MULTIPLE RECURRENCE AND CONVERGENCE FOR SEQUENCES RELATED TO THE PRIME NUMBERS

NIKOS FRANTZIKINAKIS, BERNARD HOST, AND BRYNA KRA

ABSTRACT. For any measure preserving system $(X, \mathcal{X}, \mu, T)$ and $A \in \mathcal{X}$ with $\mu(A) > 0$, we show that there exist infinitely many primes p such that $\mu\big(A \cap T^{-(p-1)}A \cap T^{-2(p-1)}A\big) > 0$ (the same holds with $p-1$ replaced by $p+1$). Furthermore, we show the existence of the limit in $L^2(\mu)$ of the associated ergodic average over the primes. A key ingredient is a recent result of Green and Tao on the von Mangoldt function. A combinatorial consequence is that every subset of the integers with positive upper density contains an arithmetic progression of length three and common difference of the form $p-1$ (or $p+1$) for some prime p.

1. INTRODUCTION

1.1. **Results about the primes.** The *von Mangoldt function* $\Lambda \colon \mathbb{Z} \to \mathbb{R}$, defined by

$$\Lambda(n) = \begin{cases} \log p & \text{if } n = p^m \text{ for some } m \in \mathbb{N} \text{ and } p \in \mathbb{P} \\ 0 & \text{otherwise} \end{cases}$$

plays a major role in understanding the distribution of the prime numbers. The classical circle method is the study of the Fourier transform of Λ restricted to an interval $[1, N]$, that is, evaluating the correlation of Λ with complex exponentials on this interval.

Green and Tao ([12], [13], [14]) generalize this method by estimating the correlations of Λ with "nilsequences" and then deduce an estimate for the third Gowers norm of a modification of Λ (a precise definition of this norm is given in Section 2). This result is the base of our study; we need some notation to state it explicitly.

Let $\mathbb{P}$ be the set of prime numbers. If w is a positive integer and $r \in \mathbb{Z}$, setting

$$W = \prod_{p \in \mathbb{P}, p < w} p \,,$$

for $n \in \mathbb{N}$ we define

$$\tilde{\Lambda}_{w,r}(n) = \frac{\phi(W)}{W} \cdot \Lambda(Wn + r) \ ,$$

where ϕ is the Euler function. It is easier to work with this modified von Mangoldt function, as it is better distributed over congruence classes than the original function.

If N is a prime number we identify $[0, N-1]$ with $\mathbb{Z}/N\mathbb{Z}$ in the natural way and consider the restriction $\tilde{\Lambda}_{w,N,r}(n)$ of $\tilde{\Lambda}_{w,r}(n)$ to this interval as a function on $\mathbb{Z}/N\mathbb{Z}$. It follows immediately from the results in [14] that (notation explained in Section 2):

Theorem 1 (Green and Tao [14]). *For $r \in \mathbb{Z}$ with $(r, W) = 1$ we have*

$$\left\| \left(\tilde{\Lambda}_{w,N,r}(n) - 1 \right) \cdot \mathbf{1}_{[0,[N/3])}(n) \right\|_{U_3(\mathbb{Z}/N\mathbb{Z})} = o_{N\to\infty;w}(1) + o_{w\to\infty}(1).$$

We use this result to derive several results in ergodic theory, on recurrence and convergence properties of some sequences related to the primes, and use them to deduce a combinatorial consequence.

1.2. **Multiple recurrence along primes.** The connection between additive combinatorics and ergodic theory began with Szemerédi's celebrated theorem [20], which states that a subset of the integers with positive upper density contains arbitrarily long arithmetic progressions. Furstenberg [8] found an equivalent formulation of Szemerédi's Theorem in terms of multiple recurrence. He proved this "correspondence principle" and showed that if $(X, \mathcal{X}, \mu, T)$ is a measure preserving system[1], $A \in \mathcal{X}$ has positive measure, and k is a positive integer, then there are infinitely many positive integers n such that

(1) $$\mu\big(A \cap T^{-n}A \cap T^{-2n}A \cap \ldots \cap T^{-kn}A\big) > 0 \ .$$

A natural question is what restrictions can be placed on n such that the measure of the intersection in (1) remains positive. This motivates the following definition:

Definition 1. Given an integer $k \geq 1$, $S \subset \mathbb{N}$ is a *set of k-recurrence* if for every measure preserving system $(X, \mathcal{X}, \mu, T)$ and $A \in \mathcal{X}$ with $\mu(A) > 0$, there exist infinitely many $n \in S$ such that

$$\mu\big(A \cap T^{-n}A \cap T^{-2n} \cap \cdots \cap T^{-kn}A\big) > 0 \ .$$

Via Furstenberg's correspondence principle, the equivalent combinatorial formulation of this definition is classical:

---

[1]A measure preserving system is a quadruple $(X, \mathcal{X}, \mu, T)$, where $(X, \mathcal{X}, \mu)$ is a probability space and $T \colon X \to X$ is a measurable map such that $\mu(T^{-1}A) = \mu(A)$ for all $A \in \mathcal{X}$. Without loss of generality we can assume that the probability space is Lebesgue.

Proposition. *Let $k \geq 1$ be an integer. The set $S \subset \mathbb{N}$ is a set of k-recurrence if and only if every subset A of integers with positive upper density contains at least one arithmetic progression of length $k + 1$ and common difference in S.*

A set of k-recurrence is also known in the number theory literature as a *k-intersective set*.

There are many known examples of sets of 1-recurrence. For example, one can take all multiples of a fixed number or $S - S$ for any infinite set $S \subset \mathbb{N}$. Furstenberg [9] and Sárközy [18] independently showed that values of a polynomial, meaning to $\{q(n) \colon n \in \mathbb{N}\}$ where $q(n)$ is an integer valued polynomial with $q(0) = 0$, form a set of 1-recurrence. Another interesting example comes from the primes. Sárközy [19] showed that the set of shifted primes $\mathbb{P} - 1$ (and the set $\mathbb{P} + 1$) form a set of 1-recurrence. Wierdl [21] reproved this result using methods from ergodic theory.

For higher k, restricting the exponents in (1) is more difficult. It is an immediate consequence of Szemerédi's theorem that one can take the set of all multiples of a fixed nonzero integer. Bergelson and Leibman [1] showed that polynomial values $\{q(n) \colon n \in \mathbb{N}\}$ where $q(n)$ is an integer valued polynomial with $q(0) = 0$ are sets of k-recurrence for all $k \geq 1$. On the other hand, Furstenberg [9] constructed an infinite set S whose set of differences $S - S$ is not a set of 2-recurrence. For general k, a set of k-recurrence but not $(k + 1)$-recurrence was constructed in [7].

The recent results of Green and Tao on the von Mangoldt function allow us to show 2-recurrence for the primes:

Theorem 2. *Let $(X, \mathcal{X}, \mu, T)$ be a measure preserving system and let $A \in \mathcal{X}$ have positive measure. There exist infinitely many $p \in \mathbb{P}$ such that*

$$\mu\big(A \cap T^{-(p-1)}A \cap T^{-2(p-1)}A\big) > 0$$

and there exist infinitely many $p \in \mathbb{P}$ such that

$$\mu\big(A \cap T^{-(p+1)}A \cap T^{-2(p+1)}A\big) > 0 \ .$$

It is easy to check that the set $\mathbb{P} + r$ is not even a set of 1-recurrence for $r \in \mathbb{Z} \setminus \{-1, 1\}$, by considering a periodic system with period r if $r \neq 0$ and with period 2 if $r = 0$. A related topological version of this question was posed by Brown, Graham, and Landman [4].

A corollary of Furstenberg's correspondence principle is that any set of integers with positive upper density contains arithmetic progressions of length 3 with common difference in $\mathbb{P} - 1$ (and also in $\mathbb{P} + 1$).

The higher order statement of Theorem 1 was conjectured to be true in [14], if it holds then the obvious generalizations of our proofs show that $\mathbb{P} - 1$ and $\mathbb{P} + 1$ are sets k-recurrence for every $k \in \mathbb{N}$. This and other generalizations are discussed in Section 7.

1.3. **Convergence of averages along primes.** A closely related question is the convergence of the associated multiple ergodic averages. In his proof

of Szemerédi's Theorem, Furstenberg proved more than the measure of the intersection in (1) is positive. He showed that given a measure preserving system $(X, \mathcal{X}, \mu, T)$ and $A \in \mathcal{X}$ of positive measure, we have

$$(2) \qquad \liminf_{N \to \infty} \frac{1}{N} \sum_{0 \le n < N} \mu\big(A \cap T^{-n}A \cap T^{-2n}A \cap \ldots \cap T^{-kn}A\big) > 0 \ .$$

A natural question is the existence of the limit of the associated ergodic average. More generally, given a measure preserving system $(X, \mathcal{X}, \mu, T)$, functions $f_1,\ f_2,\ \ldots,\ f_k \in L^\infty(\mu)$, and an increasing sequence of integers $\{s_n\}_{n \in \mathbb{N}}$, we can study the existence of the limit of

$$\frac{1}{N} \sum_{0 \le n < N} (T^{s_n}f_1 \cdot T^{2s_n}f_2 \cdot \ldots \cdot T^{ks_n}f_k)$$

or, writing $S = \{s_n \colon n \ge 1\}$, the limit of

$$(3) \qquad \frac{1}{|S \cap [0, N)|} \sum_{s \in S, s < N} (T^s f_1 \cdot T^{2s} f_2 \cdot \ldots \cdot T^{ks} f_k)$$

in $L^2(\mu)$ as $N \to \infty$.

In [15], it is shown that the limit exists for $s_n = n$ (another proof is given in [23]). If $q(n)$ is an integer valued polynomial, the existence of the limit for $\{q(n) \colon n \in \mathbb{N}\}$ is shown in [16] (see [17] for a more general result). It was a folklore theorem that for $k = 1$, the limit exists for the primes; it follows easily from results on exponential sums due to Vinogradov and was written down explicitly by Wierdl [21]. (Bourgain [3] and Wierdl [22] actually proved several stronger results on pointwise convergence along primes.)

Let $\pi(N)$ denote the number of primes less than or equal to N. Using Theorem 1 we show:

Theorem 3. *Let $(X, \mathcal{X}, \mu, T)$ be a measure preserving system and let f_1, $f_2 \colon X \to \mathbb{C}$ be bounded measurable functions. Then the limit*

$$(4) \qquad \lim_{N \to \infty} \frac{1}{\pi(N)} \sum_{p \in \mathbb{P}, p < N} (T^p f_1 \cdot T^{2p} f_2)$$

exists in $L^2(\mu)$.

Moreover, we show that for $k = 2$ a certain factor, known as the Kronecker factor, controls the limiting behavior in $L^2(\mu)$ of the corresponding multiple ergodic averages. Furthermore, in a totally ergodic system (meaning T and all its powers are ergodic), we show that the average over the primes has the same limit as the average over the set of positive integers.

To prove multiple recurrence for the shifted primes and convergence results for the primes, we compare the multiple ergodic average related to the primes with the standard multiple ergodic average over the full set of integers. Using estimates in Lemma 3 and the uniformity estimates of Green and Tao in Theorem 1, we then show that the difference of some modification of the two averages tends to zero. Combining this with an ergodic version of

Roth's theorem, we prove double recurrence for the shifted primes in Section 4. In a similar manner, we approach the convergence questions for the double average over the primes. In Section 5 we prove double convergence and in Section 6 we find a characteristic factor for this average.

2. Notation and Gowers norms

By $o_{N\to\infty;a_1,\dots,a_k}(1)$ we mean a quantity that depends on $N, a_1, \dots, a_k$ and for fixed $a_1, \dots, a_k$ converges to zero as $N \to \infty$; the convergence is assumed to be uniform with respect to all variables that are not included in the indices.

Let

$$\Lambda'(n) = \mathbf{1}_\mathbb{P}(n) \cdot \log n \ ,$$

where $\mathbf{1}_\mathbb{P}$ denotes the indicator function of the primes. We have

$$(5) \qquad \frac{1}{N} \sum_{0 \le n < N} \big(\Lambda(n) - \Lambda'(n)\big) \le \frac{1}{N} \sum_{0 \le p < \sqrt{N}} \frac{\log N}{\log p} \cdot \log p$$

$$= \frac{\log N}{N} \sum_{0 \le p < \sqrt{N}} 1 \sim \frac{2}{\sqrt{N}} \ .$$

When f is a function defined on a finite set A, let $\mathbb{E}\,(f(n) \,|\, n \in A) = \mathbb{E}_{n \in A} f(n)$ denote the average value of f on A:

$$\mathbb{E}_{n \in A} f(n) = \frac{1}{|A|} \sum_{n \in A} f(n) \ ,$$

where by $|A|$ we mean the number of elements in A. We also use a higher dimensional version of the expectation. For example, by $\mathbb{E}_{m,n \in A} f(n, m)$, we mean iteration of the one variable expectation:

$$\mathbb{E}_{m \in A}\big(\mathbb{E}_{n \in A} f(n, m)\big) \ .$$

Some of our addition and averages are taken in $\mathbb{Z}/N\mathbb{Z}$ and some in $\mathbb{Z}$. In order to distinguish between the two senses, we use the expectation notation $\mathbb{E}_{n \in \mathbb{Z}/N\mathbb{Z}}$ for an average in $\mathbb{Z}/N\mathbb{Z}$ and $\frac{1}{N} \sum_{0 \le n < N}$ for an average in $\mathbb{Z}$.

If $f \colon \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$, we inductively define:

$$\|f\|_{U_1(\mathbb{Z}/N\mathbb{Z})} = \big|\mathbb{E}(f(n) \,|\, n \in \mathbb{Z}/N\mathbb{Z})\big|$$

and

$$\|f\|_{U_{d+1}(\mathbb{Z}/N\mathbb{Z})} = \Big(\mathbb{E}\big(\big\|f_h \cdot \overline{f}\big\|^{2^d}_{U_d(\mathbb{Z}/N\mathbb{Z})} \,\big|\, h \in \mathbb{Z}/N\mathbb{Z}\big)\Big)^{1/2^{d+1}} \ ,$$

where $f_h(n) = f(n + h)$. Gowers [11] showed that for $d \ge 2$ this defines a norm on $\mathbb{Z}/N\mathbb{Z}$.

3. Some lemmas

For studying an average over the primes, it is convenient to replace this average with a certain weighted average over the integers. The next lemma enables us to do this.

Lemma 1. *If $|a_n| \leq 1$ for $n \in \mathbb{N}$, then*

$$\left| \frac{1}{\pi(N)} \sum_{p \in \mathbb{P}, p < N} a_p - \frac{1}{N} \sum_{0 \leq n < N} (\Lambda(n) \cdot a_n) \right| = o_{N \to \infty}(1)$$

where $p(N)$ denotes the greatest prime number less than or equal to N.

Proof. As noted in equation (5), we can replace $\Lambda(n)$ by $\Lambda'(n)$ making only a small error. Then we have

$$\left| \frac{1}{\pi(N)} \sum_{p \in \mathbb{P}, p < N} a_p - \frac{1}{N} \sum_{0 \leq n < N} (\Lambda'(n) \cdot a_n) \right| \leq \frac{1}{N} \sum_{p \in \mathbb{P}, p < N} |(\log N - \log p) \cdot a_p|$$

$$\leq \frac{1}{N} \sum_{p \in \mathbb{P}, p < N} \log N - \frac{1}{N} \sum_{0 \leq n < N} \Lambda'(n) .$$

Using equation (5) again, we have that up to a small error term this last difference is equal to

$$\frac{1}{N} \sum_{p \in \mathbb{P}, p < N} \log N - \frac{1}{N} \sum_{0 \leq n < N} \Lambda(n).$$

By the prime number theorem and the well known fact that Λ has mean one, this difference goes to zero as $N \to \infty$. $\square$

The next Lemma is only used in the proof of Lemma 3.

Lemma 2. *For $k \in \mathbb{N}$, let $\theta, \phi_0, \ldots, \phi_{k-1} \colon \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ be functions with $|\phi_i| \leq 1$ for $1 \leq i \leq k - 1$ and let ϕ_0 be arbitrary. Then*

$$\left| \mathbb{E}_{m,n \in \mathbb{Z}/N\mathbb{Z}} \big(\theta(n) \cdot \phi_0(m) \cdot \phi_1(m+n) \cdot \ldots \cdot \phi_{k-1}(m + (k-1)n) \big) \right|$$

$$\leq \|\theta\|_{U_k(\mathbb{Z}/N\mathbb{Z})} \cdot \|\phi_0\|_{L^2(\mathbb{Z}/N\mathbb{Z})},$$

where $\|\phi_0\|_{L^2(\mathbb{Z}/N\mathbb{Z})} = (\mathbb{E}_{n \in \mathbb{Z}/N\mathbb{Z}} |\phi_0(n)|^2)^{1/2}$.

Proof. We make use of the identity

$$\left| \mathbb{E}_{n \in \mathbb{Z}/N\mathbb{Z}} a(n) \right|^2 = \mathbb{E}_{n,h \in \mathbb{Z}/N\mathbb{Z}} (a(n+h) \cdot \overline{a(n)}) ,$$

which holds for $a \colon \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$. We use induction in k. For $k = 1$,

$$\left| \mathbb{E}_{m,n \in \mathbb{Z}/N\mathbb{Z}} (\theta(n) \cdot \phi_0(m)) \right| = \left| \mathbb{E}_{n \in \mathbb{Z}/N\mathbb{Z}} \theta(n) \right| \cdot \left| \mathbb{E}_{m \in \mathbb{Z}/N\mathbb{Z}} \phi_0(m) \right|$$

$$\leq \left| \mathbb{E}_{n \in \mathbb{Z}/N\mathbb{Z}} \theta(n) \right| \cdot \|\phi_0\|_{L^2(\mathbb{Z}/N\mathbb{Z})} .$$

Suppose the statement holds for $k = l$. We show that it also holds for $k = l+1$. Applying Cauchy-Schwarz and the previous identity we have that

$$\left| \mathbb{E}_{m,n \in \mathbb{Z}/N\mathbb{Z}} \big(\theta(n) \cdot \phi_0(m) \cdot \phi_1(m+n) \cdot \ldots \cdot \phi_l(m+ln) \big) \right|^2$$

$$\leq \mathbb{E}_{m \in \mathbb{Z}/N\mathbb{Z}} \left| \mathbb{E}_{n \in \mathbb{Z}/N\mathbb{Z}} \big(\theta(n) \cdot \phi_1(m+n) \cdot \ldots \cdot \phi_l(m+ln) \big) \right|^2 \cdot \|\phi_0\|_{L^2(\mathbb{Z}/N\mathbb{Z})}$$

$$= \mathbb{E}_{m,n,h \in \mathbb{Z}/N\mathbb{Z}} \big(\theta(n+h) \cdot \overline{\theta(n)} \cdot \phi'_{0,h}(m) \cdot \phi'_{1,h}(m+n) \cdot \ldots \cdot$$
$$\phi'_{l-1,h}(m+(l-1)n) \big) \cdot \|\phi_0\|_{L^2(\mathbb{Z}/N\mathbb{Z})},$$

where $\phi'_{i,h}(m) = \phi_{i+1}(m+h) \cdot \overline{\phi_{i+1}(m)}$ satisfies $|\phi'_{i,h}| \leq 1$ for $i = 0, \ldots, l-1$. The last average equals

$$\mathbb{E}_{h \in \mathbb{Z}/N\mathbb{Z}} \Big(\mathbb{E}_{m,n \in \mathbb{Z}/N\mathbb{Z}} \big(\theta(n+h) \cdot \overline{\theta(n)} \cdot \phi'_{0,h}(m) \cdot \phi'_{1,h}(m+n) \cdot \ldots \cdot$$
$$\phi'_{l-1,h}(m+(l-1)n) \big) \Big) .$$

By the induction hypothesis and the estimate $\left\| \phi'_{0,h} \right\|_{L^2(\mathbb{Z}/N\mathbb{Z})} \leq 1$, the last average is bounded by

$$\mathbb{E}_{h \in \mathbb{Z}/N\mathbb{Z}} \left\| \theta(n+h) \cdot \overline{\theta(n)} \right\|_{U_l(\mathbb{Z}/N\mathbb{Z})} \leq \left(\mathbb{E}_{h \in \mathbb{Z}/N\mathbb{Z}} \left\| \theta(n+h) \cdot \overline{\theta(n)} \right\|_{U_l(\mathbb{Z}/N\mathbb{Z})}^{2^l} \right)^{1/2^l}$$

$$= \|\theta\|_{U_{l+1}(\mathbb{Z}/N\mathbb{Z})}^2 .$$

This completes the induction. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

We use $[x]$ to denote the greatest integer less than or equal to x.

Lemma 3. *Let $k \geq 2$ and $N > k$ be integers, and $\theta \colon \mathbb{Z}/N\mathbb{Z} \to \mathbb{R}$ be a function that is zero for $[N/k] \leq n < N$. Let $(X, \mathcal{X}, \mu, T)$ be a measure preserving system and $f_1, \ldots, f_{k-1} \colon X \to \mathbb{C}$ be measurable functions with $\|f_i\|_\infty \leq 1$ for $i = 1, \ldots, k-1$. Then*

$$\left\| \frac{1}{[N/k]} \sum_{0 \leq n < [N/k]} \big(\theta(n) \cdot T^n f_1 \cdot \ldots \cdot T^{(k-1)n} f_{k-1} \big) \right\|_{L^2(\mu)} \leq C_k \cdot \|\theta\|_{U_k(\mathbb{Z}/N\mathbb{Z})} ,$$

for some constant $C_k > 0$.

Proof. Let $f_0 \in L^\infty(\mu)$ be arbitrary. We apply Lemma 2 for the functions $\phi_{x,i}(n) \colon \mathbb{Z}/N\mathbb{Z} \to \mathbb{R}$ defined by

$$\phi_{x,0}(n) = \begin{cases} f_0(T^n x) & \text{if } 0 \leq n < [N/k] \\ 0 & \text{if } [N/k] \leq n < N , \end{cases}$$

and $\phi_{x,i}(n) = f_i(T^n x)$ for $i = 1, \ldots, k-1$. Note that for $0 \leq m, n < N$ we have

$$\theta(n) \cdot \phi_{x,0}(m) \cdot \phi_{x,1}(m+n) \cdot \ldots \cdot \phi_{x,k}(m+(k-1)n) = 0$$

except when $0 \le m, n < [N/k]$, in which case the above expression is equal to

$$\theta(n) \cdot f_0(T^m x) \cdot f_1(T^{m+n} x) \cdot \ldots \cdot f_{k-1}(T^{m+(k-1)n} x) \ .$$

Thus by Lemma 2,

$$\left| \frac{1}{N^2} \sum_{0 \le m, n < [N/k]} \left(\theta(n) \cdot f_0(T^m x) \cdot f_1(T^{m+n} x) \cdot \ldots \cdot f_{k-1}(T^{m+(k-1)n} x) \right) \right|$$

$$\le \|\theta\|_{U_k(\mathbb{Z}/N\mathbb{Z})} \cdot \left(\frac{1}{N} \sum_{0 \le n < [N/k]} |f_0(T^n x)|^2 \right)^{1/2} .$$

Integrating over X and applying Cauchy-Schwarz to the right hand side gives

$$\left| \int f_0(x) \cdot \frac{1}{[N/k]} \sum_{0 \le n < [N/k]} \left(\theta(n) \cdot f_1(T^n x) \cdot \ldots \cdot f_{k-1}(T^{(k-1)n} x) \right) d\mu \right|$$

$$\le \left(\frac{N}{[N/k]} \right)^{3/2} \cdot \|\theta\|_{U_k(\mathbb{Z}/N\mathbb{Z})} \cdot \|f_0\|_{L^2(\mu)} \ .$$

By duality, we have the advertised estimate with $C_k = (2k)^{3/2}$. $\square$

4. RECURRENCE

Theorem 2 follows immediately from the next result, which we prove using Theorem 1 (the same statement holds for $\mathbb{P} + 1$, with the obvious modifications).

Theorem 4. *Let $(X, \mathcal{X}, \mu, T)$ be a measure preserving system and $A \in \mathcal{X}$ with $\mu(A) > 0$. Then*

$$\liminf_{N \to \infty} \frac{1}{\pi(N)} \sum_{p \in \mathbb{P}, p < N} \mu(A \cap T^{-(p-1)} A \cap T^{-2(p-1)} A) > 0 \ .$$

Proof. By Lemma 1 it suffices to show that for prime numbers N we have

$$\liminf_{N \to \infty} \frac{1}{N} \sum_{0 \le n < N} \left(\Lambda(n+1) \cdot \mu(A \cap T^{-n} A \cap T^{-2n} A) \right) > 0 \ .$$

For this, it suffices to show that for some $w \in \mathbb{N}$ we have

(6) $\displaystyle \liminf_{N \to \infty} \frac{1}{[N/3]} \sum_{0 \le n < [N/3]} \left(\tilde{\Lambda}_{w,N,1}(n) \cdot \mu(A \cap T^{-Wn} A \cap T^{-2Wn} A) \right) > 0 \ ,$

where $W = \prod_{p \in \mathbb{P}, p < w} p$. We claim that

(7) $\displaystyle \lim_{N \to \infty} \frac{1}{[N/3]} \sum_{0 \le n < [N/3]} \left((\tilde{\Lambda}_{w,N,1}(n) - 1) \cdot \mu(A \cap T^{-Wn} A \cap T^{-2Wn} A) \right)$

$$= o_{w \to \infty}(1) \ .$$

To see this, we first apply Cauchy-Schwarz to get

$$\left| \frac{1}{[N/3]} \sum_{0 \le n < [N/3]} \left((\tilde{\Lambda}_{w,N,1}(n) - 1) \cdot \mu(A \cap T^{Wn} A \cap T^{2Wn} A) \right) \right|$$

$$\le \left\| \frac{1}{[N/3]} \sum_{0 \le n \in < [N/3]} \left((\tilde{\Lambda}_{w,N,1}(n) - 1) \cdot T^{Wn} \mathbf{1}_A \cdot T^{2Wn} \mathbf{1}_A \right) \right\|_{L^2(\mu)} .$$

By Lemma 3 this last term is bounded by

$$\left\| (\tilde{\Lambda}_{w,N,1}(n) - 1) \cdot \mathbf{1}_{[0,[N/3])} \right\|_{U_3(\mathbb{Z}/N\mathbb{Z})} ,$$

and by Theorem 1, this is

$$o_{N \to \infty; w}(1) + o_{w \to \infty}(1) .$$

Letting $N \to \infty$ gives (7).

We now proceed to show (6). Let $\mu(A) = \delta$. Roth's theorem easily implies (see Theorem 2.1 in [2] for details) that for every measure preserving system $(Y, \mathcal{Y}, \nu, S)$ and $B \in \mathcal{Y}$ with $\nu(B) \ge \delta$ there exists a constant $c(\delta) > 0$ such that

$$(8) \qquad \liminf_{N \to \infty} \frac{1}{[N/3]} \sum_{0 \le n < [N/3]} \nu(B \cap S^{-n} B \cap S^{-2n} B) \ge c(\delta) .$$

Using (7) and applying (8) for the systems $(X, \mathcal{X}, T^W, \mu)$, we have

$$\liminf_{N \to \infty} \frac{1}{[N/3]} \sum_{0 \le n < [N/3]} \left(\tilde{\Lambda}_{w,N,1}(n) \cdot \mu(A \cap T^{-Wn} A \cap T^{-2Wn} A) \right)$$

$$\ge \liminf_{N \to \infty} \frac{1}{[N/3]} \sum_{0 \le n < [N/3]} \mu(A \cap T^{-Wn} A \cap T^{-2Wn} A) + o_{w \to \infty}(1)$$

$$\ge c(\delta) + o_{w \to \infty}(1) .$$

Taking w sufficiently large, we have (6), completing the proof. $\qquad \square$

5. L^2-CONVERGENCE

We now prove convergence of averages along the primes:

Proof of Theorem 3. We can assume that $\|f_i\|_\infty \le 1$ for $i = 1, 2$. By Lemma 1 it suffices to prove the corresponding results for the weighted averages

$$(9) \qquad \lim_{N \to \infty} \frac{1}{N} \sum_{0 \le n < N} \left(\Lambda(n) \cdot T^n f_1 \cdot T^{2n} f_2 \right) .$$

For $x \in X$ let $a_x(n) = f_1(T^n x) \cdot f_2(T^{2n} x)$. We claim that

$$(10) \qquad \left\| \frac{1}{[WN/3]} \sum_{0 \le n < [WN/3]} (\Lambda(n) \cdot a_x(n)) - \right.$$

$$\left. \frac{1}{\phi(W)} \sum_{\substack{0 \le r < W \\ (r,W)=1}} \frac{1}{[N/3]} \sum_{0 \le n < [N/3]} a_x(Wn+r) \right\|_{L^2(\mu)}$$

$$= o_{N\to\infty;w}(1) + o_{w\to\infty}(1) \ .$$

To prove (10) first note that

$$\frac{1}{[WN/3]} \sum_{0 \le n < [(WN)/3]} (\Lambda(n) \cdot a_x(n))$$

$$= \frac{1}{\phi(W)} \sum_{\substack{0 \le r < W \\ (r,W)=1}} \frac{1}{[N/3]} \sum_{0 \le n < [N/3]} \left(\frac{\phi(W)}{W} \cdot \Lambda(Wn+r) \cdot a_x(Wn+r) \right)$$

$$+ o_{N\to\infty;w}(1)$$

$$= \frac{1}{\phi(W)} \sum_{\substack{0 \le r < W \\ (r,W)=1}} \frac{1}{[N/3]} \sum_{0 \le n < [N/3]} \left(\tilde{\Lambda}_{w,N,r}(n) \cdot a_x(Wn+r) \right) + o_{N\to\infty;w}(1) \ ,$$

where the error terms are introduced because Λ is supported on the prime powers rather than the primes and the use of integer parts. Hence,

$$\left\| \frac{1}{[WN/3]} \sum_{0 \le n < [WN/3]} (\Lambda(n) \cdot a_x(n)) - \right.$$

$$\left. \frac{1}{\phi(W)} \sum_{\substack{0 \le r < W \\ (r,W)=1}} \frac{1}{[N/3]} \sum_{0 \le n < [N/3]]} (a_x(Wn+r)) \right\|_{L^2(\mu)}$$

$$\le \frac{1}{\phi(W)} \sum_{\substack{0 \le r < W \\ (r,W)=1}} \left\| \frac{1}{[N/3]} \sum_{0 \le n < [N/3]} \left((\tilde{\Lambda}_{w,N,r}(n) - 1) \cdot a_x(Wn+r) \right) \right\|_{L^2(\mu)}$$

$$+ o_{N\to\infty;w}(1)$$

$$\le C \cdot \frac{1}{\phi(W)} \sum_{\substack{0 \le r < W \\ (r,W)=1}} \left\| \mathbf{1}_{[0,N/3)} \cdot (\tilde{\Lambda}_{w,N,r}(n) - 1) \right\|_{U_3(\mathbb{Z}/N\mathbb{Z})} + o_{N\to\infty;w}(1) \ .$$

The last inequality follows by applying Lemma 3 with $S = T^W$ and $g_i = T^{ri} f_i$, $i = 1, 2$. Now (10) follows using Theorem 1.

We proceed to show that the sequence

$$A_x(N) = \mathbb{E}_{0 \le n < N}(\Lambda(n) \cdot a_x(n))$$

converges in $L^2(\mu)$ by showing that it is a Cauchy sequence. Let $\varepsilon > 0$ and

$$B_{x,w,r}(N) = \frac{1}{[N/3]} \sum_{0 \le n < [N/3]} a_x(Wn + r) .$$

Using (10) and the fact that $B_{x,w,r}(N)$ converges in $L^2(\mu)$ we get that for some W_0, if M, N are sufficiently large then

$$\left\| A_x((W_0 N)/3) - \frac{1}{\phi(W)} \sum_{\substack{0 \le r < W \\ (r,W)=1}} (B_{x,w_0,r}(N)) \right\|_{L^2(\mu)} \le \varepsilon/3 ,$$

$$\|B_{x,w_0,r}(N) - B_{x,w_0,r}(M)\|_{L^2(\mu)} \le \varepsilon/3 ,$$

for all $0 \le r < W_0$ with $(r, W_0) = 1$. Using this and the triangle inequality we have that if M, N are large enough then

$$\|A_x((W_0 N)/3) - A_x((W_0 M)/3)\|_{L^2(\mu)} \le \varepsilon .$$

Since

$$A_x((W_0 N)/3 + i) = A_x((W_0 N)/3) + o_{N \to \infty}(1)$$

for $0 \le i < W_0/3$, we conclude that $A_x(N)$ is Cauchy. $\qquad\square$

6. A characteristic factor for the average (4)

A *factor* of a measure preserving system $(X, \mathcal{X}, \mu, T)$ is defined to be a T-invariant sub-σ-algebra $\mathcal{Z}$ of $\mathcal{X}$. The factor $\mathcal{Z}$ is characteristic for L^2-convergence of the averages in (4) if f_1 and f_2 can be replaced by their conditional expectations $\mathbb{E}(f_1 \,|\, \mathcal{Z})$ and $\mathbb{E}(f_2 \,|\, \mathcal{Z})$ without changing the value of the limit, taken in $L^2(\mu)$. The *Kronecker factor* $\mathcal{K}$ is defined to be the smallest sub-σ-algebra of $\mathcal{X}$ with respect to which the eigenfunctions of T are measurable.

Theorem 5. *Let $(X, \mathcal{X}, \mu, T)$ be an ergodic measure preserving system and let $f_1, f_2 \colon X \to \mathbb{C}$ be bounded measurable functions. Then the Kronecker factor $\mathcal{K}$ is characteristic for $L^2(\mu)$ convergence of the average in (4). Furthermore, if $(X, \mathcal{X}, \mu, T)$ is a totally ergodic system, then the limit of the average in (4) is equal to*

$$\lim_{N \to \infty} \frac{1}{N} \sum_{0 \le n < N} (T^n f_1 \cdot T^{2n} f_2) .$$

Proof. We first show that the Kronecker factor is characteristic. It suffices to show that if either f_1 or f_2 is orthogonal to the Kronecker factor then the average in (4) converges to zero in $L^2(\mu)$. By $a_x(N)$, $A_x(N)$ and $B_{x,w,r}(N)$ we denote the sequences defined in the proof of Theorem 3. As it was shown in [8], for every w and r the sequence $B_{x,w,r}(N)$ converges to zero in $L^2(\mu)$

as $N \to \infty$. Since $\lim_{N\to\infty} A_x(N)$ exists, by (10) we have that

$$
\lim_{N\to\infty} A_x(N) = \lim_{N\to\infty} A_x(WN/3)
$$
$$
= \lim_{N\to\infty} B_{x,w,r}(N) + o_{w\to\infty}(1) = o_{w\to\infty}(1) \ ,
$$

where all the limits are taken in $L^2(\mu)$. The result follows by letting $w \to \infty$.

Next, we evaluate the limit for totally ergodic systems. For any such system we have for every $w, r \in \mathbb{N}$ and $f_1, f_2 \in L^\infty(\mu)$ that

$$
(11) \quad \lim_{N\to\infty} \frac{1}{N} \sum_{0\le n<N} T^{Wn+r} f_1 \cdot T^{2Wn+r} f_2 = \lim_{N\to\infty} \frac{1}{N} \sum_{0\le n<N} T^n f_1 \cdot T^{2n} f_2 \ ,
$$

where both limits are taken in $L^2(\mu)$. One can see this using the formula for the limit of these averages given in [8]. Equation (11) gives

$$
\lim_{N\to\infty} B_{x,w,r}(N) = \lim_{N\to\infty} \frac{1}{N} \sum_{0\le n<N} a_x(N)
$$

for every $w, r \in \mathbb{N}$. Since $A_x(N)$ converges in $L^2(\mu)$ we conclude from (10) that for every w

$$
\lim_{N\to\infty} A_x(N) = \lim_{N\to\infty} A_x((WN)/3) = \lim_{N\to\infty} \frac{1}{N} \sum_{0\le n<N} a_x(n) + o_{w\to\infty}(1) \ ,
$$

where all the limits are taken in $L^2(\mu)$. The result follows by letting $w \to \infty$. $\qquad\square$

7. Further generalizations

It is natural to ask about higher order recurrence for the shifted primes and higher order convergence along the primes. The major missing ingredient is the higher order statements of Theorem 1 with respect to the Gowers norms. If such estimates hold, then again using a uniform version of the multiple recurrence theorem of Furstenberg (contained in [2]), our proof carries over. For convergence, the proof also carries over, using the general result on convergence of linear averages in [15]. Furthermore, the generalization of Theorem 5 holds, using the description of the characteristic factors in [15] and the higher analog of identity (11) in [6].

With small modifications the argument used to prove Theorems 2 and 3 also gives analogous recurrence and convergence results for two commuting transformations. (The proofs of these are almost identical to the proofs given and so we omit them.) Namely, if T_1 and T_2 are commuting invertible measure preserving transformations of a probability space $(X, \mathcal{X}, \mu)$ and $A \in \mathcal{X}$ with $\mu(A) > 0$, then there exists $p \in \mathbb{P}$ such that

$$
\mu(A \cap T_1^{-(p-1)} A \cap T_2^{-(p-1)} A) > 0 \ .
$$

The analogous statement holds with $p + 1$ instead of $p - 1$. Replacing the role of the uniform version of Furstenberg's recurrence theorem is a uniform version of the multidimensional Szemerédi Theorem of Furstenberg

and Katznelson [10]. If one can obtain an estimate for the k-th Gowers norms of $\tilde{\Lambda} - 1$ for $k \geq 4$, analogous to the one of Theorem 1, then as for a single transformation, one would obtain the higher order commuting version.

Similarly, one can show that if $f_1, f_2 \in L^\infty(\mu)$, then the averages

$$\frac{1}{\pi(N)} \sum_{p \in \mathbb{P}, p < N} (T_1^p f_1 \cdot T_2^p f_2)$$

converge in $L^2(\mu)$ as $N \to \infty$. Again, we compare this average with the standard ergodic average for two commuting transformations, whose convergence was proven in [5]. Since convergence is not known for the standard average of $k \geq 3$ commuting transformations, even with higher order estimates on the Gowers norm of $\tilde{\Lambda} - 1$, our proof would not generalize.

REFERENCES

[1] V. Bergelson and A. Leibman, Polynomial extensions of van der Waerden's and Szemerédi's theorems. *J. Amer. Math. Soc.* **9** (1996), 725–753.

[2] V. Bergelson, B. Host, R. McCutcheon, F. Parreau, Aspects of uniformity in recurrence. *Colloq. Math.* **85** (2000), 549–576.

[3] J. Bourgain. An approach to pointwise ergodic theorems. *Geometric aspects of functional analysis (1986/87).* Lecture Notes in Math., 1317, Springer, Berlin (1988), 204–223.

[4] T. Brown, R. Graham, and B. Landman. On the set of common differences in van der Waerden's Theorem on arithmetic progressions. *Canad. Math. Bull.* **42** (1999), 25–36.

[5] J.-P. Conze and E. Lesigne. Sur un théorème ergodique pour des mesures diagonales. *C. R. Acad. Sci. Paris Série I* **306** (1988), 491–493.

[6] N. Frantzikinakis. The structure of strongly stationary systems. *J. d'Analyse Math.* **93** (2004), 359–388.

[7] N. Frantzikinakis, E. Lesigne, and M. Wierdl. Sets of k-recurrence but not $(k+1)$-recurrence. To appear, *Ann. Inst. Fourier.*

[8] H. Furstenberg. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. d'Analyse Math.* **71** (1977), 204–256.

[9] H. Furstenberg. Recurrence in ergodic theory and combinatorial number theory. *Princeton University Press*, Princeton, N.J., 1981.

[10] H. Furstenberg and Y. Katznelson. An ergodic Szemerédi theorem for commuting transformations. *J. d'Analyse Math.* **34** (1979), 275-291.

[11] T. Gowers. A new proof of Szemerédi's Theorem. *Geom. Funct. Anal.* **11** (2001), 465–588.

[12] B. Green and T. Tao. The primes contain arbitrarily long arithmetic progressions. To appear, *Annals. Math.*

[13] B. Green and T. Tao. An inverse theorem for the Gowers U^3 norm. Preprint. Available at http://arxiv.org/abs/math.NT/0606087.

[14] B. Green and T. Tao. Linear equations in primes. Preprint. Available at http://arxiv.org/abs/math.NT/0606088.

[15] B. Host and B. Kra. Nonconventional ergodic averages and nilmanifolds. *Annals. Math.* **161**, (2005), 397–488.

[16] B. Host and B. Kra. Convergence of polynomial ergodic averages. *Isr. J. Math.* **149** (2005), 1–19.

[17] A. Leibman. Convergence of multiple ergodic averages along polynomials of several variables. *Isr. J. Math.* **146** (2005), 303–316.

[18] A. Sárközy. On difference sets of sequences of integers, I. *Acta Math. Acadm. Sci. Hungar.* **31** (1978), 125–149.

[19] A. Sárközy. On difference sets of sequences of integers, III. *Acta Math. Acadm. Sci. Hungar.* **31** (1978), 355–386.

[20] E. Szemerédi, On sets of integers containing no k elements in arithmetic progression. *Acta Arith.* **27** (1975), 299–345.

[21] M. Wierdl. Almost everywhere convergence and recurrence along subsequences in ergodic theory. PhD Thesis, Ohio State University, 1989.

[22] M. Wierdl. Pointwise ergodic theorem along the prime numbers. *Israel J. Math.* **64** (1988), 315–336.

[23] T. Ziegler, Universal characteristic factors and Furstenberg averages. To appear, *J. Amer. Math. Soc.*

INSTITUTE FOR ADVANCED STUDY, 1 EINSTEIN DRIVE, PRINCETON, NJ 08540

ÉQUIPE D'ANALYSE ET DE MATHÉMATIQUES APPLIQUÉES, UNIVERSITÉ DE MARNE LA VALLÉE, 77454 MARNE LA VALLÉE CEDEX, FRANCE

DEPARTMENT OF MATHEMATICS, NORTHWESTERN UNIVERSITY, 2033 SHERIDAN ROAD, EVANSTON, IL 60208-2730, USA

E-mail address: `nikos@ias.edu`

E-mail address: `Bernard.Host@univ-mlv.fr`

E-mail address: `kra@math.northwestern.edu`