A.E. MULTIPLE RECURRENCE FOR WEAKLY MIXING COMMUTING ACTIONS

I. ASSANI

ABSTRACT. Let \((X, \mathcal{A}, \mu)\) be a probability measure space and let \(T_i, 1 \leq i \leq H\), be commuting invertible measure preserving transformations on this measure space. We prove the pointwise convergence of the averages

\[
\frac{1}{N} \sum_{n=1}^{N} f_1(T_1^n x) f_2(T_2^n x) \cdots f_H(T_H^n x)
\]

for every function \(f_i \in L^\infty(\mu)\) when each transformation is weakly mixing. This result is a consequence of a more general result for commuting measure preserving homeomorphisms on compact metric spaces.

1. INTRODUCTION

Let \((X, \mathcal{A}, \mu)\) be a probability measure space. Let \(T_i, 1 \leq i \leq H\), be invertible measure preserving transformations on this measure space. For \(f_i \in L^\infty(\mu), 1 \leq i \leq H\), we look at the well known open problem of the pointwise convergence of the nonconventional ergodic averages

\[
\frac{1}{N} \sum_{n=1}^{N} \prod_{i=1}^{H} f_i(T_i^n x).
\]

The case \(H = 1\) corresponds to the classical ergodic averages for which the pointwise convergence is known by Birkhoff ergodic theorem. In [10], H. Furstenberg asked if for a measure preserving transformation \(T\) on \((X, \mathcal{A}, \mu)\), bounded functions \(f, g\) and \(m\) a positive

Department of Mathematics, UNC Chapel Hill, NC 27599, assani@email.unc.edu.
integer \( m \neq 1 \) the averages \( \frac{1}{N} \sum_{n=1}^{N} f(T^n x)g(T^{mn} x) \) converge a.e. J. Bourgain [7] proved that this was indeed the case. The natural question then became; for any positive integer \( H \), and bounded functions \( f_1, f_2, \ldots, f_H \) do we have the pointwise convergence of the averages \( \frac{1}{N} \sum_{n=1}^{N} f_1(T^n x) \cdots f_H(T^{Hn} x) \)? Partial results were obtained in [9] for K-systems and in [1] for weakly mixing systems \( T \) for which the restriction to the Pinsker algebra had singular spectrum. The arguments in this last paper relied in part on J. Bourgain result [7]. We provided a simplification of Bourgain’s proof for a class of ergodic dynamical systems in [3] and gave some consequences of this simplification in [2]. Beyond these results little progress has been made in this direction.

For the norm convergence the situation is pretty much settled. In their initial work, J.P. Conze and E. Lesigne [8] proved the norm convergence of the averages \( \frac{1}{N} \sum_{n=1}^{N} f \circ T_1^n f_2 \circ T_2^n \) for commuting measure preserving transformations \( T_1 \) and \( T_2 \) on the same probability measure space. In [11], B. Host and B. Kra and independently T. Ziegler [13], proved the norm convergence of the averages \( \frac{1}{N} \sum_{n=1}^{N} f_1 \circ T_1^n f_2 \circ T_2^n \cdots f_H \circ T_H^n \). In [12], T. Tao extended their result by proving that for commuting measure preserving transformations \( T_i \), \( 1 \leq i \leq H \) on the same probability measure space the averages \( \frac{1}{N} \sum_{n=1}^{N} f_1 \circ T_1^n f_2 \circ T_2^n \cdots f_H \circ T_H^n \) converge in norm for every bounded function \( f_i \), \( 1 \leq i \leq H \). See also [5] for another proof. M. Walsh [15] extended Tao’s result to the case where the maps \( T_i \), \( 1 \leq i \leq H \) generate a nilpotent group. In view of the negative result provided by V. Bergelson and A. Leibman [6] for solvable groups this is the best possible case for convergence in norm.

Our goal is to present a new approach to the pointwise convergence of these non-conventional ergodic averages. This approach will enable us to prove the following results.
Theorem 1. Let \((X, A, \mu)\) be a probability measure space and let \(H\) be a positive integer. Let \(T_i, 1 \leq i \leq H\) be \(H\) commuting weakly mixing transformations on \((X, A, \mu)\). For every bounded functions \(f_i, 1 \leq i \leq H\) the averages \(\frac{1}{N} \sum_{n=1}^{N} \prod_{i=1}^{H} f_i(T_i^n x)\) converge a.e. to \(\prod_{i=1}^{H} \int f_i d\mu\).

Theorem 1 is a consequence of the following theorem.

Theorem 2. Let \(X\) be a compact metrizable space and let \(A\) its Borelian \(\sigma\)-algebra. Let \(T_i, 1 \leq i \leq H\) be commuting homeomorphims on \(X\) each preserving the same Borel measure \(\mu\). We denote by \(\phi : X^H \to X^H\) the homeomorphism given by the equation \(\phi(z) = (T_1 z_1, T_2 z_2, ..., T_H z_H)\). We denote by \(\mu_\Delta\) the diagonal measure on \((X^H, A^H)\) and by \(\nu\) the probability measure defined on \((X^H, A^H)\) by

\[
\nu(A) = \frac{1}{3} \sum_{n=-\infty}^{\infty} \frac{1}{2|n|} \mu_\Delta(\phi^{-n}(A)).
\]

Assume that

\[
(1) \quad \nu\{z \in X^H : \{\phi^n z ; n \in \mathbb{Z}\} \text{ is dense in } X^H\} = 1.
\]

Then for every function \(f_i \in L^\infty(\mu), 1 \leq i \leq H\), the averages

\[
\frac{1}{N} \sum_{n=1}^{N} f_1(T_1^n x)f_2(T_2^n x)\cdots f_H(T_H^n x)
\]

converge \(\mu\) a.e.

We first give a proof of Theorem 2. Then we will show how to derive from it Theorem 1.
2. Proof of Theorem 2 for H=2

2.1. Preliminaries- Notations. We assume that $X$ is a compact metric space, $\mathcal{A}$ is the set of Borelian subsets of $X$ and $T_1$ and $T_2$ are commuting homeomorphisms on $X$ preserving the same measure $\mu$ on $\mathcal{A}^2$.

We consider now the diagonal measure $\mu_\Delta$ as the unique measure defined on $(X \times X, \mathcal{A}^2)$ by the equation

$$\mu_\Delta(A) = \int 1_A(x, y) d\mu_\Delta = \int 1_A(x, x) d\mu$$

for any measurable subset $A$ of $X \times X$. In particular we have for each measurable function $f$ and $g$,

$$\int f(x)g(x) d\mu = \int f(x)g(y) d\mu_\Delta.$$  \hspace{1cm} (2)

We denote by $\mathcal{L}$ the algebra of finite linear combinations of product functions $f_i \otimes g_i$ defined on $X \times X$ where $f_i$ and $g_i$ are bounded and measurable on $X$. The norm convergence result for two commuting measure preserving transformations gives us an operator $R$ defined on $\mathcal{L}$ such that for all function $F \in \mathcal{L}$ and for all measurable subset $W \in \mathcal{A}^2$

$$\lim_{L} \int 1_W(x, x) \frac{1}{L} \sum_{l=0}^{L-1} F(T_1^n x, T_2^n y) d\mu = \lim_{L} \int 1_W(x, y) \frac{1}{L} \sum_{l=0}^{L-1} F(T_1^n x, T_2^n y) d\mu_\Delta$$

$$= \int 1_W(x, y) R(F)(x, y) d\mu_\Delta$$ \hspace{1cm} (3)

More can be said about the limit function $R$.

**Lemma 1.** For any two invertible commuting measure preserving transformations, $T_1$ and $T_2$ on the probability measure space $(X, \mathcal{A}, \mu)$ and any two $L^\infty(\mu)$ functions, $f_1$ and $f_2$, let us denote by $R(f_1 \otimes f_2)$ the norm limit of the averages

$$\frac{1}{N} \sum_{n=0}^{N-1} f_1 \circ T_1^n f_2 \circ T_2^n.$$
If $\mathcal{I}$ is the $\sigma$-algebra of the invariant sets for the measure transformation $T_1 \circ T_2^{-1}$ we have

$$\lim_N \int \left( \frac{1}{N} \sum_{n=0}^{N-1} f_1(T_1^n x) f_2(T_2^n x) \right) d\mu(x) = \int \mathbb{E}[f_1|\mathcal{I}] \mathbb{E}[f_2|\mathcal{I}] d\mu.$$  

So there exists a measure $\omega$ on $(X \times X, \mathcal{A}^2)$ defined by

$$(4) \quad \omega(f_1 \otimes f_2) = \int \mathbb{E}[f_1|\mathcal{I}] \mathbb{E}[f_2|\mathcal{I}] d\mu = \int R(f_1 \otimes f_2)(x, y) d\mu_{\Delta}.$$  

In particular if $T_1 \circ T_2^{-1}$ or $T_2 \circ T_1^{-1}$ is ergodic then $\omega = \mu \otimes \mu$.

**Proof.** This follows from the commuting property of the transformations $T_1$ and $T_2$ and the mean ergodic theorem as the limit is equal to

$$\lim_N \int f_1(x) \frac{1}{N} \sum_{n=0}^{N-1} f_2(T_2 \circ T_1^{-1})^n(x) d\mu = \int f_1 \mathbb{E}(f_2|\mathcal{I}) d\mu.$$  

where $\mathcal{I}$ is the $\sigma$-algebra of invariant subsets of $\mathcal{A}$ for the transformation $T_2 \circ T_1^{-1}$. The equation

$$\omega(f_1 \otimes f_2) = \int \mathbb{E}[f_1|\mathcal{I}] \mathbb{E}[f_2|\mathcal{I}] d\mu$$  

easily defines a measure on $(X \times X, \mathcal{A}^2)$. The remaining part of the lemma follows directly from the equation $(4)$. Finally if $T_1 \circ T_2^{-1}$ is ergodic then the conditional expectations $\mathbb{E}[f_1|\mathcal{I}]$ and $\mathbb{E}[f_2|\mathcal{I}]$ are respectively the integral of $f_1$ and $f_2$ with respect to the measure $\mu$. The equality $\omega = \mu \otimes \mu$ follows easily from this last remark.  

In the setting we defined above lemma 1 applies to continuous function $F$ defined on $X^2$. We have the following inequalities

$$(5) \quad \int F(x, y) d\omega = \lim_L \int \frac{1}{L} \sum_{l=0}^{L-1} F(T_1^l x, T_2^l x) d\mu.$$
Furthermore for any open subset $O$ of $X^2$ we have

$$
\int 1_O(x,y)\,d\omega \leq \liminf L \sum_{l=0}^{L-1} 1_O(T_1^lx, T_2^l)\,d\mu.
$$

The last equation follows from the fact that the characteristic function of an open set is an increasing limit of continuous functions. From now on we fix $\varepsilon$ a positive real number and $f_1$ and $f_2$ two continuous real valued functions on $X$. We denote by $F$ the function defined on $X^2$ as $f_1 \otimes f_2$ and by $M_L(F)(z)$ the averages $\frac{1}{L} \sum_{n=0}^{L-1} F(T_1^n x, T_2^n y) = \frac{1}{L} \sum_{n=0}^{L-1} F(\phi^n(z))$ where $z = (x,y)$ and $\phi^n(z) = (T_1^n x, T_2^n y)$. Our main goal is to transfer the problem of the pointwise convergence of the averages $\frac{1}{L} \sum_{n=0}^{L-1} f_1(T_1^n x)f_2(T_2^n x)$ with respect to $\mu$ to the one on $X^2$ for the averages $M_L(F)(z)$ with respect to a probability measure on $(X^2, \mathcal{A}^2)$ for which $\phi$ is nonsingular. To this end we start with the diagonal measure $\mu_\Delta$ and introduce the measure $\nu : \mathcal{A}^2 \to [0,1]$ where $\nu(A) = \frac{1}{3} \sum_{n=-\infty}^{\infty} \frac{1}{2|n|} \mu_\Delta(\phi^{-n}(A))$. It is simple to check that if $\nu(A) = 0$ then $\nu(\phi^{-1}(A)) = 0$ property which makes $\phi$ nonsingular with respect to $\nu$.

Therefore we put ourselves in the setting of the nonsingular transformation $\phi$ and the measure space $(X^2, \mathcal{A}^2, \nu)$. We note that a null set for $\nu$ is also a null set for $\mu_\Delta$ and so establishing the pointwise convergence with respect to $\nu$ will automatically imply the same result with respect to $\mu_\Delta$.

The second main new idea in our approach is the quantity $\inf_{k \in \mathbb{Z}} M^*(F)(\phi^k z)$.

**Definition 1.** Given $F$ a continuous function on $X^2$ and $M^*(F)(z) = \sup_N \frac{1}{N} \sum_{n=1}^{N} F(\phi^n z)$ we denote by $V(F) : X^2 \to \mathbb{R}$ the quantity defined pointwise by the equation $V(F)(z) = \inf_{k \in \mathbb{Z}} M^*(F)(\phi^k z)$. 
Lemma 2. Let \((Y, \mathcal{G}, \rho, S)\) be a measure preserving system and \(F\) a \(L^1(\rho)\) function. If \(\mathcal{J}\) denotes the \(\sigma\) algebra of the invariant subsets of \(\mathcal{G}\) then for a.e. \(\rho\) we have \(V(F)(y) = \inf_{k \in \mathbb{Z}} M^*F(\phi^k y) = \mathbb{E}(F|\mathcal{J})(y)\) where \(\mathbb{E}(F|\mathcal{J})(y)\) denotes the conditional expectation of \(F\) with respect to \(\mathcal{J}\) and \(M^*F(y) = \sup_N \frac{1}{N} \sum_{n=1}^N F(S^ny)\).

Proof. Without loss of generality we can assume that the map \(S\) is ergodic (by taking an ergodic decomposition of \(\rho\)). We have the inequality

\[
V(F)(y) \geq \limsup_N \frac{1}{N} \sum_{n=1}^N F(S^ny) = \int Fd\rho
\]

by the pointwise ergodic theorem.

The invariance of \(V(F)\) added to the ergodicity of \(S\) with respect to \(\rho\) gives us a constant \(\alpha\) such that \(\alpha = V(F)(y)\) for \(\rho\) a.e. We just need to show that \(\alpha \leq \int Fd\rho\). To this end we use the maximal inequality which says that

\[
\lambda \rho\{y : M^*F(y) > \lambda\} \leq \int_{\{y: M^*F(y) > \lambda\}} Fd\rho
\]

for each \(\lambda \in \mathbb{R}\). This inequality follows from the inequality \(\int_{\{M^*F(y) > \lambda\}} Fd\rho \geq 0\) and applying it to \(F - \lambda\). By taking \(\lambda = \alpha\) we obtain the inequality \(\alpha \leq \int Fd\rho\). Combining this with (7) we get the equality \(V(F)(y) = \int Fd\rho\). \(\square\)

2.2. End of the proof of Theorem 2 for \(H=2\). We consider \(F = f_1 \otimes f_2\) where \(f_i\) for \(1 \leq i \leq 2\) is a continuous function on \(X\). The measure \(\omega\) in Lemma 1 is invariant with respect to \(T_1 \times T_2\), which implies that we have

\[
V(F)(y) = \mathbb{E}[F|\mathcal{W}(y)]
\]
for \( \omega \) a.e. \( y \in X^2 \), where \( \mathcal{W} \) denotes the \( \sigma \)-algebra of functions invariant with respect to \( \phi \).

We fix \( \varepsilon > 0 \). We denote by \( A_N \) the set \( \{ z : M^*(F)(z) \leq \sup_{1 \leq n \leq N} M_n(F)(z) + \varepsilon \} \). Since the function \( M^*(F) \) is lower semicontinuous each set \( A_N \) is closed. Furthermore we have \( \bigcup_{N=1}^\infty A_N = X^2 \). By Baire’s theorem one of these sets \( A_N \) contains a non empty open ball \( B(\gamma, \delta) \). We have the inequality

\[
1_{B(\gamma, \delta)} M^*(F) \leq 1_{B(\gamma, \delta)} \left( \sup_{1 \leq n \leq N} M_n(F) + \varepsilon \right)
\]

(10)

Applying \( \phi \) to both sides we get

\[
1_{B(\gamma, \delta)} \circ \phi M^*(F) \circ \phi \leq 1_{B(\gamma, \delta)} \circ \phi \left( \sup_{1 \leq n \leq N} M_n(F) \circ \phi + \varepsilon \right).
\]

(11)

By combining both inequalities if we denote by \( M^*_N(F) \) the quantity \( \sup_{1 \leq n \leq N} M_n(F) \) we conclude that

\[
1_{B(\gamma, \delta) \cup \phi^{-1}(B(\gamma, \delta))} \min(M^*(F), M^*(F) \circ \phi) \leq 1_{B(\gamma, \delta) \cup \phi^{-1}(B(\gamma, \delta))} \left( \min(M^*_N(F), M^*_N(F) \circ \phi) + \varepsilon \right).
\]

By induction on \( k \in \mathbb{Z} \) we derive the inequality

\[
1_{\bigcup_{k \in \mathbb{Z}} \phi^k (B(\gamma, \delta))} V(F) \leq 1_{\bigcup_{k \in \mathbb{Z}} \phi^k (B(\gamma, \delta))} \left( \inf_{k \in \mathbb{Z}} M^*_N(F) \circ \phi^k + \varepsilon \right).
\]

(12)

If \( (O_n)_n \) is a countable basis for the topology on \( X^2 \), the set of points \( z \in X^2 \) having a dense orbit is equal to \( D(\phi) = \bigcap_n \bigcup_{m=-\infty}^\infty \phi^m(O_n) \). Note that \( \nu(D(\phi)) = 1 \). Noticing that each of the points \( z \) having a dense orbit must be contained in \( \mathcal{O} = \bigcup_{k \in \mathbb{Z}} \phi^k (B(\gamma, \delta)) \) we conclude that \( \nu(\mathcal{O}) = 1 \). Going back to (12) and picking a point \( z \in D(\phi) \) we have \( V(F)(z) \leq \inf_{k \in \mathbb{Z}} M^*_N(F)(\phi^k z) + \varepsilon \). If we take any point \( y \in X^2 \) the continuity of the function \( M^*_N(F) \) implies that \( V(F)(z) \leq M^*_N(F)(y) + \varepsilon \), for every \( y \in X^2 \). The point \( y \)
being arbitrary we also have $V(F)(z) \leq M^s_k(F)(\phi^k y) + \epsilon$ for each $k \in \mathbb{Z}$. As a consequence for every $z \in D(\phi)$ and every $y \in X^2$ we get the inequality

$$V(F)(z) \leq V(F)(y) + \epsilon$$

To get rid of $\epsilon$ we just need to let $\epsilon$ ranges through a countable sequence $\epsilon_r$ converging to zero. Each $\epsilon_r$ corresponds to an open set $O_r$ with $\nu$ measure 1 which contains $D(\phi)$. By taking $z$ in $D(\phi)$ and letting $\epsilon_r$ go to 0 we derive the inequality

$$V(F)(z) \leq V(F)(y)$$

for all $y \in X^2$ and $z \in D(\phi)$. Using (9) we get for every $z \in D(\phi)$ and for $\omega$ a.e $y \in X^2$, (the null set here is independent of $z$), $V(F)(z) \leq E[F|W](y)$. Noticing that the invariance of $\limsup M_n(F)$ with respect to $\phi$ implies the inequality

$$\limsup_n M_n(F)(z) \leq V(F)(z) \leq E[F|W](y).$$

So we can integrate with respect to $\nu$ and $\omega$. This gives us the inequality

$$\int \limsup_n M_n(F)(z)d\nu(z) \leq \int F(y)d\omega.$$

Finally observing that the diagonal measure $\mu_\Delta$ and $\nu$ take the same value on invariant sets we obtain

$$\int \limsup_n M_n(F)(x)d\mu \leq \int F(y)d\omega.$$

It remains to prove the reverse inequality. It is derived from the norm convergence of the averages $M_N(F)$ with respect to $\mu$ to $R(F)$ We can find a subsequence $M_{N_k}(F)$ converging
μ a.e. to \( R(F) \). We conclude that \( \int Fd\omega \leq \int \limsup_n M_n(F)(x)d\mu \). So we have shown that

\[
\int \limsup_n M_n(F)(x)d\mu = \int Fd\omega.
\]

To get the pointwise convergence of the averages it is enough to apply (16) to the function \(-F\). We get \( -\int \liminf_n M_n(F)(x)d\mu = -\int Fd\omega \) and this combined with (18) proves that \( \int \limsup_n M_n(F)(x)d\mu = \int \liminf_n M_n(F)(x)d\mu \). This concludes the proof of Theorem 2 for \( H = 2 \).

2.3. **Proof of Theorem 2 for** \( H \geq 2 \). The reader can observe that the arguments developed for the case \( H = 2 \) apply equally well to the general case.

We only sketch them. We start with the new Lemma 1 that we state in the general context of invertible measure preserving transformations as we did for the case \( H = 2 \).

**Lemma 3.** Let \( T_i, 1 \leq i \leq H \) be commuting invertible measure preserving transformations on the probability measure space \((X, \mathcal{A}, \mu)\). Let \( f_i, 1 \leq i \leq H \) be functions in \( L^\infty(\mu) \). We denote by \( R(f_1, f_2, \cdots, f_H) \) the limit in norm of the averages \( \frac{1}{N} \sum_{n=1}^{N} f_1 \circ T_1^n f_2 \circ T_2^n \cdots f_H \circ T_H^n \). Then there exists a measure \( \omega \) defined on \((X^H, \mathcal{A}^H)\) and invariant under \( T_1 \times T_2 \cdots \times T_H \) such that \( \int R(f_1, f_2, \cdots, f_H)d\mu = \int f_1 \otimes f_2 \cdots \otimes f_H d\omega \). Furthermore if each \( T_i \) is weakly mixing then the limit in norm, \( R(f_1, f_2, \cdots, f_H) \) is equal to \( \prod_{i=1}^{H} \int f_i d\mu \).

Now we assume that the transformations \( T_i \) are commuting homeomorphisms defined on the compact metric space \( X \). We assume that preserve the same probability measure on \((X, \mathcal{A})\) where denotes the \( \sigma \)-field of Borelian subsets of \( X \).

**Step 1**

We define the homeomorphism \( \phi : X^H \to X^H \) which sends \( z = (z_1, z_2, \cdots, z_H) \) to \( \phi(z) = \)
We also define the measure \( \nu : A \in \mathcal{A}^H \rightarrow [0,1] \) such that
\[
\nu(A) = \frac{1}{3} \sum_{n=-\infty}^{\infty} \frac{1}{2^n} \mu_\Delta(\phi^n(A))
\]

**Step 2**

We take \( H \) continuous functions \( f_i \) on \( X \) and denote by \( F \) the function on \( X^H \) equal to \( f_1 \otimes f_2 \otimes \cdots \otimes f_H \). By the same argument as the one given at the end of the case \( H = 2 \), for \( \varepsilon > 0 \), by Baire’s theorem, we can find an open set \( O \) invariant under \( \phi \) which contains the set \( D(\phi) \) of points \( z \in X^H \) having dense orbit in \( X^H \).

**Step 3**

This last property leads to the inequality
\[
V(z) \leq V(y)
\]
for each \( z \in D(\phi) \) and every \( y \in X^H \), where \( V(z) = \inf_{k \in \mathbb{Z}} M^*(F)(\phi^k z) \).

**Step 4**

Integrating both sides with respect to \( \mu_\Delta \) we get
\[
\int \lim \sup_n M_n(F) d\mu = \int \lim \sup_n M_n(F) d\nu \leq \int F d\omega.
\]

**Step 5**

Finally by changing \( F \) to \(-F\) in the last equation we obtain the equation
\[
\int \lim \inf_n M_n(F) d\mu = \int F d\omega = \int \lim \sup_n M_n(F) d\mu \text{ and so } M_n(F)(x) \text{ converges } \mu \text{ a.e.}
\]

### 3. Proof of Theorem 💫

In this section we want to show how to derive Theorem 💫 from Theorem 2. We just do it in the case \( H = 2 \), the general case being similar. We start then with two weakly mixing commuting invertible transformations \( T_1, T_2 \) on the probability measure space \((X, \mathcal{A}, \mu)\). Both preserve the measure \( \mu \). To avoid a trivial case we assume that \( T_1 \neq T_2 \). We can observe that the action of the group generated by \( T_1 \) and \( T_2 \) is ergodic as one member of
this group is ergodic. Therefore we can use B. Weiss’ generalization of Jewett and Krieger isomorphism theorem for $\mathbb{Z}^2$-actions. Under this isomorphism we get the following properties.

1. First the new maps that we denote again by $T_1$ and $T_2$ are commuting homeomorphisms on a compact metrizable space. These homeomorphisms are weakly mixing preserving a new measure that we also denote by $\mu$.

2. Secondly this new measure $\mu$ is such that for any not empty open set $V \subset X$ we have $\mu(V) > 0$.

3. Thirdly the measure $\omega$ given by Lemma 1 is the product measure $\mu \otimes \mu$.

In order to apply Theorem 2 we need to show that the assumption (1) is satisfied. Because the set of points $D(\phi)$ in $X^2$ having a dense orbit is invariant, its $\nu$ measure is the same as its $\mu_\Delta$ measure. We recall that if $O_n$ is a countable basis of open sets for the topology of $X^2$ the set $D(\phi)$ is equal to $\cap_m \cup_{n=-\infty}^{\infty} \phi^n(O_m)$. It is enough to show that each open set $P_m = \cup_{n=-\infty}^{\infty} \phi^n(O_m)$ has $\mu_\Delta$ measure equal to one. By the mean ergodic theorem (which is much simpler to prove in the weakly mixing case) [12], we can extract after a diagonal process a subsequence of integers $N_k$ such that for any continuous functions $F$ on $X^2$ we have $\lim_k \frac{1}{N_k} \sum_{n=1}^{N_k} F(T_1^n x, T_2^n x) = \int F d\mu \otimes \mu$. off a single null set for $\mu$. As the characteristic function of an open set is an increasing limit of continuous functions we derive that $1_{P_m}(x, x) = \limsup_k \frac{1}{N_k} 1_{P_m}(T_1^n x, T_2^n x) \geq \mu \otimes \mu(P_m) > 0$ because in this setting the measure of a non empty open set is positive. The map $T_1 \times T_2$ being ergodic with respect to $\mu \otimes \mu$ we have $\mu \otimes \mu(P_m) = 1$. We conclude that $\mu_\Delta(P_m) = 1$ and this implies that $\mu_\Delta(\cap_m P_m) = 1$. 
Remarks

The proof of Theorem 1 can be used to show that in the general setting of commuting measure preserving homeomorphisms, we have the pointwise convergence of the averages $M_n(F)$ on the $D(\phi)$ part of the space $X^H$. This is done more explicitly in [1].

References

[1] I. Assani: “Multiple recurrence and almost sure convergence for weakly mixing dynamical systems” Israel J. Math. 103 (1998), 111-124.

[2] I. Assani: “Pointwise convergence of nonconventional averages,” Colloq. Math. 102 (2005), no. 2, 245-262.

[3] I. Assani: “Wiener Wintner Dynamical Systems” Erg. Th. and Dyn. Syst. 23, 1637-1654, 2003.

[4] I. Assani: “Pointwise multiple recurrence for commuting measure preserving transformations. preprint 2013.

[5] T. Austin: “On the norm convergence of non-conventional ergodic averages,” Ergod. Th. and Dynam. Sys. 30 (2010), 321-338.

[6] V. Bergelson, A. Leibman: “A nilpotent Roth Theorem”, Invent. Math. 147 (2) (2002), 429-470.

[7] J. Bourgain: “Double recurrence and almost sure convergence,” J. Reine Angew. Math. 404 (1990), 140 161.

[8] J-P. Conze, E. Lesigne: “Theoremes ergodiques pour des mesures diagonales”, Bull. Soc. Math. France 112 (1984), no. 2, 143175.

[9] J-M. Derrien, E. Lesigne: “Un theoreme ergodique polynomial ponctuel pour les endomorphismes exacts et les K-systemes,” Ann. Inst. H. Poincare Probab. Statist. 32 (1996), no. 6, 765-778.

[10] H. Furstenberg: “Recurrence in ergodic theory and combinatorial number theory,” Princeton University Press, Princeton, (1981).
[11] **B. Host, B. Kra**: “Nonconventional ergodic averages and nilmanifolds,” Ann. of Math. (2) 161 (1) (2005), 397-488. diagonales,” Bull. Soc. Math. France 121 (1993), no. 3, 315351. Theory Dynam. Systems 23 (2003), 11731198.

[12] **T. Tao**: “Norm convergence of multiple ergodic averages for commuting transformations,” Ergodic Theory Dynam. Systems 28 (2008), no. 2, 657688.

[13] **T. Ziegler**: “Universal characteristic factors and Furstenberg averages,” J. Amer. Math. Soc. 20 (2007), 53-97.

[14] **B. Weiss**: “Strictly ergodic models for dynamical systems”. Bull. Amer. Math. Soc. (N.S.) 13 (1985), no. 2, 143146.

[15] **M. Walsh**: “Norm convergence of nilpotent ergodic averages” Annals of Math, Volume 175 (2012), Issue 3, 1667-1688 (2012).