The Multifarious Poincaré Recurrence Theorem

Vitaly Bergelson
Department of Mathematics
The Ohio State University
Columbus, OH 43210

Acknowledgement. I wish to express my sincere appreciation to Steve Cook and Sasha Leibman for their help in preparing these notes. I also would like to thank A. Dynin for useful discussions on the philosophical aspects of Poincaré recurrence theorem. This work was supported by NSF grant DMS-9706057.

1 Some history and background

The Poincaré recurrence theorem (PRT), which one can find in virtually any book on ergodic theory, is usually stated as follows:

PRT *If T is a measure-preserving transformation of a probability space $(\Omega, \mathcal{B}, \mu)$ and $A \in \mathcal{B}$ with $\mu(A) > 0$, then there exists a measurable subset $A_0 \subseteq A$ with $\mu(A_0) = \mu(A)$ such that for any $x \in A_0$ there exists an infinite sequence $(n_i)_{i=1}^{\infty}$ such that $T^{n_i}x \in A_0$ for all i.*

This theorem is considered so basic that some books do not give a reference for it, and those which do either quote [P3] (cf. [Ha] or [Ho]), or refer to the three-volume, 1000-plus-page *Les méthodes nouvelles de la mécanique céleste* [P2], usually without giving the reader any more specific directions. (For the reader's information: the version of PRT appearing in [P2] is to be found in Ch. 26, Vol. 3). Yet prior to 1899, the year of publication of the third volume of *Méthodes nouvelles*, this theorem was at the center of quite stormy discussions related to Zermelo's *Wiederkehreinwand* [1]. (See [Z1], [Z2], [Bo1], [Bo2], [Bo3].) We shall return to Zermelo's argument involving PRT later, but first we want to formulate PRT as it appeared in Poincaré's King Oscar Prize-winning memoir [P1]. (This memoir was itself a source of some controversy. See, for example, [B] or [Gor], Section 1.3.) While not resolving the (still open) problem of the stability of the solar system, this work of Poincaré was, according to Weierstrass, "of such importance that its publication will open a new era in the history of celestial mechanics." The object of our interest in this essay, PRT, is referred to by Poincaré in the introduction to [P1] in the following way:

> J'ai étudié plus spécialement un cas particulier du problème des trois corps, celui où l'une des masses est nulle et où le mouvement des deux autres est circulaire; j'ai reconnu que dans ce cas les trois corps repasseront une infinité de fois aussi près que l'on veut de leur position initiale, à moins que les conditions initiales du mouvement ne soient exceptionnelles.

[1] According to M. Moravcsik, translator of [EE], this cumbersome word means something like "objection or counter-argument based on reasoning involving a return to the same state."

> Comme on le voit, ces résultats ne nous apprennent que peu de chose sur le cas général du problème; mais ce qui peut leur donner quelque prix, c'est qu'ils sont établis avec rigueur, tandis que le problème des trois corps ne paraissait jusqu'ici abordable que par des méthodes d'approximation successive où l'on faisait bon marché de cette rigueur absolue qui est exigée dans les autres parties des mathématiques.

Here then is Poincaré's original formulation of PRT, *Théorème* I from [P1]. There are only three statements in this 270-page memoir, all of them in Section 8, which Poincaré calls *Théorème.*

> **Théorème I** *Supposons que le point P reste à distance finie, et que le volume $\int dx_1\,dx_2\,dx_3$ soit un invariant intégral; si l'on considère une région r_0 quelconque, quelque petite que soit cette région, il y aura des trajectoires qui la traverseront une infinité de fois.*

After formulating the recurrence theorem, Poincaré first establishes a combinatorial principle (See Principle P below), on the basis of which he proceeds to discuss two different approaches to the question of recurrence or, as he calls it at the beginning of Section 8 in [P1] and in the introduction to Ch. 26 in [P2], *stability in the sense of Poisson.* These two approaches are, essentially, a topological one and a probabilistic, or rather, measure preserving one; and while the modern reader may find it hard to agree with Poincaré's claim of *"rigueur absolue"*, he will undoubtedly recognize in the discussions of Section 8 of [P1] and Ch. 26 of [P2], the familiar elements of modern versions of recurrence theorems. The reader is referred to [C] for the first modern rendition of PRT. See also [Hi2] and [O] for a discussion of a category statement which, according to J. Oxtoby, "has to be read between the lines of Poincaré's discussion."

The combinatorial principle mentioned above is nothing but a "crossbreeding" between the pigeon-hole principle and the *stationarity* assumption. Enhanced further by the possibility of repeated *iterations*, this principle not only leads to PRT and some of its numerous refinements, but, as we shall try to demonstrate, provides a simplified and unified approach to many number-theoretical and combinatorial results. Here is the relevant passage from [P1]:

> En effet le point P restant à distance finie, ne sortira jamais d'une région limitée R. J'appelle V le volume de cette région R.
>
> Imaginons maintenant une région très petite r_0, j'appelle v le volume de cette région. Par chacun des points de r_0 passe une trajectoire que l'on peut regarder comme parcourue par un point mobile suivant la loi définie par nos équations différentielles. Considérons donc une infinité de points mobiles remplissant au temps 0 la région r_0 et se

mouvant ensuite conformément à cette loi. Au temps τ ils rempliront une certaine région r_1, au temps 2τ une région r_2, etc. au temps $n\tau$ une région r_n. Je puis supposer que τ est assez grand et r_0 assez petit pour que r_0 et r_1 n'aient aucun point commun.

Le volume étant un invariant intégral, ces diverse régions $r_0, r_1, \cdots, r_n$ auront même volume v. Si ces régions n'avaient aucun point commun, le volume total serait plus grand que nv; mais d'autre part toute ces régions sont intérieures à R, le volume total est donc plus petit que V. Si donc on a:

$$n > \frac{V}{v},$$

il faut que deux au moins de nos régions aient une partie commune. Soient r_p et r_q ces deux régions $(q > p)$. Si r_p et r_q ont une partie commune, il est clair que r_0 et r_{q-p} devront avoir une partie commune.

Here now is a formulation in modern terms:

Principle P *Let μ be a finitely additive probability measure defined on an algebra $\mathcal{B}$ of subsets of a set X. Assume further that the sets $A_n \in \mathcal{B}$, $n = 0, 1, 2, \cdots$ satisfy, for any $n \geq m \geq 0$, the* stationarity *condition*

$$\mu(A_n \cap A_m) = \mu(A_0 \cap A_{n-m})$$

and that $\mu(A_0) = a > 0$. Then there exists a positive integer $k \leq \left[\frac{1}{a}\right] + 1$ such that $\mu(A_0 \cap A_k) > 0$.

A natural question is: what is the best $\delta = \delta(a)$ such that for some $k > 0$ one has $\mu(A_0 \cap A_k) > \delta$? Taking any sequence of pairwise independent sets (say, $A_n = \bigcup_{i=1}^{2^{n-1}} \left[\frac{2i-2}{2^n}, \frac{2i-1}{2^n}\right] \subseteq [0,1]$) shows that $\delta(a) \leq a^2$. The following useful statement, which we shall call EPP (Enhanced Principle P), supplies quite a satisfactory answer to the question above.

EPP *Under the assumptions of Principle P, for any $0 < \lambda < 1$ there exists $c = c(a, \lambda)$ such that for some $0 < k < c$ one has $\mu(A_0 \cap A_k) > \lambda a^2$.*

Proof. We are going to utilize an idea due to Gillis ([G]). (He worked with σ-additive measures, but, as we shall see, it does not really matter.) Given a simple function $f = \sum_{i=1}^{n} \alpha_i 1_{A_i}$, write $\int f \, d\mu = \sum_{i=1}^{n} \alpha_i \mu(A_i)$. It is trivial to see that $\int f \, d\mu$ does not depend on the representation of f. What is important to us is that this limited notion of integral obeys the Cauchy-Schwartz inequality:

$$\left(\int f g \, d\mu\right)^2 \leq \int f^2 \, d\mu \int g^2 \, d\mu.$$

Remembering that $\mu(X) = 1$, we get for $g \equiv 1$

$$\Big(\int f\, d\mu\Big)^2 \le \int f^2\, d\mu.$$

That is all one needs to conclude the proof of EPP, since if no $c = c(a, \lambda)$ with the desired property exists, the following inequality would be contradictory for large enough n:

$$\begin{aligned} n^2a^2 = \Big(\int \sum_{i=1}^{n} 1_{A_i}\, d\mu\Big)^2 &\le \int \Big(\sum_{i=1}^{n} 1_{A_i}\Big)^2\, d\mu \\ &= \sum_{i=1}^{n} \mu(A_i) + 2 \sum_{1\le i<j\le n} \mu(A_i \cap A_j). \end{aligned}$$

■

Remark 1.1 The reader may wonder why we bothered to formulate Principle P and its enhanced version in terms of finitely additive rather than countably additive measures. The answer is that in many situations outside the realm of ergodic theory and dynamical systems one often does not have the luxury of countable additivity. We shall see, however, that finite additivity is quite sufficient for many applications of Principle P.

To see that Principle P is all that one needs to prove PRT as stated at the beginning of this section, let us note first that when applied to the sequence $A_n = T^{-n}A$, $n = 0, 1, 2, \cdots$, where $\mu(A) > 0$, Principle P implies the existence of $k \in \mathbb{N}$ such that $\mu(A \cap T^{-k}A) > 0$. For $n \in \mathbb{N}$, let $B_n \subseteq A$ be the set of those $x \in A$ which do not return to A under $S = T^n$. Formally, $B_n = A \cap (\bigcap_{i=1}^{\infty} S^{-i}(A^c))$. (In particular, B_n is measurable.) We claim that $\mu(B_n) = 0$. Indeed, if $\mu(B_n) > 0$ then for some $k \in \mathbb{N}$, $\mu(B_n \cap S^{-k}B_n) > 0$. But then for any $x \in B_n \cap S^{-k}B_n$ one has $S^k x = T^{kn}x \in B_n \subseteq A$ which contradicts the definition of B_n. It follows that for any $n \in \mathbb{N}$ the measurable subset $C_n \subseteq A$ defined by $C_n = \{x \in A : \exists m > n : T^m x \in A\}$ satisfies $\mu(C_n) = \mu(A)$. We are done since all the points of the set $A_0 = \bigcap_{n=1}^{\infty} C_n$ return to A infinitely many times and $\mu(A_0) = \mu(\bigcap_{n=1}^{\infty} C_n) = \mu(A)$.

It is hard not to agree with Marc Kac who, after remarking that "there are many proofs of this theorem [i.e. PRT] all of which are almost trivial," added a footnote:

> We have here another example of an important and even profound fact whose purely mathematical content is very much on the surface. ([Kac], p. 63)

The examples which we bring in the next section will provide further support to the validity of Kac's remark.

We want to conclude this introductory section with an excerpt from [Z1] in the translation of S.G. Brush ([Br], pp. 208-209). For Boltzmann's response and for the ensuing discussion, the reader is referred to [Br], and for a neat analysis of the *Wiederkehreinwand*, to [EE].

> In the second chapter of Poincaré's prize essay on the three-body problem, there is proved a theorem from which it follows that the usual description of the thermal motion of molecules, on which is based for example the kinetic theory of gases, requires an important modification in order that it be consistent with the thermodynamic law of increase of entropy. Poincaré's theorem says that *in a system of mass-points under the influence of forces that depend only on position in space, in general any state of motion (characterized by configurations and velocities) must recur arbitrarily often, at least to any arbitrary degree of approximation even if not exactly, provided that the coordinates and velocities cannot increase to infinity.* Hence, in such a system *irreversible processes are impossible* since (aside from singular initial states) no single-valued continuous function of the state variables, such as entropy, can continually increase; if there is a finite increase, then there must be a corresponding decrease when the initial state recurs. Poincaré, in the essay cited, used his theorem for astronomical discussions on the stability of sun systems; he does not seem to have noticed its applicability to systems of molecules or atoms and thus to the mechanical theory of heat...

2 Combinatorial richness of large sets in countable amenable groups

Recall that a discrete group G is called *amenable* if there exists a finitely additive measure μ on $\mathcal{P}(G)$ such that $\mu(G) = 1$ and for any $g \in G$ and $A \subseteq G$, $\mu(gA) = \mu(A)$ (i.e., μ is left-invariant). The notion of amenability may be defined in many equivalent ways. (One of them, via the Følner condition, is especially useful and will be given below.) The class of amenable groups includes solvable (in particular, abelian) groups as well as the profinite groups, such as, say, the group S_∞ of finite permutations of $\mathbb{N}$. On the other hand, the non-abelian free group $F_2 = \langle a, b\rangle$ is a classical example of a non-amenable group. We cannot resist the temptation to show that all one needs for the proof of this fact is to apply Principle P. The proof is by contradiction. Assume that μ is a left-invariant, finitely-additive probability measure on $\mathcal{P}(F_2)$. Consider the partition $F_2 = A^+ \cup A^- \cup B^+ \cup B^- \cup \{e\}$, where e is the unit of F_2 ($\equiv$ the "empty word") and the sets A^+,A^-,B^+, and

B^- consist of the (reduced) words starting, respectively, with a,a^{-1},b, and b^{-1}. Since μ is finitely additive, one of these five sets has to have positive measure. Clearly, $\mu(\{e\}) = 0$. (If $\mu(\{e\}) = c > 0$ then, by shift-invariance, $\mu(\{a^n\}) = c$ for all $n \in \mathbb{Z}$ and one gets a contradiction by taking $N \geq \frac{1}{c}$ and considering the set $B = \{a^i,\ i = 0, 1, 2, \cdots, N\}$ which has to satisfy $\mu(B) = \sum_{i=0}^{N} \mu(\{a^i\}) = (N+1)c > 1$.) So assume $\mu(A^+) = c > 0$. (The same proof will work for any set of the partition which happens to have positive measure.) Let $A_n = b^n A^+$, $n = 0, 1, 2, \cdots$. By shift-invariance we have, for any $n \geq m \geq 0$:

$$\mu(A_n \cap A_m) = \mu(b^m A^+ \cap b^n A^+) = \mu(A^+ \cap b^{n-m} A^+) = \mu(A_0 \cap A_{n-m}).$$

By Principle P there exists $k \in \mathbb{N}$ such that $\mu(A_0 \cap b^k A_0) = \mu(A^+ \cap b^k A^+) > 0$. But this is impossible since, obviously, $A^+ \cap b^k A^+ = \emptyset$. (No reduced word in F_2 can start with both a and b!)

From now on we shall, for the sake of convenience, deal exclusively with countable groups. It should be remarked, though, that many of the examples and results which we bring in this paper extend to general discrete and topological amenable semigroups.

A sequence of finite sets $\{F_n\}_{n=1}^{\infty}$ is called a *left Følner sequence*, if for any $g \in G$ one has

$$\lim_{n\to\infty} \frac{|gF_n \triangle F_n|}{|F_n|} = 0.$$

For example, in $\mathbb{Z}$, any sequence of intervals $[a_n, b_n]$ with $|b_n - a_n| \to \infty$ is a Følner sequence. Here is one more useful example. Let G be the direct sum $\mathbb{Z}_p^{\infty}$ of countably many copies of $\mathbb{Z}_p$. (It is convenient to envision $\mathbb{Z}_p^{\infty}$ as the set of infinite sequences $(a_1, a_2, \cdots)$ where $a_i \in \mathbb{Z}$, and all but finitely many $a_i = 0$, and addition is defined component-wise modulo p.) Let $F_n = \left\{(a_1, a_2, \cdots) \in \mathbb{Z}_p^{\infty} : a_i = 0\ \forall i \geq n+1\right\}$. One can easily check that $\{F_n\}$ is a Følner sequence. We shall return to this example later.

The existence of a Følner sequence in a group is tightly related to amenability: a (countable) group is amenable if and only if it has a left Følner sequence. Følner sequences are also helpful in all kinds of Ramsey-theoretical questions since they allow one to define a notion of largeness in a natural way.

Definition 2.1 Given a left Følner sequence $\{F_n\}$ and a set $E \subseteq G$, the *upper density* of E with respect to $\{F_n\}$ is defined by

$$\bar{d}_{\{F_n\}}(E) = \limsup_{n\to\infty} \frac{|E \cap F_n|}{|F_n|}$$

We shall say that a set $E \subseteq G$ is *left-large* if for some left Følner sequence $\bar{d}_{\{F_n\}}(E) > 0$, and *left-conull* if $\bar{d}_{\{F_n\}}(E) = 1$.

According to the principles of Ramsey theory (see [GRS] and [Be3]), large sets, and especially conull sets, ought to be combinatorially rich. We are going to present a few results which substantiate this claim. Before doing so, we collect some useful facts about sets of positive upper density in the following Proposition.

Proposition 2.2 *Let $\{F_n\}$ be a left Følner sequence in G. For $E \subseteq G$ one has:*

(i) $\forall g \in G$, $\bar{d}_{\{F_n\}}(gE) = \bar{d}_{\{F_n\}}(E)$

(ii) *If $\bar{d}_{\{F_n\}}(E) = c > 0$, then for any $0 < \lambda < 1$ there exists g such that $\bar{d}_{\{F_n\}}(E \cap gE) > \lambda c^2$. The element g can be chosen to lie outside of any prescribed finite set $F \subseteq G$.*

Proof. (i) trivially follows from the definition of a Følner sequence. (ii) is just an application of the Enhanced Principle P (See Remark 1.1). ∎

Definition 2.3 Given a subset $\Gamma = \{g_i\}_{i \in I} \subseteq G$ (where I is a finite or countable subset of $\mathbb{N}$), an *FP-set*, generated by Γ, is the set of all finite products of distinct elements of Γ with ascending indices. More formally, writing $\mathcal{F}(I)$ for the set of finite, non-empty subsets of I, we have

$$\mathrm{FP}(\Gamma) = \Big\{ g_{i_1} g_{i_2} \cdots g_{i_k},\ i_1 < i_2 < \cdots < i_k,\ \{i_1, \cdots, i_k\} \in \mathcal{F}(\Gamma) \Big\}.$$

Remarks 2.4 1. There exists, of course, a dual notion of FP-sets which corresponds to taking products with descending indices. Our choice was dictated by the fact that we are working with *left* Følner sequences.
2. When G is an additive abelian group, one replaces products with sums and speaks of FS-sets. FS-sets which are formed with a countable set of indices I are called, in ergodic theory and topological dynamics, *IP-sets* (a term coined by Furstenberg and Weiss in [FW]). It turns out that many familiar ergodic and dynamical results involving group actions can be extended and refined to actions of IP-sets, which brings, in particular, some strong applications to combinatorics and number theory. See, for example, [FW], [F2], [FK], and [BeM2].

As the following Proposition shows, every large set contains translates of sets of the form $\mathrm{FP}(g_i)_{i=1}^n$ with arbitrarily large n (and pairwise distinct g_i's).

Proposition 2.5 *Let $E \subseteq G$ be left-large ($\bar{d}_{\{F_n\}}(E) > 0$). Then for any n there are pairwise distinct $g_0 = e, g_1, \cdots, g_n$ such that $\bar{d}_{\{F_n\}}\Big(\bigcap_{g \in \mathrm{FP}(g_i)_{i=0}^n} g^{-1}E \Big) > 0$.*

Remark 2.6 Clearly, any $x \in \bigcap_{g \in \mathrm{FP}(g_i)_{i=0}^n} g^{-1}E$ satisfies $x\mathrm{FP}(g_i)_{i=0}^n \subseteq E$.

Proof. The proposition follows by iteration of property (ii) from Proposition 2.2 above. Let $g_1 \neq e$ be such that $\bar{d}_{\{F_n\}}(E \cap g_1^{-1}E) > 0$. Denoting $E_1 = E \cap g_1 E$, let $g_2 \notin \{e, g_1\}$ be such that $\bar{d}_{\{F_n\}}(E_1 \cap g_2^{-1}E_1) = \bar{d}_{\{F_n\}}(E \cap g_1^{-1} \cap g_2^{-1}E \cap g_2^{-1}g_1^{-1}E) > 0$. Continuing in this fashion one arrives at a sequence $g_0 = e, g_1, \cdots, g_n$ with the desired property. ■

The simple Proposition which we have just proved contains, as a quite special case, the following result, whose proof occupies more than two pages in [Hi1]. (Hilbert needed it to show that if the polynomial $p(x, y) \in \mathbb{Z}[x, y]$ is irreducible then, for some $c \in \mathbb{N}$, $p(x, c) \in \mathbb{Z}[x]$ is irreducible.)

Proposition 2.7 (Hilbert, [Hi1], pp. 104-107) *For any $k, r \in \mathbb{N}$, if $\mathbb{N} = \bigcup_{i=1}^{r} C_i$, then one of the C_i contains infinitely many translates of a set of the form $\mathrm{FS}(n_i)_{i=1}^{k}$ (where the n_i's are pairwise distinct).*

Proof. Fix, in $\mathbb{N}$, any sequence of intervals $I_n = [a_n, b_n]$ with $|I_n| = |b_n - a_n| \to \infty$ and observe that one of the C_i satisfies $\bar{d}_{\{I_n\}}(C_i) > 0$. Apply Proposition 2.5. ■

Hilbert's result is a forerunner of a much deeper modern theorem due to Hindman ([Hi3]), which claims that for any finite partition of $\mathbb{N}$, one of the cells of the partition contains an IP-set. We shall give a proof of Hindman's theorem below, but first we prove a related result about conull sets.

Proposition 2.8 *Any left-conull set $E \subseteq G$ contains an IP-set.*

Proof. The proof uses the same iterational idea as Proposition 2.5 above, but since this time we are dealing with a conull set, the iterations can be arranged in a more controlled fashion.

Let us fix a left Følner sequence with respect to which E is left-conull, and let us denote the corresponding upper density by $\bar{d}$. Choose $g_1 \in E$ arbitrarily. Clearly, $\bar{d}(E \cap g_1^{-1}E) = 1$. Pick $g_2 \in E_1 = E \cap g_1^{-1}E$ so that $g_2 \neq g_1$. Observe that $1 = \bar{d}(E_1 \cap g_2^{-1}E_1) = \bar{d}(E \cap g_1^{-1}E \cap g_2^{-1}E \cap g_2^{-1}g_1^{-1}E)$. Notice that any $g_3 \in E_2 = E_1 \cap g_2^{-1}E$ has the property that $\mathrm{FP}(g_i)_{i=1}^{3} \subseteq E$. Continuing in this fashion (and taking care to choose each successive g_k so that $g_k \notin \{g_1, g_2, \cdots, g_{k-1}\}$), we shall arrive at an infinite sequence $(g_i)_{i=1}^{\infty}$ such that for any $n \in \mathbb{N}$, $\mathrm{FP}(g_i)_{i=1}^{n} \subseteq E$. We are done. ■

Extending the notion of an FP-set, let us, given a set $\Gamma = \{g_i\}_{i \in I} \subseteq G$ and $l \in \mathbb{N}$, define an $\mathrm{FP}^{(l)}$-*set* generated by Γ as

$$\mathrm{FP}^{(l)}(\Gamma) = \left\{ g_{i_1}^{c_1} g_{i_2}^{c_2} \cdots g_{i_k}^{c_k} : i_1 < \cdots < i_k, \{i_1, \cdots, i_k\} \subseteq \mathcal{F}(I), 1 \leq c_i \leq l \right\}.$$

In other words, in $\mathrm{FP}^{(l)}$-sets bounded repetitions of generators are allowed. Of course, $\mathrm{FP}^{(1)}$-sets are just the FP-sets defined above. Note that if G is an

abelian group with uniformly bounded torsion, then for large enough l, any $\mathrm{FP}^{(l)}$-set in G is a subgroup. As before, when the notation is additive, we shall talk about $\mathrm{FS}^{(l)}$-sets, and (when the set of indices I is infinite) $\mathrm{IP}^{(l)}$-sets.

While not every conull set in, say, $\mathbb{Z}$ or $\mathbb{Z}_p^\infty$ contains an $\mathrm{FS}^{(l)}$-set for $l > 1$, the following Proposition gives a convenient criterion for a conull set to contain $\mathrm{IP}^{(l)}$-sets for any l. We remark that (ii) below was proved by M. Karpovsky and V. Milman in [KaM] (by a different method).

Proposition 2.9 (i) *If the intervals $I_n = [a_n, b_n] \subseteq \mathbb{Z}$, with $b_n - a_n \to \infty$, satisfy $[a_n, b_n] \subseteq [a_{n+1}, b_{n+1}]$ for all $n \in \mathbb{N}$, then any $E \subseteq \mathbb{Z}$ such that $\bar{d}_{\{I_n\}}(E) = 1$ contains an $\mathrm{IP}^{(l)}$-set for any $l \in \mathbb{N}$.*
(ii) *Let $G = \mathbb{Z}_p^\infty$ and $F_n = \{(a_1, a_2, \cdots) \in \mathbb{Z}_p^\infty : a_k = 0 \ \forall k > h\}$. If $\bar{d}_{\{F_n\}}(E) = 1$, then $E \cup \{0\}$ contains an infinite subgroup (which is isomorphic to $\mathbb{Z}_p^\infty$).*

Proof. We shall prove (ii) only, the proof of (i) being similar. For $k \in \mathbb{Z}\backslash\{0\}$ let

$$E/k = \{g \in \mathbb{Z}_p^\infty : kg \in E\}.$$

(This definition, of course, makes sense for any abelian group G with additive notation.) Writing $\bar{d}$ for the upper density defined by the sequence of subgroups $\{F_n\}$, observe that if $\bar{d}(E) = 1$, then for any $1 \leq k \leq p-1$, $\bar{d}(E/k) = 1$. This observation will allow us to prove the desired fact by a simple iterative process. We start with the set $E_1 = \bigcap_{k=1}^{p-1} E/k$. Note that, $\bar{d}(E_1) = 1$. If $x_1 \in E_1, x_1 \neq 0$, then

$$S(x_1) = \{ix_1, i = 0, 1, \cdots, p-1\} \subseteq E \cup \{0\},$$

and clearly, $S(x_1)$ is isomorphic to $\mathbb{Z}_p$. Now let $E_2 = \bigcap_{i_2=1}^{p-1} \bigcap_{i_1=0}^{p-1} (E - i_1 x_1)/i_2$. Then $\bar{d}(E_2) = 1$ and any $x_2 \in E_2$ such that $x_2 \notin S(x_1)$ has the property that

$$S(x_1, x_2) = \left\{i_1 x_1 + i_2 x_2 : i_1, i_2 \in \{0, 1, \cdots, p-1\}\right\} \subseteq E \cup \{0\}$$

and, in addition, $S(x_1, x_2)$ is isomorphic to $\mathbb{Z}_p \oplus \mathbb{Z}_p$. At the next step one considers the set $E_3 = \bigcap_{i_3}^{p-1} \bigcap_{i_1, i_2 = 0}^{p-1} (E - i_1 x_1 - i_2 x_2)/i_3$, picks $x_3 \in E_3$ subject to the condition $x_3 \notin S(x_1, x_2)$, and so on. It is clear that, continuing in this fashion, one arrives at an infinite sequence $(x_i)_{i=1}^\infty$ such that for any $n \geq 1$,

$$x_{n+1} \notin S(x_1, x_2, \cdots, x_n) = \big\{i_1 x_1 + i_2 x_2 + \cdots + i_n x_n : \\ i_1, \cdots, i_n \in \{0, 1, \cdots, p-1\}\big\} \subseteq E \cup \{0\}.$$

Since, for every n, $S(x_1, \cdots, x_n)$ is isomorphic to the direct sum of n copies of $\mathbb{Z}_p$, we are done. ∎

3 Principle P and ultrafilters

First, we are going to introduce an important family of finitely additive probability measures, the so-called *ultrafilters*. As we shall see, Hindman's theorem, alluded to in Section 2, follows in a natural way by repeated application of Principle P to a sequence of sets which are large with respect to a conveniently chosen ultrafilter.

We give only the minimal amount of background information on ultrafilters. The interested reader will find the missing details and much more discussion in [CN], [HiS], and [Be3].

Recall that a *filter* p on $\mathbb{N}$ is a set of subsets of $\mathbb{N}$ satisfying the following conditions:

(i) $\emptyset \notin p$.

(ii) $A \in p$ and $A \subseteq B$ imply $B \in p$.

(iii) $A \in p$ and $B \in p$ imply $A \cap B \in p$.

Now, an ultrafilter is a filter which, additionally, has the property

(iv) If $r \in \mathbb{N}$ and $\mathbb{N} = A_1 \cup A_2 \cup \cdots \cup A_r$, then $A_i \in p$ for some i, $1 \leq i \leq r$.

A rather dull class of examples of ultrafilters is provided by the so-called *principal* ones, which are defined, for any $n \in \mathbb{N}$, by $p_n = \{A \subseteq \mathbb{N} : n \in A\}$. To construct less trivial examples, one has to resort to Zorn's lemma. (One can show that this is unavoidable: see, for example, [CN], pp. 161-162.)

Assume that a family $\mathcal{C}$ of subsets of $\mathbb{N}$ satisfies conditions (i), (ii), and (iii). We claim there is an ultrafilter p such that for any $C \in \mathcal{C}$, one has $C \in p$. To see this, let

$$\tilde{\mathcal{C}} = \big\{\mathcal{D} \subseteq \mathcal{P}(\mathbb{N}) : \mathcal{D} \text{ satisfies (i), (ii) and (iii), and } \mathcal{C} \subseteq \mathcal{D}\big\}.$$

Since $\mathcal{C} \in \tilde{\mathcal{C}}$, $\tilde{\mathcal{C}} \neq \emptyset$. Also, the union of any chain in $\tilde{\mathcal{C}}$ is a member of $\tilde{\mathcal{C}}$. By Zorn's lemma, there is a maximal member p of $\tilde{\mathcal{C}}$ which, being maximal, has to satisfy (iv).

Here is a useful example. Let $I_n = [a_n, b_n]$ be a sequence of intervals in $\mathbb{N}$ satisfying $b_n - a_n \to \infty$, and let $\mathcal{C} = \{A \subseteq \mathbb{N} : \bar{d}_{\{I_n\}}(A) = 1\}$. $\mathcal{C}$ satisfies conditions (i), (ii), and (iii), and hence there exists an ultrafilter p such that $\mathcal{C} \subseteq p$. Note that if $B \in p$, then $\bar{d}_{\{I_n\}}(B) > 0$. (Otherwise, $B^c \in \mathcal{C}$.)

The set of ultrafilters on $\mathbb{N}$ is naturally identified with $\beta\mathbb{N}$, the Stone-Čech compactification of $\mathbb{N}$. The sets $\bar{A} = \{p \in \beta\mathbb{N}, A \in p\}$, where $A \subseteq \mathbb{N}$, form a basis for the open sets in $\beta\mathbb{N}$ (and a basis for the closed sets). With this topology $\beta\mathbb{N}$ is a compact Hausdorff space which is, in some respects, rather an odd object. In particular, it has a dense countable subset (namely, the set of principal ultrafilters), but has the cardinality of $\mathcal{P}(\mathcal{P}(\mathbb{N}))$ (and hence is non-metrizable).

Each ultrafilter $p \in \beta\mathbb{N}$ can be naturally identified with a finitely additive, zero-one valued probability measure μ_p on the power set $\mathcal{P}(\mathbb{N})$. Indeed, let $\mu_p(A) = 1$ if and only if $A \in p$. From now on we are going to view ultrafilters as measures, but we will prefer to write $A \in p$ instead of $\mu_p(A) = 1$.

In addition to being a compact Hausdorff space, $\beta\mathbb{N}$ has two natural algebraic structures which are induced by $(\mathbb{N}, +)$ and $(\mathbb{N}, \cdot)$. With respect to each of these, $\beta\mathbb{N}$ is a compact left-topological semigroup. We shall concentrate on the operation which comes from $(\mathbb{N}, +)$.

Definition 3.1 Given $p, q \in \beta\mathbb{N}$, define

$$p + q = \{A \subseteq \mathbb{N} : \{n \in \mathbb{N} : (A - n) \in p\} \in q\},$$

where $A - n$ is the set of all m for which $m + n \in A$.

Remarks 3.2 1. The operation just introduced is nothing but convolution of measures! The reader should find it instructive to compare it with familiar formulas for the convolution of measures μ, ν on a locally compact group G:

$$\mu * \nu(A) = \int_G \nu(x^{-1}A)\, d\mu(x) = \int_G \mu(Ay^{-1})\, d\nu(y).$$

2. One easily checks that for principal ultrafilters, the operation $+$ corresponds to addition on $\mathbb{N}$.

It is not hard to verify that $p + q \in \beta\mathbb{N}$ and that the operation $+$ is associative. However, a word of warning is in place here: $+$ is, generally speaking, not commutative. One can actually show that the center of the semigroup $(\beta\mathbb{N}, +)$ contains only the principal ultrafilters. One can also show that for any fixed p, the function $f_p(q) = p + q$ is continuous. In other words, the operation $+$ is left-continuous.

By a theorem due to Ellis ([El]), any compact semigroup with a left-continuous operation has an idempotent. It is the idempotent ultrafilters which are the key to understanding Hindman's theorem, which we are going to prove now. Before embarking on the proof, let us look more closely at the notion of an idempotent ultrafilter. Given an ultrafilter p, call a set $C \subseteq \mathbb{N}$ *p-big* if $C \in p$. Assume now that $p \in \beta\mathbb{N}$ satisfies $p + p = p$. By the definition of $+$, this means that

$$A \in p \Longleftrightarrow A \in p + p \Longleftrightarrow \{n \in \mathbb{N} : (A - n) \in p\} \in p.$$

In other words, if p is an idempotent, then A is p-big if and only if for p-many $n \in \mathbb{N}$, the shifted set $(A - n)$ is p-big. This explains why idempotent ultrafilters are often called "almost shift invariant". Indeed, A is p-big if p-almost all shifts of A are also p-big. We are now in a position to state and prove Hindman's theorem.

Theorem 3.3 (Hindman, [Hi3]) *If, for some $r \in \mathbb{N}$, $\mathbb{N} = \bigcup_{i=1}^{r} C_i$ then one of the C_i contains an* IP*-set.*

Proof. We shall prove a stronger fact: if $p \in \beta\mathbb{N}$ is idempotent, then any $C \in p$ contains $\mathrm{FS}(n_j)_{j=1}^{\infty}$ for some increasing sequence $(n_j)_{j=1}^{\infty}$. We start by observing that if a set C is a member of the idempotent ultrafilter p, then the basic conclusion of Principle P is satisfied: for some n (and actually for p-almost all n) one has $C \cap (C-n) \in p$. Since $C \cap \{n : (C-n) \in p\} \in p$ as well, we see that one can always pick $n_1 \in C$ so that $C \cap (C-n_1) \in p$. The rest of the proof is virtually identical to that of Proposition 2.8 and follows by iteration. Let $C_1 = C \cap (C-n_1)$. Pick $n_2 \in C_1$ so that $C_2 = C_1 \cap (C_1 - n_2) \in p$. Since p is an idempotent, it is a non-principal ultrafilter, and hence its members are infinite sets. This allows us always to assume that any new element chosen from a member of p lies outside any given finite set. In particular, we can assume that $n_2 > n_1$. Now, $n_2 \in C_1 = C \cap (C-n_1)$ implies that $\mathrm{FS}(n_i)_{i=1}^{2} \subseteq C$. Continuing in the same fashion, let $n_3 \in C_2$ be such that $C_2 \cap (C_2 - n_3) \in p$, and $n_3 > n_1 + n_2$. Notice that $n_3 \in C_2 = C_1 \cap (C_1 - n_2) = C \cap (C-n_1) \cap (C-n_2) \cap (C-(n_1+n_2))$ implies $\mathrm{FS}(n_i)_{i=1}^{3} \subseteq C$. And so on! The sequence $(n_i)_{i=1}^{\infty}$ created this way will have the property that for any $k \in \mathbb{N}$, $\mathrm{FS}(n_i)_{i=1}^{k} \subseteq C$. We are done. ∎

4 The law of return of large sets

As we saw in Sections 1 and 2, a typical application of Principle P is to ensure that large sets return to themselves under transformations which preserve the notion of largeness. For example, to prove PRT, one first establishes the fact that if $\mu(A) > 0$, then for some $n \in \mathbb{N}$, $\mu(A \cap T^{-n}A) > 0$. Hilbert's theorem (Proposition 2.7) hinges on a similar statement: if $\bar{d}(E) > 0$, then for some $n \neq 0$, $\bar{d}(E \cap (E-n)) > 0$. Finally, Hindman's theorem also starts with the same kind of statement: if $p \in \beta\mathbb{N}$ is idempotent and $C \in p$, then for p-many n one has $(C-n) \in p$ and hence $C \cap (C-n) \in p$.

In this Section we shall have a closer look at the phenomenon of the return of large sets and discuss some of its applications and refinements.

Given an abelian group $(G,+)$ and a set $S \subseteq G$, let $S-S = \{x-y : x,y \in S\}$. One often encounters results which can be expressed as follows: if S is large, then $S-S$ is very large. On many occasions, such statements are just simple corollaries of the law of return of large sets. Perhaps the best known result of this kind is the following useful theorem due to Steinhaus ([St]).

Theorem 4.1 *If A is a Lebesgue measurable subset of $\mathbb{R}$ with $\mu(A) > 0$, then $A - A$ contains an open interval around 0.*

Proof. All that one needs to demonstrate is a form of "local" Poincaré recurrence: if $\mu(A) > 0$, then for all small enough x, $\mu(A \cap (A-x)) > 0$.

This, in turn, follows directly from Lebesgue's theorem about points of density (which says that for almost every $x \in A$ one has $\lim_{\varepsilon \downarrow 0} \frac{\mu(A \cap (x-\varepsilon, x+\varepsilon)}{2\varepsilon} = 1$), but can also be shown to follow almost immediately from the mere definition of Lebesgue measure. The following is essentially Steinhaus' original argument. Assuming without loss of generality that $\mu(A) < \infty$, let (I_n) be a sequence of open intervals which cover A and satisfy $\sum \mu(I_n) < \frac{4}{3}\mu(A)$. It is easy to see that one of the I_n, call it I, satisfies $\mu(I \cap A) > \frac{3}{4}\mu(I)$ (otherwise, $\mu(A) \leq \sum \mu(A \cap I_n) \leq \frac{3}{4} \sum \mu(I_n) < \mu(A)$, a contradiction). But then for any x satisfying $|x| < \frac{1}{2}\mu(I)$ one has $\mu\big(A \cap (A-x)\big) \geq \mu\big((A \cap I) \cap ((A \cap I) - x)\big) > 0$ (otherwise, $\mu\big((A \cap I) \cup ((A \cap I) - x)\big) = 2\mu(A \cap I) > \frac{3}{2}\mu(I)$, which would contradict $\mu\big(I \cup (I - x)\big) < \frac{3}{2}\mu(I)$). We are done. ∎

The argument used in the above proof may be iterated to show that if $\mu(A) > 0$, then for sufficiently small $|x_i|$, $i = 1, 2, \cdots, k$, $\mu\big(A \cap (A - x_1) \cap \cdots \cap (A - x_k)\big) > 0$. This gives us the following refinement of Theorem 4.1.

Theorem 4.2 *If $A \subseteq \mathbb{R}$ is a Lebesgue measurable set of positive measure, then A contains an affine image of any finite subset of $\mathbb{R}$.*

Proof. Given $F = \{x_1, \cdots, x_k\} \subseteq \mathbb{R}$, observe that for small enough $t > 0$, one has

$$\mu\big(A \cap (A - tx_1) \cap (A - tx_2) \cap \cdots \cap (A - tx_k)\big) > 0.$$

This clearly implies that for some $a \in A$

$$a + tF = \{a + tx_i : i = 1, 2, \cdots, k\} \subseteq A.$$

∎

Here is a more recent result which deals with different notions of large and very large, but surely fits the pattern we are interested in.

Theorem 4.3 ([BeS]) *Let F be an infinite field and Γ a multiplicative subgroup of finite index in F^*. (F^* denotes the multiplicative group of the field F.) Then*

$$\Gamma - \Gamma = \{x - y : x, y \in \Gamma\} = F.$$

Theorem 4.3 has a "finitistic" version which says that if $n \in \mathbb{N}$ is fixed and a finite field F is large enough, then $\{x^n - y^n : x, y \in F\} = F$. (Note that $\{x^n, x \in F^*\}$ is a multiplicative subgroup.) As a corollary, one obtains an old result of Dickson ([D]): for fixed n and large enough prime p, the equation $x^n - y^n \equiv z^n (\mathrm{mod}\, p)$ has non-trivial solutions (i.e. solutions with $x, y, z \neq 0$).

It was Schur who, in 1916 ([Sch]), gave a simple proof of Dickson's result. Schur's proof uses the following lemma, which is a (very) special case of Hindman's theorem.

Proposition 4.4 ([Sch]) *If $r \in \mathbb{N}$ and $\mathbb{N} = \bigcup_{i=1}^{r} C_i$, then one of the C_i contains x, y, z such that $x + y = z$.*

Schur's lemma, in turn, follows from the following result on returns of large sets.

Proposition 4.5 *Let $I_n = [a_n, b_n]$, $n \in \mathbb{N}$, be a sequence of increasing intervals in $\mathbb{N}$. If $\mathbb{N} = \bigcup_{i=1}^{r} C_i$, then one of the C_i, call it C, has the property that for some $x \in C$, $\bar{d}_{\{I_n\}}(C \cap (C - x)) > 0$.*

To see that Proposition 4.5 implies Proposition 4.4, notice that if $y \in C \cap (C - x)$, then x, y, and $z = x + y$ all lie in C. For a short proof and further discussion of Proposition 4.5 the reader is referred to [Be1]. See also [BeM1] for a treatment of Schur-type theorems in general amenable groups.

Motivated by Theorem 4.3, one may ask whether any set $A \subseteq \mathbb{N}$ with $\bar{d}(A) = \limsup_{n\to\infty} \frac{|A\cap\{1,\cdots,n\}|}{n} > 0$ contains an affine image of any finite set. The answer is yes, but the corresponding result is nontrivial and is, actually, equivalent to the following deep theorem due to Szemerédi.

Theorem 4.6 ([Sz]) *If $A \subseteq \mathbb{N}$ satisfies $\bar{d}(A) > 0$, then A contains arbitrarily long arithmetic progressions.*

Corollary 4.7 *If $A \subseteq \mathbb{N}$, $\bar{d}(A) > 0$, and $F = \{x_1, \cdots, x_k\} \subseteq \mathbb{N}$, then for some $a \in A$ and $d \in \mathbb{N}$, one has $a + dF = \{a + dx_i : i = 1, 2, \cdots, k\} \subseteq A$.*

Proof. Let $m \geq \max F$. By Szemerédi's theorem, there are a and d such that $\{a, a + d, a + 2d, \cdots, a + md\} \subseteq A$. Clearly $a + dF = \{a + dx_i : i = 1, 2, \cdots, k\} \subseteq A$. ∎

Remark 4.8 One also immediately observes that Corollary 4.7, in its turn, implies Szemerédi's theorem. (Just take F to be of the form $\{1, 2, \cdots, m\}$.)

The proof of Szemerédi's theorem in [Sz] is elementary but very involved. A completely different, ergodic theoretic approach to Szemerédi's theorem and, indeed, to a variety of problems belonging to Ramsey Theory, was initiated by Furstenberg ([F1]), who derived Szemerédi's theorem from a far-reaching extension of Poincaré's recurrence theorem, which corresponds to $k = 1$ in the following.

Theorem 4.9 (Furstenberg, [F1]) *Let $(X, \mathcal{B}, \mu, T)$ be a probability measure-preserving system. For any $k \in \mathbb{N}$, and for any $A \in \mathcal{B}$ with $\mu(A) > 0$, there exists $n \in \mathbb{N}$ such that $\mu(A \cap T^{-n}A \cap T^{-2n}A \cap \cdots \cap T^{-kn}A) > 0$.*

In order to derive Szemerédi's theorem from Theorem 4.9, Furstenberg introduced a correspondence principle which allows one to translate recurrence results in ergodic theory into statements about returns of large sets.

Theorem 4.10 (Furstenberg's correspondence principle) *Given a set $E \subseteq \mathbb{Z}$ with*

$$d^*(E) = \lim_{N-M\to\infty} \frac{|E \cap \{M, M+1, \cdots, N\}|}{N-M+1} > 0,$$

there exists a probability measure-preserving system $(X, \mathcal{B}, \mu, T)$ and a set $A \in \mathcal{B}$, $\mu(A) = d^*(E)$, such that for any $k \in \mathbb{N}$ and any $n_1, n_2, \cdots, n_k \in \mathbb{Z}$ one has:

$$d^*\big(E \cap (E-n_1) \cap \cdots \cap (E-n_k)\big) \geq \mu(A \cap T^{-n_1}A \cap \cdots \cap T^{-n_k}A).$$

Remark 4.11 The quantity $d^*(E)$ featured in the formulation of Furstenberg's correspondence principle is called *upper Banach density*. Clearly, if $d^*(E) > 0$, then for some sequence of intervals $I_n = [a_n, b_n]$ with $|b_n - a_n| \to \infty$ one has $d^*(E) = \bar{d}_{\{I_n\}}(E) > 0$.

Furstenberg's seminal paper started a whole new area, Ergodic Ramsey Theory. See [F2] and [Be3] for further information. See also the recent work of Gowers ([Go1], [Go2]) for an approach to Szemerédi's theorem, which provides a strong estimate for the number $N(k, \delta)$, defined as the minimal natural number such that every subset of $\{1, 2, \cdots, n\}$ containing more than δn elements must contain a length k arithmetic progression whenever $n \geq N(k, \delta)$.

5 A generalization of Khintchine's recurrence theorem

Recall that a set S in a countable abelian group G is called *syndetic* (or, sometimes, relatively dense) if there exists a finite set $F \subseteq G$ such that $S+F = \{x+y : x \in S, y \in F\} = G$. The following result, originally proved by Khintchine ([Kh]) for measure-preserving $\mathbb{R}$-actions, is usually called Khintchine's recurrence theorem.

Theorem 5.1 *For any invertible probability measure-preserving system $(X, \mathcal{B}, \mu, T)$, any $0 < \lambda < 1$, and any $A \in \mathcal{B}$ with $\mu(A) > 0$, the set*

$$\big\{n \in \mathbb{Z} : \mu(A \cap T^n A) > \lambda \mu(A)^2\big\}$$

is syndetic.

Khintchine's recurrence theorem immediately follows from the following corollary to the classical von Neumann theorem.

Theorem 5.2 *For any probability measure-preserving system $(X, \mathcal{B}, \mu, T)$ and any $A \in \mathcal{B}$ one has:*

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \mu(A \cap T^{-n}A) \geq \mu(A)^2.$$

In this section we are going to show that similar results hold for the second iteration – that is, for the analogous expression for $\mu((A\cap T^n A)\cap T^m(A\cap T^n A))$ – as well. Namely, we are going to establish the following theorems.

Theorem 5.3 *Let $(X, \mathcal{B}, \mu, T)$ be an invertible probability measure-preserving system. Then:*

(i) *For any $f_1, f_2, f_3 \in L^\infty(X, \mathcal{B}, \mu)$*

$$\lim_{N-M\to\infty} \frac{1}{(N-M)^2} \sum_{n,m=M}^{N-1} f_1(T^n x) f_2(T^m x) f_3(T^{n+m} x)$$

exists in L^2,

(ii) *For any $A \in \mathcal{B}$ with $\mu(A) > 0$*

$$\lim_{N-M\to\infty} \frac{1}{(N-M)^2} \sum_{n,m=M}^{N-1} \mu(A \cap T^n A \cap T^m A \cap T^{n+m} A) \geq \mu(A)^4.$$

Corollary 5.4 *For any invertible measure-preserving system $(X, \mathcal{B}, \mu, T)$, any $0 < \lambda < 1$, and any $A \in \mathcal{B}$, $\mu(A) > 0$, the set*

$$\left\{(n, m) \in \mathbb{Z}^2 : \mu((A \cap T^n A) \cap T^m(A \cap T^n A)) > \lambda \mu(A)^4\right\}$$

is syndetic.

Remark 5.5 Although the original paper [Kh], as well as numerous books on ergodic theory (see, for example, [Ho],[Pa],[Pe]), derive Khintchine's recurrence theorem from a much stronger Theorem 5.2, one can give a very simple proof based on the enhanced Principle P, which, moreover, works for measure-preserving actions of arbitrary (not necessarily amenable) semi-groups. (See [Be3], Section 5 for details.) Unfortunately, this simple approach does not seem to be easily modifiable to enable one to prove the two-parameter version, Corollary 5.4.

Before embarking on the proof of Theorem 5.3 we want to make some remarks and review some facts that we are going to use.

First of all, since Theorem 5.3 is trivial when μ is atomic (and since one can treat the atomic part of μ separately), we are going to assume that

the measure μ is non-atomic. Having made this assumption, we can further assume that the measure space $(X, \mathcal{B}, \mu)$ is a Lebesgue space. To see this, note that given the measure-preserving transformation T and functions f_1, f_2, f_3 featured in the formulation of Theorem 5.3 (in part (ii) one takes $f_1 = f_2 = f_3 = 1_A$), we may restrict ourselves to a T-invariant separable sub-σ-algebra of $\mathcal{B}$, with respect to which all the functions $T^n f_i$, $n \in \mathbb{Z}$, $i = 1, 2, 3$ are measurable. Now, by Carathéodory's theorem (see [R], Ch. 15, Theorem 4), any separable atomless σ-algebra of subsets of a probability space is isomorphic to the σ-algebra $\mathcal{L}$ induced by Lebesgue measure on the unit interval. This isomorphism carries T into a Lebesgue measure-preserving isomorphism of $\mathcal{L}$ which, by a theorem due to von Neumann ([R], Ch. 15, Theorem 20), admits a realization as a point mapping. Having assumed that $(X, \mathcal{B}, \mu)$ is Lebesgue, we can (and will) further assume that T is ergodic with respect to μ. Indeed, if the measure-preserving system $(X, \mathcal{B}, \mu, T)$ is not ergodic, one considers the ergodic decomposition of μ, defined by

$$\mu(A) = \int_Y \mu_\omega(A)\, d\nu(\omega), \quad A \in \mathcal{B},$$

where μ_ω are T-invariant ergodic measures on $(X, \mathcal{B})$, indexed by elements of a Lebesgue space $(Y, \mathcal{D}, \nu)$ (where the measure ν may have atoms). It is not hard to see that the validity of Theorem 5.3 for ergodic measure-preserving systems $(X, \mathcal{B}, \mu_\omega, T)$ implies its validity for $(X, \mathcal{B}, \mu, T)$. To see that Corollary 5.4 is also implied by its validity in the ergodic case, one argues as follows. Assume that for any $\omega \in Y$ one has

$$\lim_{N-M\to\infty} \frac{1}{(N-M)^2} \sum_{n,m=M}^{N-1} \mu_\omega(A \cap T^n A \cap T^m A \cap T^{n+m} A) \geq \mu_\omega(A)^4.$$

Then we have

$$\begin{aligned}
&\lim_{N-M\to\infty} \frac{1}{(N-M)^2} \sum_{n,m=M}^{N-1} \mu(A \cap T^n A \cap T^m A \cap T^{n+m} A) \\
=&\lim_{N-M\to\infty} \frac{1}{(N-M)^2} \sum_{n,m=M}^{N-1} \int_Y \mu_\omega(A \cap T^n A \cap T^m A \cap T^{n+m} A)\, d\nu(\omega) \\
=&\int_Y \Big(\lim_{N-M\to\infty} \frac{1}{(N-M)^2} \sum_{n,m=M}^{N-1} \mu_\omega(A \cap T^n A \cap T^m A \cap T^{n+m} A) \Big)\, d\nu(\omega) \\
\geq&\int_Y \mu_\omega(A)^4\, d\nu(\omega) \geq \Big(\int_Y \mu_\omega(A)\, d\nu(\omega) \Big)^4 = \mu(A)^4.
\end{aligned}$$

We shall also need the following simple fact.

Proposition 5.6 *If the invertible measure preserving system $(X, \mathcal{B}, \mu, T)$ is ergodic, then for any $h, f, g \in L^\infty(X, \mathcal{B}, \mu)$ one has*

$$\lim_{N-M\to\infty} \frac{1}{(N-M)^2} \sum_{n,m=M}^{N-1} \int h(T^n x) f(T^m x) g(T^{n+m} x)\, d\mu(x) = \int h\, d\mu \int f\, d\mu \int g\, d\mu.$$

Proof. We show first that

$$\frac{1}{(N-M)^2} \sum_{n,m=M}^{N-1} f(T^{m-n}x) g(T^m x) \to \int f\, d\mu \int g\, d\mu$$

in L^2 as $N - M \to \infty$. To verify this assertion, observe first that without loss of generality one can assume that $\int f\, d\mu = 0$. It follows from von Neumann's ergodic theorem that for every $\varepsilon > 0$ there exists $C > 0$ such that if $N - M > C$, then

$$\Big\| \frac{1}{N-M} \sum_{n=M}^{N-1} f(T^{m-n}x) \Big\|_2 < \varepsilon$$

uniformly in m. Hence, for $N - M > C$ one has

$$\begin{aligned}
&\Big\| \frac{1}{(N-M)^2} \sum_{n,m=M}^{N-1} f(T^{m-n}x) g(T^m x) \Big\|_2 \\
=&\Big\| \frac{1}{N-M} \sum_{m=M}^{N-1} g(T^m x) \Big(\frac{1}{N-M} \sum_{n=M}^{N-1} f(T^{m-n}x) \Big) \Big\|_2 \\
\le& \frac{1}{N-M} \sum_{m=M}^{N-1} \Big\| g(T^m x) \Big(\frac{1}{N-M} \sum_{n=M}^{N-1} f(T^{m-n}x) \Big) \Big\|_2 \\
\le& \varepsilon \|g\|_\infty.
\end{aligned}$$

Since ε was arbitrary, it follows that, under our assumptions,

$$\lim_{N-M\to\infty} \frac{1}{(N-M)^2} \sum_{n,m=M}^{N-1} f(T^{m-n}x) g(T^m x) = 0,$$

which implies that, for general $f \in L^\infty(X, \mathcal{B}, \mu)$,

$$\lim_{N-M\to\infty} \frac{1}{(N-M)^2} \sum_{n,m=M}^{N-1} f(T^{m-n}x) g(T^m x) = \int f\, d\mu \int g\, d\mu.$$

We have now:

$$\begin{aligned}
&\lim_{N-M\to\infty}\frac{1}{(N-M)^2}\sum_{n,m=M}^{N-1}\int h(T^nx)f(T^mx)g(T^{n+m}x)\,d\mu(x)\\
=&\lim_{N-M\to\infty}\frac{1}{(N-M)^2}\sum_{n,m=M}^{N-1}\int h(x)f(T^{m-n}x)g(T^mx)\,d\mu(x)\\
=&\int h(x)\Big(\lim_{N-M\to\infty}\frac{1}{(N-M)^2}\sum_{n,m=M}^{N-1}f(T^{m-n}x)g(T^mx)\Big)\,d\mu(x)\\
=&\int h\,d\mu\int f\,d\mu\int g\,d\mu.
\end{aligned}$$

■

Finally, we shall need a two-parameter version of the so-called van der Corput trick, which is often helpful in dealing with multiple recurrence. (See, for example, [Be1] and [BeM2], Lemma A6.)

Proposition 5.7 *Assume that $(x_{n,m})_{n,m\in\mathbb{N}}$ is a double bounded sequence of vectors in a Hilbert space. If*

$$\lim_{H\to\infty}\frac{1}{H^2}\sum_{h_1,h_2=1}^{H}\Big|\lim_{N-M\to\infty}\frac{1}{(N-M)^2}\sum_{n,m=M}^{N-1}\langle x_{n,m},x_{n+h_1,m+h_2}\rangle\Big|=0,$$

then

$$\lim_{N-M\to\infty}\Big\|\frac{1}{(N-M)^2}\sum_{n,m=M}^{N-1}x_{n,m}\Big\|=0.$$

We are now ready to start the proof of Theorem 5.3. Let $\mathcal{H}=L^2(X,\mathcal{B},\mu)$. We are going to utilize the well-known decomposition $\mathcal{H}=\mathcal{H}_{\rm d}\oplus\mathcal{H}_{\rm wm}$, where the orthogonal subspaces $\mathcal{H}_{\rm d}$ and $\mathcal{H}_{\rm wm}$ correspond to the *discrete spectrum* and *weak mixing* of the unitary operator induced by T (i.e. $(Tf)(x):=f(Tx)$), and are defined by

$$\mathcal{H}_{\rm d}=\overline{\rm Span}\{f:\exists\lambda:Tf=\lambda f\},$$

$$\mathcal{H}_{\rm wm}=\mathcal{H}_{\rm d}^{\perp}=\big\{f\in\mathcal{H}:\forall g\ \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}|\langle T^nf,g\rangle|=0\big\}.$$

Remark 5.8 We shall actually need the following equivalent definition of $\mathcal{H}_{\rm wm}$:

$$\mathcal{H}_{\rm wm}=\big\{f\in\mathcal{H}:\forall g\ \lim_{H\to\infty}\frac{1}{H^2}\sum_{h_1,h_2=1}^{H}\big|\langle T^{h_1+h_2}f,g\rangle\big|=0\big\}.$$

Let $f_i = \phi_i + \psi_i$, $\phi_i \in \mathcal{H}_{\mathrm{d}}$, $\psi_i \in \mathcal{H}_{\mathrm{d}}^{\perp}$, $i = 1, 2, 3$, be the corresponding decompositions of f_i. Substituting into the expression

$$\frac{1}{(N-M)^2} \sum_{n,m=M}^{N-1} f_1(T^n x) f_2(T^m x) f_3(T^{n+m} x), \qquad (*)$$

we shall get a representation of $(*)$ as a sum of eight expressions of the form

$$\frac{1}{(N-M)^2} \sum_{n,m=M}^{N-1} g_1(T^n x) g_2(T^m x) g_3(T^{n+m} x),$$

where each of the g_i's lies either in $\mathcal{H}_{\mathrm{d}}$ or in $\mathcal{H}_{\mathrm{d}}^{\perp} = \mathcal{H}_{\mathrm{wm}}$. We shall show first that if at least one of the g_i's belongs to $\mathcal{H}_{\mathrm{wm}}$, then

$$\lim_{N-M\to\infty} \Big\| \frac{1}{(N-M)^2} \sum_{n,m=M}^{N-1} g_1(T^n x) g_2(T^m x) g_3(T^{n+m} x) \Big\|_2 = 0.$$

Let, for instance, $g_3 \in \mathcal{H}_{\mathrm{wm}}$ (the other six cases are verified in a similar fashion). Let $x_{n,m} = g_1(T^n x) g_2(T^m x) g_3(T^{n+m} x)$. We are going to apply Proposition 5.7. We have

$$\begin{aligned} \langle x_{n,m}, x_{n+h_1,m+h_2} \rangle &= \int T^n g_1 T^m g_2 T^{n+m} g_3 T^{n+h_1} g_1 T^{m+h_2} g_2 T^{n+m+h_1+h_2} g_3 \, d\mu \\ &= \int T^n (g_1 T^{h_1} g_1) T^m (g_2 T^{h_2} g_2) T^{n+m} (g_3 T^{h_1+h_2} g_3) \, d\mu. \end{aligned}$$

By Proposition 5.6, this expression converges to

$$\int g_1 T^{h_1} g_1 \, d\mu \int g_2 T^{h_2} g_2 \, d\mu \int g_3 T^{h_1+h_2} g_3 \, d\mu.$$

Note that since $g_3 \in \mathcal{H}_{\mathrm{wm}}$, one has, by Remark 5.8,

$$\lim_{H\to\infty} \frac{1}{H^2} \sum_{h_1,h_2=1}^{H} \big| \langle T^{h_1+h_2} g_3, g_3 \rangle \big| = 0.$$

Since $g_1, g_2 \in L^\infty$, this implies

$$\lim_{H\to\infty} \frac{1}{H^2} \sum_{h_1,h_2} \left| \int g_1 T^{h_1} g_1 \, d\mu \int g_2 T^{h_2} g_2 \, d\mu \int g_3 T^{h_1+h_2} g_3 \, d\mu \right| = 0,$$

and hence, in accordance with Proposition 5.7,

$$\begin{aligned} &\lim_{N-M\to\infty} \Big\| \frac{1}{(N-M)^2} \sum_{n,m=M}^{N-1} x_{n,m} \Big\|_2 \\ =&\lim_{N-M\to\infty} \Big\| \frac{1}{(N-M)^2} \sum_{n,m=M}^{N-1} g_1(T^n x) g_2(T^m x) g_3(T^{n+m} x) \Big\|_2 = 0. \end{aligned}$$

Now, to finish the proof of part (i) of Theorem 5.3, we have only to show that

$$\lim_{N-M\to\infty}\frac{1}{(N-M)^2}\sum_{n,m=M}^{N-1} f_1(T^n x)f_2(T^m x)f_3(T^{n+m}x)$$

exists whenever $f_1, f_2, f_3 \in \mathcal{H}_{\rm d}$. But this is almost obvious. Indeed, each $f \in \mathcal{H}_{\rm d}$ has a representation $f = \sum_\lambda a_\lambda f_\lambda$, where $f_\lambda(Tx) = \lambda f_\lambda(x)$, and it is enough to verify the convergence for finite approximations of the form $\tilde{f}_i = \sum_{k=1}^{L} a_{\lambda_k}^{(i)} f_{\lambda_k}$, for which the convergence statement is trivial.

We now turn our attention to part (ii) of Theorem 5.3. Let $A \in \mathcal{B}$ with $\mu(A) > 0$. Let $g = 1_A = f + h$, where $f \in \mathcal{H}_{\rm d}$ and $h \in \mathcal{H}_{\rm d}^{\perp} = \mathcal{H}_{\rm wm}$. Note that since g is bounded, f is bounded as well. In view of the proof of part (i) we have

$$\begin{aligned}&\lim_{N-M\to\infty}\frac{1}{(N-M)^2}\sum_{n,m=M}^{N-1} g(T^n x)g(T^m x)g(T^{n+m}x)\\ =&\lim_{N-M\to\infty}\frac{1}{(N-M)^2}\sum_{n,m=M}^{N-1} f(T^n x)f(T^m x)f(T^{n+m}x).\end{aligned}$$

It follows that

$$\begin{aligned}&\lim_{N-M\to\infty}\frac{1}{(N-M)^2}\sum_{n,m=M}^{N-1}\mu(A\cap T^n A\cap T^m A\cap T^{n+m}A)\\ =&\lim_{N-M\to\infty}\frac{1}{(N-M)^2}\sum_{n,m=M}^{N-1}\int g(x)g(T^n x)g(T^m x)g(T^{n+m}x)\,d\mu\\ =&\int g(x)\lim_{N-M\to\infty}\frac{1}{(N-M)^2}\sum_{n,m=M}^{N-1} g(T^n x)g(T^m x)g(T^{n+m}x)\,d\mu\\ =&\int (f(x)+h(x))\lim_{N-M\to\infty}\frac{1}{(N-M)^2}\sum_{n,m=M}^{N-1} f(T^n x)f(T^m x)f(T^{n+m}x)\,d\mu\\ =&\lim_{N-M\to\infty}\frac{1}{(N-M)^2}\sum_{n,m=M}^{N-1}\int f(x)f(T^n x)f(T^m x)f(T^{n+m}x)\,d\mu.\end{aligned}$$

(We used the fact that f is bounded and that the product of bounded functions from $\mathcal{H}_{\rm d}$ belongs to $\mathcal{H}_{\rm d}$.) Taking into account that the constant functions belong to $\mathcal{H}_{\rm d}$ and that $1_A = g = f + h$ with $f \in \mathcal{H}_{\rm d}$, $h \in \mathcal{H}_{\rm d}^{\perp}$ implies $\int f\,d\mu = \int g\,d\mu = \mu(A)$, we see that in order to prove (ii) it is enough to establish the following.

Proposition 5.9 *Let $(X, \mathcal{B}, \mu, T)$ be an ergodic measure-preserving system. If $f \in \mathcal{H}_{\mathrm{d}}$, where f is bounded and non-negative, then*

$$\lim_{N-M\to\infty} \frac{1}{(N-M)^2} \sum_{n,m=M}^{N-1} \int fT^n fT^m fT^{n+m} f \, d\mu \geq \left(\int f \, d\mu\right)^4.$$

Proof. We shall need some basic facts about the eigenvalues and eigenfunctions of the unitary operators induced by measure-preserving transformations. (For details see [Ha] and [CFS].) Here is a summary of what we are going to use. (Warning: some of the facts below are true for ergodic transformations only.)
(a) The set $\Gamma = \{\lambda \in \mathbb{C} : \exists f \in L^2(X, \mathcal{B}, \mu) : Tf = \lambda f\}$ is a subgroup of the unit circle. Since we are dealing with Lebesgue spaces, this group is countable. Any eigenvalue $\lambda \in \Gamma$ has multiplicity one.
(b) The eigenfunctions corresponding to distinct eigenvalues are orthogonal. They have constant modulus and will be assumed to be normalized so that each of them will satisfy the condition $|f(x)| = 1$ a.e. We shall also assume that if $\lambda_1, \lambda_2 \in \Gamma$ and $f_{\lambda_1}, f_{\lambda_2}$ are the corresponding eigenfunctions, then $f_{\lambda_1} f_{\lambda_2} = f_{\lambda_1 \lambda_2}$.

We return to the proof of Proposition 5.9. Fix a set $\{f_\lambda\}_{\lambda\in\Gamma}$ of eigenfunctions so that the conditions described in item (b) above are satisfied. Let $f = \sum_{\lambda\in\Gamma} a_\lambda f_\lambda$ be the expansion of our function f in this basis. Note that $f_1 = 1$ a.e., $a_1 = \int f \, d\mu$, and also, for any $\lambda \neq 1$, $\int f_\lambda \, d\mu = 0$. Substituting $f = \sum a_\lambda f_\lambda$ into the integral

$$\int fT^n fT^m fT^{m+n} f \, d\mu$$

and changing the order of integration and summation, we shall arrive at the sum of terms of the form (where the sum will be taken over ρ, λ, τ and ν)

$$\begin{aligned} K_{\rho\lambda\tau\nu} &= \int a_\rho f_\rho T^n (a_\lambda f_\lambda) T^m (a_\tau f_\tau) T^{n+m} (a_\nu f_\nu) \, d\mu \\ &= \lambda^n \tau^m \nu^{n+m} \int a_\rho a_\lambda a_\tau a_\nu f_\rho f_\lambda f_\tau f_\nu \, d\mu \\ &= (\lambda\nu)^n (\tau\nu)^m a_\rho a_\lambda a_\tau a_\nu \int f_{\rho\lambda\tau\nu} \, d\mu. \end{aligned}$$

The contribution of such a term to the double Cesaro limit

$$\lim_{N-M\to\infty} \frac{1}{(N-M)^2} \sum_{n,m=M}^{N-1}$$

will be non-zero only if $\lambda\nu = 1$, $\tau\nu = 1$, and $f_{\rho\lambda\tau\nu} = f_1 = 1$ (which implies $\rho\lambda\tau\nu = 1$). The three conditions on ρ,λ,τ,ν imply $\lambda = \tau$, $\rho = \nu$, and hence

$a_\rho a_\lambda a_\tau a_\nu = a_\lambda^2 a_\nu^2$. Now, since $f \geq 0$, f is equal to its complex conjugate: $\bar{f} = f$, which gives

$$\sum \bar{a}_\lambda \bar{f}_\lambda = \sum a_\lambda f_\lambda$$

or

$$\sum \bar{a}_\lambda f_{\bar{\lambda}} = \sum a_{\bar{\lambda}} f_{\bar{\lambda}}.$$

From the uniqueness of the expansion we get $a_{\bar{\lambda}} = \bar{a}_\lambda$. Since $\lambda = \bar{\nu}$, we have

$$a_\rho a_\lambda a_\tau a_\nu = a_\lambda^2 a_\nu^2 = a_{\bar{\nu}}^2 a_\nu^2 = \bar{a}_\nu^2 a_\nu^2 = |a_\nu|^4 \geq 0.$$

We showed that each time a term of the form $K_{\rho\lambda\tau\nu}$ gives a non-zero contribution to our double Cesaro limit, this contribution is non-negative. Also, at least one $K_{\rho\lambda\tau\nu}$ – namely, the one corresponding to $\rho = \lambda = \tau = \nu = 1$ – gives the contribution

$$\int a_1^4 f_1 \, d\mu = \Big(\int f \, d\mu \Big)^4 \int f_1 \, d\mu = \Big(\int f \, d\mu \Big)^4.$$

We are done. ∎

Theorem 5.3 can be easily derived from the following more general result which may be proved by a similar argument.

Theorem 5.10 *Let G be a countable abelian group and $(X, \mathcal{B}, \mu, \{T_g\}_{g\in G})$ a probability measure-preserving system. Let $\{F_n\}_{n=1}^\infty$ be a Følner sequence in $G \times G$. Then:*

(i) *For any $f_1, f_2, f_3 \in L^\infty(X, \mathcal{B}, \mu)$*

$$\lim_{n\to\infty} \frac{1}{|F_n|} \sum_{g,h\in F_n} f_1(T_g x) f_2(T_h x) f_3(T_{g+h} x)$$

exists in L^2.

(ii) *For any $A \in \mathcal{B}$ with $\mu(A) > 0$*

$$\lim_{n\to\infty} \frac{1}{|F_n|} \sum_{g,h\in F_n} \mu(A \cap T_g A \cap T_h A \cap T_{g+h} A) \geq \mu(A)^4.$$

Corollary 5.11 *For any countable abelian group G, any measure-preserving system $(X, \mathcal{B}, \mu, \{T_g\}_{g\in G})$, any $0 < \lambda < 1$, and any $A \in \mathcal{B}$ with $\mu(A) > 0$ the set*

$$\big\{(g,h) \in G \times G : \mu(A \cap T_g A \cap T_h A \cap T_{g+h} A) > \lambda \mu(A)^4\big\}$$

is syndetic.

It would be interesting to see whether Theorem 5.10 and Corollary 5.11 generalize further to higher order iterations and noncommutative group actions.

References

[B] J. Barrow-Green, *Poincaré and the Three Body Problem*, Amer. Math. Soc., Providence; London Math. Soc., London, 1997.

[Be1] V. Bergelson, A density statement generalizing Schur's theorem, *J. Combinatorial Theory (A)* **43** (1986), 338-343.

[Be2] V. Bergelson, Weakly mixing PET, *Ergodic Theory and Dynamical Systems*, **7** (1987), 337-349.

[Be3] V. Bergelson, Ergodic Ramsey Theory—an Update, Ergodic Theory of $\mathbb{Z}^d$-actions, M. Pollicott and K. Schmidt, editors, *London Math. Soc. Lecture Notes Series* **228**, Cambridge University Press (1996), 1-61.

[BeM1] V. Bergelson and R. McCutcheon, Recurrence for semigroup actions and a non-commutative Schur theorem, *Topological Dynamics and Applications* (Minneapolis, MN, 1995), *Contemporary Math.* **215**, Amer. Math. Soc., Providence, 1998, 205-222.

[BeM2] V. Bergelson and R. McCutcheon, An ergodic IP polynomial Szemerédi theorem, *Memoirs of AMS* (to appear).

[BeS] V. Bergelson and D. Shapiro, Multiplicative subgroups of finite index in a ring, *Proc. Amer. Math. Soc.* **119** (1993), 1127-1134.

[Bo1] L. Boltzmann, Entgegnung auf die wärmetheoretischen Betrachtungen des Hrn. E. Zermelo, *Annalen der Physik* **57** (1896), 773-784.

[Bo2] L. Boltzmann, Zu Hrn. Zermelo's Abhandlung "Über die mechanische Erklärung irreversibler Vorgänge", *Annalen der Physik* **59** (1896), 392-398.

[Bo3] L. Boltzmann, Über einen mechanischen Satz von Poincaré, *Wien Ber.* **106** (1897), p. 12.

[Br] S.G. Brush, *Kinetic Theory. Vol. 2. Irreversible Processes*, Pergamon Press, 1966.

[C] C. Carathéodory, Über den Wiederkehrsatz von Poincaré, *S. B. Preuss. Akad. Wiss.*, (1919), 580-584.

[CN] W. Comfort and S. Negrepontis, *The Theory of Ultrafilters*, Springer-Verlag, Berlin and New York, 1974.

[CFN] I. Cornfeld, S. Fomin, and Y. Sinai, *Ergodic Theory*, Springer-Verlag, 1982.

[D] L. Dickson, Lower limit for the number of sets of solutions of $x^l + y^l + z^l \equiv 0 (\mathrm{mod}\, p)$, *J. Reine Angew. Math.* **135** (1908), 181-188.

[EE] P. Ehrenfest and T. Ehrenfest, *The Conceptual Foundations of the Statistical Approach in Mechanics* (translation of *Begriffliche Grundlagen der statistischen Auffassung in der Mechanik*), Dover, 1990.

[El] R. Ellis, Distal transformation groups, *Pac. J. Math.* **8** (1958), 401-405.

[F1] H. Furstenberg, Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions, *J. d'Analyse Math.* **31** (1977), 204-256.

[F2] H. Furstenberg, *Recurrence in Ergodic Theory and Combinatorial Number Theory*, Princeton University Press, 1981.

[FK] H. Furstenberg and Y. Katznelson, An ergodic Szemerédi theorem for IP-systems and combinatorial theory, *J. d'Analyse Math.* **45** (1985), 117-168.

[FW] H. Furstenberg and B. Weiss, Topological dynamics and combinatorial number theory, *J. d'Analyse Math.* **34** (1978), 61-85.

[GRS] R. Graham, B. Rothschild, and J. Spencer, *Ramsey Theory*, Wiley, New York, 1980.

[Gi] J. Gillis, Note on a property of measurable sets, *J. London Math. Soc.* **11** (1936), 139-141.

[Gor] D. Goroff, Introduction to [P2].

[Gow1] W.T. Gowers, A new proof of Szemerédi's theorem for arithmetic progressions of length four, *Geom. Funct. Anal.* **8** (1998), 529-55.

[Gow2] W.T. Gowers, A new proof of Szemerédi's theorem, preprint.

[Ha] P. Halmos, *Lectures on Ergodic Theory*, Chelsea Publishing Co., New York, 1956.

[Hi1] D. Hilbert, Über die Irreducibilität ganzer rationaler Functionen mit ganzzahligen Coefficienten, *J. Math.* **110** (1892), 104-129.

[Hi2] H. Hilmy, Sur les théoremes de récurrence dans la dynamique générale, *Amer. J. Math.* **61** (1939), 149-160.

[Hi3] N. Hindman, Finite sums from sequences within cells of a partition of $\mathbf{N}$, *J. Combinatorial Theory* (Series A) **17** (1974), 1-11.

[HiS] N. Hindman and D. Strauss, Algebra in the Stone-Čech compactification. Theory and applications, *de Gruyter Expositions in Mathematics* **27**, Walter de Gruyter, 1998.

[Ho] E. Hopf, *Ergodentheorie*, Chelsea, New York, 1948.

[Kac] M. Kac, *Probability and Related Topics in the Physical Sciences*, Interscience Publishers, 1959.

[KaM] M. Karpovsky and V. Milman, On subspaces contained in subsets of finite homogeneous spaces, *Discrete Mathematics* **22**, (1978), 273-280.

[Kh] A. Y. Khintchine, Eine Verschärfung des Poincaréschen "Wiederkehrsatzes", *Comp. Math.* **1** (1934), 177-179.

[O] J. Oxtoby, *Measure and Category*, Second edition, Springer-Verlag, 1980.

[Pa] W. Parry, *Topics in Ergodic Theory*, Cambridge Univ. Press, Cambridge, 1981.

[Pe] K. Petersen, *Ergodic Theory*, Cambridge Univ. Press, 1983.

[P1] H. Poincaré, Sur le problème des trois corps et les équations de la Dynamique, *Acta Mathematica* **13** (1890), 1-270.

[P2] H. Poincaré, *New Methods of Celestial Mechanics* (translation of *Les méthodes nouvelles de la mécanique céleste* I (1892), II (1893), and III (1899)), D. Goroff, editor, Amer. Inst. of Physics, New York, 1993.

[P3] H. Poincaré, *Calcul des probabilités*, *Gauthier-Villars*, 1912.

[R] H. Royden, *Real Analysis*, Third edition, Prentice Hall, 1988.

[Sch] I. Schur, Über die Kongruenz $x^m + y^m \equiv z^m (\mathrm{mod}\, p)$, *Jahresbericht der Deutschen Math.-Ver.* **25** (1916),114-117.

[St] H. Steinhaus, Sur les distances des ensembles de mésure positive, *Fund. Math.***1** (1920), 93-104.

[Sz] E. Szemerédi, On sets of integers containing no k elements in arithmetic progression, *Acta Arith.* **27** (1975), 199-245.

[Z1] E. Zermelo, Über einen Satze der Dynamik und die mechanischen Wärmetheorie, *Annalen der Physik* **57** (1896), 485-494.

[Z2] E. Zermelo, Über die mechanische Erklärung irreversibler Vorgänge, *Annalen der Physik* **59** (1896), 793-801.