative coefficients. Since all T_e's belong to this class, the same is true for $M(\mathbf{G})_{u,v}$, and

$$\lambda(M(\mathbf{G})_{u,v}) = \sum_{e \in \mathcal{E}(\mathbf{G},u,v)} \lambda(T_e) = \sum_{e \in \mathcal{E}(\mathbf{G},u,v)} 1 = M^\circ(\mathbf{G})_{u,v}.$$

Then, consider an n-th power of the graph $\mathbf{G}$, that is, a graph $\mathbf{G}' = \mathbf{G}^n$, where $\mathcal{V}(\mathbf{G}') = \mathcal{V}(\mathbf{G})$, $\mathcal{E}(\mathbf{G}') = \{l \in \mathcal{P}(\mathbf{G}), |l| = n\}$, and $I(l) = I(e_1)$, $F(l) = F(e_n)$ for $l = e_1 \dots e_n \in \mathcal{E}(\mathbf{G}')$.

By definition, $M(\mathbf{G}^n)_{u,v} = \sum_{l \in L^{\mathbf{G}}_{u,v,n}} T_l$. It is also clear that $(M(\mathbf{G}))^n = M(\mathbf{G}^n)$, and

$$\lambda((M(\mathbf{G})^n)_{u,v}) = (M^\circ(\mathbf{G})^n)_{u,v} = \#L^{\mathbf{G}}_{u,v,n}.$$

Now if we define an operation P on the class $\mathcal{B}^+$ as $\mathsf{P}(T) = T/\lambda(T)$ if $T \neq 0$, $\mathsf{P}(0) = 0$, then we have

$$s^{\mathbf{G}}_{u,v,n} = \mathsf{P}((M(\mathbf{G}^n))_{u,v}), \quad c^{\mathbf{G}}_{u,v,N} = \frac{1}{N}\sum_{n=0}^{N-1} \mathsf{P}((M(\mathbf{G}^n))_{u,v}).$$

Similarly,

$$s^{\mathbf{G}}_{v_0,*,n} = \mathsf{P}\left(\sum_{v \in \mathcal{V}(\mathbf{G})} (M(\mathbf{G}^n))_{v_0,v}\right), \quad c^{\mathbf{G}}_{v_0,*,N} = \frac{1}{N}\sum_{n=0}^{N-1} \mathsf{P}\left(\sum_{v \in \mathcal{V}(\mathbf{G})} (M(\mathbf{G}^n))_{v_0,v}\right).$$

3 The Main Lemma

The proof of statements 1a–b of Theorem 2 is obtained through a decomposition of the graph $\mathbf{G}$ into smaller blocks. The basic (non-decomposable) situation is the case of a strongly connected graph (that is, a graph such that for any its vertices u, v there exists a path from u to v) and in this case the theorem is proven in [6]. A step of the procedure starts with a decomposition of the set $\mathcal{V}(\mathbf{G})$ into two disjoint nonempty sets V_1, V_2 with no arcs from V_2 to V_1. Then we apply Theorem 2 to the induced subgraphs with these sets of vertices (that is, a graphs $\mathbf{G}_i$, $i = 1, 2$, with $\mathcal{V}(\mathbf{G}_i) = V_i$ and $\mathcal{E}(\mathbf{G}_i) = \{e \in \mathcal{E}(\mathbf{G}_i) : I(e), F(e) \in V_i\}$), and use Lemma 1 (see below), which is the main technical statement of the paper. The statements 2a–b of Theorem 2 are deduced from the statements 1a–b using the same lemma.

Definition 2. A sequence $\{x_n\}_{n=0}^{\infty}$, $x_n \geq 0$, is called *regular* if there exists a number $q \in \mathbb{N}$ such that for each $r = 0, \ldots, q-1$ one of the following statements holds:

(a) $x_{qk+r} = 0$ for all but finite number of $k \geq 0$,

(b) $\lim\limits_{k\to\infty} \dfrac{x_{qk+r}}{ak^b c^k} = 1$ for some $a > 0$, $b \in \mathbb{N}$, $c \geq 1$.

Definition 3. A sequence $\{T_n\}_n$, $T_n \in \mathcal{B}^+$, is called *pre-convergent* if

1. the sequence $\{\lambda(T_n)\}_n$ is regular;
2. for any $\varphi \in L^p(X, \nu)$ the sequence $\left\{\dfrac{1}{N}\sum\limits_{n=0}^{N-1} \mathsf{P}(T_n)(\varphi)\right\}$ converges in $L^p(X, \nu)$ as $N \to \infty$;
3. for any $\varphi \in L^\infty(X, \nu)$ the sequence $\left\{\dfrac{1}{N}\sum\limits_{n=0}^{N-1} \mathsf{P}(T_n)(\varphi)\right\}$ converges almost everywhere as $N \to \infty$.

In these terms, Theorem 2 can be reformulated as follows.

Proposition 1. *Under conditions of Theorem 2 the following statements hold.*
1. *For any induced subgraph $\mathbf{G}'$ of the graph $\mathbf{G}$ the sequence $\{(M(\mathbf{G}')^n)_{u,v}\}_n$ is pre-convergent for any $u, v \in \mathcal{V}(\mathbf{G}')$.*
2. *The sequence*

$$\left\{\sum_{v\in\mathcal{V}(\mathbf{G})} (M(\mathbf{G})^n)_{v_0,v}\right\}_n$$

is pre-convergent for any $v_0 \in \mathcal{V}(\mathbf{G})$.

The first statement of Proposition 1 is equivalent to the statements 1a–b of Theorem 2 for all induced subgraphs of $\mathbf{G}$. This is convenient for our inductive argument. The basis for the induction is the following theorem.

Theorem 3 ([6]). *If a graph $\mathbf{G}$ is strongly connected, then the sequence $\{(M(\mathbf{G})^n)_{u,v}\}_n$ is pre-convergent for any $u, v \in \mathcal{V}(\mathbf{G})$.*

Remark. 1. Regularity of the sequence $\{\lambda((M(\mathbf{G})^n)_{u,v})\}_n = \{(M^\circ(\mathbf{G})^n)_{u,v}\}_n$ in the case of strongly connected graph follows from the Perron—Frobenius theorem.
2. Convergence of $c^{\mathbf{G}}_{u,v,N}$ in $L^1(X,\nu)$ and almost everywhere (for functions in $L^1(X,\nu)$) is shown in [6] (see Theorems 1, 2; note that strong connectivity of $\mathbf{G}$ is called irreducibility of $A = M^\circ(\mathbf{G})$ in [6]). L^p-convergence for functions in $L^p(X,\nu)$ follows immediately.

The step of the inductive procedure relies on the following lemma.

Lemma 1. *If sequences $\{F_n\}$ and $\{G_n\}$ of operators from the class $\mathcal{B}^+$ are pre-convergent, then the following ones are also pre-convergent:*

1. $\{H_n^{(1)}\}_n$, $H_n^{(1)} = F_n$ *for* $n \geq n_0$, $H_n^{(1)} \in \mathcal{B}^+$;
2. $\{H_n^{(2)} = F_{n+M}\}_n$ *for any* $M \in \mathbb{Z}$;
3. $\{H_n^{(3a)} = AF_n\}_n$, $\{H_n^{(3b)} = F_nA\}_n$, *where* $A \in \mathcal{B}^+$;
4. $\{H_n^{(4)} = F_n + G_n\}_n$;
5. $\{H_n^{(5)} = \sum_{k+m=n} F_kG_m\}_n$.

We now derive Proposition 1 from Lemma 1.

Proof of Proposition 1. 1. The proof of the first statement is by induction on the number of vertices in $\mathbf{G}'$.

(a) Any graph $\mathbf{G}'$ with $\#\mathcal{V}(\mathbf{G}') = 1$ is strongly connected, thus we can apply Theorem 3.

(b) Take any induced subgraph $\mathbf{G}'$ with k vertices and suppose that the statement holds for any induced subgraph of $\mathbf{G}$ with less than k vertices. Then there are two cases: (1) $\mathbf{G}'$ is strongly connected; (2) $\mathbf{G}'$ can be decomposed as follows: $\mathcal{V}(\mathbf{G}') = V_1 \sqcup V_2$, $V_{1,2} \neq \varnothing$, and there are no arcs from V_2 to V_1.

In the first case we may apply Theorem 3. In the second case consider graphs $\mathbf{G}_{1,2}$ that are induced subgraphs with $\mathcal{V}(\mathbf{G}_i) = V_i$. Since $\mathbf{G}_{1,2}$ have less that k vertices, the theorem holds for them.

Now consider $c^{\mathbf{G}'}_{u,v,N}$. If $u, v \in V_i$, $i = 1, 2$, a path from u to v can't leave $\mathbf{G}_i$, so $(M(\mathbf{G}')^n)_{u,v} = (M(\mathbf{G}_i)^n)_{u,v}$, hence $c^{\mathbf{G}'}_{u,v,N} = c^{\mathbf{G}_i}_{u,v,N}$, and the statement is reduced to the one for $\mathbf{G}_i$. The case $u \in V_2$, $v \in V_1$ is even simpler: there are no paths from u to v, so $c^{\mathbf{G}'}_{u,v,N} = 0$ for all N.

The only nontrivial case is $u \in V_1$, $v \in V_2$. Here

$$(M(\mathbf{G}')^n)_{u,v} = \sum_{\substack{k+m=n-1 \\ v' \in V_1, u' \in V_2}} (M(\mathbf{G}_1)^k)_{u,u'} M(\mathbf{G}')_{u',v'} (M(\mathbf{G}_2)^m)_{v',v}$$

and the statement follows from Lemma 1. Indeed, by assumption, the sequence $\{(M(\mathbf{G}_1)^n)_{u,u'}\}_n$ is pre-convergent, hence, by item 3b of this lemma, the sequence $\{(M(\mathbf{G}_1)^n)_{u,u'} M(\mathbf{G}')_{u',v'}\}_n$ is. Then, as $\{(M(\mathbf{G}_2)^n)_{v',v}\}_n$ is pre-convergent by assumption, item 5 gives us that

$$\left\{A^{u,u',v,v'}_n = \sum_{k+m=n} (M(\mathbf{G}_1)^k)_{u,u'} M(\mathbf{G}')_{u',v'} (M(\mathbf{G}_2)^m)_{v',v}\right\}_n$$

is also pre-convergent. Now $\{A^{u,u',v,v'}_{n-1}\}_n$ is pre-convergent by item 2, and, finally the sequence

$$\left\{(M(\mathbf{G}')^n)_{u,v} = \sum_{v' \in V_1, u' \in V_2} A^{u,u',v,v'}_{n-1}\right\}_n$$

is pre-convergent by item 4 of Lemma 1.

2. The second statement follows immediately from item 4 of Lemma 1. □

4 Proof of Lemma 1

The rest of the paper is devoted to the proof of Lemma 1. The proof will often use the following proposition.

Proposition 2. *Let $A \in \mathcal{B}^+$, $\varphi_n \in L^\infty(X, \nu)$, $\|\varphi_n\|_\infty \le C$, $\varphi_n(x) \to \varphi(x)$ for almost all $x \in X$. Then $(A\varphi_n)(x) \to (A\varphi)(x)$ for almost all $x \in X$.*

Proof. Clearly, it is sufficient to prove this only for $\varphi = 0$.

Further, decompose φ_n as $\varphi_n = \varphi_n^+ - \varphi_n^-$, where $\varphi_n^\pm = \max(0, \pm\varphi_n)$, $\|\varphi_n^\pm\|_\infty \le \|\varphi_n\|_\infty \le C$. Therefore, if we prove that $A\varphi_n^\pm \xrightarrow{\text{a. e.}} 0$, then $A\varphi_n = A\varphi_n^+ - A\varphi_n^- \xrightarrow{\text{a. e.}} 0$. So we can assume that $\varphi_n \ge 0$.

Now, take $\psi_n(x) = \sup\{\varphi_k(x) \mid k \ge n\}$. Then $\psi_n(x)$ is monotonically nonincreasing and tends to zero for almost all $x \in X$. Since $0 \le \varphi_n \le \psi_n$, the same is true for their images: $0 \le A\varphi_n \le A\psi_n$ and therefore, it is

sufficient to prove that $A\psi_n \xrightarrow{\text{a. e.}} 0$. But as $A\psi_n(x)$ is nonnegative and nonincreasing, there is a limit $\theta(x) = \lim_{n\to\infty} A\psi_n(x)$, and, by monotone convergence theorem,

$$\|A\psi_n - \theta\|_1 = \int_X A\psi_n(x) - \theta(x)\, d\nu(x) \to 0$$

Therefore, $A\psi_n \to \theta$ in $L^1(X,\nu)$. But A is a bounded operator in $L^1(X,\nu)$ and $\|\psi_n\|_1 \to 0$ (also due to monotone convergence theorem), so $A\psi_n \to 0$ in $L^1(X,\nu)$. Thus $\theta(x) = 0$ almost everywhere. □

Proof of Lemma 1. The plan of the proof is the following. After some preparations, we'll prove the first condition in Definition 3 for all sequences $\{H_n^{(*)}\}_n$ (here and below the asterisk $*$ denotes one of the symbols 1, 2, 3a, 3b, 4, 5), and then we'll prove the second and the third condtions of that Definition simultaneously.

1. First of all, it is sufficient to prove that this lemma holds for the sequences $\{\lambda(F_n)\}$ and $\{\lambda(G_n)\}$ satisfying Definition 2 with $q = 1$ (and that in this case the sequence $\{\lambda(H_n^{(*)})\}$ is also regular with $q = 1$).

Indeed, in general case we take q to be the least common multiple of q_F and q_G (i. e., q's from Definition 2 for the sequences $\{F_n\}_n$ and $\{G_n\}_n$). For $* \neq 5$, it is clear that for a given $r = 0, \ldots, q-1$ the sequence $\{H^{(*)}_{qs+r}\}_s$ depends in the same fashion on one of $\{F_{qs+r'}\}_s$ and $\{G_{qs+r''}\}_s$ with some r', r''.

Now consider $* = 5$. Let $k = qu + r'$, $m = qv + r''$ ($u, v \geq 0$, $0 \leq r', r'' \leq q-1$) and decompose the sum

$$\sum_{k+m=qs+r} F_k G_m$$

into q sums corresponding to all possible pairs (r', r'') (there are only q possibilities, since $r' + r'' \equiv r \pmod q$):

$$H^{(5)}_{qs+r} = \sum_{r'+r''\equiv r \,(\mathrm{mod}\, q)} S_s^{r',r''},$$

where

$$S_s^{r',r''} = \sum_{\substack{u,v\geq 0 \\ (qu+r')+(qv+r'')=qs+r}} F_{qu+r'} G_{qv+r''} = \sum_{\substack{u,v\geq 0 \\ u+v=s+\frac{r-r'-r''}{q}}} F_{qu+r'} G_{qv+r''},$$

that is, the sequence $\{S_s^{r',r''}\}_s$ is the convolution of the sequences $\{F_{qs+r'}\}_s$ and $\{G_{qs+r''}\}_s$ shifted by $\frac{r-r'-r''}{q} \in \{-1, 0\}$.

2. Let us prove that the sequences $\{\lambda(H_n^{(*)})\}_n$ are regular. For $* = 1, 2, 3a, 3b$ this is clear from the definitions. Let $* = 4$. If $\{\lambda(F_n)\}$ or $\{\lambda(G_n)\}$ contains only finitely many nonzero elements, this is clear. Otherwise, let (a_F, b_F, c_F) and (a_G, b_G, c_G) be the constants given in Definition 2 for these sequences.

If (1) $c_G < c_F$ or (2) $c_G = c_F$, $b_G < b_F$, then

$$\lim_{k\to\infty} \frac{\lambda(G_k)}{a_F k^{b_F} c_F^k} = 0,$$

so

$$\lim_{k\to\infty} \frac{\lambda(F_k + G_k)}{a_F k^{b_F} c_F^k} = 1.$$

The symmetric cases $(1')$ $c_F < c_G$; $(2')$ $c_F = c_G$, $b_F < b_G$ are similar. The only remaining case is $c_F = c_G = c$, $b_F = b_G = b$. Here

$$\lim_{k\to\infty} \frac{\lambda(F_k + G_k)}{(a_F + a_G) k^b c^k} = 1.$$

Now let $* = 5$. The case of finitely many nonzeros is again clear, otherwise we can assume that $c_F \geq c_G$. There are two cases, $c_F > c_G$ and $c_F = c_G$.

Suppose that $c_F > c_G$. Then

$$\frac{\sum\limits_{k+m=n} \lambda(F_k)\lambda(G_m)}{a_F (n+1)^{b_F} c_F^n} = \sum_{m=0}^{n} \lambda(G_m) \left(\frac{\lambda(F_{n-m})}{a_F (n+1)^{b_F} c_F^n} \right).$$

Let us prove that this sum tends to $\sum_{m=0}^{\infty} \lambda(G_m) c_F^{-m}$.

Denote[1]

$$\alpha_n = \frac{\lambda(F_n)}{a_F (n+1)^{b_F} c_F^n}, \quad \beta_n = \frac{\lambda(G_m)}{c_F^m}$$

and fix $\varepsilon > 0$. Note that the series $\sum_{m=0}^{\infty} \lambda(G_m) c_F^{-m} = \sum_{m=0}^{\infty} \beta_n$ converges absolutely, so there is m_0 such that $\sum_{m>m_0} \beta_m < \varepsilon$. Let A be an upper bound for all α_n, $n \geq 1$ (it exists since $\alpha_n \to 1$). Then

$$\left| \sum_{m=0}^{n} \beta_m \alpha_{n-m} \Big(\frac{n-m+1}{n+1} \Big)^{b_F} - \sum_{m=0}^{\infty} \beta_m \right| \leq$$

$$\leq \left| \sum_{m=0}^{m_0} \beta_m \left[\alpha_{n-m} \Big(1 - \frac{m}{n+1} \Big)^{b_F} - 1 \right] \right| +$$

$$+ \left| \sum_{m=m_0+1}^{n} \beta_m \alpha_{n-m} \Big(1 - \frac{m}{n+1} \Big)^{b_F} \right| + \left| \sum_{m=m_0+1}^{\infty} \beta_m \right|.$$

[1]We write $(n+1)^b$ in the denominator instead of n^b to have well-defined α_0. Nevertheless, α_n tends to 1.

The last term is less than ε, the second one is less than $A\varepsilon$ and, if n is sufficiently large, the first term is less than ε, hence the whole difference is less than $(2+A)\varepsilon$ for sufficiently large n. Thus, $\{\lambda(H_n^{(5)}\}$ is regular with

$$a_H = a_F \sum_{m=0}^{\infty} \lambda(G_m) c_F^{-m}, \quad b_H = b_F, \quad c_H = c_F.$$

Now let $c_F = c_G = c$. In this case we have

$$\frac{\sum_{k+m=n} \lambda(F_k)\lambda(G_m)}{a_F a_G (n+1)^{b_F+b_G+1} c^n} =$$
$$\frac{1}{n+1} \sum_{k=0}^{n} \underbrace{\frac{\lambda(F_k)}{a_F (k+1)^{b_F} c_F^k}}_{\alpha_k} \underbrace{\frac{\lambda(G_{n-k})}{a_G (n-k+1)^{b_G} c_G^{n-k}}}_{\beta_{n-k}} \underbrace{\left(\frac{k+1}{n+1}\right)^{b_F} \left(1 - \frac{k}{n+1}\right)^{b_G}}_{\gamma_{n,k}} \tag{1}$$

and denote α_k, β_{n-k} and $\gamma_{n,k}$ as it is shown here. Let us show that

$$\lim_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} \alpha_k \beta_{n-k} \gamma_{n,k} - \frac{1}{n+1} \sum_{k=0}^{n} \gamma_{n,k} = 0.$$

Indeed, by Definition 2, the sequences $\{\alpha_k\}$ $\{\beta_k\}$ tends to 1, hence there are A, B such that $\alpha_k \le A$, $\beta_k < B$ for all k. Take any $\varepsilon < 1$ and find p such that $|\alpha_k - 1| < \varepsilon$, $|\beta_k - 1| < \varepsilon$ for all $k \ge p$. Then

$$\Delta_n = \frac{1}{n+1} \sum_{k=0}^{n} \alpha_k \beta_{n-k} \gamma_{n,k} - \frac{1}{n+1} \sum_{k=0}^{n} \gamma_{n,k} =$$
$$\frac{1}{n+1} \left(\sum_{k=0}^{p-1} + \sum_{k=p}^{n-p} + \sum_{k=n-p+1}^{n} \right) (\alpha_k \beta_{n-k} - 1) \gamma_{n,k}.$$

Since $0 \le \gamma_{n,k} \le 1$, any term of the first and the last sums is bounded by $AB+1$ and any term of the middle sum is bounded by $2\varepsilon + \varepsilon^2 \le 3\varepsilon$. Therefore, we have

$$\Delta_n \le \frac{2p(AB+1) + 3\varepsilon(n+1-2p)}{n+1} \le 3\varepsilon + \frac{2(AB+1)p}{n+1}.$$

If n is large enough then the last term is less than ε, hence $\Delta_n \le 4\varepsilon$.

It remains to find the limit

$$\lim_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} \gamma_{n,k}.$$

We have

$$\frac{1}{n+1}\sum_{k=0}^{n}\gamma_{n,k}=\Big(\frac{n+2}{n+1}\Big)^{b_F+b_G+1}\bigg(\frac{1}{n+2}\sum_{j=1}^{n+1}\Big(\frac{j}{n+2}\Big)^{b_F}\Big(1-\frac{j}{n+2}\Big)^{b_G}\bigg).$$

The first multiplier tends to 1. The second one equals the Riemann sum of the function $f(x)=x^{b_F}(1-x)^{b_G}$ over the unit interval with the partition

$$\Big\{x_i=\frac{i}{n+2}\Big\}_{i=0}^{n+2},\quad \{t_i=x_i\}_{i=0}^{n+1},$$

hence it tends to $\int_0^1 f(x)\,dx=B(b_F+1,b_G+1)$. Thus, in this case $\{\lambda(H_n^{(5)}\}$ is regular with the constants

$$a_H=a_Fa_GB(b_F+1,b_G+1),\quad b_H=b_F+b_G+1,\quad c_H=c.$$

3. We proceed to the proof of the second and the third conditions in Definition 3.

For $*=1,2$ the difference between Cesàro sums satisfies the relations

$$\frac{1}{N}\sum_{n=0}^{N-1}\mathsf{P}(H_n^{(1)})-\frac{1}{N}\sum_{n=0}^{N-1}\mathsf{P}(F_n)=\frac{1}{N}\sum_{n=0}^{n_0-1}\big(\mathsf{P}(H_n^{(1)})-\mathsf{P}(F_n)\big),$$
$$\frac{1}{N}\sum_{n=0}^{N-1}\mathsf{P}(H_n^{(2)})-\frac{1}{N}\sum_{n=0}^{N-1}\mathsf{P}(F_n)=\frac{1}{N}\sum_{n=0}^{M-1}\big(\mathsf{P}(F_{N+n})-\mathsf{P}(F_n)\big),$$

whence it tends to zero even in operator norm in any $L^p(X,\nu)$, $p\in[1,\infty]$.

For $*=3\mathrm{a},3\mathrm{b}$ the conditions follows from the identities

$$\frac{1}{N}\sum_{n=0}^{N-1}\mathsf{P}(H_n^{(3\mathrm{a})})=\mathsf{P}(A)\bigg(\frac{1}{N}\sum_{n=0}^{N-1}\mathsf{P}(F_n)\bigg),$$
$$\frac{1}{N}\sum_{n=0}^{N-1}\mathsf{P}(H_n^{(3\mathrm{b})})=\bigg(\frac{1}{N}\sum_{n=0}^{N-1}\mathsf{P}(F_n)\bigg)\mathsf{P}(A).$$

The only remaining cases are $*=4,5$. Let us show that we can make "approximate" normalisations of operators instead of "precise" ones (that is, $\mathsf{P}(\,\cdot\,)$) in the second and the third conditions in Definition 3. Speaking formally, the following holds.

Claim 1. *Suppose that the sequence $\{T_n\}_n$, $T_n \in \mathcal{B}^+$, satisfies the condition*

$$\lim_{n\to\infty} \frac{\lambda(T_n)}{an^b c^n} = 1$$

with some $a > 0$, $b \in \mathbb{N}$, and $c \geq 1$. Let

$$\hat{T}_n = \frac{T_n}{a(n+1)^b c^n}.$$

Then for any $\varphi \in L^p(X,\nu)$ the sequences

$$C_N(\varphi) = \frac{1}{N}\sum_{n=0}^{N-1} \mathsf{P}(T_n)(\varphi) \quad \textit{and} \quad C'_N(\varphi) = \frac{1}{N}\sum_{n=0}^{N-1} \hat{T}_n(\varphi)$$

converge (in L^p or a. e.) simultaneously and their limits coincide.

Proof. If $\lambda(\hat{T}_n) = \gamma_n$, $\gamma_n \to 1$, then we have

$$\|C_N - C'_N\|_p = \left\| \frac{1}{N}\sum_{n=0}^{N-1} (\mathsf{P}(T_n) - \hat{T}_n) \right\|_p = \left\| \frac{1}{N}\sum_{n=0}^{N-1} \mathsf{P}(T_n)(1-\gamma_n) \right\|_p \leq$$
$$\leq \frac{1}{N}\sum_{n=0}^{N-1} \|\mathsf{P}(T_n)\|_p \cdot |1-\gamma_n| \leq \frac{1}{N}\sum_{n=0}^{N-1} |1-\gamma_n|,$$

and the latter is the Cesàro sum of $x_n = |1-\gamma_n|$, which tends to zero. Thus the difference $C_N - C'_N$ tends to zero in operator norm in any $L^p(X,\nu)$, $p \in [1,\infty]$. □

Now let $* = 4$. If one of the sequences $\{F_n\}$, $\{G_n\}$ has only finitely many nonzero terms, we can use the lemma's statement for $\{H_n^{(1)}\}_n$. Otherwise take the constants $a_F, b_F, c_F,\ a_G, b_G, c_G$ same as before and introduce operators $\hat{F}_n$, $\hat{G}_n$, $\hat{H}_n^{(4)}$ in the same way as in Claim 1.

From the previous section of the proof one can see that $\hat{H}_n^{(4)}$ is either $\frac{a_F}{a_F+a_G}\hat{F}_n + \frac{a_G}{a_F+a_G}\hat{G}_n$ (if $c_F = c_G$ and $b_F = b_G$), or $\hat{F}_n + \varepsilon_n \hat{G}_n$ with $\varepsilon_n \to 0$ (if $c_G < c_F$, or if $c_G = c_F$ and $b_G < b_F$), or $\varepsilon_n \hat{F}_n + \hat{G}_n$ (in symmetric cases). The convergence of Cesàro sums of $\hat{H}_n^{(4)}$ in the first case is obvious, in the two latter cases the term $\varepsilon_n \hat{F}_n$ (or $\varepsilon_n \hat{G}_n$) tends to zero in operator norm:

$$\|\varepsilon_n \hat{F}_n\| \leq \varepsilon_n (\|F_n\|/a_F n^{b_F} c_F^n) \to 0 \cdot 1,$$

and so does the sequence of its Cesàro averages.

Finally, suppose $* = 5$. As usual, the proof is clear if $\{F_n\}$ or $\{G_n\}$ contains finitely many nonzero terms, otherwise let $a_{F,G,H}$, $b_{F,G,H}$, $c_{F,G,H}$ be the coefficients in the regularity condition respectively for $\{\lambda(F_n)\}$, $\{\lambda(G_n)\}$, $\{\lambda(H_n^{(5)})\}$. Similarly to the case $* = 4$, we'll prove convergence for the sequence

$$\Big\{\frac{1}{N}\sum_{n=0}^{N-1}\hat{H}_n^{(5)}(\varphi)\Big\}_N.$$

There are three cases, $c_F > c_G$, $c_F < c_G$, and $c_F = c_G$. Suppose the first one. Then $c_H = c_F$, $b_H = b_F$, and for any $\varphi \in L^p(X,\nu)$ we have

$$C_N(\varphi) = \frac{1}{N}\sum_{n=0}^{N-1}\hat{H}_n^{(5)}(\varphi) = \frac{1}{N}\sum_{k+m<N}\frac{a_F}{a_H}\hat{F}_k\Big(\frac{G_m(\varphi)}{c_F^m}\Big)\cdot\Big(\frac{k+1}{k+m+1}\Big)^{b_F} = \sum_{m=0}^{N-1}\underbrace{\frac{1}{N}\sum_{k<N-m}\frac{a_F}{a_H}\hat{F}_k\Big(\frac{G_m(\varphi)}{c_F^m}\Big)\cdot\Big(\frac{k+1}{k+m+1}\Big)^{b_F}}_{S_{m,N}(\varphi)}. \quad (2)$$

Let A be chosen in such a way that $\|\hat{F}_k\| \le A$ for all k. Then we have

$$\|S_{m,N}(\varphi)\|_p \le \frac{1}{N}\sum_{k<N-m}\frac{a_F}{a_H}\|\hat{F}_k\|_p\frac{\lambda(G_m)}{c_F^m}\|\varphi\|_m\Big(\frac{k+1}{k+m+1}\Big)^{b_F} \le \frac{1}{N}\sum_{k<N-m}\frac{a_F}{a_H}\cdot A\cdot\frac{\lambda(G_m)}{c_F^m}\cdot\|\varphi\|_m\cdot 1 \le \frac{Aa_F\|\varphi\|_p}{a_H}\cdot\frac{\lambda(G_m)}{c_F^m}.$$

Since $\sum_{m=0}^{\infty}\lambda(G_m)c_F^{-m} < \infty$, we can choose m_0 such that

$$\sum_{m>m_0}\lambda(G_m)c_F^{-m} < \varepsilon\cdot\frac{a_H}{a_F\|\varphi\|_p}. \quad (3)$$

Then we have

$$\Big\|\sum_{m>m_0}S_{m,N}(\varphi)\Big\|_p \le A\varepsilon. \quad (4)$$

Further, let us find the limit of $S_{m,N}(\varphi)$ as $N\to\infty$. Denote

$$\psi_m = \frac{a_F}{a_H}\frac{G_m(\varphi)}{c_F^m},$$

then

$$S_{m,N}(\varphi) = \frac{N-m}{N}\Bigg(\frac{1}{N-m}\sum_{k<N-m}\hat{F}_k(\psi_m) - \\ - \frac{1}{N-m}\sum_{k<N-m}\hat{F}_k(\psi_m)\Big[1-\Big(\frac{k+1}{k+m+1}\Big)^{b_F}\Big]\Bigg). \quad (5)$$

Due to regularity of the sequence $\{F_k\}$, the first term in parentheses tends in L^p or a. e. to a function, which will be denoted as $F^0(\psi_m)$. Note also that the equality

$$F^0(\theta) = \lim_{n\to\infty}\frac{1}{N}\sum_{k=0}^{N-1}\mathsf{P}(F_k)(\theta) = \lim_{n\to\infty}\frac{1}{N}\sum_{k=0}^{N-1}\hat{F}_k(\theta)$$

defines a linear operator $F^0 \in \mathcal{B}^+$, with $\lambda(F^0) = 1$.

The second term in parentheses in (5) is the Cesàro average for the sequence

$$\theta_{m,k} = \hat{F}_k(\psi_m)\Big[1-\Big(\frac{k+1}{k+m+1}\Big)^{b_F}\Big],$$

which tends to zero in $L^p(X,\nu)$, $p \in [1,\infty]$, as $k \to \infty$. Thus $S_{m,N} \to F^0(\psi_m)$ in L^p or a. e. In particular, there exists N_m such that for any $N > N_m$ we have

$$\|S_{m,N}(\varphi) - F^0(\psi_m)\|_p \le \frac{\varepsilon}{m_0+1}. \quad (6)$$

Similarly, for $\varphi \in L^\infty(X,\nu)$, for almost all x there exists $N_m(x)$ such that for any $N > N_m(x)$ we have

$$|S_{m,N}(\varphi)(x) - F^0(\psi_m)(x)| \le \frac{\varepsilon}{m_0+1}. \quad (6')$$

Note also that (3) yields

$$\Big\|\sum_{m>m_0}\psi_m\Big\|_p \le \sum_{m>m_0}\frac{a_F}{a_H}\|\varphi\|_p\frac{\lambda(G_m)}{c_F^m} < \varepsilon,$$

whence, noting that $\|F^0\|_p \le 1$, we obtain

$$\Big\|\sum_{m=0}^{m_0}F^0(\psi_m) - F^0\Big(\frac{a_F}{a_H}\sum_{m=0}^{\infty}\frac{G_m(\varphi)}{c_F^m}\Big)\Big\|_p < \varepsilon. \quad (7)$$

Now we can see that

$$C_N(\varphi) - F^0\Big(\frac{a_F}{a_H}\sum_{m=0}^{\infty}\frac{G_m(\varphi)}{c_F^m}\Big) = \sum_{m>m_0} S_{m,N}(\varphi) + \sum_{m=0}^{m_0}(S_{m,N}(\varphi) - F^0(\psi_m)) + \Big(\sum_{m=0}^{m_0} F^0(\psi_m) - F^0\Big(\frac{a_F}{a_H}\sum_{m=0}^{\infty}\frac{G_m(\varphi)}{c_F^m}\Big)\Big),$$

and, if $N > \max(N_0, \dots, N_{m_0})$, the estimates (4), (6), (7) give us the inequality

$$\Big\|C_N(\varphi) - F^0\Big(\frac{a_F}{a_H}\sum_{m=0}^{\infty}\frac{G_m(\varphi)}{c_F^m}\Big)\Big\| < (2+A)\varepsilon.$$

Similarly, if $N > \max(N_0(x), \dots, N_{m_0}(x))$ for $\varphi \in L^\infty(X,\nu)$ then (4), (6′), and (7) imply

$$\Big|C_N(\varphi)(x) - F^0\Big(\frac{a_F}{a_H}\sum_{m=0}^{\infty}\frac{G_m(\varphi)}{c_F^m}\Big)(x)\Big| < (2+A)\varepsilon.$$

The second case $c_F < c_G$ is treated similarly. Namely, the sum for $C_N(\varphi)$ is decomposed into the sums

$$S_{k,N}(\varphi) = \frac{a_G}{a_H}\frac{1}{c_G^k}F_k\Big(\frac{1}{N}\sum_{m<N-k}\hat{G}_m(\varphi)\Big(\frac{m+1}{k+m+1}\Big)^{b_G}\Big). \tag{8}$$

The estimate of its norm for $k > k_0$ is the same, and the only difference is in the proof of convergence of $S_{k,N}(\varphi)$ as $N \to \infty$: the argument of F_k in (8) tends to $G^0(\varphi)$, so $S_{k,N}$ tends (in L^p or a. e.) to

$$\frac{a_G}{a_H}\frac{F_k(G^0(\varphi))}{c_G^k},$$

with Proposition 2 being used in case $\varphi \in L^\infty(X,\nu)$.

Now consider the third case $c_F = c_G = c$. Here

$$\frac{1}{N}\sum_{n=0}^{N-1}\hat{H}_n^{(5)} = \frac{1}{N}\sum_{k+m<N}\frac{1}{a_H(k+m)^{b_F+b_G+1}}\frac{F_k}{c^k}\frac{G_m}{c^m}$$

and the lemma follows from Proposition 3 for $X_n = F_n c^{-n}$, $Y_n = G_n c^{-n}$. □

Proposition 3. *Let $X_n, Y_n \in \mathcal{B}^+$ be such that*

1. *the sequences $\{\lambda(X_n)/(n+1)^u\}$ and $\{\lambda(Y_n)/(n+1)^v\}$ are bounded,*
2. *for any $\varphi \in L^p(X,\nu)$, $p \in [1,\infty)$, the sequences $\left\{\frac{1}{N}\sum_{n=0}^{N-1}\frac{X_n(\varphi)}{(n+1)^u}\right\}$ and $\left\{\frac{1}{N}\sum_{n=0}^{N-1}\frac{Y_n(\varphi)}{(n+1)^v}\right\}$ converge in $L^p(X,\nu)$ as $N\to\infty$,*
3. *for any $\varphi \in L^\infty(X,\nu)$ the sequences $\left\{\frac{1}{N}\sum_{n=0}^{N-1}\frac{X_n(\varphi)}{(n+1)^u}\right\}$ and $\left\{\frac{1}{N}\sum_{n=0}^{N-1}\frac{Y_n(\varphi)}{(n+1)^v}\right\}$ converge almost everywhere as $N\to\infty$.*

Let $Z_n = \sum_{k+m=n} X_k Y_m$, $w = u+v+1$. Then

1. *for any $\varphi \in L^p(X,\nu)$, $p \in [1,\infty)$, the sequence $\left\{\frac{1}{N}\sum_{n=0}^{N-1}\frac{Z_n(\varphi)}{(n+1)^w}\right\}$ converges in $L^p(X,\nu)$,*
2. *for any $\varphi \in L^\infty(X,\nu)$ the sequence $\left\{\frac{1}{N}\sum_{n=0}^{N-1}\frac{Z_n(\varphi)}{(n+1)^w}\right\}$ converges almost everywhere.*

Proof. 1. Let

$$X^0(\varphi) = \lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\frac{X_n(\varphi)}{(n+1)^u}, \quad Y^0(\varphi) = \lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\frac{Y_n(\varphi)}{(n+1)^v}. \tag{9}$$

These operators belong to $\mathcal{B}^+$. Indeed, the first two conditions are obvious, and, to check the remaining two, one can see that

$$\|X^0(\varphi)\|_p \le \limsup_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\frac{\|\lambda(X_n)\varphi\|_p}{(n+1)^u} \le \limsup_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\frac{\lambda(X_n)}{(n+1)^u}\|\varphi\|_p,$$

$$|X^0(\varphi)(x)| \le \limsup_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\frac{|X_n\varphi(x)|}{(n+1)^u)} \le \limsup_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\frac{\lambda(X_n)}{(n+1)^u}\|\varphi\|_\infty,$$

and note that the sequence

$$\left\{\frac{1}{N}\sum_{n=0}^{N-1}\frac{\lambda(X_n)}{(n+1)^u}\right\}_N$$

is bounded by the same bound as the sequence $\{\lambda(X_n)/(n+1)^u\}_n$.

2. Now introduce

$$X_n^* = X_n - (n+1)^u X^0, \qquad Y_n^* = Y_n - (n+1)^v Y^0.$$

These operators are bounded in any $L^p(X,\nu)$, $p \in [1,\infty]$, and the norms $\|X_n^*\|_p/(n+1)^u$, $\|Y_n^*\|_p/(n+1)^v$ are bounded uniformly on $p \in [1,\infty]$ and n (indeed, these bounds are simply twice the bounds for $\|X_n\|_p/(n+1)^u = \lambda(X_n)/(n+1)^u$, $\|Y_n\|_p/(n+1)^v = \lambda(Y_n)/(n+1)^v$). This is an analogue of the first condition of the proposition; one can see that the second and the third conditions hold for X_n^*, Y_n^* in place of X_n, Y_n.

Furthermore,

$$\begin{aligned}
&\frac{Z_n(\varphi)}{(n+1)^w} = \sum_{k+m=n} \frac{X_k Y_m(\varphi)}{(k+m+1)^w} = \\
&= \sum_{k+m=n} \frac{X_k^* Y_m^*(\varphi)}{(k+m+1)^w} - \Big(\sum_{k+m=n} \frac{(m+1)^v X_k^*}{(k+m+1)^w} \Big)(Y^0(\varphi)) - \\
&\quad - X^0 \Big(\sum_{k+m=n} \frac{(k+1)^u Y_m^*(\varphi)}{(k+m+1)^w} \Big) + \sum_{k+m=n} \frac{(k+1)^u (m+1)^v}{(k+m+1)^w} X^0 Y^0(\varphi).
\end{aligned} \tag{10}$$

To prove Proposition 3, it is sufficient to prove (L^p- and a. e.-) convergence of Cesàro averages for each term in (10).

3. For the last term in (10) the proof is simple:

$$\sum_{k+m=n} \frac{(k+1)^u (m+1)^v}{(k+m+1)^w} = \Big(\frac{n+2}{n+1}\Big)^w \cdot \Big(\frac{1}{n+2} \sum_{j=1}^{n+1} \Big(\frac{j}{n+2}\Big)^u \Big(1 - \frac{j}{n+2}\Big)^v \Big).$$

Here the first multiplier tends to 1 and the second one is the Riemann sum of $f(x) = x^u(1-x)^v$ with a partition of $[0,1]$ into $n+2$ equal intervals, so it tends to the Euler integral $B(u+1, v+1)$. Therefore, the last term tends to $B(u+1,v+1)X^0Y^0(\varphi)$ and so do its Cesàro averages.

4. To prove convergence of the second and the third terms in (10), it is sufficient to prove that Cesàro averages of

$$\sum_{k+m=n} \frac{(m+1)^v X_k^*(\varphi)}{(k+m+1)^w} \quad \text{and} \quad \sum_{k+m=n} \frac{(k+1)^u Y_m^*(\varphi)}{(k+m+1)^w} \tag{11}$$

converge to zero in $L^p(X,\nu)$ for any $\varphi \in L^p(X,\nu)$, $p \in [1,\infty)$, and a. e. for any $\varphi \in L^\infty(X,\nu)$. Indeed, for the second term we denote $\psi = Y^0(\varphi)$ and for the third one we use either boundedness of the operator X^0 in $L^p(X,\nu)$ or Proposition 2.

The expressions in (11) transform to another one when we swap $X \leftrightarrow Y$, $u \leftrightarrow v$, and $k \leftrightarrow m$, so we may deal only with the first of them.

Denote

$$A_n = \frac{1}{n}\sum_{k=0}^{n-1} \frac{X_k^*}{(k+1)^u}, \quad \varphi_n = A_n(\varphi).$$

By construction, φ_n tends to 0 in $L^p(X,\nu)$ for $\varphi \in L^p(X,\nu)$ and almost everywhere for $\varphi \in L^\infty(X,\nu)$. Further,

$$X_n^* = (n+1)^u\big((n+1)A_{n+1} - nA_n\big),$$

thus

$$C_N = \frac{1}{N}\sum_{n=0}^{N-1}\sum_{k+m=n} \frac{(m+1)^v X_k^*(\varphi)}{(k+m+1)^w} = \\ \frac{1}{N}\sum_{k+m<N} \frac{(m+1)^v(k+1)^u}{(k+m+1)^w}\big((k+1)\varphi_{k+1} - k\varphi_k\big),$$

and, rearranging the sum, we have

$$C_N = \sum_{k=1}^{N} \frac{k}{N}\left(\sum_{m=0}^{N-k} \frac{(m+1)^v k^u}{(m+k)^w} - \sum_{m=0}^{N-k-1} \frac{(m+1)^v(k+1)^u}{(m+k+1)^w}\right)\varphi_k. \tag{12}$$

Now we'll use the following statement.

Claim 2. *Let $\alpha_{N,k} \in \mathbb{R}$, $\xi_k \in \Xi$, where Ξ is a normed space. Suppose that*

1. *$\xi_k \to 0$ as $k \to \infty$,*
2. *for any fixed N, there are only finitely many k's with $\alpha_{N,k} \neq 0$,*
3. *for any fixed k, $\alpha_{N,k} \to 0$ as $N \to \infty$,*
4. *there is such C that $\sum_k |\alpha_{N,k}| < C$ for any N.*

Then $\sum_k \alpha_{N,k}\xi_k \to 0$ as $N \to \infty$.

Proof of Claim 2. Let $\|\xi_k\| < R$ for any k. Take any $\varepsilon > 0$ and choose k_0 in such a way that $\|\xi_k\| < \varepsilon$ for $k > k_0$. Since

$$\sum_{k \le k_0} |\alpha_{N,k}| \to 0 \quad \text{as} \quad N \to \infty,$$

we can choose N_0 such that for any $N > N_0$

$$\sum_{k \le k_0} |\alpha_{N,k}| < \varepsilon.$$

Therefore, for any $N > N_0$ we have

$$\Big\|\sum_k \alpha_{N,k}\xi_k\Big\| \le \sum_k |\alpha_{N,k}|\|\xi_k\| = \sum_{k\le k_0} |\alpha_{N,k}|\|\xi_k\| + \sum_{k>k_0} |\alpha_{N,k}|\|\xi_k\| \le \varepsilon R + C\varepsilon,$$

and the claim is established. □

We apply Claim 2 to (12) either with $\xi_k = \varphi_k$, $\Xi = L^p(X,\nu)$ (if $\varphi \in L^p(X,\nu)$) or with $\xi_k = \varphi_k(x)$, $\Xi = \mathbb{R}$ (if $\varphi \in L^\infty(X,\nu)$). Obviously, $\xi_k \to 0$, and we need to check conditions on $\alpha_{N,k}$, where

$$\alpha_{N,k} = \frac{k}{N}\left[\frac{(N-k+1)^v k^u}{N^w} + \sum_{m=0}^{N-k-1}(m+1)^v\left(\frac{k^u}{(m+k)^w} - \frac{(k+1)^u}{(m+k+1)^w}\right)\right] \tag{13}$$

for $k = 1,\dots,N$, otherwise $\alpha_{N,k} = 0$. The value in round brackets is of the form $f(k) - f(k+1)$ for $f(x) = x^u/(x+m)^w$, so we apply the mean value theorem to it.

There are two cases: $u > 0$ and $u = 0$. In the first case,

$$\left|\frac{k^u}{(m+k)^w} - \frac{(k+1)^u}{(m+k+1)^w}\right| = |f'(x_m)| = \frac{x_m^{u-1}|um - (v+1)x_m|}{(x_m+m)^{w+1}} \le$$
$$\le \frac{(k+1)^{u-1}}{(m+k)^w}\frac{um + (v+1)x_m}{x_m+m} \le \frac{(k+1)^{u-1}}{(m+k)^w}(u+v+1)$$

(here $x_m \in [k, k+1]$). Thus we have

$$|\alpha_{N,k}| \le \frac{k^{u+1}(N-k+1)^v}{N^{w+1}} + \frac{kw}{N}\sum_{m=0}^{N-k-1}\frac{(m+1)^v(k+1)^{u-1}}{(m+k)^w} \le$$
$$\le \frac{1}{N} + \frac{k(k+1)^{u-1}w}{N}\sum_{m=0}^{N-k-1}\frac{1}{(m+k)^{u+1}}$$

The sum $\sum_{j=k}^\infty j^{-(u+1)}$ is estimated as

$$\sum_{j=k}^\infty \frac{1}{j^{u+1}} = \frac{1}{k^{u+1}} + \sum_{j=k+1}^\infty \frac{1}{j^{u+1}} \le$$
$$\frac{1}{k^{u+1}} + \int_k^{+\infty}\frac{dx}{x^{u+1}} = \frac{1}{k^{u+1}} + \frac{1}{uk^u} \le \Big(1+\frac{1}{u}\Big)\frac{1}{k^u}.$$

Continue estimation for $|\alpha_{N,k}|$:

$$|\alpha_{N,k}| \le \frac{1}{N}\Big(1+\frac{w(1+u)}{u}\cdot\frac{k(k+1)^{u-1}}{k^u}\Big) = \\ \frac{1}{N}\Big(1+\frac{w(1+u)}{u}\cdot\Big(\frac{k+1}{k}\Big)^{u-1}\Big) \le \frac{1}{N}\Big(1+\frac{w(1+u)}{u}\cdot 2^{u-1}\Big).$$

Hence $\alpha_{N,k}\to 0$ as $N\to\infty$ for any fixed k , and

$$\sum_k|\alpha_{N,k}| \le 1+\frac{w(1+u)}{u}\cdot 2^{u-1}.$$

Thus in the case $u>0$ all conditions of Claim 2 hold.

Now let $u=0$. Here

$$\Big|\frac{1}{(m+k)^w}-\frac{1}{(m+k+1)^w}\Big| = \frac{|-w|}{(x_m+m)^{w+1}} \le \frac{w}{(m+k)^{w+1}},$$

and

$$|\alpha_{N,k}| \le \frac{1}{N}+\frac{k}{N}\sum_{m=0}^{N-k-1} w\frac{(m+1)^v}{(k+m)^{v+2}} \le \\ \frac{1}{N}+\frac{kw}{N}\sum_{m=0}^{N-k-1}\frac{1}{(k+m)^2} \le \frac{1}{N}+\frac{kw}{N}\cdot\frac{2}{k} = \frac{1+2w}{N},$$

hence $\alpha_{N,k}\to 0$ as $N\to\infty$ and $\sum_k|\alpha_{N,k}|\le 1+2w$.

5. It remains to consider the first term in (10). Denote

$$A_n=\frac{1}{n}\sum_{k=0}^{n-1}\frac{X_k^*}{(k+1)^u},\quad B_n=\frac{1}{n}\sum_{k=0}^{n-1}\frac{Y_k^*}{(k+1)^v},$$

hence

$$X_n^*=(n+1)^u\big((n+1)A_{n+1}-nA_n\big),\quad Y_n^*=(n+1)^u\big((n+1)B_{n+1}-nB_n\big).$$

Therefore, this term equals

$$\tilde C_N=\frac{1}{N}\sum_{n=0}^{N-1}\sum_{k+m=n}\frac{X_k^*Y_m^*(\varphi)}{(k+m+1)^w} = \\ =\frac{1}{N}\sum_{k+m\le N-1}\frac{(k+1)^u(m+1)^v}{(k+m+1)^w}\times \\ \times((k+1)A_{k+1}-kA_k)((m+1)B_{m+1}-mB_m)(\varphi).$$

Rearranging the terms we obtain[2]

$$\tilde{C}_N = \frac{1}{N}\sum_{k,m\ge 1}\Bigg(\frac{k^u m^v}{(k+m-1)^w}[k+m\le N+1] - \frac{(k+1)^u m^v}{(k+m)^w}[k+m\le N] - \frac{k^u (m+1)^v}{(k+m)^w}[k+m\le N] + \frac{(k+1)^u (m+1)^v}{(k+m+1)^w}[k+m\le N-1]\Bigg) kmA_k B_m(\varphi).$$

This sum $\tilde{C}_N$ is decomposed as $\tilde{C}_N = \tilde{C}_N^{(1)} + \tilde{C}_N^{(2)}$, where

$$\tilde{C}_N^{(1)} = \frac{1}{N}\sum_{\substack{k,m\ge 1\\ k+m\le N}}\Bigg(\frac{k^u m^v}{(k+m-1)^w} - \frac{(k+1)^u m^v}{(k+m)^w} - \frac{k^u (m+1)^v}{(k+m)^w} + \frac{(k+1)^u (m+1)^v}{(k+m+1)^w}\Bigg) kmA_k B_m(\varphi), \quad (14a)$$

$$\tilde{C}_N^{(2)} = \frac{1}{N}\Bigg(\sum_{\substack{k,m\ge 1\\ k+m=N+1}}\frac{k^u m^v}{(k+m-1)^w} kmA_k B_m(\varphi) - \sum_{\substack{k,m\ge 1\\ k+m=N-1}}\frac{(k+1)^u (m+1)^v}{(k+m+1)^w} kmA_k B_m(\varphi)\Bigg). \quad (14b)$$

We'll prove that both $\tilde{C}_N^{(1)}$ and $\tilde{C}_N^{(2)}$ tend to zero in $L^p(X,\nu)$ for $\varphi \in L^p(X,\nu)$, $p \in [1,\infty)$, or almost everywhere for $\varphi \in L^\infty(X,\nu)$.

Let us start with $\tilde{C}_N^{(1)}$. Denote $g(x,y) = x^u y^v/(x+y-1)^w$, then the expression in round brackets in (14a) equals

$$\big(g(k,m) - g(k+1,m)\big) - \big(g(k,m+1) - g(k+1,m+1)\big) = -g'_y(k,\mu) + g'_y(k+1,\mu) = g''_{xy}(\varkappa,\mu),$$

where $\varkappa \in (k,k+1)$, $\mu \in (m,m+1)$. (We apply the mean value theorem first to $h_1(y) = g(k,y) - g(k+1,y)$ and then to $h_2(x) = g'_y(x,\mu)$.) One can

[2]Here we use *Iverson bracket notation*: for any statement $\mathcal{A}$

$$[\mathcal{A}] = \begin{cases} 1, & \mathcal{A} \text{ is true,} \\ 0, & \mathcal{A} \text{ is false.} \end{cases}$$

see that

$$g''_{xy}(\varkappa,\mu) = uv\frac{\varkappa^{u-1}\mu^{v-1}}{(\varkappa+\mu-1)^w} - vw\frac{\varkappa^u\mu^{v-1}}{(\varkappa+\mu-1)^{w+1}} - \\ - vw\frac{\varkappa^{u-1}\mu^v}{(\varkappa+\mu-1)^{w+1}} + w(w+1)\frac{\varkappa^u\mu^v}{(\varkappa+\mu-1)^{w+2}}.$$

As $\varkappa > k \geq 1$, $\mu > m \geq 1$, we have $\varkappa, \mu \leq \varkappa+\mu-1$, so each fraction[3] is not more than $1/(\varkappa+\mu-1)^3$, thus

$$|g''_{xy}(\varkappa,\mu)| \leq \frac{uv+vw+uw+w(w+1)}{(\varkappa+\mu-1)^3} \\ \leq \frac{uv+vw+uw+w(w+1)}{(k+m-1)^3} = \frac{\Theta_{u,v}}{(k+m-1)^3}. \quad (15)$$

Now we proceed to an estimation of $A_kB_m(\varphi)$.

Claim 3. 1. *Let* $M_n = \sup_{k+m=n}\|A_kB_m(\varphi)\|_p$ *for some* $\varphi \in L^p(X,\nu)$, $p \in [1,\infty)$. *Then* $M_n \to 0$.
2. *Let* $M_n(x) = \sup_{k+m=n}|(A_kB_m(\varphi))(x)|$ *for some* $\varphi \in L^\infty(X,\nu)$. *Then* $M_n(x) \to 0$ *for almost all* $x \in X$.

Proof. 1. Let $\|A_k\|_p \leq C$ for all k. Denote $\varphi_m = B_m(\varphi)$. Since $\|\varphi_m\|_p \to 0$, for a given $\varepsilon > 0$ one can choose m_0 such that $\|\varphi_m\|_p < \varepsilon$ for all $m \geq m_0$. Then $\|A_{n-m}\varphi_m\|_p \leq C\varepsilon$ for $m \geq m_0$, so

$$M_n \leq \max(\|A_n\varphi_0\|_p, \|A_{n-1}\varphi_1\|_p, \ldots, \|A_{n-m_0}\varphi_{m_0}\|_p, C\varepsilon).$$

Since $\|A_n\varphi_m\|_p \to 0$ as $n \to \infty$ for any fixed m, there are $N_0, \ldots, N_{m_0}$ such that $\|A_{n-m}(\varphi_m)\|_p \leq \varepsilon$ for $n > N_m$, $m = 0, \ldots, m_0$. Therefore if $n \geq N = \max(N_0, \ldots, N_{m_0})$, then $M_n \leq \max(\varepsilon, C\varepsilon)$.

2. Now let $\varphi \in L^\infty(X,\nu)$. Since $\varphi_m \xrightarrow{\text{a. e.}} 0$, if we denote

$$\psi_r(x) = \max_{m\geq r}|\varphi_m(x)|,$$

then $\psi_r \xrightarrow{\text{a. e.}} 0$. Note that $\psi_r(x)$ is nonnegative and nonincreasing sequence for any $x \in X$.

The operators A_k need not belong to $\mathcal{B}^+$. But if we denote

$$A_k^+(\theta) = \frac{1}{N}\sum_{n=0}^{n-1}\frac{X_n(\theta)}{(n+1)^u},$$

[3]We cannot use this estimate when exponent $u-1$ (resp., $v-1$) is negative, but then u (resp., v) equals zero, and the estimate (15) is simply $0 \leq 0$ for this term.

then $A_k^+ \in \mathcal{B}^+$, and $A_k^+(\theta) \xrightarrow{\text{a. e.}} X^0(\theta)$ for any $\theta \in L^\infty(X,\nu)$ (by definition of X^0, see (9)). It is also clear that $A_k = A_k^+ - X^0$.

Now define the following "exceptional sets":

$$E^1 = \{x \mid X^0(\psi_r)(x) \xrightarrow{r\to\infty} 0\},$$
$$E_m^2 = \{x \mid A_k(\varphi_m)(x) \xrightarrow{k\to\infty} 0\},$$
$$E_r^3 = \{x \mid A_k(\psi_r)(x) \xrightarrow{k\to\infty} 0\},$$
$$E_k^4 = \{x \mid A_k(\varphi_m)(x) \xrightarrow{m\to\infty} 0\},$$

Their measure is zero due to Proposition 2 (for E^1, E_k^4) and since $A_k(\theta) \xrightarrow{\text{a. e.}} 0$ (for E_m^2, E_r^3). Denote $E = E^1 \cup \left(\bigcup_m E_m^2\right) \cup \left(\bigcup_r E_r^3\right) \cup \left(\bigcup_k E_k^4\right)$ and prove that $M_n(x) \to 0$ for any $x \in X \setminus E$.

Indeed, take any $\varepsilon > 0$. Choose r_0 such that $X^0(\psi_{r_0})(x) \le \varepsilon$ (here we use that $x \notin E^1$). Note that since $X^0 \in \mathcal{B}^+$, $X^0(\psi_r) \ge 0$ for any r and $X^0(\psi_r) \le X^0(\psi_{r_0}) \le \varepsilon$ for any $r \ge r_0$.

Now choose k_0 such that $|A_k(\psi_{r_0})(x)| < \varepsilon$ for any $k > k_0$ ($x \notin E_{r_0}^3$). Then all possible k's are divided into three classes, each class is estimated separately.

Case 1. Let $k = 0, \dots, k_0$. Then, since $x \notin E_k^4$, there exists $N_k^{(1)}$ such that $|A_k(\varphi_{n-k})(x)| < \varepsilon$ for any $n > N_k^{(1)}$. Choose $N^{(1)} = \max(N_0^{(1)}, \dots, N_{k_0}^{(1)})$. Then for any $n > N^{(1)}$

$$M_n^{(1)}(x) = \max_{\substack{k+m=n \\ k\le k_0}} |A_k B_m(\varphi)(x)|) \le \varepsilon.$$

Case 2. Let $k = k_0 + 1, \dots, n - r_0$. Then

$$|A_k(\varphi_{n-k})(x)| \le |A_k^+(\varphi_{n-k})(x)| + |X^0(\varphi_{n-k})(x)| \le$$
$$\le A_k^+(\psi_{r_0})(x) + X^0(\psi_{r_0})(x) \le 2X^0(\psi_{r_0})(x) + |A_k(\psi_{r_0})(x)| \le 2\varepsilon + \varepsilon = 3\varepsilon.$$

Thus,

$$M_n^{(2)}(x) = \max_{\substack{k+m=n \\ k_0<k\le n-r_0}} |A_k B_m(\varphi)(x)| \le 3\varepsilon.$$

Case 3. Let $k = n - r_0 + 1, \dots, n$. Then, since $A_n(\varphi_m)(x) \xrightarrow{n\to\infty} 0$ for any $m = 0, \dots, r_0 - 1$ (we use that $x \notin E_m^2$), one can choose $N_m^{(3)}$ such that $|A_{n-m}(\varphi_m)(x)| < \varepsilon$ for any $n > N_m^{(3)}$, $m = 0, \dots, r_0 - 1$. Thus, for any $n > N^{(3)} = \max(N_0^{(3)}, \dots, N_{r_0-1}^{(3)})$

$$M_n^{(3)}(x) = \max_{\substack{k+m=n \\ k>n-r_0}} |A_k B_m(\varphi)(x)|) \le \varepsilon.$$

Putting these estimates together, we obtain that

$$M_n(x) = \max(M^{(1)}(x), M^{(2)}(x), M^{(3)}(x)) \le 3\varepsilon$$

for $n > N = \max(N^{(1)}, N^{(3)})$. □

Combining (15) with Claim 3, we have

$$\begin{aligned}\left\|\tilde{C}_N^{(1)}\right\|_p &\le \frac{1}{N} \sum_{\substack{k,m\ge 1\\ k+m\le N}} \frac{\Theta_{u,v} km}{(k+m-1)^3} M_{k+m} \\ &\le \frac{1}{N} \sum_{\substack{k,m\ge 1\\ k+m\le N}} \frac{\Theta_{u,v}}{k+m-1} M_{k+m} = \frac{\Theta_{u,v}}{N} \sum_{n=2}^{N} \sum_{\substack{k,m\ge 1\\ k+m=n}} \frac{M_n}{n-1} = \frac{\Theta_{u,v}}{N} \sum_{n=2}^{N} M_n\end{aligned}$$

and $\frac{1}{N}\sum_{n=2}^{N} M_n \xrightarrow{N\to\infty} 0$ as Cesàro averages of the sequence $\{M_n\}$, which converges to zero. For a. e.-convergence this proof also works after substitution of $|\tilde{C}_N^{(1)}(x)|$ for $\|\tilde{C}_N^{(1)}\|$ and of $M_n(x)$ for M_n.

Now we estimate $\tilde{C}_N^{(2)}$.

$$\begin{aligned}\tilde{C}_N^{(2)} &= \frac{1}{N}\Bigg(\sum_{m=1}^{N} \tfrac{(N+1-m)^u m^{v+1}}{N^w} (N+1-m) A_{N+1-m} B_m(\varphi) \\ &\qquad - \sum_{m=1}^{N-1} \tfrac{(N-m)^u (m+1)^v m}{N^w} (N-1-m) A_{N-1-m} B_m(\varphi) \Bigg) = \\ &= \frac{1}{N} \sum_{m=1}^{N-1} \Big[\tfrac{(N+1-m)^u m^{v+1}}{N^w} - \tfrac{(N-m)^u (m+1)^v m}{N^w} \Big] (N-1-m) A_{N-1-m} B_m(\varphi) + \\ &\quad + \frac{1}{N} \sum_{m=1}^{N-1} \frac{(N+1-m)^u m^{v+1}}{N^w} \big[(N+1-m) A_{N+1-m} B_m(\varphi) \\ &\qquad\qquad - (N-1-m) A_{N-1-m} B_m(\varphi) \big] + \\ &\quad + \frac{1}{N^{u+1}} A_N B_1(\varphi)\end{aligned} \tag{16}$$

Convergence of the last term is immediate. For the first term we apply Claim 3. Indeed, the expression in square brackets is of the form

$$m(f(m) - f(m+1)),$$

and the mean value theorem yields that (here $\mu \in [m, m+1]$)

$$\left|\frac{(N+1-m)^u m^{v+1}}{N^w} - \frac{(N-m)^u (m+1)^v m}{N^w}\right| =$$
$$= \frac{m}{N^w}\left|-u(N+1-\mu)^{u-1}\mu^v + v(N+1-\mu)^u \mu^{v-1}\right| \leq$$
$$\leq m\left(\frac{u(N+1-\mu)^{u-1}\mu^v}{N^w} + \frac{v(N+1-\mu)^u\mu^{v-1}}{N^w}\right) \leq m\frac{u+v}{N^2} \leq \frac{u+v}{N},$$

whence L^p-norm of the first term is bounded by

$$\frac{1}{N}\sum_{m=1}^{N-1}\frac{u+v}{N}(N-1-m)\|A_{N-1-m}B_m(\varphi)\| \leq$$
$$\leq \frac{1}{N}\sum_{m=1}^{N-1}(u+v)M_{N-1} \leq (u+v)M_{N-1}$$

so it tends to zero. The same argument works for a.e.-convergence, with L^p-norm being replaced by absolute value of value at x and M_{N-1} being replaced by $M_{N-1}(x)$.

As for the second term in (16), the coefficient $(N+1-m)^u m^{v+1}/N^w$ is bounded by 1, and the expression in square brackets equals

$$(N+1-m)A_{N+1-m}B_m(\varphi) - (N-1-m)A_{N-1-m}B_m(\varphi) =$$
$$= \left(\frac{X^*_{N-m}}{(N+1-m)^u} + \frac{X^*_{N+1-m}}{(N+2-m)^u}\right)(B_m(\varphi)). \quad (17)$$

Denote

$$W_k = \frac{X_k}{(k+1)^u} + \frac{X_{k+1}}{(k+2)^u}$$

Then the sequence

$$\frac{1}{N}\sum_{k=0}^{N-1} W_k(\varphi)$$

tends to $2X^0(\varphi) = W^0(\varphi)$ in $L^p(X,\nu)$ (for $\varphi \in L^p(X,\nu)$) or a. e. (for $\varphi \in L^\infty(X,\nu)$), hence (17) is equal to $(W_{N-m} - W^0)(B_m(\varphi))$.

Claim 4. 1. *If $\varphi \in L^p(X,\nu)$, then*

$$S_N = \frac{1}{N}\sum_{m=1}^{N}\left\|(W_{N-m} - W^0)(B_m(\varphi))\right\|_p$$

tends to zero.

2. *If* $\varphi \in L^\infty(X,\nu)$, *then*

$$S_N(x) = \frac{1}{N}\sum_{m=1}^{N}\big|(W_{N-m} - W^0)(B_m(\varphi))(x)\big|$$

tends to zero almost everywhere.

The second term in (16) is estimated by S_N (in L^p-norm) or by $S_N(x)$ (pointwise in absolute value). Hence it remains to prove this claim to complete the proof of Proposition 3.

Proof of Claim 4. 1. Let C be a constant such that $\|W_k - W^0\|_p \le C$ for all k. Then

$$S_N \le \frac{1}{N}\sum_{m=1}^{N}\|W_{N-m} - W^0\|_p \cdot \|B_m(\varphi)\|_p \le \frac{C}{N}\sum_{m=1}^{N}\|B_m(\varphi)\|_p$$

the latter is the Cesàro average (multiplied by C) of the sequence $\|B_n(\varphi)\|_p$, which tends to zero.

2. As in Claim 3, denote $\varphi_m = B_m(\varphi)$, $\psi_r(x) = \max_{m\le r}|\varphi_m(x)|$. Let constants C and R be such that $\|W_k - W^0\|_\infty < C$ for all k and $\|\varphi_m\|_\infty \le R$ for all m. Define the following "exceptional sets"

$$E^1 = \{x \mid W^0(\psi_r)(x) \overset{r\to\infty}{\not\longrightarrow} 0\},$$
$$E^2_r = \Big\{x \;\Big|\; \frac{1}{N}\sum_{k=0}^{N-1}(W_k - W^0)(\psi_r)(x) \overset{k\to\infty}{\not\longrightarrow} 0\Big\}.$$

and let $E = E^1 \cup \Big(\bigcup_r E^2_r\Big)$.

Fix any $x \in X \setminus E$ and take any $\varepsilon > 0$. Choose r_0 such that $W^0(\psi_{r_0}) < \varepsilon$. Then

$$S_N(x) = \frac{1}{N}\left(\sum_{m=1}^{r_0-1} + \sum_{m=r_0}^{N}\right)|(W_{N-m} - W^0)(\varphi_m)(x)| \le$$
$$\le \frac{CR(r_0-1)}{N} + \frac{1}{N}\sum_{m=r_0}^{N}|W_{N-m}(\varphi_m)(x)| + |W^0(\varphi_m)(x)| \le$$
$$\le \frac{CR(r_0-1)}{N} + \frac{1}{N}\sum_{m=r_0}^{N}\big(W_{N-m}(\psi_{r_0})(x) + W^0(\psi_{r_0})(x)\big) \le$$

$$\leq \frac{CR(r_0-1)}{N} + \frac{1}{N}\sum_{m=1}^{N}\big(W_{N-m}(\psi_{r_0})(x) + W^0(\psi_{r_0})(x)\big) \leq$$

$$\leq \frac{CR(r_0-1)}{N} + 2W^0(\psi_{r_0})(x) + \frac{1}{N}\sum_{k=0}^{N-1}(W_k - W^0)(\psi_{r_0})(x).$$

Here the first term tends to zero as $N \to \infty$, the second one is less than 2ε, and the last one also tends to zero (since $x \notin E^2_{r_0}$). Hence for sufficiently large N one has $S_N(x) \leq 3\varepsilon$. □

Therefore Proposition 3 is completely proven. This completes the proofs of Lemma 1, Theorem 2, and Theorem 1. □

References

[1] V. I. Arnold and A. L. Krylov, Equidistribution of points on a sphere and ergodic properties of solutions of ordinary differential equations in a complex domain, *Dokl. Akad. Nauk SSSR* **148** (1963), 9–12.

[2] L. Bowen, Invariant measures on the space of horofunctions of a word hyperbolic group, *Ergodic Theory Dynam. Systems* **30** (2010), no. 1, 97–129.

[3] L. Bowen, A. Nevo, Geometric covering arguments and ergodic theorems for free groups, arXiv:0912.4953v2 [math.DS].

[4] A.I. Bufetov, Ergodic theorems for actions of several mappings, (Russian) Uspekhi Mat. Nauk, **54** (1999), no. 4 (328), 159–160, translation in Russian Math. Surveys, **54** (1999), no. 4, 835–836.

[5] A.I. Bufetov, Operator ergodic theorems for actions of free semigroups and groups, *Funct. Anal. Appl.* **34** (2000), 239–251.

[6] A.I. Bufetov, Markov averaging and ergodic theorems for several operators, in *Topology*, *Ergodic Theory*, *and Algebraic Geometry*, *AMS Transl.* **202** (2001), 39–50.

[7] A.I. Bufetov, Convergence of spherical averages for actions of free groups. *Ann. Math.*, **155** (2002), 929–944.

[8] A.I. Bufetov, C. Series, A pointwise ergodic theorem for Fuchsian groups, arXiv:1010.3362v1 [math.DS].

[9] J. Cannon, The combinatorial structure of cocompact discrete hyperbolic groups. *Geom. Dedicata*, **16** (1984), no. 2, 123–148.

[10] K. Fujiwara and A. Nevo, Maximal and pointwise ergodic theorems for word-hyperbolic groups, *Ergodic Theory Dynam. Systems* **18** (1998), 843–858.

[11] Sur les groupes hyperboliques d'après Mikhael Gromov. Papers from the Swiss Seminar on Hyperbolic Groups held in Bern, 1988. Edited by É. Ghys and P. de la Harpe. Progress in Mathematics, 83. Birkhäuser Boston, Inc., Boston, MA, 1990.

[12] R. I. Grigorchuk, Pointwise ergodic theorems for actions of free groups, *Proc. Tambov Workshop in the Theory of Functions*, 1986.

[13] R. I. Grigorchuk, Ergodic theorems for actions of free semigroups and groups, *Math. Notes*, **65** (1999), 654–657.

[14] R. I. Grigorchuk. An ergodic theorem for actions of a free semigroup. (Russian) *Tr. Mat. Inst. Steklova* **231** (2000), Din. Sist., Avtom. i Beskon. Gruppy, 119–133; translation in *Proc. Steklov Inst. Math.* 2000, no. 4 (231), 113–127.

[15] M. Gromov, Hyperbolic groups, in *Essays in Group Theory*, *MSRI Publ.* **8** (1987), 75–263, Springer-Verlag, New York.

[16] Y. Guivarc'h, Généralisation d'un théorème de von Neumann, *C. R. Acad. Sci. Paris Sér. A–B* **268** (1969), 1020–1023.

[17] A. Nevo, Harmonic analysis and pointwise ergodic theorems for noncommuting transformations, *J. Amer. Math. Soc.* **7** (1994), no. 4, 875–902.

[18] A. Nevo, Pointwise ergodic theorems for actions of groups, in *Handbook of dynamical systems, Vol. 1B*, 871–982, Elsevier B. V., Amsterdam, 2006.

[19] A. Nevo and E. M. Stein, A generalization of Birkhoff's pointwise ergodic theorem, *Acta Math.* **173** (1994), 135–154.

[20] D. Ornstein, On the pointwise behavior of iterates of a self-adjoint operator, *J. Math. Mech.* **18** (1968/1969) 473–477.

[21] V. I. Oseledets, Markov chains, skew-products, and ergodic theorems for general dynamical systems, *Th. Prob. App.* **10** (1965), 551–557.

[22] G.-C. Rota, An "Alternierende Verfahren" for general positive operators, *Bull. A. M. S.* **68** (1962), 95–102.

[23] C. Series, Geometrical Markov coding of geodesics on surfaces of constant negative curvature, *Ergodic Theory Dynam. Systems*, **6** (1986), 601–625.

[24] E. M. Stein, On the maximal ergodic theorem, *Proc. Nat. Acad. Sci. USA* **47** (1961), 1894–1897.