Pacific Journal of Mathematics

NONCONVENTIONAL ERGODIC AVERAGES AND MULTIPLE RECURRENCE FOR VON NEUMANN DYNAMICAL SYSTEMS

TIM AUSTIN, TANJA EISNER AND TERENCE TAO

Volume 250 No. 1 March 2011

PACIFIC JOURNAL OF MATHEMATICS
Vol. 250, No. 1, 2011

NONCONVENTIONAL ERGODIC AVERAGES AND MULTIPLE RECURRENCE FOR VON NEUMANN DYNAMICAL SYSTEMS

TIM AUSTIN, TANJA EISNER AND TERENCE TAO

The Furstenberg recurrence theorem (or equivalently Szemerédi's theorem) can be formulated in the language of von Neumann algebras as follows: given an integer $k \geq 2$, an abelian finite von Neumann algebra $(\mathcal{M}, \tau)$ with an automorphism $\alpha : \mathcal{M} \to \mathcal{M}$, and a nonnegative $a \in \mathcal{M}$ with $\tau(a) > 0$, one has $\liminf_{N\to\infty} N^{-1} \sum_{n=1}^{N} \operatorname{Re} \tau(a\alpha^n(a) \cdots \alpha^{(k-1)n}(a)) > 0$; a later result of Host and Kra shows this limit exists. In particular, $\operatorname{Re} \tau(a\alpha^n(a) \cdots \alpha^{(k-1)n}(a))$ is positive for all n in a set of positive density.

From the von Neumann algebra perspective, it is natural to ask to what remains of these results when the abelian hypothesis is dropped. All three claims hold for $k = 2$, and we show that all three claims hold for all k when the von Neumann algebra is asymptotically abelian, and that the last two claims hold for $k = 3$ when the von Neumann algebra is ergodic. However, we show that the first claim can fail for $k = 3$ even with ergodicity, the second claim can fail for $k \geq 4$ even when assuming ergodicity, and the third claim can fail for $k = 3$ without ergodicity, or $k \geq 5$ and odd assuming ergodicity. The second claim remains open for nonergodic systems with $k = 3$, and the third claim remains open for ergodic systems with $k = 4$.

Austin is supported by a fellowship from Microsoft Corporation. Eisner is supported by the European Social Fund and by the Baden-Württemberg Ministry of Science, Research and the Arts. Tao is supported by NSF grant DMS-0649473 and a grant from the Macarthur Foundation.
MSC2000: 46L55.
Keywords: Szemerédi's theorem, multiple recurrence, nonconventional ergodic averages, von Neumann algebras.

1. Introduction

1a. ***Multiple recurrence.*** Let $(X, \mathscr{X}, \mu)$ be a probability space, and let $T : X \to X$ be a measure-preserving invertible transformation on X (that is, T and T^{-1} are both measurable, and $\mu(T(A)) = \mu(A)$ for all measurable A). From the mean ergodic theorem we know that for any $f \in L^\infty(X)$, the averages $N^{-1}\sum_{n=1}^N f \circ T^{-n}$ converge in (say) $L^2(X)$ norm,[1] which implies in particular that the averages $N^{-1}\sum_{n=1}^N \int_X f_1(f_2 \circ T^{-n})\,d\mu$ converge for all $f_1, f_2 \in L^\infty(X)$. Furthermore, if $f_1 = f_2 = f$ is nonnegative with positive mean $\int_X f\,d\mu > 0$, then the Poincaré recurrence theorem implies that this latter limit is strictly positive. In particular, this implies that the mean $\int_X f(f \circ T^{-n})\,d\mu$ is positive for all natural numbers n in a set $E \subset \mathbb{N}$ of positive (lower) density (that is, the set E is a set such that $\liminf_{N\to\infty} N^{-1}\#\{1 \le n \le N : n \in E\} > 0$).

Thanks to a long effort starting with Furstenberg's ground breaking new proof [1977] of Szemerédi's theorem on arithmetic progressions [1975], it is now known that all of these single recurrence results extend to multiple recurrence:

Theorem 1.1 (abelian multiple recurrence). *Let $(X, \mathscr{X}, \mu)$ be a probability space, let $k \ge 2$ be an integer, and let $T : X \to X$ be a measure-preserving invertible transformation.*

- (Convergence in norm.) *For any $f_1, \dots, f_{k-1} \in L^\infty(X)$, the averages*
$$\frac{1}{N}\sum_{n=1}^{N}(f_1 \circ T^{-n}) \cdots (f_{k-1} \circ T^{-(k-1)n})$$
converge in $L^2(X)$ norm as $N \to \infty$.
- (Weak convergence.) *For any $f_0, f_1, \dots, f_{k-1} \in L^\infty(X)$, the averages*
$$\frac{1}{N}\sum_{n=1}^{N}\int_X f_0(f_1 \circ T^{-n}) \cdots (f_{k-1} \circ T^{-(k-1)n})\,d\mu$$
converge as $N \to \infty$.
- (Recurrence on average.) *For any nonnegative $f \in L^\infty(X)$ with $\int_X f\,d\mu > 0$, one has*

$$\liminf_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\int_X f(f \circ T^{-n}) \cdots (f \circ T^{-(k-1)n})\,d\mu > 0. \tag{1}$$

[1] The minus sign here is not of particular significance (other than to conform to some minor notational conventions) and can be ignored in the sequel if desired.

- (Recurrence on a dense set.) *For any nonnegative* $f \in L^\infty(X)$ *such that* $\int_X f \, d\mu > 0$, *one has*

$$\int_X f(f \circ T^{-n}) \cdots (f \circ T^{-(k-1)n}) \, d\mu > c > 0 \tag{2}$$

for some $c > 0$ *and all* n *in a set of natural numbers of positive lower density.*

We have called this result the "abelian" multiple recurrence theorem in order to emphasise the abelian nature of the algebra $L^\infty(X)$.

Remarks 1.2. Clearly, convergence in norm implies weak convergence; also, because the averages (2) are bounded and nonnegative, recurrence on average implies recurrence on a dense set. Using the weak convergence result, the limit inferior in (1) can be replaced with a limit, but we have retained the limit inferior in order to keep the two claims logically independent of each other.

As mentioned earlier, the $k = 2$ cases of Theorem 1.1 follow from classical ergodic theorems. Furstenberg [1977] established recurrence on average (and hence recurrence on a dense set) for all k, and observed that this result was equivalent (by what is now known as the *Furstenberg correspondence principle*) to Szemerédi's famous theorem [1975] on arithmetic progressions, thus providing an important new proof of that theorem. Convergence in norm (and hence in mean) was established for $k = 3$ by Furstenberg [1977], for $k = 4$ by Conze and Lesigne [1984; 1988a; 1988b] assuming total ergodicity and by Host and Kra [2001] in general, for $k = 5$ in some cases by Ziegler [2005], and for all k by Host and Kra [2005] and subsequently also by Ziegler [2007]. See [Kra 2006] for a survey of these results, and their relation to other topics such as dynamics of nilsequences, and arithmetic progressions in number-theoretic sets such as the primes.

There is also a multidimensional generalisation of the results above to multiple commuting shifts:

Theorem 1.3 (abelian multidimensional multiple recurrence). *Let* $(X, \mathscr{X}, \mu)$ *be a probability space, let* $k \geq 2$ *be an integer, and let* $T_0, \ldots, T_{k-1} : X \to X$ *be a commuting system of measure-preserving invertible transformations.*

- (Convergence in norm.) *For any* $f_1, \ldots, f_{k-1} \in L^\infty(X)$, *the averages*

$$\frac{1}{N} \sum_{n=1}^{N} T_0^n((f_1 \circ T_1^{-n}) \cdots (f_{k-1} \circ T_{k-1}^{-n}))$$

converge in $L^2(X)$ *norm.*

- (Weak convergence.) *For any* $f_0, f_1, \dots, f_{k-1} \in L^\infty(X)$, *the averages*

$$\frac{1}{N}\sum_{n=1}^{N}\int_X (f_0 \circ T_0^{-n})(f_1 \circ T_1^{-n}) \cdots (f_{k-1} \circ T_{k-1}^{-n})\, d\mu$$

  converge.
- (Recurrence on average.) *For any nonnegative* $f \in L^\infty(X)$ *with* $\int_X f\, d\mu > 0$, *one has*

$$\liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N}\int_X (f \circ T_0^{-n})(f \circ T_1^{-n}) \cdots (f \circ T_{k-1}^{-n})\, d\mu > 0. \tag{3}$$

- (Recurrence on a dense set.) *For any nonnegative* $f \in L^\infty(X)$ *such that* $\int_X f\, d\mu > 0$, *one has*

$$\int_X (f \circ T_0^{-n})(f \circ T_1^{-n}) \cdots (f \circ T_{k-1}^{-n})\, d\mu > c > 0 \tag{4}$$

  for some $c > 0$ *and all* n *in a set of natural numbers of positive lower density.*

Of course, Theorem 1.1 is the special case of Theorem 1.3 when $T_i := T^i$. It is often customary to normalise T_0 to be the identity transformation (by replacing each of the T_i with $T_0^{-1}T_i$).

Remarks 1.4. The $k = 2$ case is again classical. Recurrence on average (and hence on a dense set) in this theorem was established for all k by Furstenberg and Katznelson [1978], which by the Furstenberg correspondence principle implies a multidimensional version of Szemerédi's theorem, a combinatorial proof of which in full generality has only been obtained relatively recently in [Nagle et al. 2006] and [Gowers 2006]. Convergence in norm (and weak convergence) was established for $k = 3$ in [Conze and Lesigne 1984], for some special cases of $k = 4$ in [Zhang 1996], for all k assuming total ergodicity in [Frantzikinakis and Kra 2005], and for all k unconditionally in [Tao 2008], with subsequent proofs in [Towsner 2007; Austin 2010; Host 2009]. The results can fail if the shifts $T_0, \dots, T_{k-1}$ do not commute [Bergelson and Leibman 2004]. Note that noncommutativity of the shifts should *not* be confused with the noncommutativity of the underlying algebra, which is the focus of this paper.

1b. ***Noncommutative analogues.*** From the perspective of the theory of von Neumann algebras, the space $L^\infty(X)$ appearing in these theorems can be interpreted as an abelian von Neumann algebra, with a finite trace $\tau(f) := \int_X f\, d\mu$ and with an automorphism $T : L^\infty(X) \to L^\infty(X)$ defined by $Tf := f \circ T^{-1}$. It is then natural to ask whether the results can be extended to nonabelian settings. More precisely, we recall the following definitions.

Definition 1.5 (noncommutative systems). A *finite von Neumann algebra* is a pair $(\mathcal{M}, \tau)$, where $\mathcal{M}$ is a von Neumann algebra (that is, an algebra of bounded operators on a separable[2] complex Hilbert space that contains the identity 1, is closed under adjoints, and is closed in the weak operator topology), and $\tau : \mathcal{M} \to \mathbb{C}$ is a finite faithful trace (that is, a linear map with $\tau(a^*) = \overline{\tau(a)}$, $\tau(ab) = \tau(ba)$, and $\tau(a^*a) \geq 0$ for all $a, b \in \mathcal{M}$, with $\tau(a^*a) = 0$ if and only if $a = 0$ and $\tau(1) = 1$). The operator norm of an element $a \in \mathcal{M}$ is denoted $\|a\|$. We say that an element $a \in \mathcal{M}$ is *nonnegative* if one has $a = b^*b$ for some $b \in \mathcal{M}$. An element $a \in \mathcal{M}$ is *central* if one has $ab = ba$ for all $b \in \mathcal{M}$. The set of all central elements is denoted $\mathcal{Z}(\mathcal{M})$ and referred to as the *centre* of $\mathcal{M}$; the algebra $\mathcal{M}$ is *abelian* if $\mathcal{Z}(\mathcal{M}) = \mathcal{M}$.

A *shift* α on a finite von Neumann algebra $(\mathcal{M}, \tau)$ is trace-preserving $*$-automorphism, that is, α is an algebra isomorphism such that $\alpha(a^*) = \alpha(a)^*$ and $\tau(\alpha(a)) = \tau(a)$ for all $a \in \mathcal{M}$. We say that the shift is *ergodic* if the invariant algebra $\{a \in \mathcal{M} : \alpha(a) = a\}$ consists only of the constants $\mathbb{C}1$. We refer to the triple $(\mathcal{M}, \tau, \alpha)$ as a *von Neumann $\mathbb{Z}$-system*, or a *von Neumann dynamical system*. More generally, if $\alpha_0, \ldots, \alpha_{k-1}$ are k commuting shifts on M, we refer to $(\mathcal{M}, \tau, \alpha_0, \ldots, \alpha_{k-1})$ as a *von Neumann $\mathbb{Z}^k$-system*.

It is easy to verify that if $(X, \mathcal{X}, \mu)$ is a (classical) probability space with a shift $T : X \to X$, then $(L^\infty(X), \int_X \cdot\, d\mu, \circ T^{-1})$ is an (abelian example of a) von Neumann dynamical system, and more generally if $T_0, \ldots, T_{k-1} : X \to X$ are commuting shifts, then $(L^\infty(X), \int_X \cdot\, d\mu, \circ T_0^{-1}, \ldots, \circ T_{k-1}^{-1})$ is an abelian example of a von Neumann $\mathbb{Z}^k$-system. In fact, all abelian von Neumann dynamical systems arise (up to isomorphism of the algebras) as such examples; see [Kadison and Ringrose 1997, Chapter 5].

A finite von Neumann algebra $(\mathcal{M}, \tau)$ gives rise to an inner product $\langle a, b\rangle := \tau(a^*b)$ on $\mathcal{M}$; the properties of the trace ensure that this inner product is positive definite. (We use the convention for a scalar product to be conjugate linear in the first coordinate.) The Hilbert space completion of $\mathcal{M}$ with respect to this inner product will be referred to as $L^2(\tau)$. Note that α extends to a unitary transformation on $L^2(\tau)$. In the abelian case when $\mathcal{M} = L^\infty(X, \mathcal{X}, \mu)$, the space $L^2(\tau)$ can be canonically identified with $L^2(X, \mathcal{X}, \mu)$.

Inspired by Theorems 1.1 and 1.3, we now make the following definitions:

Definition 1.6 (noncommutative recurrence and convergence). Let $k \geq 2$ be an integer, $(\mathcal{M}, \tau, \alpha)$ be a von Neumann dynamical system, and $(\mathcal{M}, \tau, \alpha_0, \ldots, \alpha_{k-1})$ be a von Neumann $\mathbb{Z}^k$-system.

[2] In our applications, the hypothesis of separability can be omitted since one can always pass to the separable subalgebra generated by a finite collection $a_0, \ldots, a_{k-1}$ of elements and their shifts if desired.

- We say $(\mathcal{M}, \tau, \alpha)$ enjoys *order k convergence in norm* if for any $a_1, \ldots, a_{k-1}$ in $\mathcal{M}$, the averages

$$\frac{1}{N}\sum_{n=1}^{N}(\alpha^n(a_1))(\alpha^{2n}(a_2))\cdots(\alpha^{(k-1)n}(a_{k-1}))$$

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

- We say $(\mathcal{M}, \tau, \alpha)$ enjoys *order k weak convergence* if for any $a_0, a_1, \ldots, a_{k-1}$ in $\mathcal{M}$, the averages

$$\frac{1}{N}\sum_{n=1}^{N}\tau(a_0(\alpha^n(a_1))(\alpha^{2n}(a_2))\cdots(\alpha^{(k-1)n}(a_{k-1})))$$

converge as $N \to \infty$.

- We say $(\mathcal{M}, \tau, \alpha)$ enjoys *order k recurrence on average* if for any nonnegative $a \in \mathcal{M}$ with $\tau(a) > 0$, one has

$$\liminf_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\operatorname{Re}\tau(a(\alpha^n(a))(\alpha^{2n}(a))\cdots(\alpha^{(k-1)n}(a))) > 0. \tag{5}$$

- We say that $(\mathcal{M}, \tau, \alpha)$ enjoys *order k recurrence on a dense set* if for any nonnegative $a \in \mathcal{M}$ with $\tau(a) > 0$, one has

$$\operatorname{Re}\tau(a(\alpha^n(a))(\alpha^{2n}(a))\cdots(\alpha^{(k-1)n}(a))) > c > 0 \tag{6}$$

for some $c > 0$ and all n in a set of natural numbers of positive lower density.

- We say $(\mathcal{M}, \tau, \alpha_0, \ldots, \alpha_{k-1})$ *converges in norm* if for any $a_1, \ldots, a_{k-1} \in \mathcal{M}$, the averages

$$\frac{1}{N}\sum_{n=1}^{N}\alpha_0^{-n}((\alpha_1^n(a_1))(\alpha_2^n(a_2))\cdots(\alpha_{k-1}^n(a_{k-1})))$$

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

- We say $(\mathcal{M}, \tau, \alpha_0, \ldots, \alpha_{k-1})$ *converges weakly* if for any $a_0, a_1, \ldots, a_{k-1} \in \mathcal{M}$, the averages

$$\frac{1}{N}\sum_{n=1}^{N}\tau((\alpha_0^n(a_0))(\alpha_1^n(a_1))(\alpha_2^n(a_2))\cdots(\alpha_{k-1}^n(a_{k-1})))$$

converge as $N \to \infty$.

- We say that $(\mathcal{M}, \tau, \alpha_0, \ldots, \alpha_{k-1})$ enjoys *recurrence on average* if for any nonnegative $a \in \mathcal{M}$ with $\tau(a) > 0$, one has

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \operatorname{Re} \tau((\alpha_0^n(a))(\alpha_1^n(a)) \cdots (\alpha_{k-1}^n(a))) > 0. \tag{7}$$

- We say that $(\mathcal{M}, \tau, \alpha)$ enjoys *order k recurrence on a dense set* if for any nonnegative $a \in \mathcal{M}$ with $\tau(a) > 0$, one has

$$\operatorname{Re} \tau((\alpha_0^n(a))(\alpha_1^n(a)) \cdots (\alpha_{k-1}^n(a))) > c > 0. \tag{8}$$

  for some $c > 0$ and all n in a set of natural numbers of positive lower density.

Remark 1.7. As before, we may normalise α_0 to be the identity. Of course, the first four properties here are nothing more than the specialisations of the last four to the case $\alpha_i = \alpha^i$ for $0 \le i \le k-1$. The real part is needed in (5), (6), (7) and (8) because there is no necessity for the traces here to be real-valued (the difficulty being that the product of two nonnegative elements of a nonabelian von Neumann algebra need not remain nonnegative). In the case of (5), one can omit the real part by taking averages from $-N$ to N, since one has the symmetry

$$\begin{aligned}\overline{\tau(a(\alpha^n(a))(\alpha^{2n}(a)) \cdots (\alpha^{(k-1)n}(a)))} &= \tau((a(\alpha^n(a))(\alpha^{2n}(a)) \cdots (\alpha^{(k-1)n}(a)))^*) \\ &= \tau((\alpha^{(k-1)n}(a)) \cdots (\alpha^{2n}(a))(\alpha^n(a))a) \\ &= \tau(a(\alpha^{-n}(a)) \cdots (\alpha^{-(k-1)n}(a)))\end{aligned}$$

for any self-adjoint a.

Note however that it is quite possible for the expressions (6) or (8) to be negative even when a is nonnegative. Because of this, while recurrence on average still implies recurrence on a dense set, the converse is not true; one can have recurrence on a dense set but end up with a zero or even negative average due to the presence of large negative values of (6) or (8). We will see examples of this later.

Remark 1.8. As we said earlier, the Furstenberg correspondence principle equates recurrence results with combinatorial statements (such as Szemerédi's theorem) that can be formulated in a purely finitary fashion. However, we do not know whether the same is true for noncommutative recurrence results. Formulating a finitary statement that would imply recurrence results for some nonabelian von Neumann dynamical system probably requires some quite strong approximate embeddability of the system into finite-dimensional matrix algebras with approximate shifts, together with a recurrence assertion for such finite-dimensional systems in which the various parameters may all be chosen independent of the dimension. Since many of the results we prove below in the infinitary setting are negative anyway, we will not pursue this issue here.

These properties (and related topics) for von Neumann dynamical systems have been studied by Niculescu, Ströh and Zsidó [2003], Duvenhage [2009], Beyers, Duvenhage and Ströh [2010], and Fidaleo [2009]. A variant of these questions, in which one averages over a higher-dimensional range of shifts, was also studied in [Fidaleo 2007]. In this paper we shall develop further positive and negative results regarding these properties, which we now present.

1c. ***Positive results.*** When $k = 2$, all systems enjoy norm and weak convergence, as well as recurrence on average and on a dense set, thanks to the ergodic theorem for von Neumann algebras; see for example [Krengel 1985, Section 9.1]. Indeed, from that theorem, we know that for any von Neumann dynamical system $(\mathcal{M}, \tau, \alpha)$ and $a \in \mathcal{M}$, the averages $N^{-1} \sum_{n=1}^{N} \alpha^n(a)$ converge in $L^2(\tau)$ to the orthogonal projection of a to the invariant space $L^2(\tau)^\alpha := \{f \in L^2(\tau) : \alpha(f) = f\}$, giving the convergence results. If a is nonnegative and nonzero, this projection can be verified to have a positive inner product with a, giving the recurrence results.

Now we consider the cases $k \geq 3$. We have already seen from Theorems 1.1 and 1.3 that we have convergence and recurrence in those abelian systems arising from ergodic theory, and have recalled above that in fact these include all examples (up to isomorphism).

Proposition 1.9. *Let $k \geq 2$. If $(\mathcal{M}, \tau, \alpha)$ is an abelian von Neumann dynamical system, then $(\mathcal{M}, \tau, \alpha)$ enjoys weak convergence and convergence in norm, and recurrence on average and on a dense set.*

More generally, an abelian von Neumann $\mathbb{Z}^k$-system $(\mathcal{M}, \tau, \alpha_0, \ldots, \alpha_{k-1})$ enjoys weak convergence and convergence in norm, and recurrence on average and on a dense set.

We now generalise these results to the wider class of *asymptotically abelian* systems.

Definition 1.10 (asymptotic abelianness). A von Neumann dynamical system $(\mathcal{M}, \tau, \alpha)$ is *asymptotically abelian* if

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \|[\alpha^n(a), b]\|_{L^2(\tau)} = 0 \quad \text{for all } a, b \in \mathcal{M},$$

where $[a, b] := ab - ba$ is the commutator.

Remark 1.11. In previous literature such as [Beyers et al. 2010], a stronger version of asymptotic abelianness is assumed, in which the $L^2(\tau)$ norm is replaced by the operator norm. Variants of this type of "topological asymptotic abelianness", and their relationship with noncommutative topological weak mixing have also been considered in [Kerr and Li 2007].

Our work also singles out this case as special, since the assumption of asymptotic abelianness seems to be essential for the correct working of some the chief technical tools taken from the commutative setting (particularly the van der Corput estimate). In [Niculescu et al. 2003; Beyers et al. 2010; Duvenhage 2009], convergence and recurrence were shown for all orders k for asymptotically abelian systems under some additional assumptions such as weak mixing or compactness. Our first main result shows that in fact all asymptotically abelian systems enjoy convergence and recurrence.

Theorem 1.12. *Let $k \geq 2$. If $(\mathcal{M}, \tau, \alpha)$ is an asymptotically abelian von Neumann dynamical system, then $(\mathcal{M}, \tau, \alpha)$ enjoys weak convergence and convergence in norm, and recurrence on average and on a dense set.*

More generally, if $(\mathcal{M}, \tau, \alpha_0, \dots, \alpha_{k-1})$ is a von Neumann $\mathbb{Z}^k$-system, and the $\alpha_i \alpha_j^{-1}$ for $i \neq j$ are each individually asymptotically abelian, then this $\mathbb{Z}^k$-system enjoys weak convergence and convergence in norm, and recurrence on average and on a dense set.

Theorem 1.12 can be deduced from the genuinely abelian case (Proposition 1.9) by using two results. The first one is essentially from [Beyers et al. 2010] or [Duvenhage 2009], which considered the model case $\alpha_i = \alpha^i$; for the sake of completeness, we present a proof in Appendix A.

Theorem 1.13 (multiple ergodic averages for relatively weakly mixing extensions). *Let $(\mathcal{M}, \tau, \alpha_0, \dots, \alpha_{k-1})$ be a von Neumann $\mathbb{Z}^k$-system, and let $\mathcal{N}$ be a von Neumann subalgebra of $\mathcal{M}$ that is invariant under all of the α_i. If for any distinct $0 \leq i, j \leq k-1$ the shift $\alpha_i \alpha_j^{-1}$ is asymptotically abelian and weakly mixing relative to $\mathcal{N}$, then the associated multiple ergodic averages satisfy*

$$\left\| \frac{1}{N} \sum_{n=1}^{N} \alpha_0^{-n} \prod_{i=1}^{k-1} \alpha_i^n(a_i) - \frac{1}{N} \sum_{n=1}^{N} \alpha_0^{-n} \prod_{i=1}^{k-1} \alpha_i^n(E_{\mathcal{N}}(a_i)) \right\|_{L^2(\tau)} \to 0$$

as $N \to \infty$, where $E_{\mathcal{N}} : \mathcal{M} \to \mathcal{N}$ is the conditional expectation constructed from τ, and the products are from left to right.

We will recall the notions of relative weak mixing and conditional expectation in Section 3.

The second result, which is new and may have other applications elsewhere, can be viewed as a partial analogue for asymptotically abelian systems of the Furstenberg–Zimmer structure theorem [Furstenberg et al. 1982].

Theorem 1.14 (structure theorem for asymptotically abelian systems). *If $(\mathcal{M}, \tau, \alpha)$ is an asymptotically abelian von Neumann dynamical system, then α is weakly mixing relative to the centre $\mathfrak{Z}(\mathcal{M}) \subset \mathcal{M}$.*

Remark 1.15. In the case when $\mathcal{M}$ is a factor (that is, when the centre is trivial), results of this nature (with a slightly different notion of mixing and of asymptotic abelianness) were established in [Bratteli and Robinson 1987, Example 4.3.24].

These results quickly imply Theorem 1.12. Indeed, when studying (for instance) convergence in norm for a $\mathbb{Z}^k$-system, one can use Theorem 1.14 followed by Theorem 1.13 to replace each of the $a_0, \dots, a_{k-1}$ by their conditional expectations $E_{\mathfrak{Z}(\mathcal{M})}(a_0), \dots, E_{\mathfrak{Z}(\mathcal{M})}(a_{k-1})$ without any affect on the convergence, at which point one can apply Proposition 1.9. (Note that the centre $\mathfrak{Z}(\mathcal{M})$ does not depend on what shift $\alpha_i^{-1}\alpha_j$ one is analysing.) The other claims are similar (using Lemma 3.1 to ensure that if a is nonnegative with positive trace, then so is the conditional expectation $E_{\mathfrak{Z}(\mathcal{M})}(a)$).

Remark 1.16. The arguments above in fact show a more quantitative statement: if a is nonnegative with $\|a\| \le 1$ and $\tau(a) \ge \delta$ for some $0 \le \delta \le 1$, then one has the same lower bound $c(k, \delta) \ge 0$ for (6) as is given by Szemerédi's theorem for (1) for nonnegative functions f with $\|f\|_{L^\infty(X)} \le 1$ and $\int_X f \, d\mu \ge \delta$; in particular, one could insert the bound of Gowers [2001]. Similar remarks apply to multiple commuting shifts. We leave the details to the reader.

The proof of Theorem 1.14, given in Section 3 below, rests on noncommutative versions of several of the steps on the way to the Furstenberg–Zimmer structure theorem in the commutative world of ergodic theory [Furstenberg 1977; Zimmer 1976b; 1976a]. In particular, it rests on a version of the dichotomy between relatively weakly mixing inclusions and those containing a relatively isometric subinclusion, well known in ergodic theory from the cited work of Furstenberg and Zimmer and already generalised to the noncommutative world by Popa [2007], for applications to the study of superrigidity phenomena.

If $(\mathcal{M}, \tau, \alpha)$ is not asymptotically abelian, matters are rather more complicated, with positive results only obtaining under additional restrictions. For $k = 3$ and for ergodic shifts, we have a positive result, established in Section 5:

Theorem 1.17. *If $k = 3$ and $(\mathcal{M}, \tau, \alpha)$ is an ergodic von Neumann dynamical system, one has weak convergence and convergence in norm, as well as recurrence on a dense set.*

The weak convergence result was previously established in [Fidaleo 2009].

1d. ***Negative results.*** Recurrence on average cannot be included in Theorem 1.17.

Theorem 1.18. *Let $k = 3$. Then there exists an ergodic von Neumann dynamical system $(\mathcal{M}, \tau, \alpha)$ for which recurrence on average fails.* (*In fact one can make the average* (5) *strictly negative.*)

We establish this in Section 2b. The main tool is a sophisticated version of the Behrend set construction, combined with the crossed product construction.

Without the ergodicity assumption,[3] one also loses recurrence on a dense set:

Theorem 1.19. *Let $k = 3$. There exists a von Neumann dynamical system $(\mathcal{M}, \tau, \alpha)$ for which recurrence on a dense set fails.* (*In fact one can make the means* (6) *equal to a negative constant for all nonzero n.*)

This result, also proved in Section 2b, is simpler to prove than Theorem 1.18, and uses the original Behrend set construction and crossed product constructions.

One loses recurrence on a dense set for larger k even when ergodicity is assumed:

Theorem 1.20. *Let $k \geq 5$ be odd. There exists an ergodic von Neumann dynamical system $(\mathcal{M}, \tau, \alpha)$ for which recurrence on a dense set fails.* (*In fact one can make the means* (6) *equal to a negative constant for all nonzero n.*)

We establish this in Section 2c. This result uses a counterexample of Bergelson, Host, Kra, and Ruzsa [Bergelson et al. 2005] combined with a group theoretic construction. The restriction to odd k is mostly technical and can almost certainly be removed; however, we are unable to decide whether Theorem 1.20 can be extended to the $k = 4$ case because it was shown in [Bergelson et al. 2005] that the $k = 5$ counterexample in that paper cannot be replicated for $k = 4$.

For convergence, we have counterexamples for $k \geq 4$ even when we assume ergodicity:

Theorem 1.21. *Let $k \geq 4$. There exists an ergodic von Neumann dynamical system $(\mathcal{M}, \tau, \alpha)$ for which weak convergence and convergence in norm fail.*

We establish this in Section 2a. The main tool is a group theoretic construction.

The counterexamples above were for the single shift case, but of course they are also counterexamples to the more general situation of multiple commuting shifts. Table 1 summarises the positive and negative results (in the single shift case).

We note in particular that the following questions remain open:

Question 1.22. If $k = 3$, does weak or norm convergence hold for nonergodic von Neumann dynamical systems $(\mathcal{M}, \tau, \alpha)$?

Question 1.23. If $k = 3$, does weak or norm convergence hold for von Neumann $\mathbb{Z}^3$-systems $(\mathcal{M}, \tau, \alpha_0, \alpha_1, \alpha_2)$, (possibly after imposing suitable ergodicity hypotheses)?

Question 1.24. If $k = 4$ (or if $k \geq 6$ is even), does recurrence on a dense set hold for ergodic von Neumann dynamical systems $(\mathcal{M}, \tau, \alpha)$?

---

[3] In the commutative case, an easy application of the ergodic decomposition allows one to recover the nonergodic case of the recurrence and convergence results from the ergodic case. Unfortunately, in the noncommutative case, the ergodic decomposition is only available when the invariant factor $\mathcal{M}^\tau$ is central, which is the case in the asymptotically abelian case, but not in general.

	Conv. norm?	Conv. mean?	Recur. avg.?	Recur. dense?
$k = 2$	Yes	Yes	Yes	Yes
$k = 3$, erg.	Yes	Yes	No	Yes
$k = 3$, nonerg.	???	???	No	No
$k \geq 4$, even, erg.	No	No	No?	???
$k \geq 4$, even, nonerg.	No	No	No?	No?
$k \geq 5$, odd, erg.	No	No	No	No
$k \geq 5$, odd, nonerg.	No	No	No	No

Table 1. Positive and negative results for noncommutative convergence and recurrence of a single shift for various values of k, and for various assumptions of ergodicity. The entries marked "No?" would be expected to have a negative answer if one adopts the principle that recurrence results which fail for one value of k, should also fail for higher values of k.

We present some remarks on the first two problems in Section 6.

Notational remark. Unfortunately this paper stands between two quite unrelated uses of the word "factor", one from operator algebras and one from ergodic theory. In the hope that it may be of interest to operator algebraists, we have deferred to their usage (even though the true notion of a factor due to Murray and von Neumann is actually not essential to our work), and will refer throughout to inclusions of von Neumann algebras, even in the commutative setting where these can be identified with ergodic-theoretic "factors".

2. Counterexamples

In this section we construct various counterexamples of von Neumann systems $(\mathcal{M}, \tau, \alpha)$ that will demonstrate the negative results in Theorems 1.18-1.21. The material in this section is independent of the positive results in the rest of the paper, but may provide some cautionary intuition to keep in mind when reading the proofs of those results.

2a. ***Nonconvergence for $k \geq 4$.*** We first show that convergence results fail for $k \geq 4$, even if one assumes ergodicity. In fact the divergence is so bad that it is essentially arbitrary:

Theorem 2.1 (no convergence for $k \geq 4$). *Let $k \geq 4$ be an integer, and let $A \subset \mathbb{Z}$ be a set. Then there exist an ergodic von Neumann system $(\mathcal{M}, \tau, \alpha)$ and elements $a_0, \ldots, a_{k-1} \in \mathcal{M}$ such that*

$$\tau(a_0 \alpha^n(a_1) \cdots \alpha^{(k-1)n}(a_{k-1})) = 1_A(n) \quad \textit{for all integers } n.$$

It is clear that this implies Theorem 1.21 by choosing A appropriately (and noting that failure of weak convergence implies failure of convergence in norm, by Cauchy–Schwarz applied in the contrapositive).

Proof. It will suffice to verify the $k = 4$ case, as the higher cases follow by setting $a_j = 1$ for $j \geq 4$. We will need a group G with four distinguished elements e_0, e_1, e_2, e_3 and an automorphism $T : G \to G$ such that T^k has no fixed points other than the identity for all $k \neq 0$ and such that

$$e_0(T^r e_1)(T^{2r} e_2)(T^{3r} e_3) = \mathrm{id}$$

holds for all $r \in A$ and fails for all $r \in \mathbb{Z} \backslash A$. Constructing such a group is somewhat nontrivial and is deferred to Appendix B, and in particular to Proposition B.11.

The group algebra $\mathbb{C}G$ of formal finite linear combinations of group elements of G acts (on the left) on the Hilbert space $\ell^2(G)$ in the obvious way (arising from convolution on G) and can thus be viewed as a subspace of the von Neumann algebra $B(\ell^2(G))$; note that all the elements of G become unitary in this perspective. We can place a finite faithful trace τ on $\mathbb{C}G$ by declaring the identity element to have trace 1, and all other elements of G to have trace zero. If we then define $\mathcal{M}$ to be the closure of $\mathbb{C}G$ in the weak operator topology of $B(\ell^2(G))$, we obtain a finite von Neumann algebra, known as the *group von Neumann algebra* LG of G. The shift T leads to an algebra isomorphism α of $\mathbb{C}G$, which then easily extends to a shift α on $\mathcal{M} = LG$. Because none of the powers of T have any nontrivial fixed points, the orbit of any nonzero group element contains no repetitions, and so one can easily establish that $\alpha^n f$ converges weakly to $\tau(f)$ as $n \to \infty$ for every $f \in \mathbb{C}G$ and hence by approximation that the unitary operator on $\ell^2(G)$ associated to α has no fixed points outside $\mathbb{C}\delta_{\mathrm{id}}$. This implies that $(\mathcal{M}, \tau, \alpha)$ is ergodic, since given $a \in \mathcal{M}$ for which $\alpha(a) = a$ and $\tau(a) = 0$ it follows that $a(\delta_{\mathrm{id}}) \in \ell^2(G)$ is a fixed point for the action of T on $\ell^2(G)$, which must therefore equal $\tau(a)\delta_{\mathrm{id}} = 0$, and hence $\tau(a^*a) = \|a(\delta_{\mathrm{id}})\|_2^2 = 0$ and so $a = 0$, by the faithfulness of τ. If we now set $a_j = e_j$ for $j = 0, 1, 2, 3$, we obtain the claim. □

Remark 2.2. An inspection of the proofs of Theorem 2.1 and Proposition B.11 shows that the expression $a_0\alpha^n(a_1)\alpha^{2n}(a_2)\alpha^{3n}(a_3)$ can more generally be replaced by $\alpha^{c_0 n}(a_0)\alpha^{c_1 n}(a_1)\alpha^{c_2 n}(a_2)\alpha^{c_3 n}(a_3)$ whenever c_0, c_1, c_2, c_3 are integers such that $c_i \neq c_{i+1}$ for all $i = 0, 1, 2, 3$ (with the cyclic convention $c_{i+4} = c_i$). Thus for instance one can construct von Neumann systems for which

$$\tau(a_0(\alpha^n(a_1))a_2\alpha^n(a_3)) = 1_A(n)$$

for an arbitrary set A. We omit the details.

Remark 2.3. The examples of nonconvergence given above are not self-adjoint or positive, and the a_i are not equal to each other. However, it is not hard to modify the examples to give an example of a positive $a_i = a$ for which the averages $N^{-1}\sum_{n=1}^{N}\tau(a\alpha^n(a)\alpha^{2n}(a)\alpha^{3n}(a))$ do not converge. Indeed, one can repeat the above construction with

$$a := \mathrm{id} + \frac{1}{100}\sum_{i=0}^{3}(e_i + e_i^*);$$

this is easily seen to be positive and self-adjoint, and a modification of the above computations then shows that

$$\tau(a\alpha^n(a)\alpha^{2n}(a)\alpha^{3n}(a)) = 1 + \frac{2}{100^4}1_A(n) \quad \text{for all } n,$$

which is enough to ensure divergence by choosing A appropriately. We leave the details to the reader.

Remark 2.4. The group G constructed here can easily be shown to have infinite conjugacy classes (by the same methods used to prove Proposition B.11). This implies that the group algebra LG is a factor. See [Kadison and Ringrose 1997, Theorem 6.7.5] for details.

2b. ***Negative averages for $k = 3$.*** We now show the negativity of various triple averages. The main tool is the following Behrend-type construction of a set that avoids progressions of length three, but contains many "hexagons":

Lemma 2.5 (Behrend-type example). *Let $\varepsilon > 0$. Then for all sufficiently large d, there exists a subset F of $\mathbb{Z}/d\mathbb{Z}$ such that $|F| \geq d^{1-\varepsilon}$, but F contains no nontrivial arithmetic progressions of length three; thus $n, n+r, n+2r \in F$ can only occur if $r = 0$. On the other hand, the set*

$$\{(x, h, k) \in \mathbb{Z}/d\mathbb{Z} : x, x+h, x+k, x+k+2h, x+2k+h, x+2k+2h \in F\}$$

of "hexagons" in F has cardinality at least $d^{3-\varepsilon}$.

The first part of the lemma already follows directly from [Behrend 1946] or the earlier [Salem and Spencer 1942]. The claim about hexagons will be needed in the proof of Theorem 2.11 below, but is not needed for the simpler results in Corollary 2.7 or Theorem 2.10.

Proof. Let R be a large multiple of 400 (depending on ε). We claim that for n a large enough multiple of 4 (depending on R), the set $\{-R, \ldots, R\}^n \subset \mathbb{Z}^n$ contains a subset E of cardinality $|E| \geq e^{-O(n)}R^n$ (where the implied constant in the O notation is absolute), and which contains $\geq e^{-O(n)}R^{3n}$ hexagons

$$\{x, x+h, x+k, x+k+2h, x+2k+h, x+2k+2h\}$$

but contains no arithmetic progressions of length three. Choosing d sufficiently large, letting n be the largest integer such that $(10R)^n \leq d$, and then embedding

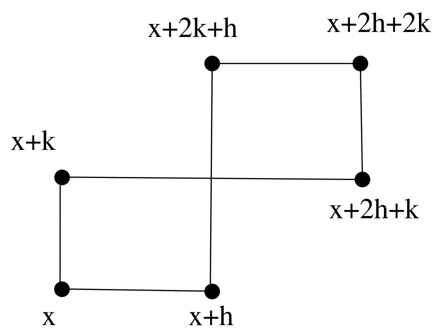

Figure 1. A hexagon. Note the absence of arithmetic progressions of length three.

$\{-R, \dots, R\}^n$ in $\mathbb{Z}/d\mathbb{Z}$ using base $10R$ (say), as in the work of Behrend or Salem and Spencer, this claim will imply the lemma (after choosing R sufficiently large depending on ε).

The claim itself remains. From the classical results on the Waring problem (see for example [Vaughan 1997]), we know that every large integer N has $\sim N^{(k-2)/2}$ representations as the sum of k squares for k large enough (one can for instance take $k = 5$, but for our purposes any fixed k will suffice). Using this, we see that for any fixed $\delta \in (0, \frac{1}{10})$, every integer r such that $\delta R^2 n \le r \le \frac{1}{10} R^2 n$ (say) will have $\ge (c_\delta R)^{n-C_\delta}$ representations as the sum of n squares of integers less than R, where $c_\delta, C_\delta > 0$ depend only on δ. In other words, the sphere

$$E_r := \{x \in \{-R, \dots, R\}^n : |x|^2 = r\}$$

has cardinality at least $(c_\delta R)^{n-C_\delta}$. On the other hand, such spheres have no nontrivial progressions of length three. Thus it will suffice (for n large enough) by the pigeonhole principle to show that there are at least $e^{-O(n)} R^{3n}$ hexagons

$$\{x, x+h, x+k, x+k+2h, x+2k+h, x+2k+2h\} \quad \text{in } \{-R, \dots, R\}^n$$

such that

$$|x|^2 = |x+h|^2 = |x+k|^2 = |x+k+2h|^2 = |x+2k+h|^2 = |x+2k+2h|^2 \le \tfrac{1}{10} R^2 n \tag{9}$$

(note that the case when $|x|^2 \le \delta R^2 n$ for sufficiently small δ can be eliminated by crude estimates).

To count the solutions to (9), we perform some elementary changes of variable to replace the constraints in (9) with simpler constraints. We begin by observing that if $a, b, c \in \{-R/100, \dots, R/100\}^n$ are such that

$$a \cdot b = b \cdot c = c \cdot a = 0 \quad \text{and} \quad c \cdot c = 3b \cdot b, \tag{10}$$

then $x := a - 2b,\ \ h := b + c,\ \ k := b - c$ can be verified to be a solution to (9), with the map $(a, b, c) \to (x, h, k)$ being injective, so it suffices to show that there are at least $e^{-O(n)} R^{3n}$ triples (a, b, c) with the properties above.

For reasons that will become clearer later, we will initially work in dimension $n/4$ rather than n. Using the Waring problem results as before, we can find at least $e^{-O(n)} R^{3n/4}$ triples $a, b, c \in \{-R/400, \dots, R/400\}^{n/4}$ such that

$$c \cdot c = 3b \cdot b.$$

This is one of the four constraints required for (10). To obtain the remaining ones, we use a pigeonholing trick followed by a tensor power trick. First, observe that if $a, b, c \in \{-R/400, \dots, R/400\}^{n/4}$, then $a \cdot b, b \cdot c, c \cdot a$ are of order $O(R^2 n) \leq e^{O(n)}$. Applying the pigeonhole principle, one can thus find $h_1, h_2, h_3 = O(R^2 n)$ such that there are $e^{-O(n)} R^{3n/4}$ triples

$$a, b, c \in \{-R/400, \dots, R/400\}^{n/4} \tag{11}$$

with

$$a \cdot b = h_1, \quad b \cdot c = h_2, \quad c \cdot a = h_3, \quad c \cdot c = 3b \cdot b. \tag{12}$$

This is an inhomogeneous version of (10) (at dimension $n/4$ rather than n), with the zero coefficients replaced by more general coefficients h_1, h_2, h_3. To eliminate these coefficients we use a tensor power trick. Let S be the set of all triples (a, b, c) obeying (11) and (12). We then observe that if $(a_i, b_i, c_i) \in S$ for $i = 1, 2, 3, 4$, then the vectors $a, b, c \in \mathbb{Z}^n$ defined by

$$a := (a_1, a_2, a_3, a_4); \quad b := (b_1, b_2, -b_3, -b_4); \quad c := (c_1, -c_2, c_3, -c_4)$$

solve (10). The map from the (a_i, b_i, c_i) to (a, b, c) is an injection from S^4 to the solution set of (10), and so we obtain at least $|S|^4 \geq e^{-O(n)} R^{3n}$ solutions to (10) as required. □

This leads to a useful matrix counterexample:

Lemma 2.6 (restricted third moment can be negative). *There exists a positive semi-definite Hermitian matrix $(A(j, k))_{1 \leq j,k \leq d}$ for which the quantity*

$$\sum_{n,r \in \mathbb{Z}/d\mathbb{Z}} A(n, n+r) A(n+r, n+2r) A(n+2r, n) \tag{13}$$

is negative, where we extend $A(i, j)$ periodically in both variables by d.

Proof. We will take d to be a multiple of 3, and $A(j, k)$ to take the form

$$A(j, k) := 1_E(j) 1_E(k) + 1_E(j) \omega^{-j} 1_E(k) \omega^k,$$

where $E \subset \mathbb{Z}/d\mathbb{Z}$ is a set to be determined later and $\omega := e^{2\pi i/3}$ is a cube root of unity. The matrix $(A(j,k))_{1\leq j,k\leq d}$ is then the sum of two rank one projections and is thus positive semidefinite and Hermitian. The expression (13) can be expanded as

$$\sum_{\substack{n,r\in\mathbb{Z}/d\mathbb{Z}:\\ n,n+r,n+2r\in E}} (1+\omega^r)(1+\omega^r)(1+\omega^{-2r}).$$

The summand can be computed to equal 8 when r is divisible by 3, and -1 otherwise. Thus, to establish the claim, it suffices to find a set E such that the set

$$\{(n,r)\in\mathbb{Z}/d\mathbb{Z} : n, n+r, n+2r \in E,\ r \neq 0 \bmod 3\}$$

is more than eight times larger than the set

$$\{(n,r)\in\mathbb{Z}/d\mathbb{Z} : n, n+r, n+2r \in E,\ r = 0 \bmod 3\};$$

thus the length three arithmetic progressions in E with spacing not divisible by 3 need to overwhelm the length three progressions with spacing divisible by 3.

To do this, we use Lemma 2.5 to get a subset $F \subset \{1,\ldots,[d/10]\}$ of cardinality $|F| \geq d^{0.99}$ that contains no arithmetic progressions of length three. We then pick three random shifts $h_0, h_1, h_2 \in \{1,\ldots,d/3\}$ uniformly at random, and consider the set

$$E := \{3(f+h_i)+i : i = 0,1,2,\ f\in F\}$$

consisting of three randomly shifted, dilated copies of F.

By construction, the only length three progressions in E with spacing divisible by 3 are the trivial progressions n, n, n with $r = 0$, so the total number of such progressions is at most d. On the other hand, for any fixed $f_0, f_1, f_2 \in F$, the numbers $3(f_i+h_i)+i$ for $i = 0,1,2$ have a probability $3/d$ of forming an arithmetic progression with spacing not divisible by 3, due to the random nature of the h_i. Thus the expected value of the total number of such progressions is at least $(d^{0.99})^3 \times 3/d = 3d^{1.97}$. For d large enough, this gives the claim. □

This gives a simple example of negative averages for nonergodic systems:

Corollary 2.7 (negative average for nonergodic system). *There exists a finite von Neumann algebra $(\mathcal{M},\tau)$ with a shift α and a nonnegative element $a\in\mathcal{M}$, such that $(2N+1)^{-1}\sum_{n=-N}^{N}\tau(a\alpha^n(a)\alpha^{2n}(a))$ converges to a negative number.*

Proof. Let $a = (A(j,k))_{1\leq j,k\leq d}$ be as in Lemma 2.6. We let $\mathcal{M}$ be the von Neumann algebra of complex $d\times d$ matrices with the normalised trace τ and with the shift

$$\alpha(B(j,k))_{1\leq j,k\leq d} := (e^{2\pi i(j-k)/d}B(j,k))_{1\leq j,k\leq d}.$$

This is easily verified to be a shift. We see that

$$\tau(a\alpha^n(a)\alpha^{2n}(a)) = \frac{1}{d} \sum_{j,k,l\in\mathbb{Z}/d\mathbb{Z}} e^{2\pi i n(k+l-2j)/d} A(j,k)A(k,l)A(l,j).$$

This expression is periodic in n with period d and has average

$$\frac{1}{d} \sum_{l,r\in\mathbb{Z}/d\mathbb{Z}} A(l,l+r)A(l+r,l+2r)A(l+2r,l)$$

and the claim then follows from Lemma 2.6. □

This shows that recurrence on average for $k=3$ can fail for nonergodic systems. However, this is not yet enough to establish either Theorem 1.18 or Theorem 1.19. To obtain these stronger results we must introduce the *crossed product construction* in von Neumann algebras. For a comprehensive introduction to this concept, see [Kadison and Ringrose 1997, Chapter 13]. We shall just recall the key properties of this construction we need here.

Suppose we have a finite von Neumann algebra $(\mathcal{M}, \tau)$, and an action U of a (discrete) group G on $\mathcal{M}$; thus for each $g \in G$ we have a shift $U(g) : \mathcal{M} \to \mathcal{M}$ such that $U(g)U(h) = U(gh)$ for all $g, h \in G$, with $U(\mathrm{id})$ being the identity. Then there exists a crossed product $(\mathcal{M} \rtimes_U G, \tau)$ that contains both the original space $(\mathcal{M}, \tau)$ and the group algebra $\mathbb{C}G$ as subalgebras. Furthermore, in this crossed product we have

$$U(g)a = gag^{-1} \tag{14}$$

for all $a \in \mathcal{M}$ and $g \in G$, and

$$\tau(ga) = \tau(ag) = 0$$

for all $a \in \mathcal{M}$ and $g \in G$ with g not equal to the identity. Finally, the span of the elements ag for $a \in \mathcal{M}$ and $g \in G$ is dense in $\mathcal{M} \rtimes_U G$.

Remark 2.8. The exact construction of the crossed product is not relevant for our applications, but for the convenience of the reader we sketch one such construction here. We first form the Hilbert space

$$\mathfrak{h} := \ell^2(G, L^2(\tau)) = \bigoplus_{g\in G} L^2(\tau)$$

consisting of tuples $(x_g)_{g\in G}$ in $L^2(\tau)$. This space has an action of $\mathcal{M}$ defined by

$$a(x_g)_{g\in G} := ((U(g^{-1})a)x_g)_{g\in G}$$

for $a \in \mathcal{M}$, and an action of G (and hence $\mathbb{C}G$) defined by

$$h(x_g)_{g\in G} := (x_{h^{-1}g})_{g\in G}.$$

One can verify that these actions combine to an action of the twisted convolution algebra $\ell^1(G, \mathcal{M})$ on $\mathfrak{h}$, defined as the space of formal sums $\sum_{h \in G} h a_h$ with $\sum_{h \in G} \|a_h\| < \infty$, and subject to the relations (14). We define a trace on such sums by the formula $\tau(\sum_{h \in G} h a_h) := \tau(a_{\mathrm{id}})$. One can then show that one can extend this to a finite trace on the weak operator topology closure of $\ell^1(G, \mathcal{M})$, viewed as a subset of $B(\mathfrak{h})$; this closure can then be denoted $\mathcal{M} \rtimes_U G$. In other words, $\mathcal{M} \rtimes_U G$ is constructed as the von Neumann algebra generated by the action of $\mathcal{M}$ and G on $\mathfrak{h}$.

Example 2.9. The group von Neumann algebra LG can be viewed as $\mathbb{C} \rtimes G$, where G acts trivially on the one-dimensional von Neumann algebra $\mathbb{C}$.

We can now get a stronger version of Corollary 2.7:

Theorem 2.10 (negative trace for nonergodic system). *There exists a von Neumann dynamical system $(\mathcal{M}, \tau, \alpha)$ and a nonnegative element $a \in \mathcal{M}$, such that $\tau(a\alpha^n(a)\alpha^{2n}(a))$ is negative (and independent of n) for all nonzero n. In particular, Theorem 1.19 holds.*

Proof. Let $(\mathcal{M}', \tau, \beta)$ be a von Neumann dynamical system to be chosen later. Using the crossed product construction, we can build an extension $\mathcal{M} := \mathcal{M}' \rtimes_U \mathbb{Z}^2$ of $\mathcal{M}'$ generated by $\mathcal{M}'$ and two commuting unitary elements u and m, such that

$$mam^{-1} = \beta(a) \tag{15}$$

and $uau^{-1} = a$ for all $a \in \mathcal{M}'$. In particular, the element u is central. It is then easy to see that we can build[4] a shift α on $\mathcal{M}$ for which

$$\alpha(a) = a, \quad \alpha(u) = u, \quad \alpha(m) = mu$$

for all $a \in \mathcal{M}'$, since the action of the group $\mathbb{Z}^2$ generated by m and u on $\mathcal{M}'$ is unchanged when one replaces m by mu.

Now let $a \in \mathcal{M}$ be an element of the form

$$a = \Big(\sum_{i \in \mathbb{Z}} f_i m^i\Big)\Big(\sum_{i \in \mathbb{Z}} f_i m^i\Big)^*,$$

where $f_i \in \mathcal{M}'$ and only finitely many of the f_i are nonzero. This is clearly nonnegative, and can be simplified by (15) to the power series

$$a = \sum_{h \in \mathbb{Z}} g_h m^h,$$

[4]To build α explicitly, we can view $\mathcal{M}$ as an algebra of operators on the Hilbert space $\mathfrak{h} := \bigoplus_{(j,k) \in \mathbb{Z}^2} L^2(\tau)$ as per Remark 2.8, and let α be the conjugation $a \mapsto WaW^*$ by the unitary operator $W : \mathfrak{h} \to \mathfrak{h}$ defined by $W(x_{(j,k)})_{(j,k) \in \mathbb{Z}^2} := (x_{(j,k-j)})_{(j,k) \in \mathbb{Z}^2}$.

where the $g_h \in \mathcal{M}'$ are the twisted autocorrelations of the f_j, given by

$$g_h = \sum_{j\in\mathbb{Z}} f_{j+h}\beta^h(f_j^*).$$

Let n be nonzero. The expression $\tau(a\alpha^n(a)\alpha^{2n}(a))$ can be expanded as

$$\sum_{h_1,h_2,h_3\in\mathbb{Z}} \tau(g_{h_1}m^{h_1}g_{h_2}(mu^n)^{h_2}g_{h_3}(mu^{2n})^{h_3}).$$

The net power of the central element u here is $n(h_2+2h_3)$, and the net power of m is $h_1+h_2+h_3$. Thus we see that the trace vanishes unless $h_2+2h_3 = h_1+h_2+h_3 = 0$, or equivalently if $(h_1,h_2,h_3) = (h,-2h,h)$ for some h. Performing this substitution and using (15), we simplify this expression to

$$\sum_{h\in\mathbb{Z}} \tau(g_h\beta^h(g_{-2h})\beta^{-h}(g_h)). \tag{16}$$

In particular, this expression is now manifestly independent of $n \neq 0$.

We now select $\mathcal{M}'$ to be the commutative von Neumann system $L^\infty(\mathbb{Z}/d\mathbb{Z})$ with the shift $\beta(f(x)) := f(x+1)$ and the normalised trace. Thus the g_h and f_h are now complex-valued functions on $\mathbb{Z}/d\mathbb{Z}$, and the expression above can be expanded explicitly as

$$\frac{1}{d}\sum_{x\in\mathbb{Z}/d\mathbb{Z}}\sum_{h\in\mathbb{Z}} g_h(x)g_{-2h}(x+h)g_h(x-h).$$

Meanwhile, the $g_h(x)$ by definition can be written as

$$g_h(x) = \sum_{j\in\mathbb{Z}} f_{j+h}(x)\overline{f_j(x+h)}.$$

We pick a large number N to be chosen later, and set

$$f_j(x) := b(x,x+j)1_{1\le j\le Nd},$$

where $b : \mathbb{Z}/d\mathbb{Z}\times\mathbb{Z}/d\mathbb{Z} \to \mathbb{C}$ is a function periodic in two variables of period d to be chosen later. Then we can compute

$$g_h(x) = \left(1-\frac{|h|}{dN}\right)_+ NA(x,x+h) + O(1),$$

where

$$A(x,y) := \sum_{z\in\mathbb{Z}/d\mathbb{Z}} b(x,z)\overline{b(y,z)} \tag{17}$$

and $O(1)$ denotes a quantity that can depend on d (and b) but is uniformly bounded in N. The expression (16) can then be computed to be

$$C\frac{N^4}{d}\sum_{x,h\in\mathbb{Z}/d\mathbb{Z}} A(x,x+h)A(x+h,x-h)A(x-h,x)+O(N^3),$$

where $C>0$ is the explicit constant $C := \int_{\mathbb{R}}(1-|h|)_+^2(1-|2h|)_+\,dh$. By the substitutions $x=m+r$ and $h=r$, we can reexpress this as

$$C\frac{N^4}{d}\sum_{m,r\in\mathbb{Z}/d\mathbb{Z}} A(m,m+r)A(m+r,m+2r)A(m+2r,m)+O(N^3). \tag{18}$$

Now, let d and $A(j,k)$ be as in Lemma 2.6. By the spectral theorem (which in particular allows one to construct self-adjoint square roots of positive definite matrices), we can find $b(x,y)$ such that (17) holds. The summand in (18) is then negative, and the claim follows by choosing N large enough depending on all other parameters. □

Of course, by Theorem 1.17, one cannot have such a result when the underlying shift α is ergodic. On the other hand, one can extend Corollary 2.7 to the ergodic case:

Theorem 2.11. *There exists an* ergodic *von Neumann system* $(\mathcal{M},\tau,\alpha)$ *and a nonnegative element* $a\in\mathcal{M}$, *such that* $(2N+1)^{-1}\sum_{n=-N}^{N}\tau(a\alpha^n(a)\alpha^{2n}(a))$ *converges to a negative number. In particular*, *Theorem 1.18 holds.*

Proof. Let d be a large odd number, and let $u := e^{2\pi i/d}$ be a primitive d-th root of unity. We will let $\mathcal{M}$ be a completion of the *noncommutative torus*. This is obtained by first forming the C*-algebra generated by two unitary generators e_1 and e_2 obeying the commutation relation

$$e_2e_1 = ue_1e_2$$

and with all of the expressions $e_1^je_2^k$ having zero trace unless $j=k=0$, in which case the trace is 1, and then completing in the weak operator topology resulting from the Gel'fand–Naimark–Segal representation on $L^2(\tau)$. One can represent this finite von Neumann algebra more explicitly by letting e_1 and e_2 act on $L^2((\mathbb{R}/\mathbb{Z})^2)$ by the maps $e_1f(x,y) := e^{2\pi ix}f(x,y)$ and $e_2f(x,y) := e^{2\pi iy}f(x+1/d,y)$, with the trace τ given by $\tau(a)=\langle\Omega,a\Omega\rangle_{L^2((\mathbb{R}/\mathbb{Z})^2)}$, where $\Omega\equiv 1$ is the identity function on $(\mathbb{R}/\mathbb{Z})^2$.

We let $\theta_1,\theta_2\in S^1$ be generic unit phases, and then define the shift α on $\mathcal{M}$ by setting

$$\alpha(e_1) := \theta_1e_1 \quad\text{and}\quad \alpha(e_2) := \theta_2e_2.$$

It is easy to see that this is a shift. If θ_1 and θ_2 are generic (so that $\theta_1^j \theta_2^k$ is not a root of unity for any $(j,k) \neq (0,0)$), this shift is easily verified to be ergodic (as one can verify the mean ergodic theorem by hand on the generators $e_1^j e_2^k$, and then argue as in the proof of Theorem 2.1 using the faithfulness of τ).

We set $a := gg^*$, where g is an element of the form $g := \sum_{k=1}^M \sum_{h \in \mathbb{Z}} c_h e_1^h e_2^k$, M is a large number (much larger than d) to be chosen later, and c_h are complex numbers to be chosen later, all but finitely many of which are zero. Clearly a is nonnegative. A computation shows that

$$a = \sum_{h,k \in \mathbb{Z}} c_{h,k} e_1^h e_2^k, \quad \text{where } c_{h,k} := M\left(1 - \frac{|k|}{M}\right)_+ \sum_{l \in \mathbb{Z}} c_{l+h} \overline{c_l} u^{kl}. \tag{19}$$

Since

$$\alpha^n(a) = \sum_{h,k \in \mathbb{Z}} c_{h,k} \theta_1^{hn} \theta_2^{kn} e_1^h e_2^k,$$

some Fourier analysis and the genericity of θ_1 and θ_2 show that the expression

$$\frac{1}{2N+1} \sum_{n=-N}^{N} \tau(a \alpha^n(a) \alpha^{2n}(a))$$

converges as $N \to \infty$ to the expression

$$\sum_{h,k} c_{h,k} c_{-2h,-2k} c_{h,k} \tau(e_1^h e_2^k e_1^{-2h} e_2^{-2k} e_1^h e_2^k).$$

The trace here simplifies to u^{3hk}. Inserting the expression for $c_{h,k}$ in (19), we can expand this expression as

$$M^3 \sum_{h,k,l_1,l_2,l_3 \in \mathbb{Z}} \phi(k/M) c_{l_1+h} \overline{c_{l_1}} c_{l_2-2h} \overline{c_{l_2}} c_{l_3+h} \overline{c_{l_3}} u^{kl_1 - 2kl_2 + kl_3 + 3hk}, \tag{20}$$

where $\phi(x) := (1-|x|)_+^2 (1-|2x|)_+$. By Poisson summation, the expression

$$\sum_k \phi(k/M) u^{kl_1 - 2kl_2 + kl_3 + 3hk}$$

can be computed to be $M \int_{\mathbb{R}} \phi(x)dx + O(1)$ if $l_1 - 2l_2 + l_3 + 3h$ is divisible by d, and $O(1)$ otherwise, where $O(1)$ denotes a quantity that can depend on d but is bounded uniformly in M. If we then assume that the c_h vanish for h outside of $\{1, \dots, M\}$ and are bounded uniformly in M, we can thus expand (20) as

$$CM^4 \sum_{\substack{h,l_1,l_2,l_3 \in \mathbb{Z}: \\ d | l_1 - 2l_2 + l_3 + 3h}} c_{l_1+h} \overline{c_{l_1}} c_{l_2-2h} \overline{c_{l_2}} c_{l_3+h} \overline{c_{l_3}} + O(M^7)$$

for some absolute constant $C > 0$.

If we now set $c_h := b(h)1_{[1,M]}(h)$, where $b : \mathbb{Z}/d\mathbb{Z} \to \mathbb{C}$ is a periodic function with period d and independent of M to be chosen later, we can express this as

$$C_d M^8 \sum_{\substack{h,l_1,l_2,l_3\in\mathbb{Z}/d\mathbb{Z}:\\ l_1-2l_2+l_3+3h=0}} b(l_1+h)\overline{b(l_1)}b(l_2-2h)\overline{b(l_2)}b(l_3+h)\overline{b(l_3)} + O(M^7)$$

for some $C_d > 0$ depending on d but independent of M. Making the substitutions $l_1 = x,\ l_2 = x+k+2h$ and $l_3 = x+2k+h$, we see that we will be done as soon as we are able to find d and b for which the expression

$$X := \sum_{x,h,k\in\mathbb{Z}/d\mathbb{Z}} \overline{b(x)}b(x+h)b(x+k)\overline{b(x+k+2h)}\,\overline{b(x+2k+h)}b(x+2k+2h)$$

is negative.

To do this, we again appeal to Lemma 2.5 to find a set $F \subset \mathbb{Z}/d\mathbb{Z}$ of size at least $d^{0.99}$ (assuming d large enough), which contains no arithmetic progressions of length three, but contains at least $d^{2.99}$ hexagons x, $x+h$, $x+k$, $x+k+2h$, $x+2k+h$, $x+2k+2h$. We then set $b(x) := \epsilon_x 1_F(x)$, where the $\epsilon_x = \pm 1$ are independent signs; thus X is now the random variable

$$X = \sum \epsilon_x\epsilon_{x+h}\epsilon_{x+k}\epsilon_{x+2h+k}\epsilon_{x+h+2k}\epsilon_{x+2h+2k},$$

where the sum is over $\{x, h, k : x, x+h, x+k, x+k+2h, x+2k+h, x+2k+2h \in F\}$. We will show (for d large enough) that the standard deviation of X exceeds its expectation, which shows that there exists a choice of signs for which X is negative.

We first compute the expectation of X. The only summands with nonzero expectation occur when all the signs cancel, which only occurs when $h = 0$ or when $k = 0$, as can be seen by an inspection of the number of ways to collapse the hexagon in Figure 1; here we need the hypothesis that d is odd. But since F contains no nontrivial arithmetic progressions, there are no summands for which only one of the h, k are zero, so we are left only with the $h = k = 0$ terms, of which there are at most d. Thus the expectation of X is at most d.

Now we compute the variance. There are at least $d^{2.99}$ hexagons in F, and all but $O(d^2)$ of them are nondegenerate in the sense that the six vertices of the hexagon are all distinct. The summands in X corresponding to nondegenerate hexagons have variance 1, and the correlation between any two summands in X is either zero or positive (the latter occurs when two summands are permutations of each other). Thus the variance of X is $\gg d^{2.99}$, so the standard deviation is $\gg d^{1.495}$, and the claim follows. □

2c. ***Negative trace for $k = 5$.*** Now we show negative traces can occur even in the ergodic case when $k = 5$.

Theorem 2.12. *There exists an* ergodic *von Neumann dynamical system* $(\mathcal{M}, \tau, \alpha)$ *and a nonnegative element* $a \in \mathcal{M}$*, such that* $\tau(a\alpha^n(a)\alpha^{2n}(a)\alpha^{3n}(a)\alpha^{4n}(a))$ *is negative for every nonzero* n*.*

This establishes the $k = 5$ case of Theorem 1.20. A similar argument holds for all larger odd values of k, which we leave to the interested reader; we restrict here to the case $k = 5$ simply for ease of notation.

To prove this theorem, our starting point is the following result of Bergelson, Host, Kra, and Ruzsa [Bergelson et al. 2005]:

Theorem 2.13. *For any* $\delta > 0$*, there is a measure-preserving system* $(X, \mathcal{X}, \mu, S)$ *and a measurable set* $A \subset X$ *with* $0 < \mu(A) < \delta$ *such that*

$$\mu(A \cap S^n(A) \cap S^{2n}(A) \cap S^{3n}(A) \cap S^{4n}(A)) \leq \mu(A)^{100}$$

(*say*) *and*

$$\mu(A \cap S^n(A)) = \mu(A)^2 \tag{21}$$

for every nonzero integer n*.*

Proof. This follows from [Bergelson et al. 2005, Theorem 1.3] (see also the remark immediately below that theorem). The property (21) is not explicitly stated in that theorem, but follows from the construction in [Bergelson et al. 2005, Section 2.3] (the system X is a torus $(\mathbb{R}/\mathbb{Z})^2$ with the skew shift $S : (x, y) \mapsto (x+\alpha, y+2x+\alpha)$, and the set A has the special form $A = (\mathbb{R}/\mathbb{Z}) \times B$ for some set B). □

We apply this theorem for some sufficiently small δ (to be chosen later) to obtain X, μ, S, A with the properties above. We will combine this with the group G, the automorphism T, and the elements e_0, e_1, e_2, e_3, e_4 arising from Proposition B.13 as follows.

First, we create the product space $L^\infty(X^G, d\mu^G)$, whose σ-algebra is generated up to negligible sets by the tensor products $\bigotimes_{g\in G} f_g$, where $f_g \in L^\infty(X, d\mu)$ is equal to 1 for all but finitely many g. This product has a unitary, trace-preserving action U of G, defined by

$$U(h) \bigotimes_{g\in G} f_g := \bigotimes_{g\in G} f_{h^{-1}g}.$$

We can therefore create the crossed product $\mathcal{M} := L^\infty(X^G, d\mu^G) \rtimes_U G$. Note that if we embed $L^\infty(X, \mu)$ into $L^\infty(X^G, d\mu^G)$ by using the identity component of X^G, we have

$$\bigotimes_{g\in G} f_g = \prod_{g\in G} U(g) f_g \tag{22}$$

(note that the $U(g) f_g$ necessarily commute with each other).

We define a shift α on $\mathcal{M}$ by requiring that

$$\alpha\Big(\bigotimes_{g\in G} f_g\Big) = \bigotimes_{g\in G} S(f_{T^{-1}g}) \quad \text{and} \quad \alpha(g) = Tg;$$

one can check that this is indeed a well-defined shift on $\mathcal{M}$.

We claim that α is ergodic. Indeed, if $a \in \mathcal{M}$ is of the form $a = fg$ for some $f \in L^\infty(X^G, d\mu^G)$ and $g \in G$ not equal to the identity, then since the powers of T have no nontrivial fixed points, the orbit $T^n g$ escapes to infinity, and the orbit $\alpha^n(a)$ converges weakly to zero. Meanwhile, if g is the identity, then it is classical that the Bernoulli system $G \circlearrowright L^\infty(X^G, d\mu^G)$ is ergodic, and so the ergodic theorem applies to a in this case. Putting the two facts together and arguing as for the ergodicity in Theorem 2.1 yields the ergodicity of α.

Note that 1_A lies in $L^\infty(X, d\mu)$, and can thus be identified with an element of $\mathcal{M}$ by the previous embedding. We set

$$a := \sum_{i=0}^{3} 1_A \cdot (2 - e_i - e_i^{-1}) \cdot 1_A.$$

Clearly a is nonnegative. Now let n be nonzero, and consider the expression

$$\tau(a\alpha^n(a)\alpha^{2n}(a)\alpha^{3n}(a)\alpha^{4n}(a)). \tag{23}$$

Expanding out a, we obtain a linear combination of terms of the form

$$\begin{aligned}\tau(&1_A g_0 1_A 1_{S^n(A)}(T^n g_1) 1_{S^n(A)} 1_{S^{2n}(A)}(T^{2n} g_2)\\ &\cdot 1_{S^{2n}(A)} 1_{S^{3n}(A)}(T^{3n} g_3) 1_{S^{3n}(A)} 1_{S^{4n}(A)}(T^{4n} g_4) 1_{S^{4n}(A)}),\end{aligned}$$

where $g_0, g_1, g_2, g_3, g_4 \in \{\mathrm{id}, e_0, e_1, e_2, e_3, e_4, e_0^{-1}, e_1^{-1}, e_2^{-1}, e_3^{-1}, e_4^{-1}\}$. This trace vanishes unless

$$g_0 T^n g_1 T^{2n} g_2 T^{3n} g_3 T^{4n} g_4 = \mathrm{id}\,. \tag{24}$$

By Proposition B.13, we conclude that g_0, g_1, g_2, g_3, g_4 are either all equal to the identity, or are a permutation of e_0, e_1, e_2, e_3, e_4, or are a permutation of $e_0^{-1}, e_1^{-1}, e_2^{-1}, e_3^{-1}, e_4^{-1}$. In the latter two cases, the contribution to (23) is either zero or negative (being negative the trace of the product of several nonnegative elements in a commutative von Neumann algebra). Here we are using the fact that 5 is odd. Discarding all of these contributions except the one where $g_{i,0} = e_{i,0}$ (which has a nontrivial contribution thanks to Proposition B.13), we conclude that (23) is at most

$$\begin{aligned}&10^5 \tau(1_A 1_{S^n(A)} 1_{S^{2n}(A)} 1_{S^{3n}(A)} 1_{S^{4n}(A)})\\ &\quad - \tau(1_A e_0 1_A 1_{S^n(A)} e_1 1_{S^n(A)} 1_{S^{2n}(A)} e_2 1_{S^{2n}(A)} 1_{S^{3n}(A)} e_3 1_{S^{3n}(A)} 1_{S^{4n}(A)} e_4 1_{S^{4n}(A)}).\end{aligned}$$

By Theorem 2.13, the first expression is at most $10^5\mu(A)^{100}$. Now consider the second expression. By Proposition B.13, we see that the partial products $e_0e_1\cdots e_i$ for $i = 0, 1, 2, 3$ are distinct. Using (22), we conclude that the trace here can be computed as

$$\mu(S^{4n}(A)\cap A)\mu(A\cap S^n(A))\mu(S^n(A)\cap S^{2n}(A)) \\ \cdot\mu(S^{2n}(A)\cap S^{3n}(A))\mu(S^{3n}(A)\cap S^{4n}(A)),$$

which by (21) is equal to $\mu(A)^{10}$. Thus the expression (23) is no more than $2^{15}\mu(A)^{100} - \mu(A)^{10}$, which is negative if the upper bound δ for $\mu(A)$ is chosen to be sufficiently small.

This concludes the proof of Theorem 2.12.

Remark 2.14. Given that the counterexample in Theorem 2.13 can be extended to any $k \geq 5$, it seems reasonable to expect that Theorem 1.20 can be extended to all $k \geq 5$ (not just the odd k), though we have not pursued this issue. On the other hand, the analogue of Theorem 2.13 fails for $k = 4$, as was shown in [Bergelson et al. 2005]. Because of this, the $k = 4$ case of Theorem 1.20 remains open; the construction given here does not work, but it is possible that some other construction would suffice instead.

3. Inclusions of finite von Neumann dynamical systems

In this section we recall some fairly well-known constructions relating to von Neumann dynamical systems and their basic properties, culminating in a treatment of Popa's [2007] noncommutative version of the Furstenberg–Zimmer dichotomy. This material will be needed to establish the structure theorem, Theorem 1.14.

Let $(\mathcal{M}, \tau)$ be a finite von Neumann algebra. As noted in the introduction, we can embed $\mathcal{M}$ into a Hilbert space $L^2(\tau)$. In order to distinguish the algebra structure from the Hilbert space structure,[5] we shall refer in this section to the embedded copy of an element $a \in \mathcal{M}$ of the algebra in $L^2(\tau)$ as $\hat{a}$ rather than a; thus for instance $\hat{\mathcal{M}} = \{\hat{a} : a \in \mathcal{M}\}$ is a dense subspace of $L^2(\tau)$.

Clearly, $L^2(\tau)$ has the structure of an $\mathcal{M}$-bimodule, formed by extending the regular bimodule structure on $\mathcal{M}$ by density; the left-representation is, of course, the classical Gel'fand–Naimark–Segal representation associated to τ. When it is necessary to denote the copy of $\mathcal{M}$ in $B(L^2(\tau))$ consisting of the members of $\mathcal{M}$ acting by multiplication on the left (respectively, right), we will denote this algebra by $\mathcal{M}_{\text{left}}$ (respectively, $\mathcal{M}_{\text{right}}$).

[5]It is tempting to ignore these distinctions and identify $\hat{\mathcal{M}}$ with $\mathcal{M}$. While this is normally quite a harmless identification, we will take some care here because we will be studying the bimodule action of $\mathcal{M}$ on $L^2(\tau)$, and keeping track of this action can become notationally confusing if the algebra elements are identified with the vectors that they act on.

The space $L^2(\tau)$ contains a distinguished vector $\hat{1}$ — the representative of the multiplicative identity 1 in $\mathcal{M}$ — with the property that $a\hat{1} = \hat{1}a = \hat{a}$ for all $a \in \mathcal{M}$. This vector will play a prominent role in the rest of this section.

Now let $(\mathcal{N}, \tau|_{\mathcal{N}})$ be a von Neumann subalgebra of $(\mathcal{M}, \tau)$ (with the inherited trace). Then we can canonically identify $L^2(\tau|_{\mathcal{N}})$ with the closed subspace

$$\overline{\{\hat{b} : b \in \mathcal{N}\}} = \overline{\mathcal{N}\hat{1}} = \overline{\hat{1}\mathcal{N}}$$

of $L^2(\tau)$ in the obvious manner.

We will make use of certain well-known properties of these constructs, which we merely recall here. A clear account of all of them can be found in [Jones and Sunder 1997, Chapters 1 and 3].

First, it is important that there is a simple necessary and sufficient condition for a vector $\xi \in L^2(\tau)$ to lie in the dense subspace $\hat{\mathcal{M}}$: this is so if and only if the linear operator $\hat{\mathcal{M}} \to L^2(\tau)$, $\hat{x} \mapsto x\xi$ is bounded for the norm $\|\cdot\|_{L^2(\tau)}$, and so extends by continuity to a bounded operator $L^2(\tau) \to L^2(\tau)$. The necessity of this conclusion is clear, and its sufficiency requires just a little argument using the fact that for a finite von Neumann algebra $(\mathcal{M}, \tau)$ we have $\mathcal{M}_{\text{right}} = \mathcal{M}''_{\text{right}}$ and $\mathcal{M}_{\text{left}} = \mathcal{M}''_{\text{left}}$; see [Jones and Sunder 1997, Theorem 1.2.4].

A simple application of this condition now shows that the orthogonal projection $e_{\mathcal{N}} : L^2(\tau) \to \overline{\mathcal{N}\hat{1}}$ maps the dense subspace $\hat{\mathcal{M}}$ into $\hat{\mathcal{N}}$, and so defines also a linear operator $E_{\mathcal{N}} : \mathcal{M} \to \mathcal{N}$. Indeed, for $a \in \mathcal{M}$ we need only to show that the map $\hat{\mathcal{M}} \to L^2(\tau)$, $\hat{x} \mapsto xe_{\mathcal{N}}(\hat{a})$ is bounded for the norm $\|\cdot\|_{L^2(\tau)}$. Since $\mathcal{N}$ is also a von Neumann algebra and $e_{\mathcal{N}}(\hat{a}) \in \overline{\mathcal{N}\hat{1}} \cong L^2(\tau|_{\mathcal{N}})$, it actually suffices to check this for $x \in \mathcal{N}$. However, since $\overline{\mathcal{N}\hat{1}}$ is an $(\mathcal{N}, \mathcal{N})$-sub-bimodule, left multiplication by x commutes with $e_{\mathcal{N}}$, and so we have, as required,

$$\|xe_{\mathcal{N}}(\hat{a})\|_{L^2(\tau)} = \|e_{\mathcal{N}}(x\hat{a})\|_{L^2(\tau)} \leq \|\hat{xa}\|_{L^2(\tau)} \leq \|a\|\|\hat{x}\|_{L^2(\tau)}.$$

The linear operator $E_{\mathcal{N}}$ is referred to as the *conditional expectation* of $\mathcal{M}$ onto $\mathcal{N}$ associated to τ, and it has the following readily verified properties:

Lemma 3.1 (properties of conditional expectation). *For all $a \in \mathcal{M}$, the operator $E_{\mathcal{N}}$ satisfies*

- (idempotence) $E_{\mathcal{N}}(E_{\mathcal{N}}(a)) = E_{\mathcal{N}}(a)$;
- (contractivity) $\|E_{\mathcal{N}}(a)\| \leq \|a\|$;
- (trace-preservation) $\tau|_{\mathcal{N}}(E_{\mathcal{N}}(a)) = \tau(a)$;
- (positivity) $E_{\mathcal{N}}(a^*a) \geq 0$ (*as a member of* $\mathcal{N}$); *and*
- (relation with $e_{\mathcal{N}}$) *for all* $\xi \in L^2(\tau)$,

$$e_{\mathcal{N}}(a(e_{\mathcal{N}}(\xi))) = E_{\mathcal{N}}(a)(e_{\mathcal{N}}(\xi)) = e_{\mathcal{N}}(E_{\mathcal{N}}(a)(\xi)).$$

Example 3.2. If $\mathcal{M} = L^\infty(X, \mathcal{X}, \mu)$ for some probability measure μ with the usual trace, and $(Y, \mathcal{Y}, \nu)$ is a factor space of $(X, \mathcal{X}, \mu)$ with a measurable factor map $\pi : X \to Y$ that pushes μ forward to ν, then $L^\infty(Y, \mathcal{Y}, \nu)$ can be identified with a subalgebra of $\mathcal{M}$, and the conditional expectation map becomes its classical counterpart from probability theory.

Together with $\mathcal{M}$, the orthogonal projection $e_\mathcal{N}$ now generates in $B(L^2(\tau))$ a larger von Neumann algebra $\langle \mathcal{M}, e_\mathcal{N}\rangle \supseteq \mathcal{M}$. In general $\langle \mathcal{M}, e_\mathcal{N}\rangle$ is no longer a finite von Neumann algebra, but it does contain the dense $*$-subalgebra

$$\mathcal{A} := \mathrm{lin}(\mathcal{M} \cup \{x e_\mathcal{N} y : x, y \in \mathcal{M}\})$$

on which we define the *lifted trace* $\bar\tau : \mathcal{A} \to \mathbb{C}$ by specifying $\bar\tau(x e_\mathcal{N} y) = \tau(xy)$. By choosing an orthonormal basis for $L^2(\tau)$ relative to the right action of $\mathcal{N}$, and consequently realising $\langle \mathcal{M}, e_\mathcal{N}\rangle$ as an amplification of $\mathcal{N}$, this linear map is seen to be nonnegative and faithful, and hence defines a semifinite normal faithful $[0, +\infty]$-valued trace (which we still denote by $\bar\tau$) on the cone $(\langle \mathcal{M}, e_\mathcal{N}\rangle)^+$ of nonnegative (and self-adjoint) elements of $\langle \mathcal{M}, e_\mathcal{N}\rangle$. This witnesses that the algebra $\langle \mathcal{M}, e_\mathcal{N}\rangle$ is semifinite (that is, any positive element of it may be approximated from below by finite-$\bar\tau$ positive elements). We will not spell out these standard manipulations here (see, for instance, [Popa 2007, Section 1.5]), but we will invoke a notion of orthonormal basis for right-$\mathcal{N}$-submodules of $L^2(\tau)$ shortly.

Remark 3.3. In case $\mathcal{N} \subset \mathcal{M}$ is a finite-index inclusion of finite II_1 factors, then we find that $\langle \mathcal{M}, e_\mathcal{N}\rangle$ is also a finite II_1 factor. Writing $\mathcal{M}_1$ for this factor, it follows that the construction above may be repeated with the inclusion $\mathcal{M} \hookrightarrow \mathcal{M}_1$ in place of $\mathcal{N} \hookrightarrow \mathcal{M}$, and indeed that it may be iterated to form an infinite tower of II_1 factors

$$\mathcal{N} \subset \mathcal{M} \subset \mathcal{M}_1 \subset \mathcal{M}_2 \subset \cdots .$$

This is Jones' *basic construction*, which underlies his famous work [1983] on the possible values of the index $[\mathcal{N} : \mathcal{M}]$, and also several more recent developments. Once again we refer the reader to [Jones and Sunder 1997] for a thorough account of its importance, and numerous further references. However, since the construction of this whole infinite tower is special to the case of II_1 factors, we will not focus on it further here.

It is easy to check that the right action of any $n \in \mathcal{N}$ commutes with any $x e_\mathcal{N} y$, and hence with any member of $\langle \mathcal{M}, e_\mathcal{N}\rangle$. In fact it can be shown that $\langle \mathcal{M}, e_\mathcal{N}\rangle' = \mathcal{N}_{\mathrm{right}}$ and hence that $\mathcal{N}'_{\mathrm{right}} = \langle \mathcal{M}, e_\mathcal{N}\rangle'' = \langle \mathcal{M}, e_\mathcal{N}\rangle$: first, if $A \in B(L^2(\tau))$ commutes with every $b \in \mathcal{M}_{\mathrm{left}}$, then it must be the right action of some $a \in \mathcal{M}$, and now if also $e_\mathcal{N}(\hat{1}a) = \hat{1}a$ then we must in fact have $a \in \mathcal{N}$; see [Jones and Sunder 1997, Proposition 3.1.2]. Let us record the following immediate but important consequence of this for our later work:

Lemma 3.4. *If $V \leq L^2(\tau)$ is a closed right-$\mathcal{N}$-submodule, then the orthogonal projection $P_V : L^2(\tau) \to V$ is a member of $\langle \mathcal{M}, e_{\mathcal{N}} \rangle$.* □

Using $\bar{\tau}$ we can also define an alternative completion of $\mathcal{A} = \operatorname{lin} \mathcal{M} e_{\mathcal{N}} \mathcal{M}$ for each $p \in [1, \infty)$ by setting $\|A\|_{p,\bar{\tau}} := \sqrt[p]{\bar{\tau}((A^*A)^{p/2})}$ for $A \in \mathcal{A}$ (where as usual the power $(A^*A)^{p/2}$ is defined using spectral theory for the self-adjoint operator A^*A, and the nonnegativity of $\bar{\tau}$ is used to show that $\bar{\tau}((A^*A)^{p/2})$ is finite even when $p/2$ is not an integer). We denote this completion by $L^p(\bar{\tau})$; it is a Hilbert space when $p = 2$. In general elements of $L^p(\bar{\tau})$ do not correspond to elements of $\langle \mathcal{M}, e_{\mathcal{N}} \rangle$, but they do give possibly unbounded but closable operators that are weakly approximable by members of this algebra, which are therefore affiliated to $\mathcal{N}_{\mathrm{right}}$. If $A \in L^p(\bar{\tau})$ is such an operator that is self-adjoint, then it admits a spectral decomposition $A = \int_{\mathbb{R}} s\, P(\mathrm{d}s)$ for some spectral measure P on $\mathbb{R}$ taking values in the projections of $\langle \mathcal{M}, e_{\mathcal{N}} \rangle \cap L^1(\bar{\tau})$, of possibly unbounded support in $\mathbb{R}$, but for which $\|A\|^p_{p,\bar{\tau}} = \int_{\mathbb{R}} |s|^p \, \bar{\tau} P(\mathrm{d}s) < \infty$.

If V is as in Lemma 3.4 then we may write that P_V has *finite lifted trace* if it corresponds to a member of $\langle \mathcal{M}, e_{\mathcal{N}} \rangle \cap L^1(\bar{\tau})$.

Now let us introduce some dynamics. Suppose that α is a shift on $\mathcal{M}$ that restricts to a shift on $\mathcal{N}$. Then, as mentioned in the introduction, α induces a unitary operator acting on $L^2(\tau)$, which we shall distinguish from α by writing it as U_α; thus for instance

$$U_\alpha \hat{a} = U_\alpha(a\hat{1}) = \alpha(a)\hat{1} = \widehat{\alpha(a)} \quad \text{for all } a \in \mathcal{M}.$$

It is clear that $\overline{\mathcal{N}\hat{1}}$ is an invariant subspace for U_α, so that U_α commutes with $e_{\mathcal{N}}$. Also, conjugation by U_α agrees with the action α on $\mathcal{M}$; thus

$$U_\alpha a U_\alpha^{-1} \xi = \alpha(a)\xi \quad \text{for all } a \in \mathcal{M} \text{ and } \xi \in L^2(\tau).$$

Thus, conjugation by U_α extends the action of α to $\langle \mathcal{M}, e_{\mathcal{N}} \rangle$.

The following special class of one-sided submodules of $L^2(\tau)$ appears here almost exactly as in the commutative setting.

Definition 3.5 (finite-rank modules). A left- (respectively, right-) $\mathcal{N}$-submodule V of $L^2(\tau)$ has *finite rank* if there are some $\xi_1, \xi_2, \dots, \xi_r \in V$ such that $V = \overline{\sum_{i=1}^r \mathcal{N}\xi_i}$ (respectively, $V = \overline{\sum_{i=1}^r \xi_i \mathcal{N}}$), and the numerical value of its *rank* is the least $r \geq 1$ for which this is possible.

Proposition 3.6 (relativised Gram–Schmidt procedure). *If $V \leq L^2(\tau)$ is a U_α-invariant right-$\mathcal{N}$-submodule of finite rank r then there are $\xi_1, \xi_2, \dots, \xi_r \in L^2(\tau)$ such that*

- *the subspaces $\overline{\xi_i \mathcal{N}} \leq L^2(\tau)$ are pairwise orthogonal, and*
- *$V = \sum_{i=1}^r \overline{\xi_i \mathcal{N}}$.*

Proof. This uses a relativised Gram–Schmidt argument much as in the commutative setting; see for example [Glasner 2003, Lemma 9.4]. We proceed by induction on r. If V has rank 1, then the result is immediate from the definition, so let us suppose that it has rank $r+1$ for some $r \geq 1$. Then given a representation

$$V = \overline{\sum_{i=1}^{r+1} \xi_i^\circ \mathcal{N}},$$

we know that any member of V may be approximated in $\|\cdot\|_{L^2(\tau)}$ by expressions of the form $\xi_1^\circ n_1 + \cdots + \xi_{r+1}^\circ n_{r+1}$ for $n_1, n_2, \ldots, n_{r+1} \in \mathcal{N}$. This, in turn, may be rewritten as

$$(\xi_1^\perp n_1 + \cdots + \xi_r^\perp n_r) + ((\xi_1^\circ - \xi_1^\perp) n_1 + \cdots + (\xi_r^\circ - \xi_r^\perp) n_r) + \xi_{r+1}^\circ n_{r+1}$$

where for each $i \leq r$ we have decomposed ξ_i° into its component $\xi_i^\perp$ orthogonal to $\overline{\xi_{r+1}\mathcal{N}}$ and the remainder $\xi_i^\circ - \xi_i^\perp \in \overline{\xi_{r+1}\mathcal{N}}$. Since $\overline{\xi_{r+1}\mathcal{N}}$ is a right-$\mathcal{N}$-submodule, it follows that the second and third inner sums in the decomposition above both lie in $\overline{\xi_{r+1}\mathcal{N}}$, and now since $\overline{\xi_{r+1}\mathcal{N}}^\perp$ is also a right-$\mathcal{N}$-submodule, we have in fact shown that

$$V = V_1 + \overline{\xi_{r+1}\mathcal{N}},$$

where $V_1 := \overline{\sum_{i=1}^r \xi_i^\perp \mathcal{N}}$ is a rank-r right-$\mathcal{N}$-submodule that is orthogonal to $\overline{\xi_{r+1}\mathcal{N}}$. Applying the inductive hypothesis to V_1 now completes the proof. □

The following definition is also drawn from the commutative world. This notion has previously been extended to the setting of noncommutative algebras by Popa in [2007], who discusses several other aspects and equivalent conditions in that paper. (See also [Niculescu et al. 2003; Duvenhage 2009; Beyers et al. 2010] for an analysis of the absolute analogue of weak mixing, in which the subalgebra $\mathcal{N}$ is the trivial algebra $\mathbb{C}1$.)

Definition 3.7 (relative weak mixing). If $(\mathcal{M}, \tau, \alpha)$ is a von Neumann dynamical system and $\mathcal{N} \subset \mathcal{M}$ is an α-invariant von Neumann subalgebra, then α is *weakly mixing relative to* $\mathcal{N}$ if for any $a \in \mathcal{M} \cap \mathcal{N}^\perp$ we have

$$\frac{1}{N} \sum_{n=1}^N \| E_{\mathcal{N}}(a^* \alpha^n(a)) \|_\tau^2 \to 0 \quad \text{as } N \to \infty.$$

The basic inverse theorem that we need, extending the idea of Furstenberg and Zimmer to the noncommutative context, is contained in the following proposition, which essentially proves again part of [Popa 2007, Lemma 2.10]:

Proposition 3.8 (lack of weak mixing implies finite trace submodule). *If α is not weakly mixing relative to $\mathcal{N}$, then there is a U_α-invariant right-$\mathcal{N}$-submodule $V \leq L^2(\tau) \ominus \overline{\mathcal{N}\hat{1}}$ such that P_V has finite lifted trace.*

Proof. Suppose that $a \in \mathcal{M} \cap \mathcal{N}^{\perp}$ is such that

$$\frac{1}{N}\sum_{n=1}^{N}\|E_{\mathcal{N}}(a^*\alpha^n(a))\|_\tau^2 \not\to 0.$$

Define $b := ae_{\mathcal{N}}a^* \in \langle \mathcal{M}, e_{\mathcal{N}}\rangle$, and now observe (using the cyclic permutability of $\bar{\tau}$ and the identity $e_{\mathcal{N}}me_{\mathcal{N}} \equiv E_{\mathcal{N}}(m)e_{\mathcal{N}}$) that for any $n \in \mathbb{N}$ we have

$$\begin{aligned}\bar{\tau}(b(U_\alpha^n bU_\alpha^{-n})) &= \bar{\tau}(ae_{\mathcal{N}}a^*U_\alpha^n(ae_{\mathcal{N}}a^*)U_\alpha^{-n}) = \bar{\tau}(ae_{\mathcal{N}}a^*\alpha^n(a)e_{\mathcal{N}}\alpha^n(a)^*)\\ &= \bar{\tau}\big(E_{\mathcal{N}}(a^*\alpha^n(a))e_{\mathcal{N}}\alpha^n(a)^*a\big) = \|E_{\mathcal{N}}(a^*\alpha^n(a))\|_\tau^2.\end{aligned}$$

Averaging in n, it follows that

$$\bar{\tau}\Big(b\frac{1}{N}\sum_{n=1}^{N}\alpha^n(b)\Big) \to \langle b, b_1\rangle_{\bar{\tau}} \neq 0,$$

where b_1 is the limit of the ergodic averages $N^{-1}\sum_{n=1}^{N}\alpha^n(b)$ in the Hilbertian completion $L^2(\bar{\tau})$, which is therefore invariant under the further extension of the unitary operator U_α to this Hilbert space.

This new element b_1 need not, in general, correspond to a member of $\langle \mathcal{M}, e_{\mathcal{N}}\rangle$ (it is easily seen to be so in the commutative setting, but for special reasons); however, as a $\|\cdot\|_{2,\bar{\tau}}$-limit of members of $\langle \mathcal{M}, e_{\mathcal{N}}\rangle = \mathcal{N}'_{\mathrm{right}}$, the element can always be identified with a closed operator on $L^2(\tau)$ that is affiliated with the right action of the algebra $\mathcal{N}$, and as such it admits a spectral decomposition $b_1 = \int_0^\infty s\, P(\mathrm{d}s)$ for some resolution of the identity P on $[0, \infty)$ whose contributing spectral projections lie in $\langle \mathcal{M}, e_{\mathcal{N}}\rangle$, and for which $\int_0^\infty s^2\bar{\tau}(P(\mathrm{d}s)) = \|b_1\|_{2,\bar{\tau}}^2 < \infty$. Hence $\bar{\tau}P(I) < \infty$ for any Borel subset $I \subseteq (0, \infty)$ bounded away from 0. Now choosing any such subset I for which $P(I) \neq 0$ gives an orthogonal projection $P(I) \in \langle M, e_{\mathcal{N}}\rangle$ of finite lifted trace that is U_α-invariant, commutes with the right-$\mathcal{N}$-action because it lies in $\langle \mathcal{M}, e_{\mathcal{N}}\rangle$, and moreover has image orthogonal to $\overline{\hat{1}\mathcal{N}}$ because we initially chose b to lie in the orthogonal complement of this subspace. □

Remark 3.9. The implication above can in fact be reversed, and these conditions shown to be equivalent to a number of others; see [Popa 2007, Lemma 2.10] for a more complete picture.

In the next section we will push the above results a little further under the additional assumption that the subalgebra $\mathcal{N}$ is central, leading to the proof of Theorem 1.14.

4. The case of asymptotically abelian systems

We now specialise to the case of an asymptotically abelian system, with the crucial additional assumption that the subalgebra $\mathcal{N}$ is *central*.

Lemma 4.1. *Suppose that $(\mathcal{M}, \tau, \alpha)$ is a von Neumann dynamical system, $\mathcal{N} \subset \mathcal{M}$ is an α-invariant central von Neumann subalgebra and $V \leq L^2(\tau)$ is a U_α-invariant right-$\mathcal{N}$-submodule of finite lifted trace. Then for any $\varepsilon > 0$ there is a further U_α-invariant right-$\mathcal{N}$-submodule $V_1 \leq V$ such that*

- $\bar{\tau}(P_V - P_{V_1}) < \varepsilon$,
- *V_1 has finite rank, say $r \geq 1$, and*
- *there are an orthogonal right-$\mathcal{N}$-basis $\xi_1, \xi_2, \dots, \xi_r$ and a unitary matrix of unitary operators $U = (u_{ji})_{1 \leq i, j \leq r} \in \mathcal{U}_{r \times r}(\mathcal{N})$ such that*

$$U_\alpha(\xi_i) = \sum_{j=1}^{r} \xi_j u_{ji} \quad \text{for all } i = 1, 2, \dots, r.$$

We refer to U as the cocycle representing the action of U_α on the basis elements ξ_i.

Proof. We will prove this invoking the picture of the representation of $\mathcal{N}$ on $L^2(\tau)$ as a direct integral coming from spectral theory. By the classical theory of direct integrals (see, for instance, [Kadison and Ringrose 1997, Chapter 14]), we can select

- a standard Borel probability space (Y, ν),
- a Borel partition $Y = \bigcup_{n \geq 1} Y_n \cup Y_\infty$,
- a collection of Hilbert spaces $\mathfrak{H}_n$ for $n \in \{1, 2, \dots, \infty\}$ with $\dim(\mathfrak{H}_n) = n$, and
- a unitary equivalence

$$\Phi : L^2(\tau) \to \mathfrak{H} := \int_Y^{\oplus} \mathfrak{H}_y \, \nu(\mathrm{d}y),$$

where we define $\mathfrak{H}_y$ to be $\mathfrak{H}_n$ when $y \in Y_n$,

such that $\mathcal{N}$ (acting on either the right or left, since these agree for a central subalgebra of $\mathcal{M}$) is identified with the algebra of functions $L^\infty(\nu)$ acting by pointwise multiplication. Explicitly, if we denote elements of $\mathfrak{H}$ as measurable sections $v : Y \to \coprod_{y \in Y} \mathfrak{H}_y$, then $f \in L^\infty(\nu)$ acts on $\mathfrak{H}$ by

$$M_f(v)(y) := f(y)v(y).$$

Moreover, in order to accommodate $\Phi(\overline{\mathcal{N}\hat{1}})$ we select a measurable section $v_0 \in \mathfrak{H}$ with $\|v_0(y)\|_{\mathfrak{H}_y} \equiv 1$, and now $\mathcal{N}\hat{1}$ is identified with

$$\{y \mapsto f(y)v_0(y) : f \in L^\infty(\mu)\},$$

so that the orthogonal projection $\Phi e_{\mathcal{N}} \Phi^{-1}$ acts by

$$\Phi e_{\mathcal{N}} \Phi^{-1}(v)(y) := \langle v(y), v_0(y) \rangle_{\mathfrak{H}_y} \cdot v_0(y).$$

The larger algebra $\mathcal{M}_{\mathrm{right}}$ is identified under Φ with a direct integral $\int_Y^{\oplus} \mathcal{M}_y\,\nu(\mathrm{d}y)$, so that elements of $\Phi(\mathcal{M})$ are expressed as measurable sections

$$T : Y \to \coprod_{y\in Y} B(\mathfrak{H}_y)$$

acting by $Tv(y) := T(y)(v(y))$ and such that $T(y) \in \mathcal{M}_y$ ν-almost surely, where $(\mathcal{M}_y)_{y\in Y}$ is a measurable field of finite von Neumann subalgebras of $B(\mathfrak{H}_y)$ for each of which the state

$$\mathcal{M}_y \to \mathbb{C} : T \mapsto \langle v_0(y), T(v_0(y))\rangle_{\mathfrak{H}_y}$$

is a faithful finite trace; overall we have

$$\tau(a) = \langle \hat{1}, a\hat{1}\rangle = \int_Y \langle v_0(y), \Phi(a)(y)(v_0(y))\rangle_{\mathfrak{H}_y}\,\nu(\mathrm{d}y) \quad \text{for } a \in \mathcal{M},$$

and so in particular if $n \in \mathcal{N}$ then $\Phi(n) \in L^\infty(\mu)$ and $\tau(n) = \int \Phi(n)\,\mathrm{d}\nu$.

Given these data, for $a, b \in \mathcal{M}$ we can compute that and

$$\begin{aligned}\Phi(ae_{\mathcal{N}}b)\Phi^{-1}v(y) &= \langle \Phi(b)(y)(v(y)), v_0(y)\rangle \cdot \Phi(a)(y)(v_0(y)),\\ \bar{\tau}(ae_{\mathcal{N}}b) = \tau(ab) &= \int_Y \langle v_0(y), \Phi(ab)(y)(v_0(y))\rangle_{\mathfrak{H}_y}\,\nu(\mathrm{d}y)\\ &= \int_Y \langle \Phi(a^*)(y)(v_0(y)), \Phi(b)(y)(v_0(y))\rangle_{\mathfrak{H}_y}\,\nu(\mathrm{d}y)\\ &= \int_Y \mathrm{tr}(\Phi(ae_{\mathcal{N}}b)\Phi^{-1}|_{\mathfrak{H}_y})\,\nu(\mathrm{d}y).\end{aligned}$$

In this representation an $\mathcal{N}$-submodule $V \leq L^2(\tau)$ corresponds to a subspace $\Phi(V) \leq \mathfrak{H}$ of the form $\int_Y^{\oplus} V_y\,\nu(\mathrm{d}y)$ for some measurable subfield of Hilbert spaces $V_y \leq \mathfrak{H}_y$, and the calculation above now shows that $\bar{\tau}(P_V) = \int_Y \dim(V_y)\,\nu(\mathrm{d}y)$, so P_V has finite lifted trace if and only if the function $y \mapsto \dim(V_y)$ is ν-integrable.

We can enhance this picture further by noting that since α preserves $\mathcal{N}$ it must correspond to some ν-preserving transformation $S \curvearrowright Y$, and that since it also preserves $\mathcal{M}$ and extends to a unitary operator on $L^2(\tau)$ it must also preserve each of the cells Y_n. Similarly, since V is U_α-invariant, the transformation S must preserve the function $y \mapsto \deg(V_y)$. It follows that the unitary operator $\Phi U_\alpha \Phi^{-1}$ on $L^2(\tau)$ is actually given by a measurable section of unitary operators $\Psi : Y \to \coprod_{y\in Y} \mathcal{U}(\mathfrak{H}_y)$ such that

$$\Phi U_\alpha \Phi^{-1} v(y) = \Psi(y)(v(S^{-1}y)).$$

Now, since $y \mapsto \deg(V_y)$ is ν-integrable, for sufficiently large $r \geq 1$ we know that

$$\int_{\{y\in Y:\deg(V_y)>r\}} \deg(V_y)\,\nu(\mathrm{d}y) < \varepsilon.$$

Define

$$W := \int_{\{y\in Y:\deg(V_y)\le r\}}^{\oplus} V_y\,\nu(\mathrm{d}y) \oplus \int_{\{y\in Y:\deg(V_y)>r\}}^{\oplus} \{0\}\,\nu(\mathrm{d}y)$$

and $V_1 := \Phi^{-1}(W)$. Clearly V_1 is still a right-$\mathcal{N}$-submodule that is U_α-invariant, and it clearly also has rank at most r (since it suffices to prove this for W, for which it follows by a relativised Gram–Schmidt construction of a fibrewise-orthonormal basis exactly as in the setting of commutative ergodic theory; see for instance [Glasner 2003, Lemma 9.4]). Also, we have

$$\bar{\tau}(P_V - P_{V_1}) = \int_{\{y\in Y:\deg(V_y)>r\}} \deg(V_y)\,\nu(\mathrm{d}y) < \varepsilon.$$

Finally, the selection of unitaries Ψ must preserve the field of subspaces V_y above the S-invariant set $\{y \in Y : \deg(V_y) = s\}$ for each $s \le r$. Choosing an abstract d-dimensional Euclidean space W_d for each $d \le r$ and adjusting each fibre of W by a unitary in order to identify each V_y for which $\dim(V_y) \le r$ with $W_{\dim(V_y)}$, we obtain a new representation of V_1 as a right-$\mathcal{N}$-submodule using these fibres W_d, so that the action of U_α is now described by a measurable family of unitaries $\Psi'(y) \in \mathcal{U}(W_{\dim(V_y)})$. Picking an orthonormal basis for each W_d, writing these unitary operators as unitary matrices in terms of these bases, noting that their individual entries are now identified with elements of $L^\infty(\mu) = \Phi(\mathcal{N})$, and carrying everything back to $L^2(\tau)$ using Φ^{-1} gives the desired expression for U_α. □

Remark 4.2. Frustratingly, both the fact that a U_α-invariant V of finite lifted trace may be approximated by a U_α-invariant V_1 of finite rank, and the fact that given such a module of finite rank the action of U_α on it may be described by a unitary element in $\mathcal{U}(M_{r\times r}(\mathcal{N}))$, seem to be difficult to prove without the assumption that $\mathcal{N}$ is central and the resulting representation of the action of $\mathcal{N}$ on $L^2(\mu)$ as the multiplication action of some $L^\infty(\nu)$ on a measurable field of Hilbert spaces. It would be interesting to settle this issue more generally:

Question 4.3. Do these conclusions hold for a finite-lifted-trace invariant submodule corresponding to an arbitrary inclusion of finite von Neumann algebras with a trace-preserving automorphism?

Before moving on let us quickly note an important difference from the setting of abelian von Neumann algebras.

Example 4.4. If $\mathcal{M}$ is abelian, then it is well known from commutative ergodic theory that all the intermediate U_α-invariant submodules $V \le L^2(\tau)$ that have finite-rank over $\mathcal{N}$ together generate an intermediate sub*algebra* between $\mathcal{N}$ and $\mathcal{M}$, and that this then corresponds to an intermediate measure-preserving system. We will see shortly that an analogous conclusion can sometimes be recovered in the asymptotically abelian setting, but it is certainly not true for general finite-rank submodules, even when the smaller algebra $\mathcal{N}$ is abelian.

Consider, for example, the inclusion $i : L\mathbb{Z} \cong L^\infty(m_\mathbb{T}) \hookrightarrow L\mathbb{F}_2$ corresponding to the embedding of $\mathbb{Z}$ as the cyclic subgroup $a^\mathbb{Z}$ of the free group $\mathbb{F}_2 = \langle a, b\rangle$. Here LG is the group von Neumann algebra of G, defined in Section 2a. In this case we can identify $L^2(\tau)$ as $\ell^2(\mathbb{F}_2)$ and $L^2(\tau|_\mathcal{N})$ as the subspace spanned by $\{\xi_{a^n}\}_{n\in\mathbb{Z}}$. Now define $\alpha \in \operatorname{Aut} L\mathbb{F}_2$ simply by lifting the group automorphism of $\mathbb{F}_2$ that fixes a and maps $b \mapsto ba$. Now the subspace $V := \overline{\operatorname{lin}}\{\xi_{ba^n} : n \in \mathbb{Z}\} \le \ell^2(\mathbb{F}_2)$ is a U_α-invariant right $\mathcal{N}$-module of rank one which is orthogonal to $L^2(\tau|_\mathcal{N})$. On the other hand, although $\xi_b \in \hat{\mathcal{M}} \cap V$, we have $\alpha^m(\xi_b^2) = \alpha^m(\xi_{b^2}) = \xi_{ba^m ba^m}$ for $m \in \mathbb{Z}$, and it is easy to see that these elements of $\mathcal{M}$ do not remain within any finite-rank right-$\mathcal{N}$-submodule.

It is true that if $L^2(\tau) \ominus L^2(\tau|_\mathcal{N})$ contains a finite-rank right-$\mathcal{N}$-submodule V, then it also contains a finite-rank left-$\mathcal{N}$-module in the form of $J(V)$, where J is the *modular automorphism* on V, defined by extending the conjugation map $a \mapsto a^*$ on $\mathcal{M} \equiv \hat{\mathcal{M}}$ by density. The point is that it can happen that $J(V) \perp V$, and that all elements of $J(V)$ are weakly mixed by U_α: it is the right-module V, and no other, that serves as the obstruction to overall relative weak mixing coming from Theorem 1.13.

Definition 4.5. A vector $\xi \in L^2(\tau)$ is *central* if $m\xi = \xi m$ for all $m \in \mathcal{M}$.

Lemma 4.6 (no nonobvious central vectors). *The closure $\overline{\mathfrak{Z}(\mathcal{M})\hat{1}} = \overline{\hat{1}\mathfrak{Z}(\mathcal{M})}$ is equal to the set of all central vectors in $L^2(\tau)$.*

Proof. Suppose that $\xi \in L^2(\tau)$ is central. Define $a_\xi : \mathcal{M}\hat{1} \to L^2(\tau)$ by $a_\xi(m\hat{1}) := \xi m$. This is a densely defined linear operator on $L^2(\tau)$, and it is closable because if $m_n\hat{1} = \hat{1}m_n \to 0$ in $\|\cdot\|_{L^2(\tau)}$ for some sequence $(m_n)_{n\ge1}$ in $\mathcal{M}$ and also $\xi m_n \to \xi'$ in $\|\cdot\|_{L^2(\tau)}$, then we have

$$\langle m'\hat{1}, \xi'\rangle = \lim_{n\to\infty}\langle m'\hat{1}, \xi m_n\rangle = \lim_{n\to\infty}\langle \hat{1}m_n^*, (m')^*\xi\rangle = 0 \quad \text{for every } m' \in \mathcal{M},$$

and so in fact we must have $\xi' = 0$. Also, we clearly have

$$a_\xi(m\hat{1}) = a_\xi(\hat{1}m) = \xi m = m\xi = (a_\xi(\hat{1}))m = m(a_\xi(\hat{1})) \quad \text{for every } m \in \mathcal{M},$$

so a_ξ is affiliated with both the right- and left-actions of $\mathcal{M}$ on $L^2(\tau)$. The same therefore holds for $a_\xi + a_\xi^*$ and $\mathrm{i}(a_\xi - a_\xi^*)$, and now these are self-adjoint and so each of them may be expressed as an unbounded spectral integral all of whose contributing spectral projections must lie in $\mathcal{M}'_{\text{left}} \cap \mathcal{M}'_{\text{right}} = \mathfrak{Z}(\mathcal{M})$. Therefore, approximating $a_\xi = \frac{1}{2}(a_\xi + a_\xi^*) + \frac{1}{2}(a_\xi - a_\xi^*)$ by a sum of two large but bounded integrals with respect to the respective resolutions of the identity, we get a sequence of elements $a_n \in \mathfrak{Z}(\mathcal{M})$ such that $a_n \to a_\xi$ pointwise on $\operatorname{dom}(\operatorname{clos}(a_\xi)) \supseteq \mathcal{M}\hat{1}$, and hence such that $a_n\hat{1} \to \xi$ in $\|\cdot\|_{L^2(\tau)}$. Hence $\xi \in \overline{\mathfrak{Z}(\mathcal{M})\hat{1}}$, as required. □

Proposition 4.7. *If $(\mathcal{M}, \tau, \alpha)$ is an asymptotically abelian von Neumann dynamical system, $\mathcal{N}$ is a shift-invariant central von Neumann subalgebra, and $V \leq L^2(\tau)$ is an α-invariant right-$\mathcal{N}$-submodule of $\mathcal{M}$ having finite lifted trace, then all elements of V are central vectors.*

Proof. Clearly it will suffice to prove this for all finite-rank approximants V_1 to V as given by Lemma 4.1. Thus we may assume that V actually has finite rank. Let $\xi_1, \xi_2, \ldots, \xi_r$ and $U = (u_{ji})_{1\leq i,j\leq r} \in \mathcal{M}_{r\times r}(\mathcal{N})$ be as given by the third part of that lemma.

Since α is asymptotically abelian, we have for any $a\hat{1} \in \mathcal{M}\hat{1}$ and $b \in \mathcal{M}$ that

$$\frac{1}{N}\sum_{n=1}^{N}\|bU_\alpha^n(a\hat{1}) - U_\alpha^n(a\hat{1})b\|_{L^2(\tau)} = \frac{1}{N}\sum_{n=1}^{N}\|b\alpha^n(a) - \alpha^n(a)b\|_{L^2(\tau)} \to 0.$$

Approximating an arbitrary $\xi \in L^2(\tau)$ by elements of $\mathcal{M}\hat{1}$, it follows that for each fixed $b \in \mathcal{M}$ and $\xi \in L^2(\tau)$, we have

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\|bU_\alpha^n(\xi) - U_\alpha^n(\xi)b\|_{L^2(\tau)} = 0.$$

On the other hand, we know that

$$U_\alpha(\xi_i) = \sum_{j=1}^{r}\xi_j u_{ji} \quad \text{for all } i = 1, 2, \ldots, r,$$

and so, writing $U^n = (u_{ji}^{(n)})_{1\leq i,j\leq r}$, we have

$$U_\alpha^{-n}(\xi_i) = \sum_{j=1}^{r}\xi_j u_{ji}^{(-n)} \quad \text{implies} \quad \xi_i = \sum_{j=1}^{r}U_\alpha^n(\xi_j)\alpha^n(u_{ji}^{(-n)}) \quad \text{for all } i = 1, 2, \ldots, r.$$

Clearly each $u_{ji}^{(-n)}$ is still a unitary, and so from this, averaging in n and the centrality of $\mathcal{N}$, we obtain

$$\begin{aligned}
\|b\xi_i - \xi_i b\|_{L^2(\tau)} &= \Big\|\frac{1}{N}\sum_{n=1}^{N}\Big(\sum_{j=1}^{r}bU_\alpha^n(\xi_j)\alpha^n(u_{ji}^{(-n)}) - \sum_{j=1}^{r}U_\alpha^n(\xi_j)\alpha^n(u_{ji}^{(-n)})b\Big)\Big\|_{L^2(\tau)} \\
&= \Big\|\frac{1}{N}\sum_{n=1}^{N}\sum_{j=1}^{r}(bU_\alpha^n(\xi_j) - U_\alpha^n(\xi_j)b)\alpha^n(u_{ji}^{(-n)})\Big\|_{L^2(\tau)} \\
&\leq \sum_{j=1}^{r}\frac{1}{N}\sum_{n=1}^{N}\|bU_\alpha^n(\xi_j) - U_\alpha^n(\xi_j)b\|_{L^2(\tau)},
\end{aligned}$$

and now since each of the summands in j tends to 0 as $N \to \infty$, it follows that we must in fact have $b\xi_i = \xi_i b$ for every $i \le r$, and hence (taking $\mathcal{N}$-linear combinations, which have central coefficients, and then a completion) that all vectors in V are central, as required. □

Corollary 4.8. *If $(\mathcal{M}, \tau, \alpha)$ is an asymptotically abelian von Neumann dynamical system, then the subalgebra $\mathcal{M}^\alpha := \{a \in \mathcal{M} : \alpha(a) = a\}$ of individually α-invariant elements is central.*

Proof. Of course, if $\alpha(a) = a$, then $\mathrm{lin}\{\hat{1}a\}$ is a rank one α-invariant submodule of $L^2(\tau)$ for the trivial central subalgebra $\mathcal{N} := \mathbb{C}\hat{1}$, and the claim follows from Proposition 4.7. This claim can also be easily verified directly from the definition of asymptotic abelianness. □

Proof of Theorem 1.14. Suppose, for the sake of contradiction, that α were not weakly mixing relative to $\mathfrak{Z}(\mathcal{M}) \subset \mathcal{M}$. Then Proposition 3.8 gives a nontrivial right-$\mathfrak{Z}(\mathcal{M})$-submodule $V \le L^2(\tau) \ominus \overline{\mathfrak{Z}(\mathcal{M})\hat{1}}$ of finite lifted trace, and now Proposition 4.7 tells us that V must consist of central vectors. However, Lemma 4.6 now gives $V \le \overline{\mathfrak{Z}(\mathcal{M})\hat{1}}$, implying a contradiction with our assumption that $V \perp \overline{\mathfrak{Z}(\mathcal{M})\hat{1}}$. □

For the results in this section it suffices to assume that for every $a \in \mathcal{M}$ there exists a sequence $\{n_j\}$ such that $\lim_{j\to\infty} \|[\alpha^{n_j}(a), b]\|_{L^2(\tau)} = 0$ for every $b \in \mathcal{M}$. We do not know whether this condition is strictly weaker than asymptotically abelianness.

Remark 4.9. A variant of Theorem 1.14 can also be deduced from the results in [Niculescu et al. 2003] (and more specifically, Theorem 4.2 and Proposition 5.5 of that paper); we thank the anonymous referee for pointing out this fact. More specifically, the result is that if α is an automorphism of a finite von Neumann algebra $\mathcal{M}$ that leaves invariant a faithful normal trace τ, and E_τ is the conditional expectation to the factor

$$\mathcal{M}_r := \overline{\mathrm{lin}}^{\mathrm{wot}}\{a \in \mathcal{M} : \alpha(a) = \lambda a \text{ for some } \lambda \in \mathbb{T}\},$$

then for any $a, b \in \mathcal{M}$ one has

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} |\langle E_\tau(a^*\alpha^n(a)) - E_\tau(a)^*\alpha^n(E_\tau(a)), b\rangle_{L^2(\tau)}| = 0;$$

in particular, for N going to infinity along a density one set of integers, the expression $E_\tau(a^*\alpha^n(a)) - E_\tau(a)^*\alpha^n(E_\tau(a))$ converges to zero in the weak operator topology. This property is weaker than the relative weak mixing property with respect to this factor (which one does not expect to hold in general, even in the abelian case), but on the other hand does not require any hypothesis of asymptotic abelianness.

5. Triple averages for nonasymptotically abelian systems

The use to which we put relative weak mixing in the preceding section is very special to asymptotically abelian systems: in general there seems to be no way to track the error term resulting from the rearrangement at the heart of the proof of Theorem 1.13 without this assumption. However, in the special case of triple averages this problem does simplify somewhat, provided we assume instead that our system $(\mathcal{M}, \tau, \alpha)$ is *ergodic*, so that $\mathcal{M}^\alpha = \mathbb{C}1$. In this case we will be able to obtain convergence weakly and in norm, as well as recurrence on a dense set (Theorem 1.17).

This assumption is not so innocuous as might be expected from its analogue in the world of commutative ergodic theory. In that setting it is possible quite generally to decompose a system (that is, more precisely, to decompose its invariant measure) into ergodic components, and then many assertions about the whole system, including multiple recurrence and the convergence of multiple averages, follow if they can be proved for each ergodic component separately. However, in the setting of a general von Neumann dynamical system, this decomposition is available only if $\mathcal{M}^\alpha$ is central in $\mathcal{M}$; otherwise the automorphism α can exhibit genuinely new phenomena precisely by virtue of having the nontrivial fixed subalgebra $\mathcal{M}^\alpha$ "move around". This was already seen in the failure of recurrence on a dense set when the ergodicity hypothesis is dropped (Theorem 1.19).

The key for convergence of triple averages is the following decomposition that is similar to the commutative case, first established (in a slightly more general setting) in [Niculescu et al. 2003] (and more specifically, from Theorem 4.2 and Proposition 5.5 in that paper); for the convenience of the reader we give a short proof of that decomposition here. The result does not require ergodicity of the system. A closely related decomposition was also used in [Fidaleo 2009].

Proposition 5.1 (decomposition of von Neumann dynamical systems). *Suppose $(\mathcal{M}, \tau, \alpha)$ is a von Neumann dynamical system. Then one has the orthogonal decomposition $\mathcal{M} = \mathcal{M}_r \oplus \mathcal{M}_s$, where*

$$\mathcal{M}_r := \overline{\operatorname{lin}}^{\mathrm{wot}}\{a \in \mathcal{M} : \alpha(a) = \lambda a \text{ for some } \lambda \in \mathbb{T}\} \quad \text{and}$$
$$\mathcal{M}_s := \Big\{a \in \mathcal{M} : \lim_{N\to\infty} N^{-1} \sum_{n=1}^N |\tau(b\alpha^n(a))| = 0 \text{ for every } b \in \mathcal{M}\Big\},$$

that is, $\mathcal{M}_r$ is the von Neumann subalgebra spanned by the eigenvectors of α and $\mathcal{M}_s$ is the subspace of the elements of $\mathcal{M}$ that are weakly mixed by α. The corresponding projection onto $\mathcal{M}_r$ is the conditional expectation of $\mathcal{M}$ onto $\mathcal{M}_r$ and in particular preserves positivity.

Proof. Since the continuation U_α of α to $L^2(\tau)$ is a unitary operator, the Jacobs–Glicksberg–de Leeuw decomposition holds for U_α (see for example [Krengel 1985, Section 2.4]), that is, $L^2(\tau) = L^2_r(\tau) \oplus L^2_s(\tau)$, where the *reversible part* $L^2_r(\tau)$ is defined as

$$L^2_r(\tau) = \overline{\text{lin}}\{x : U_\alpha(x) = \lambda x \text{ for some } \lambda \in \mathbb{T}\}$$

and the *stable part* $L^2_s(\tau)$ is defined as the space of all $x \in L^2(\tau)$ such that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} |\langle U^n_\alpha(x), y\rangle| = 0 \quad \text{for every } y \in L^2(\tau).$$

Moreover, this decomposition is orthogonal since U_α is unitary. We do not need here the Jacobs–Glicksberg–de Leeuw decomposition in full generality but only its version for unitary operators, which can be also proved via the spectral theorem.

By a result of Størmer [1974], the eigenvectors of U_α belong to $\mathcal{M}$. We thus have $\mathcal{M}_r = \mathcal{M} \cap L^2_r(\tau)$ and $\mathcal{M}_s = \mathcal{M} \cap L^2_s(\tau)$. The fact that the weak operator closure and the closure in the $L^2(\tau)$-topology coincide for self-adjoint subalgebras implies the second formula for $\mathcal{M}_r$ and thus $\mathcal{M}_r$ is a von Neumann subalgebra of $\mathcal{M}$. The conditional expectation now maps $\mathcal{M}$ onto $\mathcal{M}_r$ assuring the orthogonal decomposition $\mathcal{M} = \mathcal{M}_r \oplus \mathcal{M}_s$. □

In the remainder of this section we assume our system is ergodic.

Proposition 5.2 (convergence of triple averages). *Let* $(\mathcal{M}, \tau, \alpha)$ *be an ergodic von Neumann dynamical system. Then the averages*

$$\frac{1}{N} \sum_{n=1}^{N} \alpha^n(a)\alpha^{2n}(b) \tag{25}$$

converge in $\|\cdot\|_{L^2(\tau)}$ *as* $N \to \infty$ *for every* $a, b \in \mathcal{M}$.

Proof. By the proposition above, it suffices to assume that a and b each belong to $\mathcal{M}_r$ or $\mathcal{M}_s$. Suppose first that $a \in \mathcal{M}_r$, and fix b. The operators S_N given by

$$S_N x = \frac{1}{N} \sum_{n=1}^{N} \alpha^n(x)\alpha^{2n}(b)$$

are linear and bounded on $\mathcal{M}$ for the norm $\|\cdot\|_{L^2(\tau)}$, so we may assume that $\alpha(a) = \lambda a$ for some $\lambda \in \mathbb{T}$. Then $S_N a = (N+1)^{-1} \sum_{n=0}^{N} a(\lambda\alpha^2)^n(b)$, which converges in $L^2(\tau)$ by the mean ergodic theorem.

Suppose now that $a \in \mathcal{M}_s$. We show that the desired limit is zero. Consider $u_n := \alpha^n(a)\alpha^{2n}(b)\hat{1}$ and observe that

$$\begin{aligned}\langle u_n, u_{n+j}\rangle &= \tau(\alpha^{2n}(b^*)\alpha^n(a^*)\alpha^{n+j}(a)\alpha^{2n+2j}(b)) \\ &= \tau(\alpha^n(b^*)a^*\alpha^j(a)\alpha^{n+2j}(b)) = \tau(a^*\alpha^j(a)\alpha^n(\alpha^{2j}(b)b^*)).\end{aligned}$$

The ergodicity of the system implies

$$\gamma_j := \lim_{N\to\infty}\Big|\frac{1}{N}\sum_{n=1}^{N}\langle u_n, u_{n+j}\rangle\Big| = \Big|\tau(a^*\alpha^j(a)\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\alpha^n(\alpha^{2j}(b)b^*))\Big|$$
$$= |\tau(a^*\alpha^j(a))|\cdot|\tau(\alpha^{2j}(b)b^*)|.$$

Since $a\in\mathcal{M}_s$ and $\tau(\alpha^{2j}(b)b^*)$ are bounded in j, $\lim_{N\to\infty}N^{-1}\sum_{j=1}^{N}\gamma_j=0$, and therefore by the classical van der Corput lemma for Hilbert spaces (see for example [Furstenberg 1977] or [Bergelson 1987]), we have $\lim_{N\to\infty}N^{-1}\sum_{n=1}^{N}u_n=0$. □

Remarks 5.3. (1) For compact nonergodic systems the averages (25) converge as well, since $\mathcal{M}=\mathcal{M}_r$ in this case; this was observed in [Beyers et al. 2010].

(2) As in the commutative case we see that the Kronecker subalgebra $\mathcal{M}_r$ is characteristic for (25), that is, the limit of the averages in (25) does not change when replacing a by $E_{\mathcal{M}_r}a$ and b by $E_{\mathcal{M}_r}b$.

As was shown in Corollary 2.7, one cannot expect that

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\tau(a\alpha^n(a)\alpha^{2n}(a))>0\quad\text{for every positive } a.$$

However, a modification extending [Beyers et al. 2010, Theorem 5.13] is still true.

Proposition 5.4. *For an ergodic von Neumann system* $(\mathcal{M},\tau,\alpha)$, *one has*

$$\liminf_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}(\operatorname{Re}\tau(a\alpha^n(a)\alpha^{2n}(a)))_+>0\quad\textit{for every } 0<a\in\mathcal{M}.$$

In particular, one has recurrence on a dense set.

Proof. Decompose $a=b+c$ with $b\in\mathcal{M}_r$ and $c\in\mathcal{M}_s$ as in Proposition 5.1, with $b>0$ by Lemma 3.1. We first show that there exists a compact abelian group G, an open set $U\subset G$, and $g\in G$ such that for the 1-step Bohr set $K_U:=\{n\in\mathbb{N}:g^n\in U\}$ one has

$$\operatorname{Re}\tau(b\alpha^n(b)\alpha^{2n}(b))>\tfrac{1}{2}\tau(b^3)>0\quad\text{for every } n\in K_U. \tag{26}$$

Take $\varepsilon:=\tau(b^3)/(18\|b\|^2)$. Since $b\in\mathcal{M}_r$, we find $k\in\mathbb{N}$, $\lambda_1,\dots,\lambda_k\in\mathbb{T}$ and $b_1,\dots,b_k\in\mathcal{M}\setminus\{0\}$ such that $\alpha(b_j)=\lambda_j b_j$ for every $j=1,\dots,k$ and such that $\|b-(b_1+\cdots+b_k)\|_{L^2(\tau)}<\varepsilon$. Set now $G:=\mathbb{T}^k$, $g:=(\lambda_1,\dots,\lambda_k)$ and $U:=U_{\varepsilon/(k\max\|b_j\|)}(1)\subset\mathbb{T}^k$. Observe that for every n such that $g^n\in U$, we have

$|\lambda_j^n - 1| < \varepsilon/(k \max\|b_j\|)$ for every $j = 1, \dots, k$ and therefore

$$\begin{aligned}\|\alpha^n(b) - b\|_{L^2(\tau)} &\le \|a^n(b_1 + \cdots + b_k) - (b_1 + \cdots + b_k)\|_{L^2(\tau)} \\ &\qquad\qquad + 2\|b_1 + \cdots + b_k - b\|_{L^2(\tau)} \\ &\le \max\|b_j\|_{L^2(\tau)}(|\lambda_1^n - 1| + \cdots + |\lambda_k^n - 1|) + 2\varepsilon \\ &< \max\|b_j\| \frac{k\varepsilon}{k \max\|b_j\|} + 2\varepsilon = 3\varepsilon.\end{aligned}$$

So we can prove (26) by the Cauchy–Schwarz inequality:

$$\begin{aligned}|\tau(b\alpha^n(b)\alpha^{2n}(b)) - \tau(b^3)| &\le |\tau(b\alpha^n(b)(\alpha^{2n}(b) - b))| + |\tau(b(\alpha^n(b) - b)b)| \\ &\le \|b\|^2(\|\alpha^{2n}(b) - b\|_{L^2(\tau)} + \|\alpha^n(b) - b\|_{L^2(\tau)}) \\ &\le 3\|b\|^2\|\alpha^n(b) - b\|_{L^2(\tau)} < 9\|b\|^2\varepsilon = \tfrac{1}{2}\tau(b^3).\end{aligned}$$

Take now $V := U_{\varepsilon/(2k \max\|b_j\|)}(1) \subset U$ and a continuous function $f : G \to [0, 1]$ satisfying $\mathbf{1}_V \le f \le \mathbf{1}_U$. Then by (26), $\operatorname{Re}\tau(b\alpha^n(b)\alpha^{2n}(b))$ is positive whenever $f(g^n) \ne 0$ and therefore

$$\begin{aligned}&\liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(g^n) \operatorname{Re}\tau(b\alpha^n(b)\alpha^{2n}(b)) \\ &\qquad\qquad \ge \liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mathbf{1}_V(g^n) \operatorname{Re}\tau(b\alpha^n(b)\alpha^{2n}(b)).\end{aligned}$$

Since the set $K_V := \{n \in \mathbb{N} : g^n \in V\} \subset K_U$ is syndetic (that is, has bounded gaps) in $\mathbb{N}$, this implies by (26)

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(g^n) \operatorname{Re}\tau(b\alpha^n(b)\alpha^{2n}(b)) > 0. \tag{27}$$

Next, we show that

$$\|\cdot\|_{L^2(\tau)} - \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(g^n)\alpha^n(b)\alpha^{2n}(c) = 0. \tag{28}$$

We first consider a character $\gamma \in \hat{G}$ and define $u_n := \gamma(g^n)\alpha^n(b)\alpha^{2n}(c)\hat{1}$. We have

$$\begin{aligned}\langle u_n, u_{n+j}\rangle &= \overline{\gamma(g^n)}\gamma(g^{n+j})\gamma(\alpha^{2n}(c^*)\alpha^n(b^*)\alpha^{n+j}(b)\alpha^{2n+2j}(c)) \\ &= \gamma(g^j)\tau(\alpha^n(c^*)b^*\alpha^j(b)\alpha^{n+2j}(c)) = \gamma(g^j)\tau(b^*\alpha^j(b)\alpha^n(\alpha^{2j}(c)c^*)).\end{aligned}$$

By ergodicity of α,

$$\begin{aligned}\gamma_j &:= \lim_{N\to\infty}\left|\frac{1}{N} \sum_{n=1}^{N} \langle u_n, u_{n+j}\rangle\right| = \left|\gamma(g^j)\tau(b^*\alpha^j(b) \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \alpha^n(\alpha^{2j}(c)c^*))\right| \\ &= |\tau(b^*\alpha^j(b))| \cdot |\tau(\alpha^{2j}(c)c^*)|,\end{aligned}$$

and the assumption $c \in \mathcal{M}_s$ implies $\lim_{N\to\infty} N^{-1}\sum_{j=1}^{N}\gamma_j = 0$. By the van der Corput estimate, we thus have

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}u_n = \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\gamma(g^n)\alpha^n(b)\alpha^{2n}(c)\hat{1} = 0.$$

We have now proved (28), since the characters form a total set in $C(G)$ and the operators $S_N f := N^{-1}\sum_{n=1}^{N} f(g^n)\alpha^n(b)\alpha^{2n}(c)$ are uniformly bounded on $C(G)$. Analogously one also has

$$\begin{aligned}\|\cdot\|_{L^2(\tau)} - \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} f(g^n)\alpha^n(c)\alpha^{2n}(b)& \\ = \|\cdot\|_{L^2(\tau)} - \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} f(g^n)\alpha^n(c)\alpha^{2n}(c) &= 0.\end{aligned}$$

The Cauchy–Schwarz inequality implies now that

$$\begin{aligned}\limsup_{N\to\infty}&\Big|\frac{1}{N}\sum_{n=1}^{N} f(g^n)\tau(c\alpha^n(b)\alpha^{2n}(c))\Big| \\ &= \limsup_{N\to\infty}\Big|\tau\Big(c\frac{1}{N}\sum_{n=1}^{N} f(g^n)\alpha^n(b)\alpha^{2n}(c)\Big)\Big| \\ &\le \|c\|_{L^2(\tau)}\limsup_{N\to\infty}\Big\|\frac{1}{N}\sum_{n=1}^{N} f(g^n)\alpha^n(b)\alpha^{2n}(c)\Big\|_{L^2(\tau)} = 0.\end{aligned}$$

Analogously, the Cesàro sums of $f(g^n)\tau(c\alpha^n(c)\alpha^{2n}(b))$, $f(g^n)\tau(c\alpha^n(c)\alpha^{2n}(c))$ and $f(g^n)\tau(b\alpha^n(c)\alpha^{2n}(c))$ vanish, while

$$\tau(c\alpha^n(b)\alpha^{2n}(b)) = \tau(b\alpha^n(b)\alpha^{2n}(c)) = \tau(b\alpha^n(c)\alpha^{2n}(b)) = 0$$

follows from the orthogonality of $\mathcal{M}_r$ and $\mathcal{M}_s$ and the fact that $\mathcal{M}_r$ is an α-invariant self-adjoint subalgebra of $\mathcal{M}$.

Combining this with (27), we obtain by the linearity of τ

$$\begin{aligned}\liminf_{N\to\infty}&\frac{1}{N}\sum_{n=1}^{N}(\operatorname{Re}\tau(a\alpha^n(a)\alpha^{2n}(a)))_+ \\ &\ge \liminf_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} f(g^n)(\operatorname{Re}\tau(a\alpha^n(a)\alpha^{2n}(a)))_+ \\ &= \liminf_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} f(g^n)(\operatorname{Re}\tau(b\alpha^n(b)\alpha^{2n}(b)))_+ > 0. \qquad \square\end{aligned}$$

6. Closing remarks

We present some remarks concerning Question 1.22. By Theorem 1.17, we have a positive answer to this question when the invariant algebra $\mathcal{M}^\alpha$ is trivial. One can also extend these arguments to cover the case when the invariant algebra $\mathcal{M}^\alpha$ is central by representing $\mathcal{M}$ as a direct integral over $\mathcal{M}^\alpha$, see Kadison, Ringrose [Kadison and Ringrose 1997, Chapter 14].

It is clear that if the answer to Question 1.23 is always positive, then the same is true for Question 1.22. What is less obvious is that the converse is true; if the answer to 1.22 is always true, then the answer to 1.23 is always true. To see this, let $(\mathcal{M}, \tau)$ be a finite von Neumann algebra with two commuting shifts α_1 and α_2. We then form the infinite tensor product $\mathcal{M}^{\mathbb{Z}} := \bigotimes_{n\in\mathbb{Z}} \mathcal{M}$, which is another finite von Neumann algebra, which contains an embedded copy of $\mathcal{M}$ by using the 0 coordinate of $\mathbb{Z}$. Next, let G be the free abelian group on two generators e and f, and let U be the action of G on $\mathcal{M}^{\mathbb{Z}}$ defined by

$$U(e) \bigotimes_{n\in\mathbb{Z}} a_n := \bigotimes_{n\in\mathbb{Z}} \alpha_1^2\alpha_2^{-1}(a_n) \quad \text{and} \quad U(f) \bigotimes_{n\in\mathbb{Z}} a_n := \bigotimes_{n\in\mathbb{Z}} a_{n-1}$$

for all $a_n \in \mathcal{M}$ with all but finitely many a_n equal to 1. If we define a shift α' to $\mathcal{M}^{\mathbb{Z}}$ by the formula

$$\alpha' \bigotimes_{n\in\mathbb{Z}} a_n := \bigotimes_{n\in\mathbb{Z}} \alpha_1^{2(n+1)}\alpha_2^{-n}(a_n),$$

we then observe the identities

$$\alpha' U(e)(\alpha')^{-1} = U(e) \quad \text{and} \quad \alpha' U(f)(\alpha')^{-1} = U(fe)$$

(here we use the hypothesis that α_1 and α_2 commute). Because of this, we can define a shift α on the crossed product $\mathcal{M}^{\mathbb{Z}} \rtimes_U G$ by declaring α to equal α' on $\mathcal{M}^{\mathbb{Z}}$, and

$$\alpha(e) := e \quad \text{and} \quad \alpha(f) := fe.$$

If a_1 and a_2 lie in $\mathcal{M}^{\mathbb{Z}}$, we observe that

$$\alpha^n(a_1 f^2)\alpha^{2n}(f^{-2}a_2 f) = (\alpha')^n(a_1)((\alpha')^{2n}U(e)^{-2n}(a_2))f.$$

If we assume that a_1 and a_2 in fact lie in $\mathcal{M}$, we can simplify this as

$$\alpha_1^{2n}(a_1)\alpha_2^{2n}(a_2)f.$$

Thus, if we assume 1.22 has an affirmative answer for the system $\mathcal{M}^{\mathbb{Z}} \rtimes_U G$, we see that the averages of $\alpha_1^{2n}(a_1)\alpha_2^{2n}(a_2)f$ (and hence of $\alpha_1^{2n}(a_1)\alpha_2^{2n}(a_2)$) converge for arbitrary $a_1, a_2 \in \mathcal{M}$; from this one easily deduces (after dividing n into even and odd classes) that 1.23 has an affirmative answer for the system $\mathcal{M}$.

In particular, we see that the task of establishing Question 1.22 in the affirmative for arbitrary von Neumann dynamical systems is at least as hard as that of achieving convergence for two commuting shifts in the abelian case, a result first obtained in [Conze and Lesigne 1984].

One can also cover some other (nonergodic, nonabelian) cases of Question 1.22 by ad hoc methods. Suppose that $\mathcal{M}$ is a group von Neumann algebra LG, with shift α given by automorphisms $\alpha_1, \alpha_2 : G \to G$ of the group. Then one can affirmatively answer 1.22 as follows. First, by density and linearity we may assume that a_1 and a_2 are themselves group elements: $a_1 = g_1 \in G$ and $a_2 = g_2 \in G$. We then see that the means of $\alpha^n(g_1)\alpha^{2n}(g_2)$ will converge to zero unless there exists a group element g_0 for which

$$\alpha^n(g_1)\alpha^{2n}(g_2) = g_0 \tag{29}$$

for all n in a set of positive upper density. But such sets contain nontrivial parallelograms $n, n+h, n+k, n+h+k$ for $h, k > 0$. Applying (29) for n and $n+h$ and rearranging, one obtains

$$\alpha^n(g_2\alpha^{2h}(g_2^{-1})) = g_1^{-1}\alpha^h(g_1).$$

Similarly, applying (29) for $n+k$ and $n+h+k$, one has

$$\alpha^{n+k}(g_2\alpha^{2h}(g_2^{-1})) = g_1^{-1}\alpha^h(g_1).$$

Writing $u := g_1^{-1}\alpha^h(g_1)$, one thus has

$$\alpha^h(g_1) = g_1 u \quad \text{and} \quad \alpha^k(u) = u.$$

If we then write

$$v := g_1^{-1}\alpha^{hk}(g_1) = u\alpha^h(u)\cdots\alpha^{(k-1)h}(u),$$

we see that $\alpha^{hkn}(g_1) = g_1 v^n$ for all n, and $\alpha(v) = v$. Thus we have

$$\alpha^{hkn+j}(g_1)\alpha^{2hkn+2j}(g_2) = \alpha^j(g_1(\alpha^{2hk}(v))^n\alpha^j(g_2)) \quad \text{for any } n, j.$$

The means of this in n converge in $L^2(\tau)$ by the mean ergodic theorem. Summing over all $0 \le j < hk$, we obtain weak convergence, thus answering Question 1.22 affirmatively in this case. The same type of argument also lets one deal with crossed products of abelian systems by groups, in which the shift acts as an automorphism on the group; we omit the details.

Finally, we remark that the results on asymptotically abelian systems, while stated for $\mathbb{Z}^k$-systems, should in fact be valid for any commuting action of a general locally compact second countable (lcsc) abelian group.

Appendix A. An application of the van der Corput lemma

The purpose of this appendix is to establish Theorem 1.13. Our arguments follow [Niculescu et al. 2003, Proposition 7.4 and Theorem 7.5] closely (for another adaptation of the same argument, see also [Beyers et al. 2010, Proposition 4.4]). We may normalise α_0 to be the identity.

We induct on $k \geq 2$. When $k = 2$ we know from the usual mean ergodic theorem for von Neumann algebras (see for example [Krengel 1985, Section 9.1]) that

$$\frac{1}{N}\sum_{n=1}^{N}\alpha^n(a) \to E_{\mathcal{M}^\alpha}(a) \quad \text{in } \|\cdot\|_{L^2(\tau)},$$

and since $\mathcal{M}^\alpha \subseteq \mathcal{N}$ by the relative weak mixing assumption, we also have

$$\frac{1}{N}\sum_{n=1}^{N}\alpha^n(E_{\mathcal{N}}(a)) \to E_{\mathcal{M}^\alpha}(E_{\mathcal{N}}(a)) = E_{\mathcal{M}^\alpha}(a) \quad \text{in } \|\cdot\|_{L^2(\tau)},$$

so combining these conclusions gives the result.

Now suppose that $k \geq 3$ and that we know the desired conclusion for any similar family of $\ell < k$ automorphisms. By decomposing each a_i as $(a_i - E_{\mathcal{N}}(a_i)) + E_{\mathcal{N}}(a_i)$ and expanding out the expression $\prod_{i=1}^{k-1}\alpha_i^n(a_i)$, we find that it suffices to show that for any $i \leq k-1$,

$$a_i \perp \mathcal{N} \quad \text{implies} \quad \frac{1}{N}\sum_{n=1}^{N}\prod_{i=1}^{k-1}\alpha_i^n(a_i) \to 0 \quad \text{in } \|\cdot\|_{L^2(\tau)};$$

let us argue the case $i = 1$, the others following at once by symmetry.

By the Hilbert-space-valued version of the classical van der Corput estimate (see, for instance, [Furstenberg 1977] or [Bergelson 1987]) this will follow if we show that

$$\begin{aligned}&\frac{1}{H}\sum_{h=1}^{H}\Big|\frac{1}{N}\sum_{n=1}^{N}\Big\langle\prod_{i=1}^{k-1}\alpha_i^{n+h}(a_i), \prod_{i=1}^{k-1}\alpha_i^n(a_i)\Big\rangle_\tau\Big| \\ &= \frac{1}{H}\sum_{h=1}^{H}\Big|\frac{1}{N}\sum_{n=1}^{N}\tau\big(\alpha_{k-1}^n(\alpha_{k-1}^h(a_{k-1}^*))\cdots\alpha_1^n(\alpha_1^h(a_1^*))\cdot\alpha_1^n(a_1)\cdots\alpha_{k-1}^n(a_{k-1})\big)\Big| \to 0\end{aligned}$$

as $N \to \infty$ and then $H \to \infty$.

Let us now set $b_i := \alpha_i^n(\alpha_i^h(a_i^*))$ and $c_i := \alpha_i^n(\alpha_i^h(a_i))$ to lighten notation. Having done so, we now set ourselves up for applying the asymptotic abelianness property by observing that

$$\begin{aligned}
&b_{k-1}b_{k-2}b_{k-3}\cdots c_1c_2\cdots \\
&\qquad = (b_{k-2}b_{k-1}b_{k-3}\cdots c_1c_2\cdots) + ([b_{k-1},b_{k-2}]b_{k-3}\cdots c_1c_2\cdots) \\
&\qquad = (b_{k-2}b_{k-3}b_{k-1}b_{k-4}\cdots c_1c_2\cdots) + (b_{k-2}[b_{k-1},b_{k-3}]b_{k-4}\cdots c_1c_2\cdots) \\
&\qquad\qquad\qquad\qquad\qquad\qquad + ([b_{k-1},b_{k-2}]b_{k-3}b_{k-4}\cdots c_1c_2\cdots) \\
&\qquad \vdots \\
&\qquad = b_{k-2}b_{k-3}b_{k-4}\cdots b_1c_1c_2\cdots c_{k-2}(b_{k-1}c_{k-1}) \\
&\qquad\qquad\qquad\qquad + \sum_{j=1}^{k-2} x_j[b_{k-1},b_j]y_j + \sum_{j=1}^{k-2} u_j[b_{k-1},c_j]v_j,
\end{aligned}$$

where each x_j, y_j, u_j and v_j for $1 \le j \le k-2$ is some product of a subset of the elements $\{b_i, c_i : i \le k-2\}$.

Importantly, there is some $M > 0$ such that $\|x_j\|, \|y_j\|, \|u_j\|, \|v_j\| \le M$ for all $j \le k-2$, and not depending on n or h, while on the other hand for any $j \le k-2$ we have

$$[b_{k-1}, b_j] = [\alpha_{k-1}^n(\alpha_{k-1}^h(a_{k-1}^*)), \alpha_j^n(\alpha_j^h(a_j^*))],$$

and hence overall we have

$$\begin{aligned}
&\frac{1}{N}\sum_{n=1}^{N}\Big\|\sum_{j=1}^{k-2} x_j[b_{k-1},b_j]y_j\Big\|_{L^2(\tau)} \\
&\qquad\qquad \le M^2 \sum_{j=1}^{k-2}\frac{1}{N}\sum_{n=1}^{N}\|[b_{k-1},b_j]\|_{L^2(\tau)} \\
&\qquad\qquad = M^2 \sum_{j=1}^{k-2}\frac{1}{N}\sum_{n=1}^{N}\|[\alpha_{k-1}^h(a_{k-1}^*), (\alpha_{k-1}^{-1}\alpha_j)^n(\alpha_j^h(a_j^*))]\|_{L^2(\tau)} \to 0
\end{aligned}$$

as $N \to \infty$, by the asymptotic abelianness of $\alpha_{k-1}^{-1}\alpha_j$. The same reasoning applies to the term $\sum_{j=1}^{k-2} u_j[b_{k-1}, c_j]v_j$, and now applies again to show that in the scalar average of interest to us we may also commute b_{k-2} from the left end of our product over to be immediately on the left of c_{k-2}, and then move b_{k-3} to c_{k-3}, and so on. Overall, this shows that

$$\begin{aligned}
&\frac{1}{H}\sum_{h=1}^{H}\Big|\frac{1}{N}\sum_{n=1}^{N}\tau\big(\alpha_{k-1}^n(\alpha_{k-1}^h(a_{k-1}^*))\cdots\alpha_1^n(\alpha_1^h(a_1^*))\cdot\alpha_1^n(a_1)\cdots\alpha_{k-1}^n(a_{k-1})\big)\Big| \\
&\quad \sim \frac{1}{H}\sum_{h=1}^{H}\Big|\frac{1}{N}\sum_{n=1}^{N}\tau\big(\alpha_1^n(\alpha_1^h(a_1^*)a_1)\cdots\alpha_{k-1}^n(\alpha_{k-1}^h(a_{k-1}^*)a_{k-1})\big)\Big|
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{H}\sum_{h=1}^{H}\Big|\frac{1}{N}\sum_{n=1}^{N}\tau\Big(\alpha_1^h(a_1^*)a_1\cdot(\alpha_2\alpha_2^{-1})^n(\alpha_2^h(a_2^*)a_2) \\
&\qquad\qquad\qquad\qquad \cdots(\alpha_{k-1}\alpha_1^{-1})^n(\alpha_{k-1}^h(a_{k-1}^*)a_{k-1})\Big)\Big| \\
&= \frac{1}{H}\sum_{h=1}^{H}\Big|\tau\Big(\alpha_1^h(a_1^*)a_1\cdot\Big(\frac{1}{N}\sum_{n=1}^{N}(\alpha_2\alpha_1^{-1})^n(\alpha_2^h(a_2^*)a_2) \\
&\qquad\qquad\qquad\qquad \cdots(\alpha_{k-1}\alpha_1^{-1})^n(\alpha_{k-1}^h(a_{k-1}^*)a_{k-1})\Big)\Big)\Big|
\end{aligned}
$$

as $N\to\infty$ and then $H\to\infty$. However, now we notice that the inner average of operators with respect to N here is precisely of the form hypothesized by the theorem, but involving only the $k-1$ automorphisms $\alpha_j\alpha_1^{-1}$ for $j=1,2,\dots,k-1$, which still satisfy the necessary hypotheses of relative weak mixing and asymptotic abelianness. Hence this operator average asymptotically agrees with

$$
\begin{aligned}
&\frac{1}{H}\sum_{h=1}^{H}\Big|\tau\Big(\alpha_1^h(a_1^*)a_1\cdot\Big(\frac{1}{N}\sum_{n=1}^{N}(\alpha_2\alpha_1^{-1})^n(E_{\mathcal{N}}(\alpha_2^h(a_2^*)a_2)) \\
&\qquad\qquad\qquad\qquad \cdots(\alpha_{k-1}\alpha_1^{-1})^n(E_{\mathcal{N}}(\alpha_{k-1}^h(a_{k-1}^*)a_{k-1}))\Big)\Big| \\
&= \frac{1}{H}\sum_{h=1}^{H}\Big|\tau\Big(E_{\mathcal{N}}(\alpha_1^h(a_1^*)a_1)\cdot\Big(\frac{1}{N}\sum_{n=1}^{N}(\alpha_2\alpha_1^{-1})^n(E_{\mathcal{N}}(\alpha_2^h(a_2^*)a_2)) \\
&\qquad\qquad\qquad\qquad \cdots(\alpha_{k-1}\alpha_1^{-1})^n(E_{\mathcal{N}}(\alpha_{k-1}^h(a_{k-1}^*)a_{k-1}))\Big)\Big)\Big|,
\end{aligned}
$$

where the second equality holds because the operator average in the inner brackets now lies in $\mathcal{N}$, and so we apply the usual identity for conditional expectations $\tau(aE_{\mathcal{N}}(b))=\tau(E_{\mathcal{N}}(aE_{\mathcal{N}}(b)))=\tau(E_{\mathcal{N}}(a)E_{\mathcal{N}}(b))$.

Writing

$$
s_N := \frac{1}{N}\sum_{n=1}^{N}(\alpha_2\alpha_1^{-1})^n(E_{\mathcal{N}}(\alpha_2^h(a_2^*)a_2))\cdots(\alpha_{k-1}\alpha_1^{-1})^n(E_{\mathcal{N}}(\alpha_{k-1}^h(a_{k-1}^*)a_{k-1})),
$$

we see that $\|s_N\|\le C$ for some fixed C and all $N\in\mathbb{N}$, and now combining this bound with the Cauchy–Schwarz inequality we obtain

$$
\begin{aligned}
\frac{1}{H}\sum_{h=1}^{H}|\tau(E_{\mathcal{N}}(\alpha_1^h(a_1^*)a_1)\cdot s_n)| &= \frac{1}{H}\sum_{h=1}^{H}|\langle s_n^*\hat{1},(E_{\mathcal{N}}(\alpha_1^h(a_1^*)a_1)\hat{1}\rangle_{L^2(\tau)}| \\
&\le \frac{1}{H}\sum_{h=1}^{H}C\cdot\|E_{\mathcal{N}}(\alpha_1^h(a_1^*)a_1\|_{L^2(\tau)}.
\end{aligned}
$$

Finally, it follows that this tends to 0 as $H\to\infty$ by the our assumption that $a_1\perp\mathcal{N}$ and the relative weak mixing hypothesis. This completes the proof of Theorem 1.13.

Appendix B. A group theory construction

The purpose of this appendix is to explicitly describe a certain type of group, which we shall term a *square group*, generated by relations involving quadruples of generators. In particular, we will be able to solve the equality problem for such groups. Our arguments here are motivated by an observation of Grothendieck that groups can be identified with the sheaf of their flat connections on simplicial complexes, and experts will be able to detect the ideas of sheaf theory lurking beneath the surface of the material here, although we will not use that theory explicitly.

Definition B.1 (square groups). A *square base* $\square = (H \cup V, \square)$ consists of the following data:

- A set $H \cup V$ of generators, partitioned into a subset H of *horizontal generators* and a subset V of *vertical generators*.
- A set $\square \subset (H \times V \times H \times V) \cup (V \times H \times V \times H)$ of quadruples (e_0, e_1, e_2, e_3) of alternating orientation (thus if e_0 is horizontal then e_1 must be vertical, and so forth).

Furthermore, we require the two axioms on the set $\square$:

- (Cyclic symmetry.) If $(e_0, e_1, e_2, e_3) \in \square$, then $(e_1, e_2, e_3, e_0) \in \square$.
- (Unique continuation.) If $e_0, e_1 \in H \cup V$, then there is at most one quadruple $(e_0, e_1, e_2, e_3) \in \square$ with the first two components e_0 and e_1.

If $\square$ is a square base, we define the *square group* $G_\square$ associated to that base to be the group generated by the generators $H \cup V$, subject to the relations $e_0e_1e_2e_3 = \mathrm{id}$ for all $(e_0, e_1, e_2, e_3) \in \square$. We define the *alphabet* of the square base (or square group) to be the set $H \cup V \cup H^{-1} \cup V^{-1}$ consisting of the horizontal and vertical generators and their formal inverses.

To describe square groups explicitly, we shall need some notation of a combinatorial and geometric nature. Let $\mathbb{N} := \{0, 1, 2, \dots\}$ denote the natural numbers.

Definition B.2 (monotone paths and regions). A *monotone path* is a finite path in the discrete quadrant $\mathbb{N}^2$ from $(0, 0)$ to some endpoint (n, m) that consists only of rightward edges $(i, j) \to (i + 1, j)$ and upward edges $(i, j) \to (i, j + 1)$ (in particular, the path will have length $n + m$). Given a monotone path γ from $(0, 0)$ to (n, m), the *shadow* of γ is defined to be all the pairs $(i, j) \in \mathbb{N}^2$ such that $(i, j') \in \gamma$ for some $j' \geq j$. We say that one monotone path γ' *lies above* another monotone path γ with the same endpoint (n, m) if the shadow of γ' contains the shadow of γ. In such cases, we refer to the set-theoretic difference between the two shadows as a *monotone region* from $(0, 0)$ to (n, m), with γ' and γ referred to as the *upper boundary* and *lower boundary* of the region, respectively.

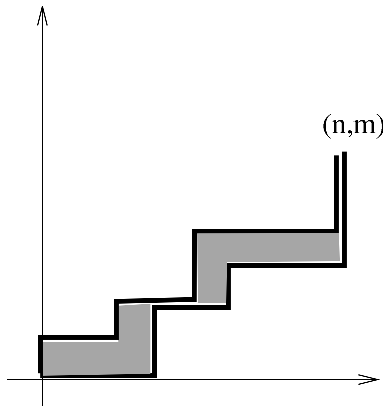

Figure 2. A monotone region, bounded above and below by two monotone paths. Note the horizontal and vertical convexity of the monotone region.

We will also consider a monotone path as a degenerate example of a monotone region. Monotone regions are horizontally and vertically convex: if two endpoints of a horizontal or vertical line segment in $\mathbb{N}^2$ lie in a monotone region, then the interior of that segment does also.

Definition B.3 (flat connections). Fix a square base $\square$, and let $\Omega \subset \mathbb{N}^2$ be a set. A *connection* Γ on Ω is an assignment $\Gamma((i, j) \to (i+1, j)) \in H \cup H^{-1}$ of a horizontal element of the alphabet to every horizontal edge $(i, j), (i+1, j) \in \Omega$, and an assignment $\Gamma((i, j) \to (i, j+1)) \in V \cup V^{-1}$ of a vertical element of the alphabet to every vertical edge $(i, j) \mapsto (i, j+1) \in \Omega$. We adopt the convention that

$$\Gamma((i+1, j) \to (i, j)) := \Gamma((i, j) \to (i+1, j))^{-1},$$
$$\Gamma((i, j+1) \to (i, j)) := \Gamma((i, j) \to (i, j+1))^{-1},$$

where $(e^{-1})^{-1} := e$ for $e \in H \cup V$ of course.

We say that the connection Γ is *flat* if for every square $(i, j), (i+1, j), (i, j+1), (i+1, j+1)$ in Ω, there exists an oriented loop f_0, f_1, f_2, f_3 of horizontal and vertical edges around the square (in either orientation) such that

$$(\Gamma(f_0), \Gamma(f_1), \Gamma(f_2), \Gamma(f_3)) \in \square.$$

We call a flat connection on a monotone region from $(0, 0)$ to (n, m) *maximal* if it cannot be extended to any strictly larger monotone region with the same endpoints. It is *reduced* if there does not exist a triple $(i, j), (i+1, j), (i+2, j)$ or $(i, j), (i, j+1), (i, j+2)$ in Ω such that $\Gamma((i, j) \to (i+1, j))\Gamma((i+1, j) \to (i+2, j)) = \mathrm{id}$ or $\Gamma((i, j+1) \to (i, j))\Gamma((i, j+1) \to (i, j+2)) = \mathrm{id}$.

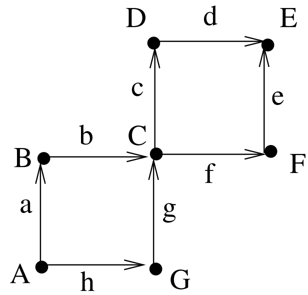

Figure 3. A monotone region $\{A, B, C, D, E, F, G\}$ (with $A = (0, 0)$, $B = (0, 1)$, and so on) with a connection Γ defined by the group elements $a, b, c, d, e, f, g, h \in G_\Box$; thus for instance $\Gamma(B \to C) = b$ and $\Gamma(C \to B) = b^{-1}$. If say (a, b, g^{-1}, h^{-1}) and (f, e, d^{-1}, c^{-1}) are in $\Box$, then this connection is flat.

In the degenerate case when Ω is just a monotone path, every connection is automatically flat, as there are no squares.

Let Γ be a flat connection on a monotone region Ω. Then one can *integrate* this connection to produce a map $\Phi_\Gamma : \Omega \to G_\Box$ by setting $\Phi_\Gamma(0, 0) := \mathrm{id}$ and $\Phi_\Gamma(v) = \Phi_\Gamma(u)\Gamma(u \to v)$ for all horizontal and vertical edges $(u \to v)$ in Ω. From the flatness of Γ and the "connected" nature of Ω it is easy to see that Φ_Γ exists and is unique. In particular, we can define the *definite integral* $|\Gamma|$ of Γ to be the group element $|\Gamma| := \Phi_\Gamma(n, m)$, where (n, m) is the endpoint of Ω.

Example B.4. The definite integral of the flat connection in Figure 3 is equal to $abcd = abfe = hgcd = hgfe$.

Every group element g in $G_\Box$ can arise as a definite integral of some flat connection, simply by expressing g as a word in the alphabet $H \cup V \cup H^{-1} \cup V^{-1}$, and creating an associated monotone path and connection for that word. Later on we shall see that the definite integral will provide a one-to-one correspondence between group elements and maximal reduced flat connections (Corollary B.10).

Lemma B.5. *Let $\Box$ be a square base, and let $(n, m) \in \mathbb{N}^2$.*

- (Unique continuation.) *If Ω is a monotone region from $(0, 0)$ to (n, m), and γ is a path from $(0, 0)$ to (n, m) in Ω, then any flat connection on Ω is uniquely determined by its restriction to γ. In other words, if Γ and Γ' are two flat connections on Ω that agree on γ, then they agree on all of Ω.*
- (Maximality.) *If Ω_0 is a monotone region from $(0, 0)$ to (n, m), and Γ is a flat connection on Ω_0, then there exists a unique extension of Γ to a maximal flat connection on a monotone region Ω from $(0, 0)$ to (n, m) containing Ω_0.*

Proof. We first establish unique continuation. This is best explained visually. The key observation is that if two flat connections on a square agree on two adjacent sides of a square, then they must agree on the whole square. This is ultimately a consequence of the unique continuation property of the square base $\Box$, and can be verified by a routine case check. Thus, if Γ and Γ' are two connections on Ω that agree on γ, they also agree on any perturbation of γ in Ω formed by taking an adjacent pair of horizontal and vertical edges in γ and "popping" them by replacing them by the other two edges of the square that they form; note that this retains the property of being a monotone path. One can check that after a sufficient number of upward and downward "popping" operations one can cover the upper and lower boundaries of Γ, and everything in between, and the claim follows.

Example B.6. We continue working with Figure 3. Suppose two flat connections Γ and Γ' on the indicated region agree on the upper boundary $ABCDE$, with the indicated connection values a, b, c, d. By unique continuation of $\Box$, the only possible values available for Γ and Γ' on the remaining two edges CF, FE of the square $CDEF$ are f and e. Thus we may "pop" the upper square and obtain that Γ and Γ' also agree on the monotone path $ABCFE$. After popping the lower square also we obtain that Γ and Γ' agree on the entire monotone region.

To prove the second claim, we simply observe that if Γ can be extended to two monotone regions Ω and Ω' containing Ω_0, then by unique continuation they agree on the intersection $\Omega \cap \Omega'$ (which is also a monotone region), and can thus be glued to form a flat connection on the union $\Omega \cup \Omega'$ (which is also a monotone region[6]). Since there are only finitely many monotone regions from $(0, 0)$ to (n, m), the claim then follows from the greedy algorithm. $\Box$

Definition B.7 (concatenation). Let Γ be a maximal reduced flat connection on some monotone region Ω from $(0, 0)$ to (n, m), and let $x \in H \cup V \cup H^{-1} \cup V^{-1}$ be a symbol in the alphabet. We define the *concatenation* $\Gamma \cdot x$ of Γ with x to be the maximal flat connection $\Gamma' = \Gamma \cdot x$ on a monotone region Ω' from $(0, 0)$ to (n', m') generated by the following rule.

- (Collapse.) If x is horizontal (that is, $x \in H \cup H^{-1}$), if $(n-1, m)$ lies in Ω, and if $\Gamma((n-1, m) \to (n, m)) = x^{-1}$, then one sets $(n', m') := (n-1, m)$, sets Ω' to be the restriction of Ω to the region $\{(i, j) \in \mathbb{N}^2 : i \leq n-1\}$ (that is, one deletes the rightmost column of Ω), and sets Γ' to be the restriction of Γ to Ω'.
- (Extension.) If x is horizontal, and either $(n-1, m)$ lies outside of Ω or $\Gamma((n-1, m) \to (n, m)) \neq x^{-1}$, then one sets $(n', m') := (n+1, m)$, and

[6] One way to see this is to rotate the plane by 45 degrees, so that monotone paths become graphs of discrete Lipschitz functions with Lipschitz constant 1, and monotone regions become the regions between two such functions.

extends Γ to $\Omega \cup \{(n+1, m)\}$ by setting $\Gamma((n, m) \to (n+1, m)) := x$; note that this is still flat because it does not create any squares. One then extends Γ further by the second part of Lemma B.5 to create the maximal flat connection Γ' on Ω' that extends Γ.

- If x is vertical instead of horizontal, one follows the analogue of the above rules but with the roles of n and m reversed.

Example B.8. Imagine one concatenated a horizontal edge x to the flat connection in Figure 3, which we shall assume to be maximal reduced. If x is not equal to d^{-1}, then the concatenated connection would thus extend one unit to the right of E to the endpoint $(3, 2)$, and may possibly extend also to the square to the right of EF if there is an appropriate tuple in $\square$ to achieve this extension. If instead x was equal to d^{-1}, then the connection would collapse to the region $\{A, B, C, D, G\}$, so that the endpoint is now $D = (1, 2)$.

This definition gives a representation of $G_\square$:

Lemma B.9. *Let $\square$ be a square base and Γ a maximal reduced flat connection.*

- (Preservation of reducibility.) $\Gamma \cdot x$ *is reduced for any* $x \in H \cup V \cup H^{-1} \cup V^{-1}$.
- (Invertibility.) *We have* $(\Gamma \cdot x) \cdot x^{-1} = \Gamma$ *for any* $x \in H \cup V \cup H^{-1} \cup V^{-1}$.
- (Square relations.) *We have* $(((\Gamma \cdot e_0) \cdot e_1) \cdot e_2) \cdot e_3 = \Gamma$ *for any* $(e_0, e_1, e_2, e_3) \in \square$.

In particular, the group $G_\square$ acts on the space $\mathbb{O}$ of maximal reduced flat connections in a unique manner, sending Γ to $\Gamma \cdot g$ for any $\Gamma \in \mathbb{O}$ and $g \in G_\square$.

Proof. We begin with the preservation of reducibility claim. If $\Gamma \cdot x$ is formed by collapsing Γ, the claim is clear, so suppose instead that $\Gamma \cdot x$ is formed by extension. By symmetry we may assume that x is horizontal. Let (n, m) denote the endpoint of Γ, and let Ω' be the domain of $\Gamma \cdot x$ (which then has endpoint $(n+1, m)$).

Assume for contradiction that $\Gamma \cdot x$ is not reduced. Since Γ was reduced, there are only two possibilities: either one has a vertical degeneracy

$$\Gamma((n+1, j) \to (n+1, j+1))\Gamma((n+1, j+1) \to (n+1, j+2)) = \mathrm{id} \tag{30}$$

for some $(n+1, j), (n+1, j+1), (n+1, j+2) \in \Omega'$, or else one has a horizontal degeneracy

$$\Gamma((n-1, j) \to (n, j))\Gamma((n, j) \to (n+1, j)) = \mathrm{id} \tag{31}$$

for some $(n-1, j), (n, j), (n+1, j) \in \Omega'$.

Suppose first that one has a vertical degeneracy (30). Consider the restrictions Γ_1 and Γ_2 of the connection Γ on the adjacent squares

$$((n, j), (n, j+1), (n+1, j), (n+1, j+1)) \quad \text{and}$$
$$((n, j+1), (n, j+2), (n+1, j+1), (n+1, j+2)).$$

By construction Γ_1 and Γ_2 agree on their common edge $((n, j+1) \to (n+1, j+1))$, and $\Gamma_1((n+1, j+1) \to (n+1, j))$ is equal to $\Gamma_2((n+1, j+1) \to (n+1, j+2))$. By the unique continuation property of $\Box$, this implies that Γ_1 and Γ_2 are reflections of each other; in particular $\Gamma_1((n, j+1) \to (n, j))$ equals $\Gamma_2((n, j+1) \to (n, j+2))$. But this implies that Γ is not reduced, a contradiction.

Suppose instead that one has a horizontal degeneracy (31). From Definition B.7 we know that j cannot equal m, otherwise we would have collapsed rather than extended Γ. Let $0 \leq j < m$ be the largest j for which (31) holds. By repeating the argument in the previous paragraph, we see that the restrictions of Γ to the adjacent squares

$$\begin{aligned}&((n-1, j), (n, j), (n-1, j+1), (n, j+1)) \quad \text{and}\\ &((n, j), (n+1, j), (n, j+1), (n+1, j+1))\end{aligned}$$

are reflections of each other, which implies that (31) also holds for $j+1$, contradicting the maximality of j. This establishes the preservation of reducibility.

Now we establish the invertibility. Again, by symmetry we may assume that x is horizontal.

If $\Gamma \cdot x$ is a (horizontal) extension of Γ, then it is easy to see from Definition B.7 that $(\Gamma \cdot x) \cdot x^{-1}$ will be the (horizontal) collapse of $\Gamma \cdot x$, which is Γ. Conversely, if $\Gamma \cdot x$ is the (horizontal) collapse of Γ, then $(\Gamma \cdot x) \cdot x^{-1}$ will be the (horizontal) extension (because Γ was reduced), which will equal Γ again (by uniqueness of maximal extension).

Finally, we establish the square relations. From cyclic symmetry and invertibility we may assume that e_0 and e_2 are horizontal and e_1 and e_3 are vertical. From invertibility again, it suffices to show that

$$(\Gamma \cdot e_0) \cdot e_1 = (\Gamma \cdot e_3^{-1}) \cdot e_2^{-1}$$

for any maximal reduced flat connection Γ. We denote the endpoint of Γ by (n, m).

We divide into four cases. Suppose first that $\Gamma \cdot e_0$ is an extension of Γ, and that $(\Gamma \cdot e_0) \cdot e_1$ is an extension of $\Gamma \cdot e_0$. Then we claim that $\Gamma \cdot e_3^{-1}$ is an extension of Γ. If this were not the case, then $\Gamma((n, m-1) \to (n, m))$ must equal e_3, but then since $(\Gamma \cdot e_0)((n, m) \to (n+1, m))$ equals e_0 by construction, the domain of $\Gamma \cdot e_0$ must include the square $(n, m-1)$, (n, m), $(n+1, m-1)$, $(n+1, m)$ with

$$(\Gamma \cdot e_0)((n+1, m-1) \to (n+1, m)) = e_1^{-1},$$

causing $(\Gamma \cdot e_0) \cdot e_1$ to be a collapse rather than an extension, a contradiction. Thus $\Gamma \cdot e_3^{-1}$ extends Γ. A similar argument shows that $(\Gamma \cdot e_3^{-1}) \cdot e_2^{-1}$ extends $\Gamma \cdot e_3^{-1}$ (otherwise $\Gamma((n-1, m) \to (n, m))$ would equal e_0^{-1}, causing $\Gamma \cdot e_0$ to be a collapse rather than an extension). It is then easy to verify that $(\Gamma \cdot e_3^{-1}) \cdot e_2^{-1}$ and $(\Gamma \cdot e_0) \cdot e_1$

are the same since they glue together to form a flat connection on Γ and on the square (n, m), $(n+1, m)$, $(n, m+1)$, $(n+1, m+1)$.

Now suppose that $\Gamma \cdot e_0$ is an extension of Γ, but that $(\Gamma \cdot e_0) \cdot e_1$ is a collapse of $\Gamma \cdot e_0$. Arguing as before, we conclude that $\Gamma((n, m-1) \to (n, m))$ equals e_3, and so $\Gamma \cdot e_3^{-1}$ is a collapse of Γ; similarly, $(\Gamma \cdot e_3^{-1}) \cdot e_2^{-1}$ cannot be a collapse of $\Gamma \cdot e_3^{-1}$ (this would force $\Gamma \cdot e_0$ to be a collapse also) and so is an extension. It is again easy to verify that $(\Gamma \cdot e_3^{-1}) \cdot e_2^{-1}$ and $(\Gamma \cdot e_0) \cdot e_1$ are the same.

The remaining two cases (when $\Gamma \cdot e_0$ is a collapse of Γ, and $(\Gamma \cdot e_0) \cdot e_1$ is either an extension or collapse of $\Gamma \cdot e_0$) are similar to the preceding two, and are left to the reader. $\square$

This gives us a satisfactory explicit description of a square group:

Corollary B.10. *Let $\square$ be a square group. Then the definite integral map $\Gamma \mapsto |\Gamma|$ is a bijection from $\mathbb{O}$ to $G_\square$; thus every group element has a unique representation as the definite integral of a maximal reduced flat connection.*

Proof. The surjectivity of this map was already established in the discussion after Definition B.3, so it suffices to establish the injectivity. We will establish this via the identity $\Gamma = \varnothing \cdot |\Gamma|$ for all for all $\Gamma \in \mathbb{O}$, where $\varnothing$ is the trivial flat connection over the monotone region $\{(0, 0)\}$ from $(0, 0)$ to $(0, 0)$. This identity shows that Γ can be reconstructed from $|\Gamma|$, demonstrating injectivity.

Let Ω be the domain of Γ, which by definition is a monotone region from $(0, 0)$ to some point (n, m). Let γ be some monotone path in Ω from $(0, 0)$ to (n, m) (for example, one could take γ to be the upper or lower boundary of Ω). We label the vertices of γ in order as $(0, 0) = (i_0, j_0), (i_1, j_1), \ldots, (i_{n+m}, j_{n+m}) = (n, m)$. From definition of $|\Gamma|$, we see that

$$|\Gamma| = \Gamma((i_0, j_0) \to (i_1, j_1))\Gamma((i_1, j_1) \to (i_2, j_2)) \cdots \Gamma((i_{n+m-1}, j_{n+m-1}) \to (i_{n+m}, j_{n+m})).$$

For each $0 \leq k \leq n + m$, defined Ω_k to be the portion of Ω that is in the region $\{(i, j) : i \leq i_k, j \leq j_k\}$; thus Ω_k is a monotone region from $(0, 0)$ to (i_k, j_k) that is increasing in k. Let Γ_k be the restriction of Γ to Ω_k. Since Γ was maximal and reduced, each of the Γ_k is also. Since $\Gamma_{n+m} = \Gamma$, it will suffice to establish that

$$\Gamma_k = \varnothing \cdot \Gamma((i_0, j_0) \to (i_1, j_1))\Gamma((i_1, j_1) \to (i_2, j_2)) \cdots \Gamma((i_{k-1}, j_{k-1}) \to (i_k, j_k))$$

for all $0 \leq k \leq n+m$. But this is easily established by induction (the reduced nature of the Γ_k is necessary to avoid the collapse case in Definition B.7). $\square$

As a consequence of this corollary, we can distinguish any two elements in $G_\square$ from each other as long as we can express them as the definite integrals of distinct maximal reduced flat connections.

Applications. We now specialise the abstract group-theoretic machinery above to the application at hand. We begin with a proposition that will be used to show nonconvergence of quadruple recurrence (Theorem 2.1).

Proposition B.11 (independence of AP4 relations). *Let $A \subset \mathbb{Z}$ be a (possibly infinite) set of integers. Then there exist a group G with elements e_0, e_1, e_2, e_3, together with an automorphism $T : G \to G$, such that for $r \in \mathbb{N}$, the relation*

$$e_0(T^r e_1)(T^{2r} e_2)(T^{3r} e_3) = \mathrm{id} \tag{32}$$

holds if and only if $r \in A$. Furthermore, no power T^k of T with $k \neq 0$ has any fixed points other than the identity element id.

Remark B.12. Informally, this proposition asserts that the algebraic relations (32) for various $r \in \mathbb{Z}$ are independent of each other. In contrast, with progressions of length three (that is, in the case $k = 3$) the analogous relations are highly degenerate. Indeed, suppose that

$$e_0(T^r e_1)(T^{2r} e_2) = \mathrm{id} \tag{33}$$

for all $r \in A$. Then if $r, r + h$ lie in A, we have

$$e_0(T^r e_1)(T^{2r} e_2) = e_0(T^r T^h e_1)(T^{2r} T^{2h} e_2),$$

which we can rearrange as $(T^h e_1^{-1})e_1 = T^r((T^{2h} e_2)e_2^{-1})$. If $r, r + h, r', r' + h$ lie in A, we thus have

$$T^r((T^{2h} e_2)e_2^{-1}) = T^{r'}((T^{2h} e_2)e_2^{-1}).$$

Assuming that $T^{r'-r}$ has no fixed points, we conclude that $(T^{2h} e_2)e_2^{-1}$ is the identity; assuming that T^{2h} has no fixed points either, we conclude that e_2 is the identity. Similar arguments can be used to show that e_0 and then e_1 are also the identity. Thus the relations (33) and the no-fixed-points hypothesis lead to a total collapse of the group generated by e_0, e_1, e_2 as soon as A contains even a single nontrivial parallelogram $r, r + h, r', r' + h$. (A variant of this argument also shows that if (33) is obeyed for r and $r + h$, then it is also obeyed for $r + 2h$ even without the fixed point hypothesis.) This algebraic distinction between triple recurrence and quadruple recurrence can be viewed as the primary reason why recurrence and convergence results continue to hold for triple products, but not for quadruple products even under the assumption of ergodicity (which is reflected here in the no-fixed-points assumption).

Proof. We let G be the group generated by the generators $e_{i,n}$ for $i = 0, 1, 2, 3$ and $n \in \mathbb{Z}$, subject to the relations

$$e_{0,n} e_{1,n+r} e_{2,n+2r} e_{3,n+3r} = \mathrm{id} \quad \text{for all } n \in \mathbb{Z} \text{ and } r \in A.$$

Since the set of such relations is invariant under the shift $e_{i,n} \mapsto e_{i,n+1}$, we see that we can define an automorphism $T : G \to G$ by setting $Te_{i,n} := e_{i,n+1}$. If we then set $e_i := e_{i,0}$, it is clear that (32) holds for all $r \in A$.

To see that (32) fails for $r \notin A$, we note that G can be viewed as a square group, with horizontal generators $\{e_{i,n} : i = 0, 2; n \in \mathbb{Z}\}$ and vertical generators $\{e_{i,n} : i = 1, 3; n \in \mathbb{Z}\}$ and square relations $\Box$ made of $(e_{0,n}, e_{1,n+r}, e_{2,n+2r}, e_{3,n+3r})$ and its cyclic permutations for all $n \in \mathbb{Z}$ and $r \in A$; note that the crucial unique continuation property follows from the basic observation that an arithmetic progression is determined by any two of its elements ("two points determine a line"). If $n \in \mathbb{Z}$ and $r \notin A$, one sees that the connection on the path of length four from $(0, 0)$ to $(2, 2)$ associated to the word $e_{0,n}e_{1,n+r}e_{2,n+2r}e_{3,n+3r}$ is already a maximal reduced flat connection (as none of the three squares that share two edges with the path can be completed to a square from $\Box$) and so by Corollary B.10, its definite integral $e_{0,n}e_{1,n+r}e_{2,n+2r}e_{3,n+3r}$ is not equal to the identity, as required.

Finally, to show that T^k has no nontrivial fixed points, one simply observes that T^k will shift any nontrivial maximal reduced flat connection to a different maximal reduced flat connection, and then invokes Corollary B.10 again. $\Box$

Next, we establish a variant that is useful for showing negative averages for quintuple recurrence (Theorem 2.12).

Proposition B.13 (independence of AP5 relations). *There exists a group G with distinct elements e_0, e_1, e_2, e_3, e_4, together with an automorphism $T : G \to G$, such that the relation*

$$e_0(T^r e_1)(T^{2r} e_2)(T^{3r} e_3)(T^{4r} e_4) = \mathrm{id} \tag{34}$$

holds for all $r \in \mathbb{Z}$. Furthermore, no power T^k of T with $k \neq 0$ has any fixed points other than the identity element id. *Finally, if $r \in \mathbb{Z}$ is nonzero, and*

$$g_0, g_1, g_2, g_3, g_4 \in \{\mathrm{id}, e_0, e_1, e_2, e_3, e_4, e_0^{-1}, e_1^{-1}, e_2^{-1}, e_3^{-1}, e_4^{-1}\}$$

are such that

$$g_0(T^r g_1)(T^{2r} g_2)(T^{3r} g_3)(T^{4r} g_4) = \mathrm{id}, \tag{35}$$

then g_0, g_1, g_2, g_3, g_4 are either equal to the identity, or are a permutation of $\{e_0, e_1, e_2, e_3, e_4\}$ or of $\{e_0^{-1}, e_1^{-1}, e_2^{-1}, e_3^{-1}, e_4^{-1}\}$.

Proof. For each $i = 0, 1, 2, 3, 4$, we define $G^{(i)}$ to be the group generated by the generators $e^{(i)}_{j,n}$ for $j \in \{0, 1, 2, 3, 4\} \setminus \{i\}$ and $n \in \mathbb{Z}$ subject to the relations

$$e^{(i)}_{0,n} e^{(i)}_{1,n+r} e^{(i)}_{2,n+2r} e^{(i)}_{3,n+3r} e^{(i)}_{4,n+4r} = \mathrm{id} \quad \text{for all } n, r \in \mathbb{Z}, \tag{36}$$

with the convention that $e^{(i)}_{i,n} = \mathrm{id}$ for all n. This group has an automorphism $T^{(i)} : G^{(i)} \to G^{(i)}$ that maps $e^{(i)}_{j,n}$ to $e^{(i)}_{j,n+1}$ for all n.

We now set G to be the product group $G := G^{(0)} \times G^{(1)} \times \cdots \times G^{(4)}$, and set

$$e_j := (e^{(0)}_{j,0}, e^{(1)}_{j,0}, \ldots, e^{(4)}_{j,0}) \quad \text{for } j = 0, 1, 2, 3, 4.$$

We also set

$$T(g^{(0)}, g^{(1)}, \ldots, g^{(4)}) := (T^{(0)} g^{(0)}, T^{(1)} g^{(1)}, \ldots, T^{(4)} g^{(4)});$$

thus T is an automorphism on G. By construction it is clear that (34) holds. Also, by the arguments in Proposition B.11, no nonzero power of $T^{(i)}$ has any nontrivial fixed points, and so the same is also true of T.

Now we establish the final claim of the proposition. Suppose $g_0, \ldots, g_4$ obey the stated properties. Let $i = 0, 1, 2, 3, 4$, and let $g^{(i)}_j$ be the $G^{(i)}$ component of g_j for $j = 0, 1, 2, 3, 4$; thus

$$g^{(i)}_0((T^{(i)})^r g^{(i)}_1)((T^{(i)})^{2r} g^{(i)}_2)((T^{(i)})^{3r} g^{(i)}_3)((T^{(i)})^{4r} g^{(i)}_4) = \mathrm{id}\,. \tag{37}$$

From the construction of $G^{(i)}$, we see that for any distinct $j, k \in \{0, 1, 2, 3, 4\} \setminus \{i\}$, there is a homomorphism $\phi^{(i)}_{j,k} : G^{(i)} \to \mathbb{Z}$ to the additive group $\mathbb{Z}$ mapping $e^{(i)}_{j,n}$ to $+1$, $e^{(i)}_{k,n}$ to -1, and all other $e^{(i)}_{l,n}$ to zero for $n \in \mathbb{Z}$ and $l \in \{0, 1, 2, 3, 4\} \setminus \{i, j, k\}$ (note that these requirements are compatible with the defining relations (36)). This homomorphism is $T^{(i)}$ invariant. Applying this homomorphism to (37), we obtain $\sum_{l=0}^{4} \phi^{(i)}_{j,k}(g^{(i)}_l) = 0$.

In other words, the number of times g_l for $l = 0, 1, 2, 3, 4$ equals e_j, minus the number of times it equals e_j^{-1}, is equal to the number of times g_l equals e_k, minus the number of times it equals e_k^{-1}. Letting j, k, i vary, we thus see that this number is independent of j. It is easy to see that this number cannot exceed 1 in magnitude, and if it is equal to $+1$ or -1, then g_0, g_1, g_2, g_3, g_4 is a permutation of $\{e_0, e_1, e_2, e_3, e_4\}$ or of $\{e_0^{-1}, e_1^{-1}, e_2^{-1}, e_3^{-1}, e_4^{-1}\}$, respectively. (Note that this argument also ensures that e_0, e_1, e_2, e_3 and e_4 are distinct.) The remaining possibility to eliminate is when this number is zero, thus each e_i occurs in g_0, g_1, g_2, g_3, g_4 as often as e_i^{-1}. Suppose for instance that g_0, g_1, g_2, g_3, g_4 contains one occurrence each of $e_0, e_0^{-1}, e_1, e_1^{-1}$. Applying (37) with $i = 4$ (say), and then applying the homomorphism that maps $e^{(4)}_{0,n}$ to zero, $e^{(4)}_{1,n}$ to n, $e^{(4)}_{2,n}$ to $-2n$, and $e^{(4)}_{3,n}$ to n (here we use the identity $(n+r)-2(n+2r)+(n+3r) = 0$ to ensure consistency with (36)), we obtain a contradiction. We argue similarly if g_0, g_1, g_2, g_3, g_4 contains any other combination of one or two distinct pairs e_j, e_j^{-1}. The remaining case to eliminate is if g_0, g_1, g_2, g_3, g_4 contains e_j and e_j^{-1} twice each for some j, say $j = 0$. Applying (37) with $i = 4$ again, we can use Corollary B.10 to contradict (37), since the right side is a definite integral of a maximal flat connection on a horizontal path of length four. We argue similarly for other values of j, and the claim follows. □

Acknowledgements

Our thanks go to Sorin Popa for several helpful discussions, Francesco Fidaleo and David Kerr for references, and to Ezra Getzler for explaining Grothendieck's interpretation of a group via its sheaf of flat connections. The authors are indebted to the anonymous referee for careful comments and suggestions.

References

[Austin 2010] T. Austin, "On the norm convergence of non-conventional ergodic averages", *Ergodic Theory Dynam. Systems* **30**:2 (2010), 321–338. MR 2599882 Zbl 05702784

[Behrend 1946] F. A. Behrend, "On sets of integers which contain no three terms in arithmetical progression", *Proc. Nat. Acad. Sci. U. S. A.* **32** (1946), 331–332. MR 8,317d Zbl 0060.10302

[Bergelson 1987] V. Bergelson, "Weakly mixing PET", *Ergodic Theory Dynam. Systems* **7**:3 (1987), 337–349. MR 89g:28022 Zbl 0645.28012

[Bergelson and Leibman 2004] V. Bergelson and A. Leibman, "Failure of the Roth theorem for solvable groups of exponential growth", *Ergodic Theory Dynam. Systems* **24**:1 (2004), 45–53. MR 2005e:37006 Zbl 1115.37003

[Bergelson et al. 2005] V. Bergelson, B. Host, and B. Kra, "Multiple recurrence and nilsequences", *Invent. Math.* **160**:2 (2005), 261–303. MR 2007i:37009 Zbl 1087.28007

[Beyers et al. 2010] C. Beyers, R. Duvenhage, and A. Ströh, "The Szemerédi property in ergodic W^*-dynamical systems", *J. Operator Theory* **64**:1 (2010), 35–67. MR 2669426 Zbl 05776146

[Bratteli and Robinson 1987] O. Bratteli and D. W. Robinson, *Operator algebras and quantum statistical mechanics, I: C^*- and W^*-algebras, symmetry groups, decomposition of states*, 2nd ed., Springer, New York, 1987. MR 88d:46105 Zbl 0905.46046

[Conze and Lesigne 1984] J.-P. Conze and E. Lesigne, "Théorèmes ergodiques pour des mesures diagonales", *Bull. Soc. Math. France* **112**:2 (1984), 143–175. MR 86i:28019 Zbl 0595.28018

[Conze and Lesigne 1988a] J.-P. Conze and E. Lesigne, "Sur un théorème ergodique pour des mesures diagonales", *C. R. Acad. Sci. Paris Sér. I Math.* **306**:12 (1988), 491–493. MR 89e:22012 Zbl 0641.28010

[Conze and Lesigne 1988b] J.-P. Conze and E. Lesigne, "Sur un théorème ergodique pour des mesures diagonales", pp. 1–31 in *Probabilités*, Publ. Inst. Rech. Math. Rennes **1987**, Univ. Rennes I, Rennes, 1988. MR 90a:28021 Zbl 0654.28012

[Duvenhage 2009] R. Duvenhage, "Bergelson's theorem for weakly mixing C^*-dynamical systems", *Studia Math.* **192**:3 (2009), 235–257. MR 2010d:46093 Zbl 1181.46052

[Fidaleo 2007] F. Fidaleo, "On the entangled ergodic theorem", *Infin. Dimens. Anal. Quantum Probab. Relat. Top.* **10**:1 (2007), 67–77. MR 2008d:37007 Zbl 1112.37004

[Fidaleo 2009] F. Fidaleo, "An ergodic theorem for quantum diagonal measures", *Infin. Dimens. Anal. Quantum Probab. Relat. Top.* **12**:2 (2009), 307–320. MR 2010h:46108 Zbl 05592933

[Frantzikinakis and Kra 2005] N. Frantzikinakis and B. Kra, "Convergence of multiple ergodic averages for some commuting transformations", *Ergodic Theory Dynam. Systems* **25**:3 (2005), 799–809. MR 2007b:37009 Zbl 1140.37300

[Furstenberg 1977] H. Furstenberg, "Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions", *J. Analyse Math.* **31** (1977), 204–256. MR 58 #16583 Zbl 0347.28016

[Furstenberg and Katznelson 1978] H. Furstenberg and Y. Katznelson, "An ergodic Szemerédi theorem for commuting transformations", *J. Analyse Math.* **34** (1978), 275–291. MR 82c:28032 Zbl 0426.28014

[Furstenberg et al. 1982] H. Furstenberg, Y. Katznelson, and D. Ornstein, "The ergodic theoretical proof of Szemerédi's theorem", *Bull. Amer. Math. Soc.* (*N.S.*) **7**:3 (1982), 527–552. MR 84b:28016

[Glasner 2003] E. Glasner, *Ergodic theory via joinings*, Mathematical Surveys and Monographs **101**, American Mathematical Society, Providence, RI, 2003. MR 2004c:37011 Zbl 1038.37002

[Gowers 2001] W. T. Gowers, "A new proof of Szemerédi's theorem", *Geom. Funct. Anal.* **11**:3 (2001), 465–588. MR 2002k:11014 Zbl 1028.11005

[Gowers 2006] W. T. Gowers, "Quasirandomness, counting and regularity for 3-uniform hypergraphs", *Combin. Probab. Comput.* **15**:1-2 (2006), 143–184. MR 2008b:05175 Zbl 1082.05081

[Host 2009] B. Host, "Ergodic seminorms for commuting transformations and applications", *Studia Math.* **195**:1 (2009), 31–49. MR 2010h:37003 Zbl 05615215

[Host and Kra 2001] B. Host and B. Kra, "Convergence of Conze–Lesigne averages", *Ergodic Theory Dynam. Systems* **21**:2 (2001), 493–509. MR 2002d:28007 Zbl 0995.37003

[Host and Kra 2005] B. Host and B. Kra, "Nonconventional ergodic averages and nilmanifolds", *Ann. of Math.* (2) **161**:1 (2005), 397–488. MR 2007b:37004 Zbl 1077.37002

[Jones 1983] V. F. R. Jones, "Index for subfactors", *Invent. Math.* **72**:1 (1983), 1–25. MR 84d:46097 Zbl 0508.46040

[Jones and Sunder 1997] V. Jones and V. S. Sunder, *Introduction to subfactors*, London Mathematical Society Lecture Note Series **234**, Cambridge University Press, 1997. MR 98h:46067 Zbl 0903.46062

[Kadison and Ringrose 1997] R. V. Kadison and J. R. Ringrose, *Fundamentals of the theory of operator algebras, I: Elementary theory*, Graduate Studies in Mathematics **15**, American Mathematical Society, Providence, RI, 1997. MR 98f:46001a Zbl 0888.46039

[Kerr and Li 2007] D. Kerr and H. Li, "Independence in topological and C^*-dynamics", *Math. Ann.* **338**:4 (2007), 869–926. MR 2009a:46126 Zbl 1131.46046

[Kra 2006] B. Kra, "From combinatorics to ergodic theory and back again", pp. 57–76 in *International Congress of Mathematicians*, vol. 3, edited by M. Sanz-Solé et al., Eur. Math. Soc., Zürich, 2006. MR 2007m:37014 Zbl 1118.37010

[Krengel 1985] U. Krengel, *Ergodic theorems*, de Gruyter Studies in Mathematics **6**, de Gruyter, Berlin, 1985. MR 87i:28001 Zbl 0575.28009

[Nagle et al. 2006] B. Nagle, V. Rödl, and M. Schacht, "The counting lemma for regular k-uniform hypergraphs", *Random Structures Algorithms* **28**:2 (2006), 113–179. MR 2007d:05084 Zbl 1093.05045

[Niculescu et al. 2003] C. P. Niculescu, A. Ströh, and L. Zsidó, "Noncommutative extensions of classical and multiple recurrence theorems", *J. Operator Theory* **50**:1 (2003), 3–52. MR 2004k:46123 Zbl 1036.46053

[Popa 2007] S. Popa, "Cocycle and orbit equivalence superrigidity for malleable actions of w-rigid groups", *Invent. Math.* **170**:2 (2007), 243–295. MR 2008f:37010 Zbl 1131.46040

[Salem and Spencer 1942] R. Salem and D. C. Spencer, "On sets of integers which contain no three terms in arithmetical progression", *Proc. Nat. Acad. Sci. U. S. A.* **28** (1942), 561–563. MR 4,131e Zbl 0060.10301

[Størmer 1974] E. Størmer, "Spectra of ergodic transformations", *J. Functional Analysis* **15** (1974), 202–215. MR 51 #13715

[Szemerédi 1975] E. Szemerédi, "On sets of integers containing no k elements in arithmetic progression", *Acta Arith.* **27** (1975), 199–245. MR 51 #5547 Zbl 0303.10056

[Tao 2008] T. Tao, "Norm convergence of multiple ergodic averages for commuting transformations", *Ergodic Theory Dynam. Systems* **28**:2 (2008), 657–688. MR 2009k:37012 Zbl 1181.37004

[Towsner 2007] H. P. Towsner, "Convergence of Diagonal Ergodic Averages", preprint, 2007. arXiv 0711.1180

[Vaughan 1997] R. C. Vaughan, *The Hardy–Littlewood method*, 2nd ed., Cambridge Tracts in Mathematics **125**, Cambridge University Press, 1997. MR 98a:11133 Zbl 0868.11046

[Zhang 1996] Q. Zhang, "On convergence of the averages $(1/N)\sum_{n=1}^{N} f_1(R^n x) f_2(S^n x) f_3(T^n x)$", *Monatsh. Math.* **122**:3 (1996), 275–300. MR 97k:28041 Zbl 0911.28012

[Ziegler 2005] T. Ziegler, "A non-conventional ergodic theorem for a nilsystem", *Ergodic Theory Dynam. Systems* **25**:4 (2005), 1357–1370. MR 2006d:37009 Zbl 1099.37002

[Ziegler 2007] T. Ziegler, "Universal characteristic factors and Furstenberg averages", *J. Amer. Math. Soc.* **20**:1 (2007), 53–97. MR 2007j:37004 Zbl 1198.37014

[Zimmer 1976a] R. J. Zimmer, "Ergodic actions with generalized discrete spectrum", *Illinois J. Math.* **20**:4 (1976), 555–588. MR 54 #2924 Zbl 0349.28011

[Zimmer 1976b] R. J. Zimmer, "Extensions of ergodic group actions", *Illinois J. Math.* **20**:3 (1976), 373–409. MR 53 #13522 Zbl 0334.28015

Received December 30, 2009. Revised July 20, 2010.

TIM AUSTIN
DEPARTMENT OF MATHEMATICS
BROWN UNIVERSITY
151 THAYER ST, BOX 1917
PROVIDENCE, RI 02912
UNITED STATES
timaustin@math.brown.edu
http://www.math.brown.edu/~timaustin

TANJA EISNER
KORTEWEG-DE VRIES INSTITUTE FOR MATHEMATICS
UNIVERSITY OF AMSTERDAM
P.O. BOX 94248
1090 GE AMSTERDAM
THE NETHERLANDS
t.eisner@uva.nl
http://staff.science.uva.nl/~eisner

TERENCE TAO
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA AT LOS ANGELES
405 HILGARD AVENUE
LOS ANGELES, CA 90095-1555
UNITED STATES
tao@math.ucla.edu
http://www.math.ucla.edu/~tao/

PACIFIC JOURNAL OF MATHEMATICS

http://www.pjmath.org

Founded in 1951 by
E. F. Beckenbach (1906–1982) and F. Wolf (1904–1989)

See inside back cover or www.pjmath.org for submission instructions.

The subscription price for 2011 is US $420/year for the electronic version, and $485/year for print and electronic.
Subscriptions, requests for back issues from the last three years and changes of subscribers address should be sent to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163, U.S.A. Prior back issues are obtainable from Periodicals Service Company, 11 Main Street, Germantown, NY 12526-5635. The Pacific Journal of Mathematics is indexed by Mathematical Reviews, Zentralblatt MATH, PASCAL CNRS Index, Referativnyi Zhurnal, Current Mathematical Publications and the Science Citation Index.

The Pacific Journal of Mathematics (ISSN 0030-8730) at the University of California, c/o Department of Mathematics, 969 Evans Hall, Berkeley, CA 94720-3840, is published monthly except July and August. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices. POSTMASTER: send address changes to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163.

PJM peer review and production are managed by EditFlow™ from Mathematical Sciences Publishers.

PUBLISHED BY PACIFIC JOURNAL OF MATHEMATICS
at the University of California, Berkeley 94720-3840
A NON-PROFIT CORPORATION
Typeset in LaTeX

PACIFIC JOURNAL OF MATHEMATICS

Volume 250 No. 1 March 2011

0030-8730(201103)250:1;1-H