

# INFINITE SUMSETS IN SETS WITH POSITIVE DENSITY

BRYNA KRA, JOEL MOREIRA, FLORIAN K. RICHTER, AND DONALD ROBERTSON

## CONTENTS

1. Introduction	1
2. Preliminaries	4
3. A discussion of $B_1 + B_2 \subset A$	8
4. Erdős cubes	20
5. Cube systems and pronilfactors	24
6. A continuous ergodic decomposition	31
7. Proof of Theorem 5.9	37
8. Open questions	42

## 1. INTRODUCTION

Szemerédi [28], settling a longstanding conjecture of Erdős and Turán, proved that every set of natural numbers with positive upper density contains arbitrarily long arithmetic progressions. Since that work, a variety of generalizations have appeared, guaranteeing the existence of various finite patterns in sufficiently large sets of integers. Among the notable achievements are the multidimensional generalization of Furstenberg and Katznelson [14], the polynomial generalization of Bergelson and Leibman [3], and the groundbreaking theorem of Green and Tao [16] on arithmetic progressions in primes.

Erdős and Graham [7] conjectured that sets of natural numbers with positive upper density contain not only finite arithmetic configurations, but various infinite ones as well. In particular, Erdős conjectured that any set with positive upper density contains a sumset  $B_1 + B_2 = \{b_1 + b_2 : b_1 \in B_1, b_2 \in B_2\}$  of two infinite sets  $B_1, B_2 \subset \mathbb{N}$ . Under the stronger assumption that the density exceeds  $1/2$ , this conjecture was proven in [5]. In full generality, it was resolved in [26], using a combination of combinatorial analysis, non-standard analysis, and an ergodic structural result. A more streamlined and purely ergodic proof was subsequently given by Host in [17].

We are concerned here with a natural generalization of this result to higher orders. To be precise, we focus on the question of whether sets with positive upper density must contain for each  $k \in \mathbb{N}$  a sumset

$$B_1 + \cdots + B_k = \{b_1 + \cdots + b_k : b_1 \in B_1, \dots, b_k \in B_k\},$$

---

Received by the editors June 3, 2022, and, in revised form, March 28, 2023.

2020 *Mathematics Subject Classification*. Primary 05D10, 11B13, 37A05; Secondary 11B30.

The first author was supported by National Science Foundation grant DMS-205464. The fourth author was supported by EPSRC grant V050362.

where  $B_1, \dots, B_k \subset \mathbb{N}$  are infinite. Our main result gives a positive answer to this question under the more general assumption of positive upper Banach density, which we phrase in terms of Følner sequences. Recall that a *Følner sequence* in  $\mathbb{N}$  is any sequence  $\Phi = (\Phi_N)_{N \in \mathbb{N}}$  of finite subsets of  $\mathbb{N}$  satisfying

$$(1.1) \quad \lim_{N \rightarrow \infty} \frac{|(\Phi_N - t) \cap \Phi_N|}{|\Phi_N|} = 1$$

for all  $t \in \mathbb{N}$ , where  $\Phi_N - t$  denotes the shift  $\{x \in \mathbb{N} : x + t \in \Phi_N\}$ . For example

$$\Phi_N = \{L_N, L_N + 1, \dots, M_N - 1\}$$

defines a Følner sequence whenever  $M_N - L_N \rightarrow \infty$  as  $N \rightarrow \infty$ . A set  $A \subset \mathbb{N}$  has *positive upper Banach density* if there is a Følner sequence  $\Phi$  such that

$$(1.2) \quad \lim_{N \rightarrow \infty} \frac{|A \cap \Phi_N|}{|\Phi_N|} > 0.$$

With this definition, we are ready to state our main theorem, which answers [26, Question 6.3].

**Theorem 1.1.** *If  $A \subset \mathbb{N}$  has positive upper Banach density then for every  $k \in \mathbb{N}$  there are infinite sets  $B_1, \dots, B_k \subset \mathbb{N}$  with  $B_1 + \dots + B_k \subset A$ .*

We believe that this result and its antecedents [5, 17, 26, 27] are the first to guarantee infinite arithmetic configurations in all sets of positive upper Banach density. The latter papers [5, 17, 26] all relied on a method introduced in [5] (see [5, Proof of Theorem 3.2] and [26, Proposition 2.5]) to construct two-fold sumsets  $B_1 + B_2$  within positive density sets using a variant of an *intersectivity lemma* of Bergelson [1, Theorem 1.1]. This technique played a particularly important role in [17, 26], where it was used to transfer the combinatorial problem of finding sumsets in sets of positive density into an ergodic theoretic problem of set recurrence in measure preserving dynamical systems. Although this approach works well when dealing with two-fold sumsets  $B_1 + B_2$ , it is not clear how to extend it to find  $B_1 + \dots + B_k$  in sets of positive density when  $k \geq 3$ . For this reason, it is not clear how to generalize the proofs in [17, 26] to produce a proof of Theorem 1.1. We overcome this obstacle by introducing a new approach that is distinct from existing methods. The remainder of this section is dedicated to outlining this approach and summarizing its key elements.

The central idea behind our proof of Theorem 1.1 is to connect  $k$ -fold sumsets in the integers with points in certain  $2^k$ -fold joinings of measure preserving systems that have highly structured recurrence properties. These joinings – known as cube systems – were introduced in [18] to give an algebraic characterization of the limiting behavior of the multiple ergodic averages introduced by Furstenberg [12] in his proof of Szemerédi’s theorem via ergodic theory. Moreover, they are used to define the dynamical counterpart of the Gowers uniformity norms, making them indispensable tools for studying higher order recurrence in dynamical systems.

The density assumption (1.2) provides, via a classical construction of Furstenberg, a dynamical system  $(X, \mu, T)$ , a clopen set  $E \subset X$  satisfying  $\mu(E) > 0$ , and a point  $a \in X$  generic for  $\mu$  such that

$$(1.3) \quad A = \{n \in \mathbb{N} : T^n a \in E\}.$$

Writing  $\llbracket k \rrbracket = \{0, 1\}^k$ , we consider the  $2^k$ -fold self-product  $X^{\llbracket k \rrbracket}$  equipped with the  $2^k$ -fold product transformation  $T^{\llbracket k \rrbracket}$ . We connect  $k$ -fold sumsets to certain special

points in  $X^{[k]}$ . We refer to such points as *k-dimensional Erdős cubes*, and they form the bridge between additive combinatorics and ergodic theory in our work. In the case  $k = 2$  the definition is relatively easy to state.

**Definition 1.2.** A point  $(x_{00}, x_{01}, x_{10}, x_{11}) \in X^{[2]}$  is a *2-dimensional Erdős cube* if, with respect to the transformation  $T \times T$ , the forward orbit of  $(x_{00}, x_{01})$  visits every neighborhood of  $(x_{10}, x_{11})$  infinitely often, and the forward orbit of  $(x_{00}, x_{10})$  visits every neighborhood of  $(x_{01}, x_{11})$  infinitely often.

For  $k \geq 3$  the definition becomes more technical, requiring additional notation, and is deferred to Definition 4.1 in Section 4. If  $(x_{00}, x_{01}, x_{10}, x_{11})$  is a 2-dimensional Erdős cube, it can be shown that for any neighborhood  $U$  of  $x_{11}$  the set  $\{n \in \mathbb{N} : T^n x_{00} \in U\}$  contains  $B_1 + B_2$  for infinite sets  $B_1, B_2 \subset \mathbb{N}$ . Thus, to prove Theorem 1.1 for  $k = 2$  it suffices to find a 2-dimensional Erdős cube  $(x_{00}, x_{01}, x_{10}, x_{11})$  with  $x_{00} = a$  and  $x_{11} \in E$ , where  $a$  and  $E$  are as in the above construction of Furstenberg, as then  $A$  contains  $B_1 + B_2$  for infinite sets  $B_1, B_2 \subset \mathbb{N}$  by (1.3). This approach leads to a new proof of Erdős's conjecture, different from those given either in [26] or in [17].

For larger  $k$ , the story is the same: if  $(x_{\vec{0}}, \dots, x_{\vec{1}}) \in X^{[k]}$  is a *k-dimensional Erdős cube* then for any neighborhood  $U$  of  $x_{\vec{1}}$  the set  $\{n \in \mathbb{N} : T^n x_{\vec{0}} \in U\}$  contains  $B_1 + \dots + B_k$  for infinite sets  $B_1, \dots, B_k \subset \mathbb{N}$ . The main idea behind our proof of Theorem 1.1 is to build for every  $k \in \mathbb{N}$  a *k-dimensional Erdős cube* with prescribed first and last coordinates. In realizing it, the theory of cube systems plays an important role.

Cube systems, which we review in Section 5, enjoy several auspicious properties we exploit in our proof. Foremost, they provide for each  $k \in \mathbb{N}$  a  $2^k$ -fold joining  $\mu^{[k]}$  of  $\mu$  called the *k-dimensional cubic measure*, which possesses exceptional symmetry: there is a natural group of symmetries acting on  $X^{[k]}$  that fixes the first coordinate of  $X^{[k]}$ , commutes with  $T^{[k]}$ , and preserves the measure  $\mu^{[k]}$ . Using these symmetries it is not hard to prove that  $\mu^{[k]}$ -almost every point is an Erdős cube. To additionally prescribe the first and last coordinates, we use a disintegration  $t \mapsto \sigma_t^{[k]}$  of the cubic measure  $\mu^{[k]}$  over the projection from  $X^{[k]}$  to its first coordinate. From this perspective, if  $\sigma_a^{[k]}$ -almost every point is an Erdős cube then we have an abundance of Erdős cubes with first coordinate equal to  $a$ , and this would suffice for proving Theorem 1.1.

Unfortunately, the general theory of disintegrations only provides a measure  $\sigma_t^{[k]}$  for  $\mu$ -almost every point  $t \in X$ , and so a priori there is no guarantee that  $\sigma_a^{[k]}$  is well-defined for a given point  $a \in X$ . By passing to an extension of the measure preserving system  $(X, \mu, T)$  we can assume that the factor maps defining certain structural factors, the *pronilfactors*, are not just measurable, but are actually continuous. This allows us to make use of topological properties of these factors. The bulk of our technical work is showing that this additional topological input implies the existence of an ergodic decomposition  $x \mapsto \lambda_x^{[k]}$  of the cubic measure  $\mu^{[k]}$  that is not just measurable but actually continuous. We use this decomposition to define  $\sigma_t^{[k]}$  for every point  $t \in X$  in such a way that  $\sigma_t^{[k]}$  is invariant under the above-mentioned symmetries of  $\mu^{[k]}$  and pushes forward to  $\mu$  on the last coordinate. This allows us to prove that  $\sigma_a^{[k]}$ -almost every point is an Erdős cube whenever  $a$  is generic for  $\mu$ . The existence of a *k-dimensional Erdős cube*  $(x_{\vec{0}}, \dots, x_{\vec{1}})$  with  $x_{\vec{0}} = a$  and  $x_{\vec{1}} \in E$  now follows readily, yielding a proof of Theorem 1.1.

The relevance of cube systems to finding sumsets in positive density sets highlights some surprising parallels between dynamical approaches to Szemerédi's theorem and our approach to proving Theorem 1.1 finding  $(k + 1)$ -term arithmetic progressions and finding  $k$ -fold sumsets in a set of positive upper density rely on the same algebraic structures provided by the  $k$ -step pronilfactors that arise in the structure theory of measure preserving systems.

*Structure of the paper:* Section 2 is a preparatory section that covers preliminaries from ergodic theory needed in the proof of Theorem 1.1. Depending on the reader's background in ergodic theory, this section may be skimmed through or skipped.

In Section 3 we carry out our proof of Theorem 1.1 in the special case  $k = 2$ . We treat this case separately to illustrate our method for the approach to the general case. The remainder of the paper is devoted to the generalization of this proof to higher orders.

In Section 4 we formulate our main technical result, Theorem 4.4 which states that  $k$ -dimensional Erdős cubes with first coordinate prescribed and last coordinate in any fixed set of positive measure exist. The deduction of Theorem 1.1 from Theorem 4.4 is carried out in Section 4.2.

In Section 5 we recall the theory of cubic measures and pronilfactors, and prove every measure preserving system has an extension with continuous factor maps to its pronilfactors. We also state Theorem 5.9 and show how it implies Theorem 4.4. In Section 5.4 we also explain the relation between our new notion of Erdős cubes and the already established notion of dynamical cubes introduced in [18, 21].

The proof of Theorem 5.9 is carried out in Sections 6 and 7. The first step is to show the existence – using crucially the continuous factor maps to pronilfactors in the hypothesis of Theorem 5.9 – of an ergodic decomposition  $x \mapsto \lambda_x^{[k]}$  of the measure  $\mu^{[k]}$  that is continuous, and this is carried out in Theorem 6.5. This decomposition is in turn used in Section 7.1 to construct a disintegration of  $\mu^{[k]}$  over the projection to the first coordinate. In Section 7.2 these components are assembled to complete the proof of Theorem 5.9.

The use of cube systems to find infinite patterns in sets of positive upper Banach density opens new avenues of research, and we discuss some open questions and potential extensions of our work in Section 8.

For reference, we include a summary of the chain of implications used in the proof of Theorem 1.1.

Theorem 7.5  $\xrightarrow{\text{Sec. 7.1}}$  Theorem 5.9  $\xrightarrow{\text{Sec. 5.2}}$  Theorem 4.4  $\xrightarrow{\text{Sec. 4.2}}$  Theorem 1.1  
 Theorem 7.5 is proved in Section 7.

## 2. PRELIMINARIES

In this section we collect preliminary definitions and results from ergodic theory that we use throughout the paper.

*Topological systems.* By a *topological system*, we mean a pair  $(X, T)$  where  $X$  is a compact metric space and  $T: X \rightarrow X$  is a homeomorphism. We write  $\mathcal{O}(x, T) = \{T^n x : n \in \mathbb{Z}\}$  for the *orbit* of  $x \in X$  under  $T$  and

$$\omega(x, T) = \bigcap_{n \in \mathbb{N}} \overline{\{T^j(x) : j \geq n\}}$$

for the set of forward limit points of  $x$  under  $T$ . In other words,  $y \in \omega(x, T)$  means that there is a strictly increasing sequence  $c: \mathbb{N} \rightarrow \mathbb{N}$  with  $T^{c(n)}(x) \rightarrow y$  as  $n \rightarrow \infty$ . Note that  $x$  need not belong to  $\omega(x, T)$ . The topological system  $(X, T)$  is *transitive* when there is a point  $a \in X$ , called a *transitive point*, whose orbit  $\mathcal{O}(a, T)$  is dense in  $X$ . A topological system  $(X, T)$  is *minimal* if every point is transitive, and the system is *distal* if

$$\inf\{d(T^n x, T^n y) : n \in \mathbb{N}\} > 0$$

for all  $x \neq y \in X$ , where  $d$  is the metric on  $X$ .

*Measure preserving systems.* By a *system*, we mean a triple  $(X, \mu, T)$  where  $(X, T)$  is a topological system and  $\mu$  is a  $T$ -invariant Borel probability measure on  $X$ . This means  $\mu(T^{-1}D) = \mu(D)$  for all Borel sets  $D \subset X$ . We stress that all our measure preserving systems have an underlying topological structure. Whenever  $f: X \rightarrow Y$  is measurable and  $\mu$  is a Borel measure on  $X$  the *push forward* of  $\mu$  under  $f$  is the measure  $f\mu$  on  $Y$  defined by  $(f\mu)(D) = \mu(f^{-1}D)$ .

*Factor maps.* We make use of two different types of factor maps between systems.

**Definition 2.1** (Measurable factor map). For a system  $(X, \mu, T)$  we say that the system  $(Y, \nu, S)$  is a *measurable factor* of  $(X, \mu, T)$  if there is a measurable map  $\pi: X \rightarrow Y$ , the *measurable factor map*, such that  $\pi\mu = \nu$  and  $(S \circ \pi)(x) = (\pi \circ T)(x)$  for  $\mu$ -almost every  $x \in X$ .

**Definition 2.2** (Continuous factor map). A system  $(Y, \mu, S)$  is a *continuous factor* of a system  $(X, \mu, T)$  if there is a continuous and surjective map  $\pi: X \rightarrow Y$ , the *continuous factor map*, with  $\pi\mu = \nu$  and  $(S \circ \pi)(x) = (\pi \circ T)(x)$  for all  $x \in X$ .

We caution the reader that the distinction between Definitions 2.1 and 2.2 is important throughout the sequel, and reassure the reader that we always specify measurable or continuous, as appropriate.

*Generic points and support of a measure.* We write  $C(X)$  for the space of complex valued continuous functions on the compact metric space  $X$ , and write  $\mathcal{M}(X)$  for the space of Borel probability measures on  $X$  equipped with the weak\* topology. The *support* of a Borel probability measure  $\mu$  on a compact metric space  $X$  is the smallest closed, full measure subset  $\text{supp}(\mu)$  of  $X$ .

Recall that a *Følner sequence* is a sequence  $\Phi = (\Phi_N)_{N \in \mathbb{N}}$  of finite subsets of  $\mathbb{N}$  satisfying (1.1). We stress that all Følner sequences used in this paper are in  $\mathbb{N}$ , and not in  $\mathbb{Z}$ , despite the fact that all our systems are invertible.

**Definition 2.3** (Generic points). For a system  $(X, \mu, T)$  and a Følner sequence  $\Phi$ , the point  $x \in X$  is *generic for  $\mu$  with respect to  $\Phi$* , written  $x \in \text{gen}(\mu, \Phi)$ , if

$$\lim_{N \rightarrow \infty} \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} f(T^n x) = \int_X f \, d\mu$$

for every  $f \in C(X)$ . We write  $x \in \text{gen}(\mu)$  and say  $x$  is *generic for  $\mu$*  if

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N f(T^n x) = \int_X f \, d\mu$$

for all  $f \in C(X)$ .

**Lemma 2.4.** *Fix a system  $(X, \mu, T)$  and a Følner sequence  $\Phi$ . If  $x \in \text{gen}(\mu, \Phi)$  and  $y \in \text{supp}(\mu)$ , then  $y \in \omega(x, T)$ .*

*Proof.* Fix a compatible metric on  $X$  and write  $B(y, r)$  for the open ball centered at  $y$  of radius  $r$ . For every  $\varepsilon > 0$ , there exists a non-negative  $f \in C(X)$  with  $f = 1$  on the ball  $B(y, \varepsilon/2)$  and  $f = 0$  outside  $B(y, \varepsilon)$ . Now  $y \in \text{supp}(\mu)$  implies

$$\int_X f d\mu > 0$$

and since  $x \in \text{gen}(\mu, \Phi)$  there are infinitely many  $n \in \mathbb{N}$  with  $T^n x \in B(y, \varepsilon)$ .  $\square$

*Conditional expectation.* Fix a system  $(X, \mu, T)$ . Whenever  $\mathcal{A}$  is a sub- $\sigma$ -algebra of the Borel  $\sigma$ -algebra on  $X$ , every  $f \in L^2(X, \mu)$  has a *conditional expectation* on  $\mathcal{A}$  which we write  $\mathbb{E}(f | \mathcal{A})$  and define to be the orthogonal projection in  $L^2(X, \mu)$  of  $f$  on the closed subspace of  $\mathcal{A}$ -measurable functions. We most often apply this when  $\mathcal{A}$  is the  $\sigma$ -algebra of  $T$ -invariant sets, which we denote  $\mathcal{I}$ .

We also need to condition on a factor. Let  $(Y, \nu, S)$  be a measurable factor of  $(X, \mu, T)$  via a measurable factor map  $\pi$ . Put  $\mathcal{A} = \{\pi^{-1}(F) : F \subset Y \text{ Borel}\}$ . Given  $f \in L^2(X, \mu)$  we write  $\mathbb{E}(f | Y)$  for the conditional expectation  $\mathbb{E}(f | \mathcal{A})$ . and with the usual identification we may think of  $\mathbb{E}(f | Y)$  as an element of  $L^2(Y, \nu)$ . The following properties of the conditional expectation are standard and are made use of occasionally.

- $\mathbb{E}(Tf | Y) = \mathbb{E}(f | Y) \circ T$  for every  $f \in L^2(X, \mu)$ .
- $\int_X \mathbb{E}(f | Y) \cdot \mathbb{E}(g | Y) d\mu = \int_X \mathbb{E}(f | Y) \cdot g d\mu$  for all  $f, g \in L^2(X, \mu)$ .

We refer to [9, Chapter 5] for further background.

*Disintegrations.* We make significant use of disintegrations of measures over factors. Fix a system  $(X, \mu, T)$  and a probability space  $(\Omega, \nu)$ . Whenever we have a map  $\omega \mapsto \mu_\omega$  from  $\Omega$  to  $\mathcal{M}(X)$  with the properties

- $\omega \mapsto \mu_\omega(F)$  is measurable for every Borel  $F \subset X$
- $\int_X f d\mu = \int_\Omega \int_X f d\mu_\omega d\nu(\omega)$  for all measurable and bounded  $f: X \rightarrow \mathbb{C}$

we call  $\omega \mapsto \mu_\omega$  a *disintegration* of  $\mu$ . The following result says every measurable factor map gives rise to a disintegration that agrees with the conditional expectation on the factor.

**Theorem 2.5** (Disintegrations over factor maps, see for example [9, Theorem 5.14]). *If  $\pi: (X, \mu, T) \rightarrow (Y, \nu, S)$  is a measurable factor map, then there exists a disintegration  $y \mapsto \mu_y$  of  $\mu$  defined on  $(Y, \nu)$  such that*

$$\mathbb{E}(f | Y)(y) = \int_X f d\mu_y$$

*for  $\nu$ -almost every  $y \in Y$ , whenever  $f: X \rightarrow \mathbb{C}$  is measurable and bounded. Furthermore, if  $y \mapsto \eta_y$  is another such disintegration then  $\eta_y = \mu_y$  for  $\nu$ -almost every  $y \in Y$ .*

*Ergodicity and ergodic decompositions.* Given a topological system  $(X, T)$  we say that a  $T$ -invariant Borel probability measure  $\mu$  on  $X$  is *ergodic* if every Borel set  $D$  satisfying  $T^{-1}(D) = D$  has  $\mu(D) \in \{0, 1\}$ . A system  $(X, \mu, T)$  is *ergodic* if  $\mu$  is ergodic for the topological system  $(X, T)$ .

We make use of two versions of the pointwise ergodic theorem. Assume  $(X, \mu, T)$  is a system. Write  $\mathcal{I}$  for the  $\sigma$ -algebra of  $T$ -invariant Borel sets. For every  $f$  in  $L^1(X, \mu)$  the convergence

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N f(T^n x) = \mathbb{E}(f \mid \mathcal{I})(x)$$

holds for  $\mu$ -almost every  $x \in X$ . Moreover (see, for example [24] Section 1]), every Følner sequence has a subsequence  $\Phi$  with the following property: for every  $f$  in  $L^1(X, \mu)$  the convergence

$$\lim_{N \rightarrow \infty} \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} f(T^n x) = \mathbb{E}(f \mid \mathcal{I})(x)$$

holds for  $\mu$ -almost every  $x \in X$ . We make use of the following standard corollary of the pointwise ergodic theorem.

**Corollary 2.6.** *Assume  $(X, \mu, T)$  is an ergodic system. Every Følner sequence has a subsequence  $\Phi$  such that  $\mu(\text{gen}(\mu, \Phi)) = 1$ .*

There is a sense in which every system can be decomposed into ergodic systems.

**Definition 2.7** (Ergodic decomposition). Let  $(X, \mu, T)$  be a system. An *ergodic decomposition* of  $\mu$  is a disintegration  $x \mapsto \mu_x$  of  $\mu$  defined on  $(X, \mu)$  such that for every measurable and bounded  $f: X \rightarrow \mathbb{C}$ ,

$$\int_X f \, d\mu_x = \mathbb{E}(f \mid \mathcal{I})(x)$$

holds for  $\mu$ -almost every  $x \in X$ .

**Theorem 2.8** (Existence of ergodic decompositions, see for example [9] Theorem 6.2]). *Every system  $(X, \mu, T)$  has an ergodic decomposition  $x \mapsto \mu_x$ , satisfying the following properties.*

- If  $x \mapsto \nu_x$  is another ergodic decomposition of  $(X, \mu, T)$ , then  $\mu_x = \nu_x$  for  $\mu$ -almost every  $x \in X$ .
- We have

$$\int_X f \, d\mu = \int_X \left( \int_X f \, d\mu_x \right) d\mu(x)$$

for every bounded and measurable  $f: X \rightarrow \mathbb{C}$ .

- For  $\mu$ -almost every  $x \in X$ , the measure  $\mu_x$  is  $T$ -invariant and ergodic.

Corollary 2.9 is an immediate consequence of the pointwise ergodic theorem and Theorem 2.8

**Corollary 2.9.** *Fix an ergodic decomposition  $x \mapsto \mu_x$  of a system  $(X, \mu, T)$ . Then  $\mu$ -almost every  $x \in X$  is generic for  $\mu_x$ .*

*Furstenberg's correspondence principle.* The proof of our main theorem uses a general method for transferring combinatorial problems into dynamical ones known as *Furstenberg's correspondence principle*. It first appeared in Furstenberg's ergodic proof of Szemerédi's theorem [12][13], and has since formed the basis for numerous applications of ergodic theory to combinatorics and number theory. For our purposes we need a variant of Furstenberg's original method, which we state and prove (see, for example, [2] Proposition 3.1]).



**Theorem 2.10.** *Let  $A \subset \mathbb{N}$  and suppose  $\Phi$  is a Følner sequence in  $\mathbb{N}$  such that the limit*

$$(2.1) \quad \delta = \lim_{N \rightarrow \infty} \frac{|A \cap \Phi_N|}{|\Phi_N|}$$

*exists. There exist an ergodic system  $(X, \mu, T)$ , a clopen set  $E \subset X$ , a Følner sequence  $\Psi$  in  $\mathbb{N}$ , and a point  $a \in \text{gen}(\mu, \Psi)$  such that  $\mu(E) \geq \delta$  and*

$$(2.2) \quad A = \{n \in \mathbb{N} : T^n a \in E\}.$$

*Proof.* Let  $X = \{0, 1\}^{\mathbb{Z}}$  denote the set of all bi-infinite binary sequences, endowed with the product topology, and write  $T: \{0, 1\}^{\mathbb{Z}} \rightarrow \{0, 1\}^{\mathbb{Z}}$  for the shift map  $(Tx)(n) = x(n+1)$ . We can associate to the subset  $A \subset \mathbb{N}$  a point  $a \in \{0, 1\}^{\mathbb{Z}}$  that represents  $A$  via

$$a(n) = \begin{cases} 1 & \text{if } n \in A, \\ 0 & \text{otherwise.} \end{cases}$$

The set  $E = \{x \in X : x(0) = 1\}$  is a clopen subset of  $X$  and, by construction, (2.2) is satisfied. Let  $\mu'$  be any weak\* accumulation point of the sequence

$$N \mapsto \mu_N = \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} \delta_{T^n a}$$

of measures, where  $\delta_x$  denotes the Dirac measure at  $x$ . Observe that

$$\mu_N(E) = \frac{|A \cap \Phi_N|}{|\Phi_N|}$$

and hence  $\mu'(E) = \delta$  by (2.1). Although  $\mu'$  is not necessarily ergodic with respect to  $T$ , since it is  $T$ -invariant and by the ergodic decomposition there exists an ergodic  $T$ -invariant Borel probability measure  $\mu$  on  $X$  with  $\mu(E) \geq \mu'(E) = \delta$ . Without loss of generality, we can assume that  $\mu$  is supported on the orbit closure of  $a$  because  $\mu'$  is. By [13, Proposition 3.9], there exists a Følner sequence  $\Psi$  in  $\mathbb{N}$  such that  $a \in \text{gen}(\mu, \Psi)$ , finishing the proof.  $\square$

### 3. A DISCUSSION OF $B_1 + B_2 \subset A$

In this section we present our proof of Theorem 1.1 in the case  $k = 2$ . This section is not logically necessary for the full proof of Theorem 1.1 but we hope this separate, unadorned treatment of our argument clearly introduces the main ideas behind our approach as preparation for the proof in the general case. For convenience, let us formulate the statement that we aim to prove as a separate theorem.

**Theorem 3.1.** *If  $A \subset \mathbb{N}$  has positive upper Banach density then there are infinite sets  $B_1, B_2 \subset \mathbb{N}$  with  $B_1 + B_2 \subset A$ .*

The first step in our proof of Theorem 3.1 is to recast the problem in dynamical terms using Theorem 2.10. Doing so paves the way for the language and tools of ergodic theory on which our argument relies.

**Theorem 3.2** (Dynamical reformulation). *Let  $(X, \mu, T)$  be an ergodic system and let  $E \subset X$  be a clopen set with  $\mu(E) > 0$ . If  $a \in \text{gen}(\mu, \Phi)$  for some Følner sequence  $\Phi$ , then there exist strictly increasing sequences  $c_1, c_2: \mathbb{N} \rightarrow \mathbb{N}$  such that*

$$(3.1) \quad \lim_{j \rightarrow \infty} \lim_{m \rightarrow \infty} T^{c_1(j) + c_2(m)} a \in E \quad \text{and} \quad \lim_{m \rightarrow \infty} \lim_{j \rightarrow \infty} T^{c_1(j) + c_2(m)} a \in E.$$



*Proof that Theorem 3.2 and Theorem 3.1 are equivalent.* First we deduce Theorem 3.2 from Theorem 3.1. Fix an ergodic system  $(X, \mu, T)$ , a clopen set  $E$  with  $\mu(E) > 0$ , and  $a \in \text{gen}(\mu, \Phi)$  for some Følner sequence  $\Phi$ . The assumption that  $a$  is generic for  $\mu$  along  $\Phi$  lets us deduce that the set

$$\{n \in \mathbb{N} : T^n a \in E\}$$

has positive upper Banach density. It then follows from Theorem 3.1 that there are infinite sets  $B_1, B_2 \subset \mathbb{N}$  with  $B_1 + B_2 \subset \{n \in \mathbb{N} : T^n a \in E\}$ . We draw from these sets strictly increasing sequences  $c_1, c_2 : \mathbb{N} \rightarrow \mathbb{N}$  with  $T^{c_1(j)+c_2(m)} a \in E$  for all  $j, m \in \mathbb{N}$  so that, after passing to suitable subsequences such that all the limits involved exist, (3.1) follows immediately.

To deduce Theorem 3.1 from Theorem 3.2 fix  $A \subset \mathbb{N}$  with positive upper Banach density. Theorem 2.10 yields an ergodic system  $(X, \mu, T)$ , a point  $a \in \text{gen}(\mu, \Phi)$  for some Følner sequence  $\Phi$ , and a clopen set  $E$  with  $\mu(E) > 0$  such that

$$\{n \in \mathbb{Z} : T^n a \in E\} = A.$$

It follows from Theorem 3.2 that there are strictly increasing sequences  $c_1, c_2 : \mathbb{N} \rightarrow \mathbb{N}$  such that (3.1) holds.

We now apply alternately the limits in (3.1) to construct subsequences  $b_1, b_2$  of  $c_1, c_2$  such that  $b_1(j) + b_2(m) \in \{n \in \mathbb{N} : T^n a \in E\}$  for all  $j, m \in \mathbb{N}$ . Begin by selecting  $b_1(1)$  in  $\{c_1(j) : j \in \mathbb{N}\}$  such that

$$\lim_{m \rightarrow \infty} T^{b_1(1)+c_2(m)} a \in E.$$

Then let  $b_2(1)$  from  $\{c_2(m) : m \in \mathbb{N}\}$  satisfy  $b_1(1) + b_2(1) \in \{n \in \mathbb{N} : T^n a \in E\}$  and

$$\lim_{j \rightarrow \infty} T^{c_1(j)+b_2(1)} a \in E.$$

Next, we take  $b_1(2) > b_1(1)$  to be any element in  $\{c_1(n) : n \in \mathbb{N}\}$  with the property that  $b_1(2) + b_2(1) \in \{n \in \mathbb{N} : T^n a \in E\}$  and

$$\lim_{m \rightarrow \infty} T^{b_1(2)+c_2(m)} a \in E.$$

Thereafter, we take  $b_2(2)$  from  $\{c_2(n) : n \in \mathbb{N}\}$  with  $b_1(1) + b_2(2), b_1(2) + b_2(2) \in \{n \in \mathbb{N} : T^n a \in E\}$  and

$$\lim_{j \rightarrow \infty} T^{c_1(j)+b_2(2)} a \in E.$$

Continuing this procedure by induction yields strictly increasing sequences  $b_1, b_2 : \mathbb{N} \rightarrow \mathbb{N}$  with  $T^{b_1(j)+b_2(m)} a$  belonging to  $E$  for all  $j, m \in \mathbb{N}$ . Since  $A = \{n \in \mathbb{Z} : T^n a \in E\}$ , we conclude  $b_1(j) + b_2(m) \in A$  for all  $j, m \in \mathbb{N}$ , finishing the proof.  $\square$

*Remark 3.3.* When  $a \in \text{supp}(\mu)$  there is a strictly increasing sequence  $b : \mathbb{N} \rightarrow \mathbb{N}$  and some  $c \in \mathbb{N}$  with

$$\{b(i_1) + \cdots + b(i_n) + c : i_1 < \cdots < i_n \in \mathbb{N}, n \in \mathbb{N}\} \subset A,$$

which is to say that  $A$  contains a shift of an IP set, a significantly stronger conclusion than that of Theorem 3.1. We stress that there is no reason to expect that  $a$  belongs to  $\text{supp}(\mu)$ , and in fact it is known that not every set with positive upper density contains a shift of an IP set. In such cases, necessarily  $a \notin \text{supp}(\mu)$ , and this introduces new complications in our proofs.

The remainder of this section is dedicated to proving Theorem 3.2 using the techniques of ergodic theory. In fact, we prove a stronger result: under the hypotheses of Theorem 3.2 there exist strictly increasing sequences  $c_1, c_2: \mathbb{N} \rightarrow \mathbb{N}$  such that the double limits in (3.1) not only belong to  $E$  but are also equal to one another, that is,

$$(3.2) \quad \lim_{j \rightarrow \infty} \lim_{m \rightarrow \infty} T^{c_1(j)+c_2(m)} a = \lim_{m \rightarrow \infty} \lim_{j \rightarrow \infty} T^{c_1(j)+c_2(m)} a.$$

To better analyze this transposition of limits, it is convenient to keep track of the intermediate limit points. For this purpose, when (3.2) holds we introduce the notation

$$\begin{aligned} x_{00} &= a, \\ x_{01} &= \lim_{m \rightarrow \infty} T^{c_2(m)} a, \\ x_{10} &= \lim_{j \rightarrow \infty} T^{c_1(j)} a, \\ x_{11} &= \lim_{j \rightarrow \infty} \lim_{m \rightarrow \infty} T^{c_1(j)+c_2(m)} a = \lim_{m \rightarrow \infty} \lim_{j \rightarrow \infty} T^{c_1(j)+c_2(m)} a. \end{aligned}$$

In terms of this notation (3.2) is equivalent to the following identities:

$$(3.3) \quad \begin{aligned} \lim_{j \rightarrow \infty} (T \times T)^{c_1(j)}(x_{00}, x_{01}) &= (x_{10}, x_{11}), \\ \lim_{m \rightarrow \infty} (T \times T)^{c_2(m)}(x_{00}, x_{10}) &= (x_{01}, x_{11}). \end{aligned}$$

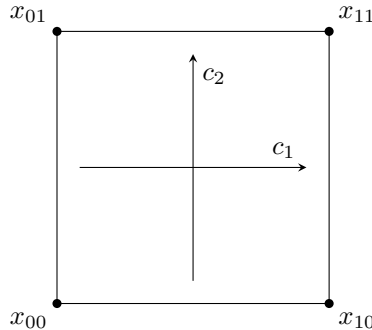
Following the notation introduced in [18], we write  $\llbracket 1 \rrbracket = \{0, 1\}$  and  $\llbracket 2 \rrbracket = \{00, 01, 10, 11\}$ . Tuples  $(x_{00}, x_{01}, x_{10}, x_{11}) \in X^{\llbracket 2 \rrbracket}$  for which (3.3) holds are central to our approach and we formalize them as follows.

**Definition 3.4** (2-Dimensional Erdős cubes). Given a topological system  $(X, T)$  we call any point  $x = (x_{00}, x_{01}, x_{10}, x_{11}) \in X^{\llbracket 2 \rrbracket}$  satisfying (3.3) for some strictly increasing sequences  $c_1, c_2: \mathbb{N} \rightarrow \mathbb{N}$  a *2-dimensional Erdős cube*.

Observe that Definition 3.4 matches Definition 1.2. Indeed, (3.3) is equivalent to the assertion that the forward orbit of  $(x_{00}, x_{01})$  visits every neighborhood of  $(x_{10}, x_{11})$  infinitely often, and that the forward orbit of  $(x_{00}, x_{10})$  visits every neighborhood of  $(x_{01}, x_{11})$  infinitely often. This can also be written as

$$(3.4) \quad (x_{10}, x_{11}) \in \omega((x_{00}, x_{01}), T \times T) \quad \text{and} \quad (x_{01}, x_{11}) \in \omega((x_{00}, x_{10}), T \times T).$$

We think of the coordinates  $x_{00}, x_{01}, x_{10}, x_{11}$  in a 2-dimensional Erdős cube as forming the vertices of a square



and the sequences  $c_1, c_2: \mathbb{N} \rightarrow \mathbb{N}$  as describing iterations of  $T \times T$  at which the left and bottom sides approximate the right and top sides respectively.

We can now state the main dynamical theorem of this section, which is a strengthening of Theorem 3.2

**Theorem 3.5** (Existence of 2-dimensional Erdős cubes). *Assume  $(X, \mu, T)$  is an ergodic system,  $a \in \text{gen}(\mu, \Phi)$  for some Følner sequence  $\Phi$ , and  $E \subset X$  is an open set with  $\mu(E) > 0$ . There exists a 2-dimensional Erdős cube  $(x_{00}, x_{01}, x_{10}, x_{11}) \in X^{[2]}$  with  $x_{00} = a$  and  $x_{11} \in E$ .*

Although it can already be inferred from our discussion, we include a quick and self-contained proof of the fact that Theorem 3.5 implies Theorem 3.2

*Proof that Theorem 3.5 implies Theorem 3.2* If  $(x_{00}, x_{01}, x_{10}, x_{11})$  is a 2-dimensional Erdős cube then, by definition, there exist increasing sequences  $c_1, c_2: \mathbb{N} \rightarrow \mathbb{N}$  such that

$$\lim_{n \rightarrow \infty} T^{c_1(n)} x_{00} = x_{10}, \quad \text{and} \quad \lim_{m \rightarrow \infty} T^{c_2(m)} x_{10} = x_{11}.$$

Since  $x_{00} = a$  and  $x_{11} \in E$ , we conclude that

$$\lim_{m \rightarrow \infty} \lim_{n \rightarrow \infty} T^{c_1(n) + c_2(m)} a \in E.$$

Similarly, we have

$$\lim_{m \rightarrow \infty} T^{c_2(m)} x_{00} = x_{01} \quad \text{and} \quad \lim_{n \rightarrow \infty} T^{c_1(n)} x_{01} = x_{11},$$

which implies

$$\lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} T^{c_1(n) + c_2(m)} a \in E.$$

This completes the proof of (3.1).  $\square$

Before embarking on the proof of Theorem 3.5 which involves finding an Erdős cube with prescribed first and last coordinates, we explain why there is a natural abundance of Erdős cubes when these extra restrictions are omitted. A key role is played by the *cubic measures*, which are special measures introduced in [18] to analyze multiple ergodic averages. We recall the definition of one- and two-dimensional cubic measures (the definition for higher dimensions is given in Section 5.1).

**Definition 3.6** (1-Dimensional cubic measure). Let  $(X, \mu, T)$  be an ergodic system. The measure  $\mu^{[1]} = \mu \times \mu$  on  $X^{[1]}$  is the *1-dimensional cubic measure* of  $(X, \mu, T)$ . It gives rise to the *1-dimensional cube system*  $(X^{[1]}, \mu^{[1]}, T^{[1]})$ , where  $T^{[1]} = T \times T$ .

Fix an ergodic decomposition  $y \mapsto (\mu^{[1]})_y$  of  $\mu^{[1]}$  using Theorem 2.8. As stipulated in Section 2 we view ergodic decompositions as disintegrations over the invariant factor, which means  $y \mapsto (\mu^{[1]})_y$  is a  $\mu^{[1]}$ -almost everywhere defined map on  $X^{[1]}$  satisfying

$$(3.5) \quad \mu^{[1]} = \int_{X^{[1]}} (\mu^{[1]})_y d\mu^{[1]}(y),$$

and there is a full  $\mu^{[1]}$ -measure set of  $y \in X^{[1]}$  for which  $(\mu^{[1]})_y$  is  $T^{[1]}$ -invariant and ergodic and  $y$  is generic for  $(\mu^{[1]})_y$ .

**Definition 3.7** (2-Dimensional cubic measure). Given an ergodic system  $(X, \mu, T)$  and an ergodic decomposition  $y \mapsto (\mu^{[1]})_y$  of  $\mu^{[1]}$ , the *2-dimensional cubic measure* of  $(X, \mu, T)$ , denoted by  $\mu^{[2]}$ , is the measure on  $X^{[2]}$  defined by

$$(3.6) \quad \mu^{[2]} = \int_{X^{[1]}} (\mu^{[1]})_y \times (\mu^{[1]})_y d\mu^{[1]}(y).$$

Writing  $T^{[2]} = T \times T \times T \times T$ , the system  $(X^{[2]}, \mu^{[2]}, T^{[2]})$  is the *2-dimensional cube system* associated to  $(X, \mu, T)$ .

In general, the measure  $\mu^{[2]}$  is not equal to the product measure  $\mu^{[1]} \times \mu^{[1]}$ , but it is still invariant under the diagonal transformation  $T^{[2]} = T \times T \times T \times T$ . Particularly important to our discussion is the fact that  $\mu^{[2]}$  possesses special symmetries. Consider the permutation on the set  $[2]$  given by  $00 \mapsto 00$ ,  $01 \mapsto 10$ ,  $10 \mapsto 01$ , and  $11 \mapsto 11$ . This permutation naturally induces a map  $\phi: X^{[2]} \rightarrow X^{[2]}$  defined by

$$(3.7) \quad \phi(x_{00}, x_{01}, x_{10}, x_{11}) = (x_{00}, x_{10}, x_{01}, x_{11}).$$

One can show that  $\phi$  is always an automorphism of the 2-dimensional cube system  $(X^{[2]}, \mu^{[2]}, T^{[2]})$ , which means that  $\phi$  commutes with  $T^{[2]}$  and preserves the measure  $\mu^{[2]}$ , i.e.,

$$(3.8) \quad \phi(\mu^{[2]}) = \mu^{[2]}.$$

A proof of this *permutation symmetry* of  $\mu^{[2]}$  is given in [20, Proposition 8.8].

Next, we describe the connection between cubic measures and Erdős cubes. We begin with the combinatorially uninteresting case of 1-dimensional Erdős cubes. These are the points  $x \in X^{[1]}$  for which there is a strictly increasing sequence  $c_1: \mathbb{N} \rightarrow \mathbb{N}$  with

$$\lim_{n \rightarrow \infty} T^{c_1(n)}(x_0) = x_1.$$

In view of Lemma 2.4 if  $x_0 \in \text{gen}(\mu, \Phi)$  and  $x_1 \in \text{supp}(\mu)$  then the pair  $(x_0, x_1)$  is a 1-dimensional Erdős cube. After passing to a subsequence of  $\Phi$  if necessary, the set  $\text{gen}(\mu, \Phi) \times \text{supp}(\mu)$  has full measure with respect to  $\mu^{[1]}$  by Corollary 2.6. This proves that  $\mu^{[1]}$ -almost every point in  $X^{[1]}$  is a 1-dimensional Erdős cube.

The extra symmetry provided by (3.8) allows us to supplement the previous argument to produce 2-dimensional Erdős cubes.

**Proposition 3.8.** *Let  $(X, \mu, T)$  be an ergodic system. Then  $\mu^{[2]}$ -almost every  $x \in X^{[2]}$  is a 2-dimensional Erdős cube.*

*Proof.* Fix an ergodic decomposition  $y \mapsto (\mu^{[1]})_y$  of  $\mu^{[1]}$ . Since  $(\mu^{[1]})_y$  is ergodic for  $\mu^{[1]}$ -almost every  $y \in X^{[1]}$  we can repeat the previous argument for finding 1-dimensional Erdős cubes to show that  $(\mu^{[1]})_y \times (\mu^{[1]})_y$  gives full measure to

$$(3.9) \quad \left\{ (x_{00}, x_{01}, x_{10}, x_{11}) \in X^{[2]} : (x_{10}, x_{11}) \in \omega((x_{00}, x_{01}), T \times T) \right\}$$

for  $\mu^{[1]}$ -almost every  $y \in X^{[1]}$ . It is then immediate from (3.6) that  $\mu^{[2]}$  gives full measure to (3.9). We now use (3.8). It follows that (3.9) and its inverse image under  $\phi$  both have full measure with respect to  $\mu^{[2]}$ . Since  $x \in X^{[2]}$  is a 2-dimensional Erdős cube if and only if  $x$  and  $\phi(x)$  belong to (3.9), we conclude that  $\mu^{[2]}$ -almost every point is a 2-dimensional Erdős cube.  $\square$

We now embark on the discussion of the more delicate issue of producing Erdős cubes with prescribed first coordinate. Once this is in hand, the proof of Theorem 3.5 is straightforward.

In view of Proposition 3.8 we can attempt to obtain Erdős cubes with a prescribed first coordinate by disintegrating  $\mu^{[2]}$  with respect to the projection  $(x_{00}, x_{01}, x_{10}, x_{11}) \mapsto x_{00}$  from  $X^{[2]}$  to its first coordinate. Since the coordinate projection  $x \mapsto x_{00}$  is a measurable factor map from  $(X^{[2]}, \mu^{[2]}, T^{[2]})$  to  $(X, \mu, T)$ , we can apply Theorem 2.5 to obtain a disintegration of  $\mu^{[2]}$  over this map: a collection of measures  $\sigma_t^{[2]}$  on  $X^{[2]}$  defined for  $\mu$ -almost every  $t \in X$ . Any such disintegration satisfies

$$\mu^{[2]} = \int_X \sigma_t^{[2]} d\mu(t),$$

which, combined with Proposition 3.8, shows  $\sigma_t^{[2]}$ -almost every point is a 2-dimensional Erdős cube for  $\mu$ -almost every  $t \in X$ . In fact

$$\sigma_t^{[2]}(\{x \in X^{[2]} : x_{00} = t\}) = 1$$

for  $\mu$ -almost every  $t$ , and we conclude that for  $\mu$ -almost every  $t \in X$  there are 2-dimensional Erdős cubes  $x \in X^{[2]}$  with  $x_{00} = t$ . It can also be shown that the push forward of  $\sigma_t^{[2]}$  to the *last* coordinate via the map  $x \mapsto x_{11}$  is equal to  $\mu$  for  $\mu$ -almost every  $t$ , and therefore one can find for  $\mu$ -almost every  $t \in X$  an Erdős cube  $x$  with  $x_{00} = t$  and  $x_{11} \in E$ .

Unfortunately, this argument is not enough to prove Theorem 3.5 because Theorem 2.5 only defines  $\sigma_t^{[2]}$  for  $\mu$ -almost every  $t \in X$ : the measure  $\sigma_a^{[2]}$  might not be defined. Thus, to proceed this way we must first define  $\sigma_t^{[2]}$  for *every* point  $t \in X$ . We are able to do this when the system  $(X, \mu, T)$  admits a continuous factor map to a certain measurable factor of the ergodic system  $(X, \mu, T)$  called its Kronecker factor. Informally, it is the largest factor of  $(X, \mu, T)$  that is isomorphic to a rotation on a compact abelian group.

**Definition 3.9** (Group rotation). A system  $(Z, m, R)$  is a *group rotation* if  $Z$  is a compact abelian group,  $m$  is its Haar measure, and the transformation  $R$  has the form  $R(z) = z + \alpha$  for some fixed  $\alpha \in Z$ .

**Theorem 3.10.** *Fix an ergodic system  $(X, \mu, T)$ . There is an ergodic group rotation  $(Z, m, R)$  that is a measurable factor of  $(X, \mu, T)$  with the property that, for every  $f, g$  in  $L^\infty(X, \mu)$ , we have*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N f(T^n x_0) \cdot g(T^n x_1) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N T^n \mathbb{E}(f \mid Z)(x_0) \cdot T^n \mathbb{E}(g \mid Z)(x_1)$$

for  $\mu \times \mu$ -almost every  $(x_0, x_1) \in X \times X$ .

*Proof.* This is a consequence of the pointwise ergodic theorem applied to  $T \times T$  together with the description [13] Lemma 4.21] of the projection onto the invariant factor in a product system.  $\square$

We refer to the group rotation factor  $(Z, m, R)$  of  $(X, \mu, T)$  in Theorem 3.10 as the *Kronecker factor* of  $(X, \mu, T)$ . Theorem 3.10 gives only a measurable factor map to the Kronecker factor, but to define  $\sigma_t^{[2]}$  for all  $t \in X$  it turns out that we need this factor map to be continuous. Fortunately, we can always pass to an extension

of our ergodic system having that property. We prove this at the end of the section, continuing for the moment under the assumption that  $(X, \mu, T)$  has its Kronecker factor  $(Z, m, R)$  as a continuous factor via a continuous factor map  $\pi$ . Under this assumption we are able to improve upon Theorem 2.8: we give a disintegration of  $\mu^{[1]}$  that is defined on all of  $X^{[1]}$ , continuous, and an ergodic decomposition of  $\mu^{[1]}$ . This result is the main technical step in the proof of Theorem 3.5.

**Proposition 3.11.** *Let  $(X, \mu, T)$  be an ergodic system and assume there is a continuous factor map  $\pi$  to its Kronecker factor  $(Z, m, R)$ . There exists a continuous map  $x \mapsto \lambda_x^{[1]}$  from  $X^{[1]}$  to  $\mathcal{M}(X^{[1]})$  with the following properties.*

- (i) *The map  $x \mapsto \lambda_x^{[1]}$  is an ergodic decomposition of  $\mu^{[1]}$  as in Definition 2.7.*
- (ii) *We have that  $\lambda_x^{[1]} = \lambda_y^{[1]}$  whenever  $(\pi(x_0), \pi(x_1)) = (\pi(y_0), \pi(y_1))$ .*
- (iii) *For all  $x \in X^{[1]}$ , we have that  $T^{[1]} \lambda_x^{[1]} = \lambda_{T^{[1]}(x)}^{[1]} = \lambda_x^{[1]}$ .*
- (iv) *For all  $x \in X$  and all Borel  $F \subset X$ , we have that  $\lambda_x^{[1]}(X \times F) = \mu(F)$ .*

*Remark 3.12.* Note that any ergodic decomposition of  $\mu^{[1]}$  satisfies properties (ii) through (iv) for almost all  $x, y \in X^{[1]}$ ; the fact that  $x \mapsto \lambda_x^{[1]}$  satisfies these properties for all  $x, y \in X^{[1]}$  is what makes it special. We refer to  $x \mapsto \lambda_x^{[1]}$  as a *continuous ergodic decomposition*, and discuss the general case in Section 6.

*Proof of Proposition 3.11.* Apply Theorem 2.5 to get a disintegration  $z \mapsto \eta_z$  of  $\mu$  over the continuous factor map  $\pi$  from  $(X, \mu, T)$  to its Kronecker factor  $(Z, m, R)$ . Define

$$(3.10) \quad \lambda_x^{[1]} = \int_Z \eta_{z+\pi(x_0)} \times \eta_{z+\pi(x_1)} \, dm(z)$$

for every  $x \in X^{[1]}$ .

We first note that for each  $x \in X^{[1]}$  the measures  $\eta_{z+\pi(x_0)}$  and  $\eta_{z+\pi(x_1)}$  are defined for  $m$ -almost every  $z \in Z$  and therefore (3.10) is well-defined. To prove that  $x \mapsto \lambda_x^{[1]}$  is continuous first note that uniform continuity implies

$$(v, w) \mapsto \int_Z f(z+w) \cdot g(z+v) \, dm(z)$$

from  $Z$  to  $\mathbb{C}$  is continuous whenever  $f, g: Z \rightarrow \mathbb{C}$  are continuous. An approximation argument then gives continuity for every  $f, g \in L^2(Z, m)$ . In particular,

$$x \mapsto \int_Z \mathbb{E}(f \mid Z)(z + \pi(x_0)) \cdot \mathbb{E}(g \mid Z)(z + \pi(x_1)) \, dm(z)$$

from  $X^{[1]}$  to  $\mathbb{C}$  is continuous whenever  $f, g: X \rightarrow \mathbb{C}$  are continuous, which in turn implies continuity of (3.10).

To prove that  $x \mapsto \lambda_x^{[1]}$  is an ergodic decomposition we calculate

$$\begin{aligned} & \int_{X^{[1]}} \int_Z \eta_{z+\pi(x_0)} \times \eta_{z+\pi(x_1)} \, dm(z) \, d\mu^{[1]}(x) \\ &= \int_Z \int_X \eta_{z+\pi(x_0)} \, d\mu(x_0) \times \int_X \eta_{z+\pi(x_1)} \, d\mu(x_1) \, dm(z), \end{aligned}$$

which is equal to  $\mu^{[1]}$  because both inner integrals are equal to  $\mu$ . We conclude that

$$(3.11) \quad \mu^{[1]} = \int_{X^{[1]}} \lambda_x^{[1]} \, d\mu^{[1]}(x),$$

which shows  $x \mapsto \lambda_x^{[1]}$  is a disintegration of  $\mu^{[1]}$ .

We are left with verifying that

$$\int_{X^{[1]}} F d\lambda_x^{[1]} = \mathbb{E}(F | \mathcal{I})(x)$$

for  $\mu^{[1]}$ -almost every  $x \in X^{[1]}$  whenever  $F: X^{[1]} \rightarrow \mathbb{C}$  is measurable and bounded. Recall that  $\mathcal{I}$  denotes the  $\sigma$ -algebra of  $T^{[1]}$ -invariant sets. Fix such an  $F$ . It follows from the pointwise ergodic theorem that

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N F(T^n x_0, T^n x_1) = \mathbb{E}(F | \mathcal{I})(x)$$

for  $\mu^{[1]}$ -almost every  $x = (x_0, x_1) \in X^{[1]}$ . We therefore wish to prove that

$$\int_{X^{[1]}} F d\lambda_x^{[1]} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N F(T^n x_0, T^n x_1)$$

holds for  $\mu^{[1]}$ -almost every  $x \in X^{[1]}$ .

By an approximation argument it suffices to verify that

$$(3.12) \quad \int_{X^{[1]}} f \otimes g d\lambda_x^{[1]} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N f(T^n x_0) \cdot g(T^n x_1)$$

holds for  $\mu^{[1]}$ -almost every  $x \in X^{[1]}$  whenever  $f, g$  belong to  $L^\infty(X, \mu)$ .

In fact, in view of Theorem 3.10 only the conditional expectations of  $f$  and  $g$  on the Kronecker factor contribute to the right-hand side. We therefore have for any  $f, g \in L^2(X, \mu)$  that

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N f(T^n x_0) \cdot g(T^n x_1) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mathbb{E}(f | Z)(T^n x_0) \cdot \mathbb{E}(g | Z)(T^n x_1)$$

for  $\mu^{[1]}$ -almost every  $x \in X^{[1]}$ . Now

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \phi(T^n z_0) \cdot \psi(T^n z_1) = \int_Z \phi(s + z_0) \cdot \psi(s + z_1) dm(s)$$

for any  $\phi, \psi$  in  $L^2(Z, m)$  using properties of rotations on a compact abelian group. Taking  $\phi = \mathbb{E}(f | Z)$  and  $\psi = \mathbb{E}(g | Z)$  gives

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mathbb{E}(f | Z)(T^n x_0) \cdot \mathbb{E}(g | Z)(T^n x_1) = \int_{X^{[1]}} f \otimes g d\lambda_x^{[1]}$$

for  $\mu^{[1]}$ -almost all  $x \in X^{[1]}$ . As a consequence we have for every  $f, g \in L^2(X, \mu)$  that

$$\int_{X^{[1]}} f \otimes g d\lambda_x^{[1]} = \mathbb{E}(f \otimes g | \mathcal{I})(x)$$

for  $\mu^{[1]}$ -almost every  $x \in X^{[1]}$ , proving property (i)

It is immediate from (3.10) that  $\pi(x_0) = \pi(y_0)$  and  $\pi(x_1) = \pi(y_1)$  together imply the measures  $\lambda_x^{[1]}$  and  $\lambda_y^{[1]}$  are equal, verifying property (ii)



Property (iii) follows from Theorem 2.8 and the continuity of  $x \mapsto \lambda_x^{[1]}$ . Finally, property (iii) follows from

$$\lambda_x^{[1]}(X \times F) = \int_Z \eta_{z+\pi(x_0)}(X) \cdot \eta_{z+\pi(x_1)}(F) \, dm(z) = \int_Z \eta_z(F) \, dm(z) = \mu(F)$$

because  $z \mapsto \eta_z$  is a disintegration of  $\mu$ .  $\square$

We next use Proposition 3.11 to define the measures  $\sigma_t^{[1]}$  directly for *every* point  $t \in X$ , dispensing with our direct application of Theorem 2.5

**Definition 3.13.** Fix an ergodic system  $(X, \mu, T)$  with a continuous factor map to its Kronecker factor and let  $\lambda^{[1]}: X^{[1]} \rightarrow \mathcal{M}(X^{[1]})$  be the disintegration given by Proposition 3.11. We define a measure  $\sigma_t^{[2]}$  on  $X^{[2]}$  by

$$(3.13) \quad \sigma_t^{[2]} = \int_X \delta_{(t,x)} \times \lambda_{(t,x)}^{[1]} \, d\mu(x)$$

for all  $t \in X$ .

For the rest of the section all references to measures  $\sigma_t^{[2]}$  are made with respect to Definition 3.13. The following properties of the measures  $\sigma_t^{[2]}$  are essential for producing Erdős cubes with prescribed coordinates. Writing  $F^*: X^{[2]} \rightarrow X^{[2]} \setminus \{00\}$  for the projection

$$(3.14) \quad F^*(x_{00}, x_{01}, x_{10}, x_{11}) = (x_{01}, x_{10}, x_{11}),$$

note in particular the last property, which allows us to make certain statements about  $\sigma_t^{[2]}$ -almost every point even when  $t$  does not lie in the support of  $\mu$ .

**Theorem 3.14.** Let  $(X, \mu, T)$  be an ergodic system and assume there is a continuous factor map  $\pi$  to its Kronecker factor. Then the map  $t \mapsto \sigma_t^{[2]}$  is continuous, gives full measure to  $\{x \in X^{[2]} : x_{00} = t\}$ , satisfies

$$(3.15) \quad \int_X \sigma_t^{[2]} \, d\mu(t) = \mu^{[2]}$$

and the following properties.

- (i) The push forward  $T^{[2]} \sigma_t^{[2]}$  equals  $\sigma_{T(t)}^{[2]}$  for every  $t \in X$ .
- (ii) The push forwards  $F^* \sigma_t^{[2]}$  and  $F^* \sigma_s^{[2]}$  are equal whenever  $\pi(t) = \pi(s)$ .

*Remark 3.15.* In analogy to Remark 3.12 let us point out that any decomposition of  $\mu^{[2]}$  with respect to the first coordinate satisfies properties (i) and (ii) for almost all  $t \in X$ . The fact that  $t \mapsto \sigma_t^{[2]}$  satisfies these properties for all  $t \in X$  is what distinguishes it from a generic disintegration.

*Proof of Theorem 3.14.* The fact that the set of points in  $X^{[2]}$  whose first coordinate equals  $t$  has full measure with respect to  $\sigma_t^{[2]}$  is immediate from (3.13). To

prove (3.15) we calculate

$$\begin{aligned}
\int_X \sigma_t^{[2]} d\mu(t) &= \int_X \int_X \delta_{(t,s)} \times \lambda_{(t,s)}^{[1]} d\mu(t) d\mu(s) \\
&= \int_{X^{[1]}} \delta_{(t,s)} \times \lambda_{(t,s)}^{[1]} d\mu^{[1]}(t, s) \\
&= \int_{X^{[1]}} \left( \int_{X^{[1]}} \delta_{(u,v)} \times \lambda_{(u,v)}^{[1]} d\lambda_{(t,s)}^{[1]}(u, v) \right) d\mu^{[1]}(t, s) \\
&= \int_{X^{[1]}} \left( \int_{X^{[1]}} \delta_{(u,v)} \times \lambda_{(t,s)}^{[1]} d\lambda_{(t,s)}^{[1]}(u, v) \right) d\mu^{[1]}(t, s) \\
&= \int_{X^{[1]}} \lambda_{(t,s)}^{[1]} \times \lambda_{(t,s)}^{[1]} d\mu^{[1]}(t, s) = \mu^{[2]},
\end{aligned}$$

where the fourth equality follows from

$$\lambda_{(t,s)}^{[1]}(\{(u, v) : \lambda_{(t,s)}^{[1]} = \lambda_{(u,v)}^{[1]}\}) = 1$$

for  $\mu^{[1]}$ -almost every  $(t, s)$ , which in turn is a consequence of part (iii) of Proposition 3.11 and ergodicity of  $\lambda_{(t,s)}^{[1]}$  for  $\mu^{[1]}$ -almost every  $(t, s)$ .  $\square$

We have reduced the proof of Theorem 3.5 to the statement that  $\sigma_a^{[2]}$ -almost every point is an Erdős cube. The proof of Proposition 3.8 made essential usage of the fact that  $\mu^{[2]}$  was symmetric with respect to the permutation map (3.7). Our penultimate ingredient in the proof of Theorem 3.5 is that the measures  $\sigma_t^{[2]}$  have the same symmetries.

To prove that *all* of the measures  $\sigma_t^{[2]}$  have the desired symmetries we need to be able to deduce that a property holds for all  $\sigma_t^{[2]}$  if it holds for  $\mu$ -almost every  $t \in X$ . It is not enough for us that the map  $t \mapsto \sigma_t^{[2]}$  is continuous, as we are particularly interested in  $\sigma_a^{[2]}$  and in the most challenging situations the point  $a$  does not belong to  $\text{supp}(\mu)$ . Indeed, when  $a$  does belong to the support of  $\mu$  one can prove Theorem 3.5 relatively easily as in that case one can verify  $(a, a, a, a)$  is a 2-dimensional Erdős cube. Fortunately, the projections  $F^* \sigma_t^{[2]}$  only depend on the value of  $\pi(t)$ . We thus are able to apply Lemma 3.16 to make deductions about  $\sigma_t^{[2]}$  for *all* points  $t \in X$ .

**Lemma 3.16.** *Let  $\pi: (X, \mu, T) \rightarrow (Y, \nu, S)$  be a continuous factor map and assume that  $\text{supp}(\nu) = Y$ . Then the only set  $F \subset X$  that is closed, has full measure and satisfies*

$$a \in F \text{ and } \pi(b) = \pi(a) \implies b \in F$$

*is  $F = X$ .*

*Proof.* Since  $F = \pi^{-1}(\pi(F))$  we conclude that the compact set  $\pi(F)$  has full measure. But  $\nu$  has full support, so  $\pi(F) = Y$ .  $\square$

Recall the maps  $\phi: X^{[2]} \rightarrow X^{[2]}$  of (3.7) and  $F^*$  of (3.14).

**Corollary 3.17.** *Let  $(X, \mu, T)$  be an ergodic system and assume there is a continuous factor map  $\pi$  to its Kronecker factor. For every  $t \in X$  the push forward  $\phi \sigma_t^{[2]}$  is equal to  $\sigma_t^{[2]}$ .*

*Proof.* By continuity, the set

$$H = \{t \in X : \phi\sigma_t^{[2]} = \sigma_t^{[2]}\}$$

is closed. We wish to prove it has full measure. To do so we show that  $t \mapsto \phi\sigma_t^{[2]}$  is a disintegration of  $\mu^{[2]}$  and appeal to uniqueness in Theorem 2.5. To see this, calculate

$$\begin{aligned} & \int_X \int_{X^{[2]}} f_{00} \otimes f_{01} \otimes f_{10} \otimes f_{11} \, d\phi\sigma_t^{[2]} \, d\mu(t) \\ &= \int_X \int_{X^{[2]}} f_{00} \otimes f_{10} \otimes f_{01} \otimes f_{11} \, d\sigma_t^{[2]} \, d\mu(t) \\ &= \int_{X^{[2]}} f_{00} \otimes f_{10} \otimes f_{01} \otimes f_{11} \, d\mu^{[2]}(t) \end{aligned}$$

and note that  $\phi\mu^{[2]} = \mu^{[2]}$  by (3.8). Since the map  $\phi$  commutes with the projection  $(x_{00}, x_{01}, x_{10}, x_{11}) \mapsto x_{00}$  onto the first coordinate we certainly have

$$(\phi\sigma_t^{[2]})(\{x \in X^{[2]} : x_{00} = t\}) = 1$$

for every  $t \in X$ . It follows from uniqueness in Theorem 2.5 that  $\mu(H) = 1$ .

Since  $H$  has full measure, it follows that the closed set

$$H' = \{t \in X : F^*(\phi\sigma_t^{[2]}) = F^*(\sigma_t^{[2]})\}$$

is also of full measure. Since  $F^*$  and  $\phi$  commute we conclude that whenever  $t \in H'$  and  $\pi(t) = \pi(s)$  one also has  $s \in H'$ . Applying Lemma 3.16 gives  $H' = X$ . Since  $\sigma_t^{[2]} = \delta_t \times F^*\sigma_t^{[2]}$  for all  $t$  we are done.  $\square$

Our last ingredient is that  $\sigma_t^{[2]}$ -almost every point  $x \in X^{[2]}$  is generic for  $\lambda_x^{[2]}$  along some Følner sequence. It is a special case of Theorem 7.6, which we state and prove in Section 7.2. The proofs of the two results are essentially the same, relying only on the continuity and equivariance of  $t \mapsto \sigma_t^{[2]}$  and basic facts from measure theory and ergodic theory. To avoid repetition, we omit the proof of the following version, which is all we need for Theorem 3.5, and refer the reader to the proof of Theorem 7.6 in Section 7.2.

**Lemma 3.18.** *Let  $(X, \mu, T)$  be an ergodic system and assume there is a continuous factor map to its Kronecker factor. Fix  $a \in \text{gen}(\mu, \Phi)$  for some Følner sequence  $\Phi$ . Then there exists a Følner sequence  $\Psi$  such that for  $\mu$ -almost every  $x \in X$  the point  $(a, x)$  is generic for  $\lambda_{(a,x)}^{[1]}$  along  $\Psi$ .*

We are now ready to prove Theorem 3.5 under the additional assumption that  $(X, \mu, T)$  has a continuous factor map to its Kronecker factor  $(Z, m, R)$ . In fact, all we need to do is emulate the proof of Proposition 3.8 but with  $\mu^{[2]}$  replaced by  $\sigma_a^{[2]}$ .

*Proof of Theorem 3.5 assuming a continuous factor map.* We must produce an Erdős cube in  $X^{[2]}$  whose first coordinate is  $a$  and whose last coordinate belongs to  $E$ .

From Proposition 3.11 we know each of the measures  $\lambda_x^{[1]}$  has  $\mu$  as its second marginal. In particular  $\lambda_x^{[1]}(X \times E) = \mu(E)$  for all  $x \in X^{[1]}$ . It follows that

$$\sigma_t^{[2]}(X \times X \times X \times E) = \mu(E) > 0$$

for all  $t \in X$ . It is also immediate that  $\sigma_t^{[2]}$ -almost every  $x \in X^{[2]}$  has  $t$  as its first coordinate. It therefore suffices to prove that  $\sigma_a^{[2]}$ -almost every point is an Erdős cube, as then there must be at least one whose first coordinate equals  $a$  and whose last coordinate lies in  $E$ .

Corollary 3.17 tells us that  $\sigma_a^{[2]}$  and its push forward  $\phi\sigma_a^{[2]}$  are equal. In light of this symmetry, it suffices to show that for  $\sigma_a^{[2]}$ -almost every point  $x \in X^{[2]}$  there exists an increasing sequence  $c: \mathbb{N} \rightarrow \mathbb{N}$  such that

$$\lim_{n \rightarrow \infty} T^{c(n)}(x_{00}, x_{01}) = (x_{10}, x_{11}),$$

because then it follows from  $\phi\sigma_a^{[2]} = \sigma_a^{[2]}$  that  $\sigma_a^{[2]}$ -almost every  $(x_{00}, x_{01}, x_{10}, x_{11})$  is an Erdős cube. This is equivalent to showing that for  $\mu$ -almost every  $x \in X$  and  $\delta_{(a,x)} \times \lambda_{(a,x)}^{[1]}$ -almost every  $(x_{00}, x_{01}, x_{10}, x_{11}) \in X^{[2]}$  there is some  $c: \mathbb{N} \rightarrow \mathbb{N}$  with

$$\lim_{n \rightarrow \infty} T^{c(n)}(x_{00}, x_{01}) = (x_{10}, x_{11})$$

due to the description of  $\sigma_a^{[2]}$  given in Theorem 3.14. Note that the set of  $x \in X$  for which  $(a, x)$  is generic for  $\lambda_{(a,x)}^{[1]}$  along some Følner sequence has full measure with respect to  $\mu$ , due to Lemma 3.18. For any such  $x$ , the orbit of the point  $(a, x)$  under  $T \times T$  accumulates at any point in  $\text{supp}(\lambda_{(a,x)}^{[1]})$ . Since the set  $\{(a, x)\} \times \text{supp}(\lambda_{(a,x)}^{[1]})$  has full measure with respect to  $\delta_{(a,x)} \times \lambda_{(a,x)}^{[1]}$ , this finishes the proof.  $\square$

To complete the proof of Theorem 3.5, and hence the proof of Theorem 3.1, we address the assumption that  $(X, \mu, T)$  has a continuous factor map to its Kronecker factor. In doing so, the following result is vital.

**Proposition 3.19.** *Assume  $(X, \mu, T)$  is an ergodic system and  $a \in X$  is transitive. Assume  $(Z, m, R)$  is a group rotation and that there is a measurable factor map  $\rho$  from  $(X, \mu, T)$  to  $(Z, m, R)$ . Then there are a point  $\tilde{z} \in Z$  and a Følner sequence  $\Psi$  such that*

$$\lim_{N \rightarrow \infty} \frac{1}{|\Psi_N|} \sum_{n \in \Psi_N} f(T^n a) g(R^n \tilde{z}) = \int_X f \cdot (g \circ \rho) d\mu$$

for all  $f \in C(X)$  and all  $g \in C(Z)$ .

*Proof.* This is a special case of [19 Proposition 6.1] restated in our terminology.  $\square$

The existence of the point  $\tilde{z}$  guaranteed by Proposition 3.19 allows us to produce continuous factor maps to structured factors that a priori only have measurable factor maps. The analog of this construction for  $k$ -fold sumsets – in which Kronecker factors are replaced with pronilfactors – is carried out in Section 5.2

**Proposition 3.20.** *Assume  $(X, \mu, T)$  is an ergodic system and that  $a \in \text{gen}(\mu, \Phi)$  for some Følner sequence  $\Phi$  and that  $a$  is transitive. There are an ergodic system  $(\tilde{X}, \tilde{\mu}, \tilde{T})$ , a Følner sequence  $\Psi$ , a point  $\tilde{a} \in \text{gen}(\mu, \Psi)$ , and a continuous factor map  $\pi_1: \tilde{X} \rightarrow X$  with  $\pi_1(\tilde{a}) = a$  such that  $(\tilde{X}, \tilde{\mu}, \tilde{T})$  has a continuous factor map to its Kronecker factor.*

*Proof.* Let  $\pi: X \rightarrow Z$  be a measurable factor map from  $(X, \mu, T)$  to its Kronecker factor  $(Z, m, R)$ . Consider the transformation  $\tilde{T} = T \times R$  on the product  $\tilde{X} = X \times Z$ . Let  $\pi_1(x, z) = x$  and  $\pi_2(x, z) = z$  denote the two coordinate projections. Let  $\tilde{\mu}$

be the push forward of  $\mu$  under the map  $\rho: X \rightarrow \tilde{X}$  defined by  $\rho(x) = (x, \pi(x))$ . Note that  $(X, \mu, T)$  and  $(\tilde{X}, \tilde{\mu}, \tilde{T})$  are isomorphic via the map  $\rho$  and hence have isomorphic Kronecker factors. Thus  $\pi_2$  is a continuous factor map from  $(\tilde{X}, \tilde{\mu}, \tilde{T})$  to its Kronecker factor. There exist, by Proposition 3.19, a point  $\tilde{z} \in Z$  and a Følner sequence  $\Psi$  such that  $\tilde{a} = (a, \tilde{z})$  is generic for  $\tilde{\mu}$  along  $\Psi$ . Since  $\pi_1: \tilde{X} \rightarrow X$  is continuous and  $\pi_1(\tilde{a}) = a$ , we are done.  $\square$

*Proof of Theorem 3.5 (in full generality).* Suppose  $(X, \mu, T)$ ,  $E \subset X$ , and  $a \in X$  are as in the hypothesis of Theorem 3.5. We first replace  $X$  with the closure  $X'$  of  $\{T^n(a) : n \in \mathbb{Z}\}$ . Since  $a \in \text{gen}(\mu, \Phi)$ , we have that  $\mu(X') = 1$  and  $\mu(E \cap X') > 0$ . Thus  $(X', \mu, T)$  is an ergodic system in which the point  $a$  is transitive and the set  $E' = X' \cap E$  is open in  $X'$  and has positive measure. We thus proceed assuming  $a$  is transitive in  $(X, \mu, T)$ .

Let  $(\tilde{X}, \tilde{\mu}, \tilde{T})$ ,  $\pi_1: \tilde{X} \rightarrow X$ , and  $\tilde{a} \in \tilde{X}$  be as guaranteed by Proposition 3.20. Define  $\tilde{E} = \pi_1^{-1}(E)$ . Since  $(\tilde{X}, \tilde{\mu}, \tilde{T})$  admits a continuous factor map onto its Kronecker factor and Theorem 3.5 has already been proven for such systems, there exists a 2-dimensional Erdős cube  $\tilde{x} \in \tilde{X}^{[2]}$  with  $\tilde{x}_{00} = \tilde{a}$  and  $\tilde{x}_{11} \in \tilde{E}$ . Since  $\pi_1: \tilde{X} \rightarrow X$  is continuous, the quadruple

$$x = (\pi_1(\tilde{x}_{00}), \pi_1(\tilde{x}_{01}), \pi_1(\tilde{x}_{10}), \pi_1(\tilde{x}_{11}))$$

is an Erdős cube in  $X^{[2]}$  with  $x_{00} = a$  and  $x_{11} \in E$ , finishing the proof.  $\square$

#### 4. ERDŐS CUBES

In the previous section we defined 2-dimensional Erdős cubes, which played via Theorem 3.5 a vital role in our proof of the  $k = 2$  case of Theorem 1.1. The purpose of this section is to generalize the notion of Erdős cubes to higher dimensions. We then formulate the main dynamical result of this paper, Theorem 4.4 which is a natural extension of Theorem 3.5. At the end of this section we prove that Theorem 4.4 implies Theorem 1.1.

**4.1. Defining Erdős cubes.** For  $k \in \mathbb{N}$ , define the  $k$ -dimensional cube to be  $\{0, 1\}^k$  and denote it by  $[k]$ . An element of  $[k]$  is written as  $\epsilon = \epsilon_1 \cdots \epsilon_k$ , usually without any commas or parentheses other than when this may create ambiguity. We enumerate the elements of  $[k]$  in lexicographic order. For our purposes, the most relevant feature of the ordering is that the first element is  $\vec{0} = 0 \cdots 0$  and the last element is  $\vec{1} = 1 \cdots 1$ . For example, the elements of  $[3]$  listed in order are: 000, 001, 010, 011, 100, 101, 110, 111.

For any set  $X$ , a point  $x \in X^{[k]}$  is written as  $x = (x_\epsilon : \epsilon \in [k])$ , again enumerating the indices in lexicographic order. For clarity, when we write  $x \in X^{[k]}$  as a  $2^k$ -tuple indexed by  $[k]$ , we separate the entries with commas. By convention, when we write  $X^{[0]}$ , we mean just  $X$ .

A *face* of  $[k]$  is a subset of  $[k]$  determined by fixing a single coordinate. These play a particular role in our analysis, making use of the elementary observation that the complement of such a face is also a face, and we refer to this as the *opposite face*. There are  $k$  pairs of opposite faces in  $[k]$ . For each  $1 \leq i \leq k$  write  $F^i x$  for the  $i$ th upper face of  $x \in X^{[k]}$  and  $F_i x$  for the  $i$ th lower face of  $x$ . Formally, these

belong to  $X^{\llbracket k-1 \rrbracket}$  and are defined for  $x \in X^{\llbracket k \rrbracket}$  as follows:

$$\begin{aligned} F^i x &= (x_\epsilon : \epsilon_i = 1) \quad (i^{\text{th}} \text{ upper face of } x), \\ F_i x &= (x_\epsilon : \epsilon_i = 0) \quad (i^{\text{th}} \text{ lower face of } x). \end{aligned}$$

Write also  $F^* x$  for the point

$$(4.1) \quad F^* x = (x_\epsilon : \epsilon \neq \vec{0})$$

in  $X^{\llbracket k \rrbracket} \setminus \{\vec{0}\}$ .

For example, if  $k = 2$  then a point  $x = (x_{00}, x_{01}, x_{10}, x_{11}) \in X^{\llbracket 2 \rrbracket}$  has two lower faces,  $F_1 x = (x_{00}, x_{01})$  and  $F_2 x = (x_{00}, x_{10})$ , and two upper faces,  $F^1 x = (x_{10}, x_{11})$  and  $F^2 x = (x_{01}, x_{11})$ . We also have  $F^*(x) = (x_{01}, x_{10}, x_{11})$ . When  $k = 3$  the lower faces of

$$x = (x_{000}, x_{001}, x_{010}, x_{011}, x_{100}, x_{101}, x_{110}, x_{111}) \in X^{\llbracket 3 \rrbracket}$$

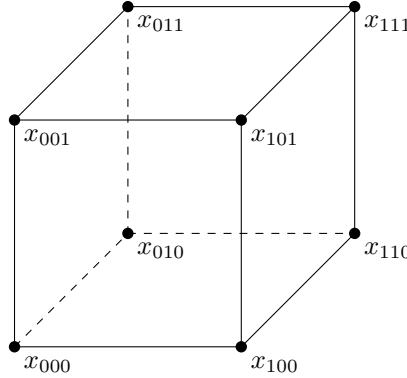
are

$$\begin{aligned} F_1 x &= (x_{000}, x_{001}, x_{010}, x_{011}), \\ F_2 x &= (x_{000}, x_{001}, x_{100}, x_{101}), \\ F_3 x &= (x_{000}, x_{010}, x_{100}, x_{110}), \end{aligned}$$

and the corresponding upper faces are

$$\begin{aligned} F^1 x &= (x_{100}, x_{101}, x_{110}, x_{111}), \\ F^2 x &= (x_{010}, x_{011}, x_{110}, x_{111}), \\ F^3 x &= (x_{001}, x_{011}, x_{101}, x_{111}), \end{aligned}$$

respectively. As a visual aid, we can think of a point in  $X^{\llbracket 3 \rrbracket}$  as being represented by a 3-dimensional cube:



This geometric interpretation of points in  $X^{\llbracket 3 \rrbracket}$  motivates our use of terminology.

Whenever  $T$  is a homeomorphism of  $X$ , we write  $T^{\llbracket k \rrbracket}$  for the homeomorphism of  $X^{\llbracket k \rrbracket}$  defined by

$$(T^{\llbracket k \rrbracket} x)_\epsilon = T(x_\epsilon)$$

for all  $x \in X^{\llbracket k \rrbracket}$  and  $\epsilon \in \llbracket k \rrbracket$ . Note that if  $x \in X^{\llbracket k \rrbracket}$  we can apply  $T^{\llbracket k-1 \rrbracket}$  to any of the upper and lower faces of  $x$ .

A permutation  $\phi$  of  $\{1, \dots, k\}$  induces a map  $(\phi\epsilon)_i = \epsilon_{\phi(i)}$  from  $\llbracket k \rrbracket$  to  $\llbracket k \rrbracket$  that, by an abuse of notation, we also denote by  $\phi$ . With a further abuse of notation we also use  $\phi$  to denote the map  $X^{\llbracket k \rrbracket} \rightarrow X^{\llbracket k \rrbracket}$  it induces via

$$(4.2) \quad (\phi x)_\epsilon = x_{\phi(\epsilon)}$$

for all  $\epsilon \in \llbracket k \rrbracket$ . See (3.7) for an example of a two dimensional permutation. For each  $1 \leq i \leq k$  write  $\phi_i$  for the permutation that exchanges 1 with  $i$  and fixes all other points. One then has

$$(4.3) \quad F^1(\phi_i x) = F^i x \quad \text{and} \quad F_1(\phi_i x) = F_i x$$

for all  $1 \leq i \leq k$ . The maps  $\phi$  are referred to as *permutations of digits* in [20, Chapter 6].

Mimicking the case  $k = 2$ , a  $k$ -dimensional Erdős cube is a point  $x \in X^{\llbracket k \rrbracket}$  thought of as forming the vertices of a  $k$ -dimensional cube, with the property that its lower faces dynamically approximate their corresponding upper faces. The formal definition is as follows.

**Definition 4.1** (Erdős cubes). Fix a compact metric space  $X$  and a homeomorphism  $T$  of  $X$ . A point  $x \in X^{\llbracket k \rrbracket}$  is a  *$k$ -dimensional Erdős cube* if

$$(4.4) \quad F^i x \in \omega(F_i x, T^{\llbracket k-1 \rrbracket})$$

for each  $1 \leq i \leq k$ .

Note that (4.4) simply says there exists a strictly increasing sequence  $c_i : \mathbb{N} \rightarrow \mathbb{N}$  such that

$$\lim_{n \rightarrow \infty} (T^{\llbracket k-1 \rrbracket})^{c_i(n)}(F_i x) = F^i x$$

holds. Even though we are working with an invertible map  $T : X \rightarrow X$ , we require in the definition of an Erdős cube that  $F^i x$  is an accumulation point of the *forward* iterates of  $F_i x$  under  $T^{\llbracket k-1 \rrbracket}$ .

We refer to them as cubes because of their connection to cube systems and since they form a natural subclass of the *dynamical cubes* (see Definition 5.10) introduced in [21] in the study of topological pronilfactors. In Section 5.4 we describe the relations among various notions of cubes, including examples and non-examples of these objects.

Lemmas 4.2 and 4.3 are useful properties that follow immediately from the definitions of Erdős cubes and the permutations in (4.3).

**Lemma 4.2.** *Let  $(X, T)$  be a topological system and let  $k \in \mathbb{N}$ . A point  $x \in X^{\llbracket k \rrbracket}$  is a  $k$ -dimensional Erdős cube if and only if*

$$\phi_i x \in \{y \in X^{\llbracket k \rrbracket} : F^1 y \in \omega(F_1 y, T^{\llbracket k-1 \rrbracket})\}$$

for all  $1 \leq i \leq k$ .

**Lemma 4.3.** *Assume  $(X, T)$  is a topological system and that  $x \in X^{\llbracket k \rrbracket}$  is a  $k$ -dimensional Erdős cube. Then  $F^i(x)$  and  $F_i(x)$  are  $(k-1)$ -dimensional Erdős cubes for all  $1 \leq i \leq k$ .*

Theorem 4.4 is the main dynamical result of this paper. It extends Theorem 3.5 and, as we show in Section 4.2, implies our main combinatorial result Theorem 1.1 for multiple sumsets.



**Theorem 4.4.** *Assume  $(X, \mu, T)$  is an ergodic system, that  $a \in \text{gen}(\mu, \Phi)$  for some Følner sequence  $\Phi$  in  $\mathbb{N}$ , and that  $E \subset X$  satisfies  $\mu(E) > 0$ . For every  $k \in \mathbb{N}$  there is a  $k$ -dimensional Erdős cube  $x \in X^{[k]}$  with  $x_{\vec{0}} = a$  and  $x_{\vec{1}} \in E$ .*

**4.2. Theorem 4.4 implies Theorem 1.1.** In this subsection we use Theorem 4.4 to prove Theorem 1.1. Our method streamlines and generalizes the proof that Theorem 3.5 implies Theorem 3.1 that we have seen in Section 3. We need to set up some terminology. Given  $\epsilon = (\epsilon_1, \dots, \epsilon_k) \in [k]$ , denote by  $\bar{\epsilon} \in [k]$  the complement  $(1 - \epsilon_1, \dots, 1 - \epsilon_k)$ . Let  $x \in X^{[k]}$  be an Erdős cube and let  $E \subset X$  be an open set such that  $x_{\vec{1}} \in E$ . Let  $B_1, \dots, B_k \subset \mathbb{N}$ . We say that  $(B_1, \dots, B_k)$  is  $(x, E)$ -acceptable if for any  $b = (b_1, \dots, b_k) \in B_1 \times \dots \times B_k$  and any  $\epsilon = (\epsilon_1, \dots, \epsilon_k) \in [k]$ ,

$$(4.5) \quad T^{\bar{\epsilon} \cdot b} x_\epsilon \in E,$$

where  $\bar{\epsilon} \cdot b = \bar{\epsilon}_1 b_1 + \dots + \bar{\epsilon}_k b_k$ . Note in particular that

$$B_1 + \dots + B_k \subset \{n : T^n x_{\vec{0}} \in E\}$$

whenever  $(B_1, \dots, B_k)$  is  $(x, E)$ -acceptable by taking  $\epsilon = \vec{0}$  in (4.5). To prove Theorem 1.1 we make use of Lemma 4.5 as an iterative step.

**Lemma 4.5.** *Let  $k \in \mathbb{N}$ , let  $(X, T)$  be a topological dynamical system, let  $x \in X^{[k]}$  be an Erdős cube, and let  $E \subset X$  be an open set with  $x_{\vec{1}} \in E$ . Assume  $(B_1, \dots, B_k)$  consists of finite sets and is  $(x, E)$ -acceptable. Fix  $i \in \{1, \dots, k\}$ . There exists  $c \in \mathbb{N} \setminus B_i$  such that, letting  $B'_i = B_i \cup \{c\}$  and  $B'_j = B_j$  for all  $j \neq i$ , the tuple  $(B'_1, \dots, B'_k)$  is still  $(x, E)$ -acceptable.*

*Proof.* Fix  $i \in \{1, \dots, k\}$ . Given  $\epsilon \in [k]$  with  $\epsilon_i = 0$ , let  $\eta(\epsilon) \in [k]$  be the element that differs from  $\epsilon$  only on the  $i$ th coordinate. Define

$$H_\epsilon = \bigcap_{b \in B_1 \times \dots \times B_k} T^{-\overline{\eta(\epsilon)} \cdot b} E$$

for each  $\epsilon \in [k]$  with  $\epsilon_i = 0$ . In view of (4.5), we have that  $x_{\eta(\epsilon)} \in H_\epsilon$  whenever  $\epsilon_i = 0$ . Since  $x$  is an Erdős cube,  $F^i x \in \omega(F_i x, T^{[k-1]})$ , and so in particular there exists  $c \in \mathbb{N} \setminus B_i$  such that

$$(4.6) \quad \text{for all } \epsilon \in [k], \epsilon_i = 0 \Rightarrow T^c x_\epsilon \in H_\epsilon.$$

We claim that with this choice of  $c$  the tuple  $(B'_1, \dots, B'_k)$  in the statement of the lemma is  $(x, E)$ -acceptable. Let  $\epsilon \in [k]$  and  $b \in B'_1 \times \dots \times B'_k$ . If  $\epsilon_i = 1$  or if  $b_i \neq c$ , then (4.5) holds automatically. If  $\epsilon_i = 0$  and  $b_i = c$ , then (4.6) and the definition of  $H_\epsilon$  imply  $T^c x_\epsilon \in T^{-\overline{\eta(\epsilon)} \cdot b} E$  as the  $i$ th coordinate of  $\overline{\eta(\epsilon)}$  is 0. In other words,  $T^{c + \overline{\eta(\epsilon)} \cdot b} x_\epsilon \in E$ , but since  $c + \overline{\eta(\epsilon)} \cdot b = \bar{\epsilon} \cdot b$ , this is precisely (4.5).  $\square$

**Corollary 4.6.** *Let  $(X, T)$  be a topological dynamical system, let  $x \in X^{[k]}$  be an Erdős cube, and let  $E \subset X$  be an open set with  $x_{\vec{1}} \in E$ . Then there exist infinite sets  $B_1, \dots, B_k \subset \mathbb{N}$  such that  $(B_1, \dots, B_k)$  is  $(x, E)$ -acceptable.*

*Proof.* The tuple  $(\emptyset, \dots, \emptyset)$  is vacuously  $(x, E)$ -acceptable for any  $x, E$ . Iterating Lemma 4.5 gives the corollary.  $\square$

We are now ready to prove Theorem 1.1

*Proof of Theorem 1.1* Let  $A \subset \mathbb{N}$  have positive Banach upper density. By Theorem 2.10 there exist an ergodic system  $(X, \mu, T)$ , a point  $a \in X$ , a clopen set  $E \subset X$ , and a Følner sequence  $\Phi$  such that  $\mu(E) > 0$ ,  $A = \{n \in \mathbb{N} : T^n a \in E\}$ , and  $a \in \text{gen}(\mu, \Phi)$ . By Theorem 4.4 there exists an Erdős cube  $x \in X^{\llbracket k \rrbracket}$  with  $x_{\vec{0}} = a$  and  $x_{\vec{1}} \in E$ . By Corollary 4.6 there exist infinite sets  $B_1, \dots, B_k \subset \mathbb{N}$  such that  $(B_1, \dots, B_k)$  is  $(x, E)$ -acceptable. Thus

$$B_1 + \dots + B_k \subset \{n \in \mathbb{N} : T^n x_{\vec{0}} \in E\} = A$$

by taking  $\epsilon = \vec{0}$  in (4.5).  $\square$

## 5. CUBE SYSTEMS AND PRONILFACTORS

Our main result - Theorem 4.4 - is that Erdős cubes with prescribed first coordinate exist under certain assumptions. In fact, we are only able to prove this directly when the system  $(X, \mu, T)$  has continuous factor maps to certain structured factors known as pronilfactors. The proof of Theorem 4.4 then follows in general from the fact that every ergodic system  $(X, \mu, T)$  has an extension with such factor maps. In this section we recall what we need from the theory of pronilfactors and prove that every ergodic system has an extension as described above. We conclude with some discussion on the relationship between Erdős cubes and other types of cubes in the dynamics literature. The material of Section 5.4 is not needed for the proof of Theorem 4.4.

**5.1. Cubic measures and structured factors.** Let  $(X, \mu, T)$  be an ergodic system. It was important in Section 3 to have an ergodic decomposition of  $\mu^{\llbracket 1 \rrbracket} = \mu \times \mu$ . As discussed in Section 1 the appropriate generalization of  $\mu \times \mu$  for the proof of Theorem 1.1 in general is the cubic measures  $\mu^{\llbracket k \rrbracket}$  whose definition we recall. We remark that while we refer to these measures and other objects as various types of cubes and use associated geometric terminology, such as faces or sides, these objects should more properly be referred to as parallelograms in two dimensions or as parallelepipeds in higher dimensions.

For each  $k \in \mathbb{N}$ , the  $k$ -dimensional cube system  $(X^{\llbracket k \rrbracket}, \mu^{\llbracket k \rrbracket}, T^{\llbracket k \rrbracket})$  is the system defined on the product space  $X^{\llbracket k \rrbracket}$  endowed with the transformation  $T^{\llbracket k \rrbracket} = T \times \dots \times T$ , the product taken  $2^k$  times, and equipped with the measure  $\mu^{\llbracket k \rrbracket}$  defined inductively by

$$(5.1) \quad \mu^{\llbracket k+1 \rrbracket} = \int_{X^{\llbracket k \rrbracket}} (\mu^{\llbracket k \rrbracket})_x \times (\mu^{\llbracket k \rrbracket})_x d\mu^{\llbracket k \rrbracket}(x),$$

where  $x \mapsto (\mu^{\llbracket k \rrbracket})_x$  is an ergodic decomposition of  $\mu^{\llbracket k \rrbracket}$ . Note that

$$(X^{\llbracket 1 \rrbracket}, \mu^{\llbracket 1 \rrbracket}, T^{\llbracket 1 \rrbracket}) = (X \times X, \mu \times \mu, T \times T)$$

by our assumption that  $(X, \mu, T)$  is ergodic, but when  $k > 1$  the measure  $\mu^{\llbracket k \rrbracket}$  is generally not equal to the product measure  $\mu^{\llbracket k-1 \rrbracket} \times \mu^{\llbracket k-1 \rrbracket}$ . Lemma 5.1 states that  $\mu^{\llbracket k \rrbracket}$  is invariant under the maps  $\phi: X^{\llbracket k \rrbracket} \rightarrow X^{\llbracket k \rrbracket}$  induced by permutations  $\phi$  of  $\{1, \dots, k\}$  as defined in (4.2).

**Lemma 5.1** ([20] Proposition 8, Chapter 8]. *Fix  $k \in \mathbb{N}$ . For every permutation  $\phi$  of  $\{1, \dots, k\}$  the push forward  $\phi\mu^{\llbracket k \rrbracket}$  is equal to  $\mu^{\llbracket k \rrbracket}$ .*

We also need later the following description of  $\mu^{\llbracket k \rrbracket}$ .

**Lemma 5.2.** *For every  $k \in \mathbb{N}$ , we have*

$$(5.2) \quad \mu^{\llbracket k+1 \rrbracket} = \int_{X^{\llbracket k \rrbracket}} \delta_x \times (\mu^{\llbracket k \rrbracket})_x d\mu^{\llbracket k \rrbracket}(x).$$

*Proof.* Using the ergodic decomposition

$$\mu^{\llbracket k \rrbracket} = \int_{X^{\llbracket k \rrbracket}} (\mu^{\llbracket k \rrbracket})_x d\mu^{\llbracket k \rrbracket}(x),$$

we have that

$$\int_{X^{\llbracket k \rrbracket}} \delta_x \times (\mu^{\llbracket k \rrbracket})_x d\mu^{\llbracket k \rrbracket}(x) = \int_{X^{\llbracket k \rrbracket}} \int_{X^{\llbracket k \rrbracket}} \delta_y \times (\mu^{\llbracket k \rrbracket})_y d\mu_x^{\llbracket k \rrbracket}(y) d\mu^{\llbracket k \rrbracket}(x).$$

Since  $x \mapsto (\mu^{\llbracket k \rrbracket})_x$  is the ergodic decomposition of  $\mu^{\llbracket k \rrbracket}$ , for  $\mu^{\llbracket k \rrbracket}$ -almost every  $x \in X^{\llbracket k \rrbracket}$  and  $(\mu^{\llbracket k \rrbracket})_x$ -almost every  $y \in X^{\llbracket k \rrbracket}$ , we have that  $(\mu^{\llbracket k \rrbracket})_y = (\mu^{\llbracket k \rrbracket})_x$ . Thus, for  $\mu^{\llbracket k \rrbracket}$ -almost every  $x \in X^{\llbracket k \rrbracket}$  it follows that

$$\int_{X^{\llbracket k \rrbracket}} \delta_y \times (\mu^{\llbracket k \rrbracket})_y d(\mu^{\llbracket k \rrbracket})_x(y) = \int_{X^{\llbracket k \rrbracket}} \delta_y \times (\mu^{\llbracket k \rrbracket})_x d(\mu^{\llbracket k \rrbracket})_x(y) = (\mu^{\llbracket k \rrbracket})_x \times (\mu^{\llbracket k \rrbracket})_x$$

and so (5.1) and (5.2) define the same measure.  $\square$

Our proof of Proposition 3.11 needs the system  $(X, \mu, T)$  to have a continuous factor map to its Kronecker factor, and it is essential in the proof that functions on the Kronecker factor describe the invariant  $\sigma$ -algebra of the system  $(X^{\llbracket 1 \rrbracket}, \mu^{\llbracket 1 \rrbracket}, T^{\llbracket 1 \rrbracket})$ . The appropriate generalization of Kronecker factor is the  $k$ -step pronilfactor of  $(X, \mu, T)$  and the structure theory of measure preserving systems tells us that functions on pronilfactors can be used to describe the invariant  $\sigma$ -algebra of  $(X^{\llbracket k \rrbracket}, \mu^{\llbracket k \rrbracket}, T^{\llbracket k \rrbracket})$ . To formally state this in the form we need, we recall some details about nilsystems.

If  $G$  is a  $k$ -step nilpotent Lie group and  $\Gamma$  is a discrete cocompact subgroup of  $G$ , then the compact manifold  $X = G/\Gamma$  is a  $k$ -step nilmanifold. The group  $G$  acts on  $X$  by left translation, and there is a unique Borel probability measure  $\mu$  on  $X$  that is invariant under this action, called the *Haar measure*.

If  $g \in G$  is a fixed element,  $T: X \rightarrow X$  is the map  $x \mapsto gx$ , and  $\mu$  is the Haar measure, then  $(X, \mu, T)$  is a  $k$ -step nilsystem. It is a classical result that for all  $k$ -step nilsystems  $(X, \mu, T)$  the properties (i) minimal; (ii) transitive; (iii) ergodic; (iv) uniquely ergodic are equivalent. (See [20] Theorem 11, Chapter 11] for a summary and discussion.)

The connection between  $\mu^{\llbracket k \rrbracket}$  and  $k$ -step nilsystems is phrased in terms of certain seminorms  $\|\cdot\|_{k+1}$  on  $L^\infty(X, \mu)$  introduced by Host and Kra [18]. For our purposes we only need the connection below between these seminorms and the measure  $\mu^{\llbracket k \rrbracket}$ . The theory of these seminorms can be found in [20]. To fully explain the connection we need to discuss inverse limits of systems. Suppose for each  $j \in \mathbb{N}$  we have a system  $(X_j, \mu_j, T_j)$ , and suppose there are continuous factor maps  $\pi_j: X_j \rightarrow X_{j-1}$  for all  $j > 1$ . Then there exists a unique *inverse limit*  $(X, \mu, T)$  which is a system admitting continuous factor maps  $\psi_j: X \rightarrow X_j$  such that  $\pi_j \circ \psi_j = \psi_{j-1}$  for every  $j > 1$ , and such that

$$\bigcup_{j \in \mathbb{N}} C(X_j) \circ \psi_j$$

is dense in  $C(X)$ . Additionally, whenever each of the systems  $(X_j, \mu_j, T_j)$  is ergodic, so is the inverse limit  $(X, \mu, T)$ . (See for example [20] Section 6, Chapter 3] for the details.)

**Theorem 5.3.** *Fix an ergodic system  $(X, \mu, T)$ . For every  $k \in \mathbb{N}$  there is a system  $(Z_k, m_k, T)$  with all the following properties.*

- *The system  $(Z_k, m_k, T)$  is an inverse limit of ergodic  $k$ -step nilsystems.*
- *There is a measurable factor map  $\rho_k$  from  $(X, \mu, T)$  to  $(Z_k, m_k, T)$ .*
- *For any  $f \in L^\infty(X, \mu)$  we have  $\|f\|_{k+1} = 0$  if and only if  $f$  and  $L^\infty(Z_k, m_k) \circ \rho_k$  are orthogonal in  $L^2(X, \mu)$ .*

*Proof.* This is a combination of Theorem 2 in Chapter 16, Theorem 7 in Chapter 9, and the discussion at the beginning of Section 3.1 in Chapter 13 of [20].  $\square$

**Definition 5.4.** Whenever  $(X, \mu, T)$  is ergodic, we call the system  $(Z_k, m_k, T)$  satisfying the conclusion of Theorem 5.3 the  $k$ -step pronilfactor of  $(X, \mu, T)$ .

The role of the  $k$ -step pronilfactor in the proof of Theorem 4.4 is analogous to the role of the Kronecker factor in the proof of Theorem 3.5.

**5.2. Continuous factor maps to pronilfactors.** It is crucial for the proof of Proposition 3.11 that the system  $(X, \mu, T)$  has a continuous factor map to its Kronecker factor. In this section we explain how to arrive at this situation via an extension procedure that holds in the more general setting of pronilfactors.

**Definition 5.5.** Given an ergodic system  $(X, \mu, T)$  and  $k \in \mathbb{N}$ , let  $(Z_k, m_k, T)$  be its  $k$ -step pronilfactor. We say that  $(X, \mu, T)$  has *topological pronilfactors* if for every  $k \in \mathbb{N}$  there is a continuous factor map  $\pi_k: X \rightarrow Z_k$ .

*Remark 5.6.* It follows from a careful analysis of our proof of Theorem 1.1 that, to find  $k$ -dimensional Erdős cubes, it suffices to have a continuous factor map to the  $k$ -step pronilfactor. However, there is no loss in assuming that we have topological pronilfactors for all  $k \in \mathbb{N}$ , and this simplifies the terminology.

Note that not every system has topological pronilfactors. In fact, there are systems (see [22] or [6]) with non-trivial Kronecker factor but no continuous eigenfunctions, which precludes the existence of a continuous factor map to a non-trivial Kronecker system. We handle such systems using Proposition 5.7 which allows us to associate with every ergodic system a larger system that has topological pronilfactors and which plays a role analogous to that played by Proposition 3.19 in Section 3.

**Proposition 5.7.** *Assume  $(X, \mu, T)$  is an ergodic system and  $a \in X$  is transitive. Assume  $(Z, S)$  is a topological system that is distal and that  $m$  is an  $S$ -invariant measure on  $Z$ . Assume also that there is a measurable factor map  $\rho$  from  $(X, \mu, T)$  to  $(Z, m, S)$ . Then there are a point  $\tilde{z} \in Z$  and a Følner sequence  $\Psi$  such that*

$$(5.3) \quad \lim_{N \rightarrow \infty} \frac{1}{|\Psi_N|} \sum_{n \in \Psi_N} f(T^n a) g(S^n \tilde{z}) = \int_X f \cdot (g \circ \rho) d\mu$$

for all  $f \in C(X)$  and all  $g \in C(Z)$ .

*Proof.* This is a restatement of [19] Proposition 6.1] in our terminology.  $\square$

**Lemma 5.8.** *Fix an ergodic system  $(X, \mu, T)$  and a transitive point  $a \in X$ . There are an ergodic system  $(\tilde{X}, \tilde{\mu}, \tilde{T})$  with topological pronilfactors, a Følner sequence  $\Psi$ , a point  $\tilde{a} \in \text{gen}(\tilde{\mu}, \Psi)$  and a continuous factor map  $\pi: \tilde{X} \rightarrow X$  with  $\pi(\tilde{a}) = a$ . Moreover  $(X, \mu, T)$  and  $(\tilde{X}, \tilde{\mu}, \tilde{T})$  are measure theoretically isomorphic systems.*

*Proof.* For each  $k \in \mathbb{N}$ , we apply Theorem 5.3 to  $(X, \mu, T)$  to obtain the  $k$ -step pronilfactor  $(Z_k, m_k, T)$  of  $(X, \mu, T)$ . It is an inverse limit of ergodic  $k$ -step nilsystems and there is a measurable factor map  $\rho_k: (X, \mu, T) \rightarrow (Z_k, m_k, T)$ . Define  $(Z, m, S)$  to be the system where

$$Z = \prod_{k \in \mathbb{N}} Z_k,$$

$m$  is the push forward of  $\mu$  under the measurable map  $\rho$  defined by

$$\rho(x) = (\rho_k(x) : k \in \mathbb{N})$$

for all  $x \in X$ , and  $S$  is the product transformation. Note that  $S$  is distal on  $Z$  and that  $\rho$  is a measurable factor map from  $(X, \mu, T)$  to  $(Z, m, S)$ . We can therefore apply Proposition 5.7 to obtain some point  $\tilde{z} \in Z$  and a Følner sequence  $\Psi$  such that (5.3) holds for all  $f \in C(X)$  and all  $g \in C(Z)$ . Write  $\tilde{a} = (a, \tilde{z})$  and  $\tilde{T} = T \times S$ . Note that (5.3) implies  $\tilde{a} \in \text{gen}(\tilde{\mu}, \Psi)$ , where  $\tilde{\mu}$  is the push forward of  $\mu$  under the map  $x \mapsto (x, \rho(x))$ . Take  $\tilde{X}$  to be the orbit closure of  $\tilde{a}$  in  $X \times Z$  and let  $\pi: \tilde{X} \rightarrow X$  be the projection on the first coordinate.

The maps  $x \mapsto (x, \rho(x))$  and  $(x, z) \mapsto x$  are almost sure inverses of one another and therefore the systems  $(X, \mu, T)$  and  $(\tilde{X}, \tilde{\mu}, \tilde{T})$  are measure theoretically isomorphic. In particular, it follows that  $(\tilde{X}, \tilde{\mu}, \tilde{T})$  is ergodic and that its  $k$ -step pronilfactor is isomorphic to  $(Z_k, m_k, T)$  for each  $k \in \mathbb{N}$ . Thus the map  $(x, z) \mapsto z_k$  is a continuous factor map from  $(\tilde{X}, \tilde{\mu}, \tilde{T})$  to its  $k$ -step pronilfactor.  $\square$

The upshot is that Theorem 5.9 implies Theorem 4.4

**Theorem 5.9.** *Assume  $(X, \mu, T)$  is an ergodic system with topological pronilfactors, that  $a \in \text{gen}(\mu, \Phi)$  for some Følner sequence  $\Phi$  and that  $E \subset X$  satisfies  $\mu(E) > 0$ . For every  $k \in \mathbb{N}$  there is a  $k$ -dimensional Erdős cube  $x \in X^{[k]}$  with  $x_{\vec{0}} = a$  and  $x_{\vec{1}} \in E$ .*

*Proof that Theorem 5.9 implies Theorem 4.4.* Assume  $(X, \mu, T)$  is an ergodic system, that  $a \in \text{gen}(\mu, \Phi)$  for some Følner sequence  $\Phi$ , and that  $E \subset X$  has positive measure. Fix  $k \in \mathbb{N}$ . We wish to apply Lemma 5.8 and for this we need the point  $a$  to be transitive. We therefore replace  $X$  with the closure  $X'$  of  $\{T^n(a) : n \in \mathbb{Z}\}$ . Since  $a \in \text{gen}(\mu, \Phi)$ , we must have  $\mu(X') = 1$  and  $\mu(E \cap X') > 0$ . Thus  $(X', \mu, T)$  is an ergodic system in which the point  $a$  is transitive and the set  $X' \cap E$  has positive measure. Apply Lemma 5.8 to  $(X', \mu, T)$  to get a system  $(\tilde{X}, \tilde{\mu}, \tilde{T})$ , a Følner sequence  $\Psi$ , a point  $\tilde{a} \in \text{gen}(\tilde{\mu}, \Psi)$ , and a continuous factor map  $\pi$  from  $(\tilde{X}, \tilde{\mu}, \tilde{T})$  to  $(X', \mu, T)$ . Set  $\tilde{E} = \pi^{-1}(X' \cap E)$ .

Now apply Theorem 5.9 to get an Erdős cube  $\tilde{x} \in \tilde{X}^{[k]}$  with  $\tilde{x}_{\vec{0}} = \tilde{a}$  and  $\tilde{x}_{\vec{1}} \in \tilde{E}$ . It is immediate that  $\pi(\tilde{x})$  is an Erdős cube whose first coordinate is  $a$  and whose last coordinate belongs to  $E$ .  $\square$

**5.3. Nilcubes.** We now turn to an issue in the proof of Theorem 1.1 for  $k \geq 3$  that is not apparent in the case  $k = 2$  discussed in Section 3. To prove Proposition 3.11 we defined for every point  $x \in X^{[1]}$  a measure  $\lambda_x^{[1]}$  on  $X \times X$  with various properties.

In the next section we define measures  $\lambda_x^{[k]}$  on  $X^{[k]}$  to prove Theorem 1.1 in general. However, even when  $(X, \mu, T)$  has topological pronilfactors, we are not able to define  $\lambda_x^{[k]}$  for all  $x \in X^{[k]}$ . In preparation for the next section, we define the space that is the domain of the map  $x \mapsto \lambda_x^{[k]}$ . To do so, we recall the definition of dynamical cubes from [21].

**Definition 5.10** (Dynamical cubes). For a topological system  $(X, T)$ , we define the set of  $k$ -dimensional dynamical cubes  $Q^{[k]}(X, T)$  to be the orbit closure of the diagonal points under the face transformations. Formally

$$Q^{[k]}(X, T) = \overline{\{(T^{\epsilon \cdot n} x : \epsilon \in [k]) : x \in X \text{ and } n \in \mathbb{Z}^k\}}$$

with the closure taken in  $X^{[k]}$  and  $\epsilon \cdot n = \epsilon_1 n_1 + \cdots + \epsilon_k n_k$  for  $\epsilon \in [k]$  and  $n \in \mathbb{Z}^k$ .

In our setting we only have a single transformation on a given topological system, so we write  $Q^{[k]}(X)$  instead of the more cumbersome  $Q^{[k]}(X, T)$ . For example,  $Q^{[2]}(X)$  is the closure of

$$\{(x, T^n x, T^m x, T^{m+n} x) : m, n \in \mathbb{Z} \text{ and } x \in X\}$$

in  $X^{[2]}$ .

**Theorem 5.11** ([20, Theorem 8, Chapter 12]). Assume  $(Z, m, T)$  is an ergodic  $k$ -step nilsystem. Then  $(Q^{[k]}(Z), T^{[k]}, m^{[k]})$  is a  $k$ -step nilsystem, and in particular  $m^{[k]}(Q^{[k]}(Z)) = 1$ .

**Definition 5.12** (Nilpotent cubes). For an ergodic system  $(X, \mu, T)$  with topological pronilfactors, we define the set of  $k$ -dimensional nilcubes to be

$$N^{[k]}(X) = (\pi_k^{[k]})^{-1}(Q^{[k]}(Z_k)),$$

where  $(Z_k, m_k, T)$  is the  $k$ -step pronilfactor of  $(X, \mu, T)$  and  $\pi_k : X \rightarrow Z_k$  is the associated continuous factor map.

It is the set  $N^{[k]}(X)$  that plays the role of the domain for the map  $x \mapsto \lambda_x^{[k]}$  we introduce in Section 6. The following basic properties of  $N^{[k]}(X)$  are used frequently.

**Lemma 5.13.** Assume the ergodic system  $(X, \mu, T)$  has topological pronilfactors. Then  $N^{[1]}(X) = X^{[1]}$ .

*Proof.* Fix a continuous factor map  $\pi$  from  $(X, \mu, T)$  to its Kronecker factor  $(Z, m, T)$ . Since the Kronecker factor is an ergodic group rotation every point therein has dense orbit, and  $Q^{[1]}(Z) = Z^{[1]}$ . Thus  $N^{[1]}(X) = (\pi^{[1]})^{-1}(Q^{[1]}(Z_k)) = (\pi^{[1]})^{-1}(Z^{[1]}) = X^{[1]}$ .  $\square$

**Lemma 5.14.** Assume the ergodic system  $(X, \mu, T)$  has topological pronilfactors. Then  $N^{[k]}(X)$  is closed,

$$(5.4) \quad Q^{[k]}(X) \subset N^{[k]}(X)$$

for all  $k \in \mathbb{N}$ , and  $\mu^{[k]}(N^{[k]}(X)) = 1$ .

*Proof.* Fix  $k \in \mathbb{N}$ . Write  $\pi$  for the continuous factor map from  $(X, \mu, T)$  to its  $k$ -step pronilfactor  $(Z, m, T)$ . The first property is immediate, as  $\pi^{[k]}$  is continuous

and  $Q^{\llbracket k \rrbracket}(Z)$  is closed. The containment (5.4) holds for the following reason: if  $x \in Q^{\llbracket k \rrbracket}(X)$  is a limit point of the sequence

$$n \mapsto (x_n, T^{a(n)}x_n, T^{b(n)}x_n, T^{a(n)+b(n)}x_n),$$

then  $\pi(x)$  is a limit point of

$$n \mapsto (\pi(x_n), T^{a(n)}(\pi(x_n)), T^{b(n)}(\pi(x_n)), T^{a(n)+b(n)}(\pi(x_n))).$$

The calculation

$$\mu^{\llbracket k \rrbracket}(N^{\llbracket k \rrbracket}(X)) = \mu^{\llbracket k \rrbracket}((\pi^{\llbracket k \rrbracket})^{-1}Q^{\llbracket k \rrbracket}(Z)) = m^{\llbracket k \rrbracket}(Q^{\llbracket k \rrbracket}(Z)) = 1$$

gives the last property by Theorem 5.11  $\square$

**5.4. Comparing cubes.** We compare Erdős cubes, nilcubes, and dynamical cubes. The material of this section is not needed later on, but is included to clarify the main objects studied.

**Definition 5.15.** For a topological system  $(X, T)$ , we define the *set of  $k$ -dimensional Erdős cubes*  $E^{\llbracket k \rrbracket}(X)$  to be the collection of all  $k$ -dimensional Erdős cubes  $x \in X^{\llbracket k \rrbracket}$  (see Definition 4.1).

**Lemma 5.16.** Assume that  $(X, T)$  is a topological system. Then for all  $k \in \mathbb{N}$ , we have that  $E^{\llbracket k \rrbracket}(X) \subset Q^{\llbracket k \rrbracket}(X)$ .

*Proof.* We proceed by induction on  $k$ . When  $k = 1$ , an Erdős cube is of the form  $(x, y)$  with  $T^{c(n)}x \rightarrow y$  as  $n \rightarrow \infty$  for an increasing sequence  $c: \mathbb{N} \rightarrow \mathbb{N}$ . Since  $(x, T^{c(n)}x) \in Q^{\llbracket 1 \rrbracket}(X)$  for every  $n \in \mathbb{N}$ , it is immediate that  $(x, y) \in Q^{\llbracket 1 \rrbracket}(X)$ .

For the inductive step, fix an Erdős cube  $x \in X^{\llbracket k \rrbracket}$ . Recall that  $F_1x \in X^{\llbracket k-1 \rrbracket}$  is the projection of  $x$  onto its first lower face. Since  $F_1x$  is an Erdős cube (see Lemma 4.3), we have that  $F_1x \in Q^{\llbracket k-1 \rrbracket}(X)$  and therefore  $(F_1x, (T^{\llbracket k-1 \rrbracket})^n F_1x) \in Q^{\llbracket k \rrbracket}(X)$  for all  $n \in \mathbb{N}$ . Since  $F^1x$  is a limit point of  $\{(T^{\llbracket k-1 \rrbracket})^n F_1x : n \in \mathbb{N}\}$  and  $Q^{\llbracket k \rrbracket}(X)$  is closed, we are done.  $\square$

Note that it does not follow from the definition that the collection of Erdős cubes is closed, while it is incorporated into the definition of dynamical cubes that this collection is closed. Moreover, an Erdős cube is defined in terms of forward orbits under the appropriate cubic transformations, while the definition of a dynamical cube allows both forward and backward orbits. This difference is motivated by the intended application of the existence of Erdős cubes to find patterns in a set of positive upper density in the positive (and not all) integers. However, even were we to extend the definition of an Erdős cube to allow two sided orbits, they would still not define the same object as the dynamical cubes. We include some examples to illustrate these differences.

**Example 5.17.** Set  $X = \{0, 1\}^{\mathbb{Z}}$  endowed with the shift map  $T$ . There exists  $x \in X$  such that

$$(5.5) \quad (x, \mathbf{1}_{2\mathbb{Z}}, \mathbf{1}_{3\mathbb{Z}}, \mathbf{1}_{5\mathbb{Z}})$$

belongs to  $Q^{\llbracket 2 \rrbracket}(X)$  but not to  $E^{\llbracket 2 \rrbracket}(X)$ . Take  $a(n) = 8^{10n}$  and  $b(n) = 9^{10n}$ . Let

$$x(i) = \begin{cases} 1_{2\mathbb{Z}}(i) & \text{if } i \in [a(n), a(n) + n] \text{ for some } n, \\ 1_{3\mathbb{Z}}(i) & \text{if } i \in [b(n), b(n) + n] \text{ for some } n, \\ 1_{5\mathbb{Z}}(i) & \text{if } i \in [a(n) + b(n), a(n) + b(n) + n] \text{ for some } n, \\ 0 & \text{otherwise.} \end{cases}$$



Then  $(x, 1_{2\mathbb{Z}}, 1_{3\mathbb{Z}}, 1_{5\mathbb{Z}})$  is a limit of  $(x, T^{a(n)}x, T^{b(n)}x, T^{a(n)+b(n)}x)$ . However (5.5) is not an Erdős cube, as  $1_{5\mathbb{Z}}$  does not lie in the orbit closure of  $1_{2\mathbb{Z}}$ .

Even in well-behaved, familiar systems, such as an algebraic skew product, we may not have equality between the two notions of cubes.

**Example 5.18.** Consider the skew-product  $T: (x, y) \mapsto (x + \alpha, y + x)$  on  $X = \mathbb{T}^2$ , where  $\alpha$  is irrational. Then  $((x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4)) \in X^4$  belongs to  $\mathbb{Q}^{\llbracket 2 \rrbracket}(X)$  if and only if  $x_1 + x_4 = x_2 + x_3$  (this can be seen by computing the orbit of a point  $(x', y_1)$  with  $x'$  close to  $x_1$  and rationally independent from  $\alpha$ ). However, the point  $((0, 0), (0, 0), (0, 0), (0, 1/2)) \in X^4$  is not in  $\mathbb{E}^{\llbracket 2 \rrbracket}(X)$ .

We have the containments

$$\mathbb{E}^{\llbracket k \rrbracket}(X) \subset \mathbb{Q}^{\llbracket k \rrbracket}(X) \subset \mathbb{N}^{\llbracket k \rrbracket}(X)$$

for all  $k \in \mathbb{N}$  whenever  $(X, \mu, T)$  is an ergodic system with topological pronilfactors. Examples 5.17 and 5.18 show that the first containment is proper in general, even if we were to adjust the definition to make use of the full (backwards and forwards) orbit. For ergodic nilsystems, and more generally for distal systems, it turns out that the closure of the  $k$ -dimensional Erdős cubes equals the set of  $k$ -dimensional dynamical cubes.

**Theorem 5.19.** *Let  $(X, T)$  be a distal topological system and let  $k \in \mathbb{N}$ . Then  $\mathbb{Q}^{\llbracket k \rrbracket}(X) = \overline{\mathbb{E}^{\llbracket k \rrbracket}(X)}$ .*

*Proof.* Define

$$P = \{(x, y) : x \in \mathbb{Q}^{\llbracket k-1 \rrbracket}(X), y \in \omega(x, T^{\llbracket k-1 \rrbracket})\}$$

and note that Lemma 4.2 implies

$$\mathbb{E}^{\llbracket k \rrbracket}(X) \supset \bigcap_{i=1}^k \phi_i(P),$$

where  $\phi_i$  is the permutation of  $\{1, \dots, k\}$  that exchanges 1 and  $i$ .

We use induction to check that for any  $x \in X$  and  $n \in \mathbb{Z}^k$ , the point  $(T^{\epsilon \cdot n}x : \epsilon \in \llbracket k \rrbracket)$  is in  $P$ . Indeed, when  $k = 1$ , since  $(X, T)$  is distal it is a disjoint union of minimal systems by [11] Theorem 3.2]. Restricting to one of these minimal systems, we have that  $(a, T^n a) \in P$  for all  $a \in X$  and all  $n \in \mathbb{Z}$ , as the forward orbit of any  $a$  is dense. The inductive step follows because each power  $(X^{\llbracket k \rrbracket}, T^{\llbracket k \rrbracket})$  of a distal system is also distal. It follows that  $\mathbb{Q}^{\llbracket k \rrbracket}(X) \subset \overline{P}$  and hence  $\mathbb{Q}^{\llbracket k \rrbracket}(X) \subset \overline{\phi_i(P)}$  for any  $i \in \{1, \dots, k\}$ , as  $\mathbb{Q}^{\llbracket k \rrbracket}(X)$  is invariant under  $\phi_i$ . Therefore

$$\mathbb{Q}^{\llbracket k \rrbracket}(X) \subset \bigcap_{i=1}^k \overline{\phi_i(P)} \subset \overline{\mathbb{E}^{\llbracket k \rrbracket}(X)} \subset \mathbb{Q}^{\llbracket k \rrbracket}(X)$$

and we conclude that  $\mathbb{Q}^{\llbracket k \rrbracket}(X) = \overline{\mathbb{E}^{\llbracket k \rrbracket}(X)}$ .  $\square$

In general, an ergodic system  $(X, \mu, T)$  having topological nilfactors need not have  $\mathbb{Q}^{\llbracket k \rrbracket}(X) = \mathbb{N}^{\llbracket k \rrbracket}(X)$ . Examples illustrating this are given in [29] Example 3.6] and [4] Section 3.2].

## 6. A CONTINUOUS ERGODIC DECOMPOSITION

The rest of the paper is dedicated to the proof of Theorem 5.9. The first step is to generalize Proposition 3.11 to higher dimensions by describing an ergodic decomposition for the cube systems  $(X^{\llbracket k \rrbracket}, \mu^{\llbracket k \rrbracket}, T^{\llbracket k \rrbracket})$  that is compatible with the topology.

**Definition 6.1.** Given a system  $(X, \mu, T)$ , we say that a map  $x \mapsto \mu_x$  from  $X$  to  $\mathcal{M}(X)$  is a *continuous ergodic decomposition* if it is continuous and is an ergodic decomposition in the sense of Definition 2.7.

Note that, in view of Theorem 2.8, a continuous ergodic decomposition, if it exists, is unique on  $\text{supp}(\mu)$ . However, a continuous ergodic decomposition can be extended from  $\text{supp}(\mu)$  to the whole space  $X$  in an arbitrary manner (as long as continuity is preserved).

**Lemma 6.2.** *A system  $(X, \mu, T)$  admits a continuous ergodic decomposition if and only if for every  $f \in C(X)$ , the conditional expectation  $\mathbb{E}(f \mid \mathcal{I})$  on the  $\sigma$ -algebra of  $T$ -invariant sets agrees  $\mu$ -almost everywhere with a continuous function.*

*Proof.* If  $x \mapsto \mu_x$  is a continuous ergodic decomposition, then for any  $f \in C(X)$  the function

$$x \mapsto \int_X f \, d\mu_x$$

is continuous, and

$$\int_X f \, d\mu_x = \mathbb{E}(f \mid \mathcal{I})(x)$$

for  $\mu$ -almost every  $x \in X$ , by definition. Conversely, if for every  $f \in C(X)$  there exists  $g \in C(X)$  such that  $\mathbb{E}(f \mid \mathcal{I})$  and  $g$  agree  $\mu$ -almost everywhere, then the measures  $\mu_x$  defined by  $\int_X f \, d\mu_x := g(x)$  define a continuous ergodic decomposition.  $\square$

We now give some examples to clarify the notion of a continuous ergodic decomposition.

**Example 6.3.** Take  $X = \mathbb{R}^2/\mathbb{Z}^2$  and define  $T$  by  $T(x, y) = (x, y + x)$ . The Haar measure  $\mu$  on  $X$  is  $T$ -invariant but is not ergodic. In this system, every point is generic for some ergodic measure, but the function that associates to each point the ergodic measure it is generic for is not a continuous function. To remedy this issue, let  $m$  denote the Lebesgue measure on  $[0, 1)$  and, for each  $(x, y) \in X$ , define the measure  $\mu_{(x, y)} = \delta_x m$ , where  $\delta_x$  denotes as usual the Dirac point mass at  $x$ .

The map  $(x, y) \mapsto \mu_{(x, y)}$  from  $X$  to the space  $\mathcal{M}(X)$  of Borel probability measures on  $X$  is continuous. Furthermore,  $\mu_{(x, y)}$  is ergodic for  $T$  whenever  $x$  is irrational, in particular it is ergodic for  $\mu$ -almost every  $(x, y) \in X$ . Finally, since  $\int_X \mu_{(x, y)} \, d\mu(x, y) = \mu$ , the assignment  $(x, y) \mapsto \mu_{(x, y)}$  is a continuous ergodic decomposition of  $\mu$ .

**Example 6.4.** For our second example fix  $\alpha \in \mathbb{R} \setminus \mathbb{Q}$  and define  $T$  on  $X = \mathbb{R}^4/\mathbb{Z}^4$  by

$$T(x, y, w, z) = (x + \alpha, y + x, w + \alpha, z + w)$$

for all  $(x, y, w, z) \in X$ . The Haar measure  $\mu$  on  $\mathbb{T}^4$  is  $T$ -invariant but not ergodic, as the set

$$X_\beta = \{(x, y, w, z) \in X : w - x = \beta\}$$

is  $T$ -invariant for every  $\beta \in \mathbb{R}/\mathbb{Z}$ . Note that  $\psi_\beta: \mathbb{R}^3/\mathbb{Z}^3 \rightarrow X_\beta$  by

$$(x, y, z) \mapsto (x, y, x + \beta, z)$$

defines an isomorphism between  $(X_\beta, T)$  and  $(\mathbb{T}^3, S_\beta)$ , where

$$S_\beta(x, y, z) = (x + \alpha, y + x, z + x + \beta)$$

on  $\mathbb{R}^3/\mathbb{Z}^3$ . It can be checked that the Haar measure  $m$  on  $\mathbb{T}^3$  is ergodic for  $S_\beta$  for almost every  $\beta \in \mathbb{R}/\mathbb{Z}$ . The map  $(x, y, w, z) \mapsto \mu_{(x, y, w, z)} := \psi_{w-x}(m)$  assigning to  $(x, y, w, z)$  the push forward of  $m$  under  $\psi_{w-x}$  is a continuous map from  $X$  to  $\mathcal{M}(X)$ . Moreover,  $\mu_{(x, y, w, z)}$  is  $T$ -ergodic for  $\mu$ -almost every  $(x, y, w, z) \in X$  and forms a disintegration of  $\mu$ , showing that  $(x, y, w, z) \mapsto \mu_{(x, y, w, z)}$  is a continuous ergodic decomposition of  $\mu$ .

Note that  $T$  is a self-product of the map  $(x, y) \mapsto (x + \alpha, y + x)$  on  $\mathbb{R}^2/\mathbb{Z}^2$  and that the Haar measure  $\nu$  on  $\mathbb{T}^2$  is ergodic for  $T$ . This disintegration is therefore an example of a continuous ergodic decomposition of  $\nu^{[1]} = \nu \times \nu$ .

Recall that, given a system  $(X, \mu, T)$  with topological pronilfactors, we denote by  $Z_k$  the  $k$ -step pronilfactor and by  $\pi_k: X \rightarrow Z_k$  the (continuous) factor map. Recall also that  $\mathbf{N}^{[k]}(X) = (\pi_k^{[k]})^{-1}(\mathbf{Q}^{[k]}(Z_k))$  is closed and  $T^{[k]}$ -invariant, and that  $\mu^{[k]}(\mathbf{N}^{[k]}(X)) = 1$ , so that  $(\mathbf{N}^{[k]}(X), \mu^{[k]}, T^{[k]})$  is a system. In particular, the support of  $\mu^{[k]}$  is contained in  $\mathbf{N}^{[k]}(X)$ .

**Theorem 6.5.** *Assume that  $(X, \mu, T)$  is an ergodic system with topological pronilfactors. For every  $k \in \mathbb{N}$  there is a unique continuous ergodic decomposition  $x \mapsto \lambda_x^{[k]}$  of the system  $(\mathbf{N}^{[k]}(X), \mu^{[k]}, T^{[k]})$  such that*

$$(6.1) \quad \pi_k^{[k]}(x) = \pi_k^{[k]}(y) \Rightarrow \lambda_x^{[k]} = \lambda_y^{[k]}$$

for all  $x, y \in \mathbf{N}^{[k]}(X)$ .

We noted that it follows from Theorem 2.8 that continuous ergodic decompositions are uniquely defined on the support of the measure being decomposed. Yet in the most interesting situations, our point  $a$  is outside the support of  $\mu$  and therefore all points  $x \in X^{[k]}$  whose first coordinate equals  $a$  are outside  $\text{supp}(\mu^{[k]})$ . However, the ergodic decomposition in Theorem 6.5 is defined on  $\mathbf{N}^{[k]}(X)$ , which contains points with any prescribed first coordinate (this follows from the fact that our system has topological pronilfactors).

The remainder of this section is devoted to proving Theorem 6.5. We start by applying a theorem of Leibman to establish Theorem 6.5 for ergodic nilsystems, and then make use of inverse limits and the Host-Kra structure theorem to prove the general case.

**Theorem 6.6** (Leibman [23] Theorem 2.2). *Let  $(Z, m, T)$  be a nilsystem with  $Z = G/\Gamma$  and  $\pi: G \rightarrow Z$  the natural map. Let  $V$  be a connected subnilmanifold of  $Z$  and let  $K$  be a connected component of  $\pi^{-1}(V)$ . There is a subnilmanifold  $Y$  of  $Z$  with the following property: for  $m$ -almost every  $z \in V$  we have  $\{T^n z : n \in \mathbb{Z}\} = aY$  whenever  $\pi(a) = z$  and  $a \in K$ .*

Using Theorem 6.6 we deduce that every nilsystem admits a continuous ergodic decomposition.

**Lemma 6.7.** *Every nilsystem  $(Z, m, T)$  admits a continuous ergodic decomposition  $\xi: Z \rightarrow \mathcal{M}(Z)$ .*

*Proof.* Let us first assume that  $Z$  is connected. Once we have established the claim for connected  $Z$ , we then derive the general case using basic properties of nilsystems.

Write  $Z = G/\Gamma$  where  $G$  is a nilpotent Lie group and  $\Gamma$  a discrete and co-compact subgroup of  $G$ , and let  $\pi: G \rightarrow Z$  denote the natural factor map from  $G$  onto  $Z$ . Let  $G^\circ$  denote the identity component of  $G$  and note that  $G^\circ\Gamma = G$  because  $Z = G/\Gamma$  is connected. Applying Theorem 6.6 with  $V = Z$  and  $K = G^\circ$  gives a subnilmanifold  $Y$  of  $Z$  and a set  $W \subset G^\circ$  of full measure in  $G^\circ$  with the following property: for every  $g \in W$  we have

$$\overline{\{T^n \pi(g) : n \in \mathbb{Z}\}} = gY.$$

In particular, we have that  $g_1 Y = g_2 Y$  whenever  $g_1, g_2 \in W$  with  $\pi(g_1) = \pi(g_2)$ . Define the map  $\tilde{\xi}: G^\circ \rightarrow \mathcal{M}(Z)$  as  $\tilde{\xi}(g) = \mu_{gY}$ , where  $\mu_{gY}$  denotes the Haar measure of the subnilmanifold  $gY$ . Note that  $(gY, T)$  is a transitive nilsystem whenever  $g \in W$ , and hence  $\mu_{gY}$  is ergodic. Moreover  $\pi(g)$  is generic for  $\tilde{\xi}(g)$ . We claim that

$$(6.2) \quad \tilde{\xi}(g\gamma) = \tilde{\xi}(g), \quad \text{for all } g \in G^\circ, \text{ for all } \gamma \in \Gamma^\circ,$$

where  $\Gamma^\circ = G^\circ \cap \Gamma$ . By the conclusion of Theorem 6.6 it follows that  $\tilde{\xi}(g\gamma) = \tilde{\xi}(g)$  for any  $g \in W$  and any  $\gamma \in \Gamma^\circ$  with  $g\gamma \in W$ , because  $\pi(g) = \pi(g\gamma)$  and hence  $gY = g\gamma Y$ . In other words, we have that (6.2) holds for all  $\gamma \in \Gamma^\circ$  and all  $g \in W \cap W\gamma^{-1}$ . Since  $\tilde{\xi}: G^\circ \rightarrow \mathcal{M}(Z)$  is a continuous map and  $W \cap W\gamma^{-1}$  is a dense subset of  $G^\circ$  (because  $W \cap W\gamma^{-1}$  has full measure), we conclude that (6.2) holds for all  $g \in G^\circ$  and all  $\gamma \in \Gamma^\circ$  as desired. Finally, (6.2) implies that the map  $\tilde{\xi}$  descends to a map  $\xi$  on the quotient  $G^\circ/\Gamma^\circ$ . Since  $G^\circ/\Gamma^\circ$  and  $G/\Gamma$  are isomorphic, we can view  $\xi$  as a map on  $Z = G/\Gamma$  and we are done.

We are left to deal with the case when  $Z$  is not connected. In this case, by compactness, the nilmanifold  $Z$  has finitely many connected components, which we denote by  $Z_1, \dots, Z_r$ . Let  $m_i$  denote the Haar measure on  $Z_i$  and observe that  $m = \frac{1}{r}(m_1 + \dots + m_r)$ .

Since  $T$  maps connected components to connected components, there exists a permutation  $\tau: \{1, \dots, r\} \rightarrow \{1, \dots, r\}$  such that  $TZ_i = Z_{\tau(i)}$ . Since any permutation on  $r$  symbols satisfies  $\tau^{r!} = \text{Id}$ , we have  $T^{r!}Z_i = Z_i$  for every  $i \in \{1, \dots, r\}$ . In particular, the triple  $(Z_i, m_i, T^{r!})$  is a connected nilsystem. Hence, there exists a continuous map  $\xi_i: Z_i \rightarrow \mathcal{M}(Z_i)$  such that  $z \mapsto \xi_{i,z}$  is an ergodic decomposition of  $m_i$  with respect to the transformation  $T^{r!}$ , and  $m_i$ -almost every  $z \in Z_i$  is generic for  $\xi_{i,z}$  for all  $i \in \{1, \dots, r\}$ . Let  $U_i$  denote the set of points in  $Z_i$  that are generic for  $\xi_{i,z}$  (under  $T^{r!}$ ) and let  $U = U_1 \cup \dots \cup U_r$ ; then let  $Y \subset Z$  be the set of points whose entire orbit is contained in  $U$  and note that  $m(Y) = 1$ . For  $z \in Y \cap Z_i$  we have an explicit description of  $\xi_{i,z}$  as

$$\xi_{i,z} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \delta_{T^{nr!}z}$$

and in particular  $T(\xi_{i,z}) = \xi_{\tau(i), Tz}$ .

For every  $z \in Z$ , let  $i_z$  be the unique number in  $\{1, \dots, r\}$  such that  $z \in Z_{i_z}$ . Consider the map  $z \mapsto \xi_z$  given by

$$\xi_z = \frac{1}{r!} \sum_{j=1}^{r!} T^j(\xi_{i_z, z})$$

and note that it is continuous. Moreover, each  $z \in Y$  is generic for  $\xi_z$  under  $T$ , which immediately implies that  $\xi_z$  is  $T$ -invariant, ergodic, and that  $\int_Z \xi_z dm = m$ .  $\square$

It now follows that ergodic nilsystems satisfy the conclusion of Theorem 6.5.

**Lemma 6.8.** *Let  $(Z, m, T)$  be an ergodic nilsystem. For every  $k \in \mathbb{N}$  there is a continuous ergodic decomposition  $z \mapsto \xi_z^{[k]}$  of the system  $(Q^{[k]}(Z), m^{[k]}, T^{[k]})$ .*

*Proof.* Fix  $k \in \mathbb{N}$ . By Theorem 5.11 it follows that  $(Q^{[k]}(Z), T^{[k]})$  is a nilsystem and that the measure  $m^{[k]}$  is equal to the Haar measure on  $Q^{[k]}(Z)$ . Thus the conclusion follows from Lemma 6.7.  $\square$

Next, we need a version of Lemma 6.8 for topological inverse limits of nilsystems.

**Lemma 6.9.** *Suppose for each  $j \in \mathbb{N}$  we have a system  $(X_j, \mu_j, T_j)$  admitting a continuous ergodic decomposition, and suppose there are continuous factor maps  $\pi_j: X_{j+1} \rightarrow X_j$ . Then the inverse limit  $(X, \mu, T)$  admits a continuous ergodic decomposition.*

*Proof.* Recall that we have continuous factor maps  $\psi_j: X \rightarrow X_j$  satisfying  $\pi_j \circ \psi_j = \psi_{j-1}$ . In view of Lemma 6.2 for each  $j \in \mathbb{N}$  and  $f \in C(X_j)$ , the function  $\mathbb{E}(f | \mathcal{I}_j)$  is  $\mu_j$ -almost everywhere continuous, and hence the function  $\mathbb{E}(f \circ \psi_j | \mathcal{I}) = \mathbb{E}(f | \mathcal{I}_j) \circ \psi_j$  is  $\mu$ -almost everywhere continuous. Since

$$\bigcup_{j \in \mathbb{N}} C(X_j) \circ \psi_j$$

is dense in  $C(X)$ , using the converse direction of Lemma 6.2 we conclude that  $(X, \mu, T)$  admits a continuous ergodic decomposition.  $\square$

**Corollary 6.10.** *Let  $k \in \mathbb{N}$  and let  $(Z, \mu, T)$  be an inverse limit of ergodic  $k$ -step nilsystems. Then the system  $(Q^{[k]}(Z), \mu^{[k]}, T^{[k]})$  admits a continuous ergodic decomposition.*

*Proof.* Suppose that  $(Z, \mu, T)$  is the inverse limit of a sequence of ergodic,  $k$ -step nilsystems  $(Z_j, m_j, T_j)$ . Then a routine check shows that  $(Q^{[k]}(Z), \mu^{[k]}, T^{[k]})$  is the inverse limit of the sequence  $(Q^{[k]}(Z_j), m_j^{[k]}, T_j^{[k]})$ . The conclusion now follows by combining Lemmas 6.8 and 6.9.  $\square$

**Lemma 6.11.** *Let  $(X, \mu, T)$  and  $(Y, \nu, S)$  be systems and suppose  $\chi: X \rightarrow Y$  is a factor map. If  $\nu$  is ergodic and  $x \mapsto \mu_x$  is an ergodic decomposition of  $\mu$ , then  $\chi(\mu_x)$  is equal to  $\nu$  for  $\mu$ -almost every  $x \in X$ .*

*Proof.* As  $\chi$  is a factor map and almost-every  $\mu_x$  is  $T$ -invariant, almost every  $\chi(\mu_x)$  is  $S$ -invariant. The fact that

$$\int_X \mu_x d\mu(x) = \mu$$

implies

$$\int_X \chi(\mu_x) d\mu(x) = \nu$$

is an ergodic decomposition of  $\mu$ . Ergodicity of  $\nu$  then implies that  $\chi(\mu_x) = \nu$  for  $\mu$ -almost every  $x \in X$ .  $\square$

We have now assembled the tools to prove Theorem 6.5.

*Proof of Theorem 6.5* Fix a system  $(X, \mu, T)$  with topological pronilfactors and fix  $k \in \mathbb{N}$ . To keep notation simple, and just throughout this proof, we denote by  $(Z, m, T)$  the  $k$ -step pronilfactor of  $(X, \mu, T)$ , and let  $\pi: X \rightarrow Z$  be the continuous factor map. We also write  $\eta = \pi^{[k]} = \pi_k^{[k]}$  for the map  $\eta: X^{[k]} \rightarrow Z^{[k]}$  that applies  $\pi$  coordinatewise.

We begin by noting that if a continuous ergodic decomposition of the system  $(\mathbb{N}^{[k]}(X), \mu^{[k]}, T^{[k]})$  satisfying (6.1) exists, then by Lemma 3.16 applied to the factor map  $\eta$ , it must be unique.

Apply Corollary 6.10 to get a continuous ergodic decomposition  $z \mapsto \xi_z^{[k]}$  from  $\mathbb{Q}^{[k]}(Z)$  to  $\mathcal{M}(\mathbb{Q}^{[k]}(Z))$ . For each  $x \in \mathbb{N}^{[k]}(X)$  define a measure  $\lambda_x^{[k]}$  on  $X^{[k]}$  by

$$(6.3) \quad \int_{X^{[k]}} \bigotimes_{\epsilon \in [k]} f_\epsilon d\lambda_x^{[k]} = \int_{Z^{[k]}} \bigotimes_{\epsilon \in [k]} \mathbb{E}(f_\epsilon | Z) d\xi_{\eta(x)}^{[k]}$$

whenever  $f = (f_\epsilon)_{\epsilon \in [k]} \in C(X)^{[k]}$ . Since the right hand side of (6.3) depends on  $x$  only through  $\eta(x)$ , it is clear that (6.1) holds.

We are left with showing that  $x \mapsto \lambda_x^{[k]}$  is a continuous ergodic decomposition.

Write  $\mathcal{I}^{[k]}$  for the  $\sigma$ -algebra of Borel measurable and  $T^{[k]}$  invariant subsets of  $X^{[k]}$ .

*Claim 1.* For any  $f \in C(X)^{[k]}$  we have

$$(6.4) \quad \mathbb{E}\left(\bigotimes_{\epsilon \in [k]} f_\epsilon | \mathcal{I}^{[k]}\right)(x) = \mathbb{E}\left(\bigotimes_{\epsilon \in [k]} \mathbb{E}(f_\epsilon | Z) \circ \pi | \mathcal{I}^{[k]}\right)(x)$$

for  $\mu^{[k]}$ -almost every  $x \in X^{[k]}$ .

To see this, note that the difference between the two sides of (6.4) can be written as a sum of  $2^k$  terms of the form  $\mathbb{E}(\bigotimes_{\epsilon \in [k]} g_\epsilon | \mathcal{I}^{[k]})$  where at least one of the  $g_\epsilon$  satisfies  $\mathbb{E}(g_\epsilon | Z) = 0$ .

Applying the Cauchy-Schwarz-Gowers inequality [20, Theorem 13, Chapter 8] gives

$$\begin{aligned} \left\| \mathbb{E}\left(\bigotimes_{\epsilon \in [k]} g_\epsilon | \mathcal{I}^{[k]}\right) \right\|_{\mu^{[k]}}^2 &= \left| \int_{X^{[k]}} \mathbb{E}\left(\bigotimes_{\epsilon \in [k]} g_\epsilon | \mathcal{I}^{[k]}\right) \cdot \mathbb{E}\left(\bigotimes_{\epsilon \in [k]} \overline{g_\epsilon} | \mathcal{I}^{[k]}\right) d\mu^{[k]} \right| \\ &= \left| \int_{X^{[k+1]}} \bigotimes_{\epsilon \in [k]} g_\epsilon \otimes \bigotimes_{\epsilon \in [k]} \overline{g_\epsilon} d\mu^{[k+1]} \right| \leq \prod_{\epsilon \in [k]} \|g_\epsilon\|_{k+1}^2 \end{aligned}$$

and note that  $\|g\|_{k+1} = 0$  if and only if  $\mathbb{E}(g | Z) = 0$  by Theorem 5.3 because  $(Z, m, T)$  is the  $k$ -step pronilfactor of  $(X, \mu, T)$ . This completes the proof of Claim 1.

*Claim 2.* Letting  $\mathcal{I}_Z^{[k]}$  denote the  $\sigma$ -algebra of  $T^{[k]}$ -invariant subsets of  $\mathbb{Q}^{[k]}(Z)$ , we have that

$$\mathbb{E}\left(\bigotimes_{\epsilon \in [k]} \mathbb{E}(f_\epsilon | Z) \circ \pi | \mathcal{I}^{[k]}\right) = \mathbb{E}\left(\bigotimes_{\epsilon \in [k]} \mathbb{E}(f_\epsilon | Z) | \mathcal{I}_Z^{[k]}\right) \circ \eta.$$

To check this, note that for any  $g \in L^2(\mathbb{Q}^{[k]}(Z), m^{[k]})$ , the pointwise ergodic theorem and the fact that  $\eta$  is a factor map imply that

$$\mathbb{E}(g \mid \mathcal{I}_Z^{[k]}) \circ \eta = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N g \circ T^n \circ \eta = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N g \circ \eta \circ T^n = \mathbb{E}(g \circ \eta \mid \mathcal{I}^{[k]}).$$

If

$$g = \bigotimes_{\epsilon \in [k]} \mathbb{E}(f_\epsilon \mid Z),$$

then

$$g \circ \eta = \bigotimes_{\epsilon \in [k]} \mathbb{E}(f_\epsilon \mid Z) \circ \pi,$$

completing the proof of Claim [2](#)

Since  $z \mapsto \xi_z^{[k]}$  is an ergodic decomposition of  $m^{[k]}$ , combining Claim [2](#) with Claim [1](#) it follows that the definition in [\(6.3\)](#) of  $\lambda_x^{[k]}$  defines an ergodic decomposition of  $\mu^{[k]}$ .

We are left with showing that the mapping  $x \mapsto \lambda_x^{[k]}$  is continuous, and to do so it suffices to show that for any  $f \in C(X)^{[k]}$  the function

$$x \mapsto \int_{X^{[k]}} \bigotimes_{\epsilon \in [k]} f_\epsilon d\lambda_x^{[k]}$$

is continuous.

*Claim 3.* For  $p = 2^k$ , the function  $\theta: L^p(Z, m)^{[k]} \rightarrow L^\infty(Z, m)$  given by

$$(6.5) \quad \theta(g)(z) = \int_{Z^{[k]}} \bigotimes_{\epsilon \in [k]} g_\epsilon d\xi_z^{[k]}$$

is continuous.

To see this, since  $z \mapsto \xi_z^{[k]}$  is an ergodic decomposition of  $m^{[k]}$ , Lemma [6.11](#) (applied to the projections  $Z^{[k]} \rightarrow Z$ ) together with Lemma [3.16](#) implies that every projection of  $\xi_z^{[k]}$  equals  $m$  for every  $z \in \mathbb{Q}^{[k]}(Z)$ .

Now given  $g, \tilde{g} \in L^p(Z, m)^{[k]}$ , if there exists  $\omega \in [k]$  such that  $g_\epsilon = \tilde{g}_\epsilon$  for all  $\epsilon \in [k] \setminus \{\omega\}$ , then by Holder's inequality

$$\|\theta(g) - \theta(\tilde{g})\|_{L^\infty(m)} \leq \prod_{\epsilon \neq \omega} \|g_\epsilon\|_{L^p(m)} \cdot \|g_\omega - \tilde{g}_\omega\|_{L^p(m)}.$$

This implies that  $\theta$  is continuous, completing the proof of Claim [3](#)

Using the definition [\(6.3\)](#), the fact that  $\mathbb{E}(f_\epsilon \mid Z) \in L^\infty(Z, m) \subset L^p(Z, m)$  and that  $\eta: \mathbb{N}^{[k]}(X) \rightarrow \mathbb{Q}^{[k]}(Z)$  is continuous, it suffices to show that for any  $g \in L^p(Z, m)^{[k]}$  the function  $\theta(g)$  defined by [\(6.5\)](#) is continuous.

Note that for  $g \in C(Z)^{[k]}$  we have  $\theta(g) \in C(Z)$ , since the map  $z \mapsto \xi_z^{[k]}$  is continuous. Therefore, continuity of  $\theta$ , the fact that  $C(Z)$  is a closed subset of  $L^\infty(Z, m)$  and the fact that  $C(Z)^{[k]}$  is dense  $L^p(Z, m)^{[k]}$ , together implies that  $\theta(g) \in C(Z)$  for any  $g \in L^p(Z, m)^{[k]}$ .  $\square$



**Corollary 6.12.** *Let  $(X, \mu, T)$  be an ergodic system with topological pronilfactors, let  $k \in \mathbb{N}$  and let  $x \mapsto \lambda_x^{[k]}$  be the continuous ergodic decomposition of  $(\mathbb{N}^{[k]}(X), \mu^{[k]}, T^{[k]})$  provided by Theorem 6.5. Then*

$$\lambda_{T^{[k]}x}^{[k]} = \lambda_x^{[k]} = T^{[k]}(\lambda_x^{[k]})$$

for every  $x \in \mathbb{N}^{[k]}(X)$ .

*Proof.* We claim that both assignments  $x \mapsto \lambda_{T^{[k]}x}^{[k]}$  and  $x \mapsto T^{[k]}(\lambda_x^{[k]})$  are continuous ergodic decompositions of  $(\mathbb{N}^{[k]}(X), \mu^{[k]}, T^{[k]})$  satisfying (6.1). In view of the almost everywhere uniqueness of ergodic decompositions we can then invoke Lemma 3.16 to conclude that  $\lambda_{T^{[k]}x}^{[k]} = \lambda_x^{[k]} = T^{[k]}(\lambda_x^{[k]})$  for any  $x \in \mathbb{N}^{[k]}(X)$ .

To prove the claim, note that for any continuous function  $f \in C(\mathbb{N}^{[k]}(X))$  we have

$$\int_{\mathbb{N}^{[k]}(X)} f d\lambda_{T^{[k]}x}^{[k]} = \mathbb{E}(f | \mathcal{I}^{[k]})(T^{[k]}x) = \mathbb{E}(f | \mathcal{I}^{[k]})(x)$$

for  $\mu^{[k]}$ -almost every  $x \in X^{[k]}$ , where the last equality follows from the fact that  $\mathcal{I}^{[k]}$  is the  $\sigma$ -algebra of  $T^{[k]}$ -invariant sets. This implies that  $x \mapsto \lambda_{T^{[k]}x}^{[k]}$  is an ergodic decomposition of  $\mu^{[k]}$ . Similarly,

$$\int_{\mathbb{N}^{[k]}(X)} f dT^{[k]}(\lambda_x^{[k]}) = \mathbb{E}(T^{[k]}f | \mathcal{I}^{[k]})(x) = \mathbb{E}(f | \mathcal{I}^{[k]})(x),$$

showing that  $x \mapsto T^{[k]}(\lambda_x^{[k]})$  is also an ergodic decomposition. Continuity follows from the fact that  $x \mapsto \lambda_x^{[k]}$  is continuous and (6.1) follows from the fact that  $\pi: X \rightarrow Z$  is a factor map. This finishes the proof.  $\square$

**Corollary 6.13.** *For every  $x \in \mathbb{N}^{[k]}(X)$  and every coordinate projection  $X^{[k]} \rightarrow X$ , the push forward of  $\lambda_x^{[k]}$  is the measure  $\mu$ .*

In other words, Corollary 6.13 says that for every  $x \in \mathbb{N}^{[k]}(X)$  the measure  $\lambda_x^{[k]}$  is a  $2^k$ -fold self-joining of  $\mu$ .

*Proof.* By Theorem 6.5 we have that  $x \mapsto \lambda_x^{[k]}$  is an ergodic decomposition, and so Lemma 6.11 (applied to the projections  $X^{[k]} \rightarrow X$ ) gives us the statement for  $\mu^{[k]}$ -almost every  $x \in X^{[k]}$ . Combining this fact with (6.1) and using Lemma 3.16 it follows that the conclusion holds for every  $x \in \mathbb{N}^{[k]}(X)$ .  $\square$

## 7. PROOF OF THEOREM 5.9

Throughout this section, whenever we are given an ergodic system  $(X, \mu, T)$  with topological pronilfactors, we use  $\lambda_x^{[k]}$  to denote the measures constructed in Theorem 6.5.

**7.1. An explicit disintegration over the first coordinate.** In this section we use the continuous ergodic decomposition obtained in the previous section to construct a concrete disintegration of  $\mu^{[k]}$  with respect to the projection  $x \mapsto x_{\vec{0}}$  from  $X^{[k]}$  to the first coordinate. The results in this section constitute general versions of Theorem 3.14 and Corollary 3.17. Recall the map  $F^*: X^{[k]} \rightarrow X^{[k]} \setminus \{\vec{0}\}$  defined in (4.1) that forgets the  $\vec{0}$  coordinate, in the sense that  $x = (x_{\vec{0}}, F^*x)$  for every  $x \in X^{[k]} = X \times X^{[k]} \setminus \{\vec{0}\}$ .

**Theorem 7.1.** *Assume that  $(X, \mu, T)$  is an ergodic system with topological pronilfactors. For every  $k \in \mathbb{N}$  and every  $t \in X$ , there exists a measure  $\sigma_t^{[k]}$  on  $X^{[k]}$  such that the following hold.*

(S1). *For all  $t \in X$ , we have that  $\text{supp}(\sigma_t^{[k]}) \subset N^{[k]}(X)$ .*

(S2). *The measures satisfy*

$$(7.1) \quad \sigma_t^{[1]} = \delta_t \times \mu; \quad \sigma_t^{[k+1]} = \int_{N^{[k]}(X)} \delta_x \times \lambda_x^{[k]} d\sigma_t^{[k]}(x)$$

*for all  $t \in X$ .*

(S3). *The map  $t \mapsto \sigma_t^{[k]}$  is both continuous and a disintegration of the measure  $\mu^{[k]}$ .*

(S4). *For all  $t \in X$ , we have that  $\sigma_t^{[k]}(\{x \in X^{[k]} : x_0 = t\}) = 1$ .*

(S5). *For all  $t, s \in X$ , whenever  $\pi_k(t) = \pi_k(s)$  then we have that  $F^* \sigma_t^{[k]} = F^* \sigma_s^{[k]}$ .*

For the remainder of the section, whenever  $(X, \mu, T)$  is an ergodic system with topological pronilfactors, we let  $\sigma_t^{[k]}$  denote the measures that are guaranteed to exist by Theorem 7.1

*Proof.* The proof is by induction on  $k$ . We begin by defining  $\sigma_t^{[1]} = \delta_t \times \mu$  for all  $t \in X$ . It is straightforward to verify  $t \mapsto \sigma_t^{[1]}$  is a continuous map from  $X$  to  $\mathcal{M}(X^{[1]})$  and a disintegration of  $\mu^{[1]} = \mu \times \mu$ . We also have  $x_0 = t$  for  $\sigma_t^{[1]}$ -almost every point and  $F^* \sigma_t^{[1]} = \mu$  for all  $t \in X$ . It follows from Lemma 5.13 that  $N^{[1]}(X) = X^{[1]}$  for all  $t \in X$ . Thus (S1) through (S5) hold when  $k = 1$ .

Suppose by induction that for some  $k \geq 1$  we have a measure  $\sigma_t^{[k]} \in \mathcal{M}(X^{[k]})$  for every  $t \in X$  satisfying properties (S1) through (S5). We may then define

$$\sigma_t^{[k+1]} = \int_{N^{[k]}(X)} \delta_x \times \lambda_x^{[k]} d\sigma_t^{[k]}(x)$$

for every  $t \in X$ . This establishes (S2) for  $k + 1$ , and we are left with verifying the other properties.

Continuity of  $x \mapsto \lambda_x^{[k]}$  and  $t \mapsto \sigma_t^{[k]}$  gives continuity of  $t \mapsto \sigma_t^{[k+1]}$ . It is a disintegration of  $\mu^{[k+1]}$  by the calculation

$$\begin{aligned} \int_X \sigma_t^{[k+1]} d\mu(t) &= \int_X \int_{N^{[k]}(X)} \delta_x \times \lambda_x^{[k]} d\sigma_t^{[k]}(x) d\mu(t) \\ &= \int_{N^{[k]}(X)} \delta_x \times \lambda_x^{[k]} d\mu^{[k]}(x) \\ &= \mu^{[k+1]} \end{aligned}$$

which uses induction for the second equality and Lemma 5.2 for the third. This shows that (S3) holds.

To prove (S4) note that, since  $\sigma_t^{[k]}$ -almost every point  $x$  satisfies  $x_0 = t$ , we can write

$$\sigma_t^{[k+1]} = \delta_t \times \int_{N^{[k]}(X)} \delta_{F^*x} \times \lambda_x^{[k]} d\sigma_t^{[k]}(x)$$

which implies  $\sigma_t^{[k+1]}$ -almost every point  $x$  satisfies  $x_0 = t$ .

Next we establish (S5). Since  $\sigma_t^{[k]}$ -almost every point  $x \in \mathbb{N}^{[k]}(X)$  can be written as  $x = (t, F^*x)$ , we have

$$F^*\sigma_t^{[k+1]} = \int_{\mathbb{N}^{[k]}(X)} \delta_{F^*x} \times \lambda_{(t, F^*x)}^{[k]} d\sigma_t^{[k]}(x) = \int_{F^*(\mathbb{N}^{[k]}(X))} \delta_y \times \lambda_{(t, y)}^{[k]} d(F^*\sigma_t^{[k]})(y).$$

Using (6.1), it follows that if  $\pi_k(t) = \pi_k(s)$  then  $\lambda_{(t, y)}^{[k]} = \lambda_{(s, y)}^{[k]}$  whenever  $y \in X^{[k]} \setminus \{\vec{0}\}$  and  $(t, y) \in \mathbb{N}^{[k]}(X)$ . Using the induction hypothesis it follows that  $F^*\sigma_t^{[k+1]} = F^*\sigma_s^{[k+1]}$ , establishing (S5).

Finally, we need to verify that (S1) holds. Since  $\mu^{[k+1]}(\mathbb{N}^{[k+1]}(X)) = 1$  by Lemma 5.14, the fact that  $t \mapsto \sigma_t^{[k+1]}$  is a disintegration of  $\mu^{[k+1]}$  gives that  $\sigma_t^{[k]}(\mathbb{N}^{[k+1]}(X)) = 1$  for  $\mu$ -almost every  $t$ . Now Lemma 5.14 also states  $\mathbb{N}^{[k+1]}(X)$  is closed so

$$\{\nu \in \mathcal{M}(X^{[k+1]}) : \nu(\mathbb{N}^{[k+1]}(X)) = 1\}$$

is closed. Continuity of  $t \mapsto \sigma_t^{[k+1]}$  then gives  $\sigma_t^{[k]}(\mathbb{N}^{[k+1]}(X)) = 1$  for every  $t \in \text{supp}(\mu)$ . We conclude that the set

$$\{t \in X : \sigma_t^{[k+1]}(\mathbb{N}^{[k+1]}(X)) = 1\}$$

satisfies the hypothesis of Lemma 3.16 and is therefore equal to  $X$ . This establishes (S1).  $\square$

**Corollary 7.2.** *Assume  $(X, \mu, T)$  is an ergodic system with topological pronilfactors. For every  $t \in X$  and every coordinate projection  $X^{[k]} \rightarrow X$  other than projection to  $\vec{0}$ , the push forward of  $\sigma_t^{[k]}$  equals  $\mu$ .*

*Proof.* This follows by combining the definition (7.1) of  $\sigma_t^{[k]}$  with Corollary 6.13.  $\square$

**Corollary 7.3.** *Assume  $(X, \mu, T)$  is an ergodic system with topological pronilfactors. For every  $k \in \mathbb{N}$ , every permutation  $\phi$  on  $\{1, \dots, k\}$ , and every  $t \in X$  we have  $\phi\sigma_t^{[k]} = \sigma_t^{[k]}$  where  $\phi\sigma_t^{[k]}$  is the push forward of  $\sigma_t^{[k]}$  under the map defined in (4.2).*

*Proof.* Fix  $k \in \mathbb{N}$  and fix a permutation  $\phi$  of  $k$  digits. Let  $\tilde{\sigma}_t^{[k]} := \phi\sigma_t^{[k]}$  for every  $t \in X$ . We claim that the measures  $\tilde{\sigma}_t^{[k]}$  satisfy properties (S3) through (S5) of Theorem 7.1.

Recall from Lemma 5.1 that  $\phi\mu^{[k]} = \mu^{[k]}$ ; therefore  $t \mapsto \tilde{\sigma}_t^{[k]}$  is a disintegration of  $\mu^{[k]}$  and we obtain (S3) for  $\tilde{\sigma}_t^{[k]}$ . Since the set  $\{x \in \mathbb{N}^{[k]}(X) : x_{\vec{0}} = t\}$  is invariant under  $\phi$ , we immediately obtain (S4) for  $\tilde{\sigma}_t^{[k]}$  as well. Finally, since the map  $\phi: X^{[k]} \rightarrow X^{[k]}$  does not involve the first coordinate, we have that  $F^*\rho = F^*\tilde{\rho}$  if and only if  $F^*\phi\rho = F^*\phi\tilde{\rho}$  for any measures  $\rho$  and  $\tilde{\rho}$  on  $X^{[k]}$ . Thus in particular we deduce that the measures  $\tilde{\sigma}_t^{[k]}$  also satisfy (S5).

In view of conditions (S3) and (S4), both  $t \mapsto \sigma_t^{[k]}$  and  $t \mapsto \tilde{\sigma}_t^{[k]}$  satisfy Theorem 2.5. From the uniqueness in that theorem it follows that  $\sigma_t^{[k]} = \tilde{\sigma}_t^{[k]}$  for  $\mu$ -almost every  $t \in X$ . Using continuity from (S3) and combining it with (S5) we can now apply Lemma 3.16 to conclude that  $\sigma_t^{[k]} = \tilde{\sigma}_t^{[k]}$  for all  $t \in X$ .  $\square$

We also need equivariance of the disintegration  $t \mapsto \sigma_t^{[k]}$  with respect to  $T^{[k]}$ .

**Corollary 7.4.** *Assume  $(X, \mu, T)$  is an ergodic system with topological pronilfactors. For every  $k \in \mathbb{N}$  and every  $t \in X$ , we have  $T^{\llbracket k \rrbracket} \sigma_t^{\llbracket k \rrbracket} = \sigma_{T(t)}^{\llbracket k \rrbracket}$ .*

*Proof.* The case  $k = 1$  is immediate as

$$T^{\llbracket 1 \rrbracket} \sigma_t^{\llbracket 1 \rrbracket} = T^{\llbracket 1 \rrbracket}(\delta_t \times \mu) = \delta_{T(t)} \times \mu = \sigma_{T(t)}^{\llbracket 1 \rrbracket}$$

for all  $t \in X$ . The proof in general is by induction, using the calculation

$$\begin{aligned} T^{\llbracket k+1 \rrbracket} \sigma_t^{\llbracket k+1 \rrbracket} &= \int_{\mathbb{N}^{\llbracket k+1 \rrbracket}(X)} T^{\llbracket k \rrbracket} \delta_x \times T^{\llbracket k \rrbracket} \lambda_x^{\llbracket k \rrbracket} d\sigma_t^{\llbracket k \rrbracket}(x) \\ &= \int_{\mathbb{N}^{\llbracket k+1 \rrbracket}(X)} \delta_{T^{\llbracket k \rrbracket}(x)} \times \lambda_{T^{\llbracket k \rrbracket}(x)}^{\llbracket k \rrbracket} d\sigma_t^{\llbracket k \rrbracket}(x) \\ &= \int_{\mathbb{N}^{\llbracket k+1 \rrbracket}(X)} \delta_x \times \lambda_x^{\llbracket k \rrbracket} d(T^{\llbracket k \rrbracket} \sigma_t^{\llbracket k \rrbracket})(x) = \sigma_{T(t)}^{\llbracket k+1 \rrbracket} \end{aligned}$$

which employs Corollary 6.12  $\square$

**7.2. The existence of Erdős cubes.** We start by reformulating Theorem 5.9 in terms of the measure  $\sigma_a^{\llbracket k \rrbracket}$ .

**Theorem 7.5.** *Assume that  $(X, \mu, T)$  is an ergodic system with topological pronilfactors and that  $a \in \text{gen}(\mu, \Phi)$  for some Følner sequence  $\Phi$ . For every  $k \in \mathbb{N}$ ,  $\sigma_a^{\llbracket k \rrbracket}$ -almost every  $x \in \mathbb{N}^{\llbracket k \rrbracket}(X)$  is a  $k$ -dimensional Erdős cube.*

*Proof that Theorem 7.5 implies Theorem 5.9.* Fix an ergodic system  $(X, \mu, T)$  with topological pronilfactors, let  $a \in \text{gen}(\mu, \Phi)$  for some Følner sequence  $\Phi$ , let  $E \subset X$  with  $\mu(E) > 0$  and let  $k \in \mathbb{N}$ . By Theorem 7.5 we have that  $\sigma_a^{\llbracket k \rrbracket}$ -almost every  $x \in \mathbb{N}^{\llbracket k \rrbracket}(X)$  is a  $k$ -dimensional Erdős cube. Since  $\sigma_a^{\llbracket k \rrbracket}$ -almost every point satisfies  $x_{\vec{0}} = a$  by (S4) and  $\sigma_a^{\llbracket k \rrbracket}$  pushes forward to  $\mu$  on the last coordinate by Corollary 7.2 we can in particular choose a  $k$ -dimensional Erdős cube with  $x_{\vec{0}} = a$  and  $x_{\vec{1}} \in E$ .  $\square$

The remainder of this section is devoted to the proof of Theorem 7.5 and for this we need one last ingredient: that  $\sigma_a^{\llbracket k \rrbracket}$ -almost every point  $x$  is generic for  $\lambda_x^{\llbracket k \rrbracket}$  along some Følner sequence. This is the only place in the proof of Theorem 4.4 where the assumption  $a \in \text{gen}(\mu, \Phi)$  is used.

**Theorem 7.6.** *Assume that  $(X, \mu, T)$  is an ergodic system with topological pronilfactors and that  $a \in \text{gen}(\mu, \Phi)$  for some Følner sequence  $\Phi$ . For each  $k \in \mathbb{N}$ , there exists a Følner sequence  $\Psi$  such that*

$$\sigma_a^{\llbracket k \rrbracket}(\{x \in \mathbb{N}^{\llbracket k \rrbracket}(X) : x \in \text{gen}(\lambda_x^{\llbracket k \rrbracket}, \Psi)\}) = 1$$

*holds.*

*Proof.* Using Corollary 2.9 and Property (S3) of Theorem 7.1, we get

$$(7.2) \quad \sigma_b^{\llbracket k \rrbracket}(\{x \in \mathbb{N}^{\llbracket k \rrbracket}(X) : x \in \text{gen}(\lambda_x^{\llbracket k \rrbracket})\}) = 1,$$

for  $\mu$ -almost every  $b \in X$ . In particular, we can fix  $b \in \text{supp}(\mu)$  satisfying (7.2). By Lemma 2.4 we can choose a sequence  $(s(n))_{n \in \mathbb{N}}$  of integers such that  $T^{s(n)} a \rightarrow b$  as  $n \rightarrow \infty$ . As the map  $t \mapsto \sigma_t^{\llbracket k \rrbracket}$  is continuous, it follows that we have the convergence of the measures

$$(7.3) \quad (T^{\llbracket k \rrbracket})^{s(n)} \sigma_a^{\llbracket k \rrbracket} \rightarrow \sigma_b^{\llbracket k \rrbracket} \quad \text{as } n \rightarrow \infty$$

via Corollary [7.4](#)

Fix a sequence of functions  $G_1, G_2, \dots \in C(X^{\llbracket k \rrbracket})$  that is dense in  $C(X^{\llbracket k \rrbracket})$  with respect to the supremum norm. From [\(7.2\)](#) the set of points  $x \in \mathbb{N}^{\llbracket k \rrbracket}(X)$  that are generic for  $\lambda_x^{\llbracket k \rrbracket}$  has full  $\sigma_b^{\llbracket k \rrbracket}$  measure. By the monotone convergence theorem, for every  $m \in \mathbb{N}$  there exists  $N(m) \in \mathbb{N}$  such that the set

$$A_m^{(1)} = \left\{ x \in \mathbb{N}^{\llbracket k \rrbracket}(X) : \max_{1 \leq j \leq m} \left| \frac{1}{N(m)} \sum_{n=1}^{N(m)} G_j((T^{\llbracket k \rrbracket})^n x) - \int_{X^{\llbracket k \rrbracket}} G_j d\lambda_x^{\llbracket k \rrbracket} \right| \leq \frac{1}{m} \right\}$$

satisfies  $\sigma_b^{\llbracket k \rrbracket}(A_m^{(1)}) \geq 1 - 1/2^m$ .

Let  $\varepsilon_m > 0$  and take  $A_m^{(2)}$  to be the set of points in  $X^{\llbracket k \rrbracket}$  whose distance from some point in  $A_m^{(1)}$  is strictly less than  $\varepsilon_m$ . Choosing  $\varepsilon_m$  sufficiently small, it follows that

$$\max_{1 \leq j \leq m} \left| \frac{1}{N(m)} \sum_{n=1}^{N(m)} G_j((T^{\llbracket k \rrbracket})^n x) - \int_{X^{\llbracket k \rrbracket}} G_j d\lambda_x^{\llbracket k \rrbracket} \right| \leq \frac{2}{m}$$

for all  $x \in A_m^{(2)}$ . Since  $A_m^{(1)}$  is closed and  $A_m^{(2)}$  is open (both as subsets of  $X^{\llbracket k \rrbracket}$ ), by Urysohn's lemma there exists a continuous function  $\phi_m: X^{\llbracket k \rrbracket} \rightarrow [0, 1]$  satisfying

$$\phi_m(x) = \begin{cases} 1 & \text{for all } x \in A_m^{(1)}, \\ 0 & \text{for all } x \notin A_m^{(2)}. \end{cases}$$

Passing to a subsequence of  $(s(n))_{n \in \mathbb{N}}$  if needed and using the convergence of the measures in [\(7.3\)](#), we can further assume that

$$\left| \int_{X^{\llbracket k \rrbracket}} (T^{\llbracket k \rrbracket})^{s(n)} \phi_m d\sigma_a^{\llbracket k \rrbracket} - \int_{X^{\llbracket k \rrbracket}} \phi_m d\sigma_b^{\llbracket k \rrbracket} \right| \leq \frac{1}{2^m}.$$

It follows that

$$\begin{aligned} \sigma_a^{\llbracket k \rrbracket}((T^{\llbracket k \rrbracket})^{-s(m)} A_m^{(2)}) &\geq \int_{X^{\llbracket k \rrbracket}} (T^{\llbracket k \rrbracket})^{s(m)} \phi_m d\sigma_a^{\llbracket k \rrbracket} \geq \int_{X^{\llbracket k \rrbracket}} \phi_m d\sigma_b^{\llbracket k \rrbracket} - \frac{1}{2^m} \\ &\geq \sigma_b^{\llbracket k \rrbracket}(A_m^{(1)}) - \frac{1}{2^m} \geq 1 - \frac{1}{2^{m-1}}. \end{aligned}$$

Let  $Y$  denote the set of all points in the support of  $\sigma_a^{\llbracket k \rrbracket}$  that belong to all but finitely many  $(T^{\llbracket k \rrbracket})^{-s(m)} A_m^{(2)}$ ,  $m \in \mathbb{N}$ . By the Borel-Cantelli lemma,  $Y$  has full  $\sigma_a^{\llbracket k \rrbracket}$ -measure. For any  $x \in Y$  and cofinitely many  $m \in \mathbb{N}$ , we have

$$\max_{1 \leq j \leq m} \left| \frac{1}{N(m)} \sum_{n=1}^{N(m)} G_j((T^{\llbracket k \rrbracket})^{n+s(m)} x) - \int_{X^{\llbracket k \rrbracket}} G_j d\lambda_{(T^{\llbracket k \rrbracket})^{s(m)} x}^{\llbracket k \rrbracket} \right| \leq \frac{3}{m}.$$

Since  $\lambda_{(T^{\llbracket k \rrbracket})^{s(m)} x}^{\llbracket k \rrbracket} = \lambda_x^{\llbracket k \rrbracket}$  by Corollary [6.12](#) the conclusion follows with the Følner sequence  $\Psi_m = \{s(m), s(m) + 1, \dots, s(m) + N(m)\}$ .  $\square$

We can now prove Theorem [7.5](#). In the language of Section [5.4](#) we show that  $\sigma_a^{\llbracket k \rrbracket}(\mathbb{E}^{\llbracket k \rrbracket}(X)) = 1$ . Although this may appear similar to [\(S1\)](#) in Theorem [7.1](#) we cannot prove it so simply because the set  $\mathbb{E}^{\llbracket k \rrbracket}(X)$  is in general not closed.

*Proof of Theorem [7.5](#)* We must show that  $\sigma_a^{\llbracket k \rrbracket}$ -almost every point  $x$  is a  $k$ -dimensional Erdős cube. In view of Lemma [4.2](#) and Corollary [7.3](#) it suffices to show that for  $\sigma_a^{\llbracket k \rrbracket}$ -almost every  $(x, y) \in \mathbb{N}^{\llbracket k \rrbracket}(X)$  we have  $y \in \omega(x, T^{\llbracket k-1 \rrbracket})$ . Using Lemma [2.4](#)

the proof is finished if we show that  $\sigma_a^{[k]}$ -almost every  $(x, y) \in \mathbb{N}^{[k]}(X)$  satisfies both  $y \in \text{supp}(\lambda_x^{[k-1]})$  and  $x \in \text{gen}(\lambda_x^{[k-1]}, \Psi)$  for a Følner sequence  $\Psi$ .

Since  $\lambda_x^{[k-1]}(\text{supp}(\lambda_x^{[k-1]})) = 1$  for every  $x \in \mathbb{N}^{[k-1]}$  it follows that the set

$$A_1 = \{(w, y) \in \mathbb{N}^{[k]}(X) : y \in \text{supp}(\lambda_w^{[k-1]})\}$$

satisfies  $(\delta_x \times \lambda_x^{[k-1]})(A_1) = 1$  for every  $x \in \mathbb{N}^{[k-1]}(X)$ . From the construction in (7.1) of  $\sigma_a^{[k]}$  we conclude that  $\sigma_a^{[k]}(A_1) = 1$ , and hence that  $\sigma_a^{[k]}$ -almost every  $(x, y) \in \mathbb{N}^{[k]}(X)$  satisfies  $y \in \text{supp}(\lambda_x^{[k-1]})$ .

Finally, we apply Theorem 7.6 (with  $k-1$  in the place of  $k$ ) to find a Følner sequence  $\Psi$  such that the set

$$A_2 = \{x \in \mathbb{N}^{[k-1]}(X) : x \in \text{gen}(\lambda_x^{[k-1]}, \Psi)\}$$

has  $\sigma_a^{[k-1]}(A_2) = 1$ . Using again (7.1) we conclude that  $\sigma_a^{[k]}(A_2 \times X^{[k-1]}) = 1$ , and hence that  $\sigma_a^{[k]}$ -almost every  $(x, y) \in \mathbb{N}^{[k]}(X)$  satisfies  $x \in \text{gen}(\lambda_x^{[k-1]}, \Psi)$ . This concludes the proof.  $\square$

## 8. OPEN QUESTIONS

**8.1. Combinatorics.** Our first question is a natural extension of our main result.

**Question 8.1.** Given  $A \subset \mathbb{N}$  with positive upper Banach density, are there infinite sets  $B_1, B_2, B_3, \dots$  such that for every  $k \in \mathbb{N}$  we have  $B_1 + \dots + B_k \subset A$ ?

We note that the same question with each  $B_i$  being finite (even if arbitrarily large) is easy: simply iterate the fact that whenever  $d(A) > 0$ , for any  $m \in \mathbb{N}$  we can find integers  $n_1, \dots, n_m$  such that  $d((A - n_1) \cap \dots \cap (A - n_m)) > 0$ .

On the other hand, it is important that we only consider the initial sums  $B_1 + \dots + B_k$ , as parity issues prevent more general sums, such as  $B_i + \dots + B_k$ , to being all contained in  $A$  (this can be readily seen by taking  $A$  to be the set of odd numbers and each  $B_j$  to be a singleton). While this simple example illustrates this phenomenon convincingly, one can also fabricate examples which are aperiodic, and in general it is not clear what are the exact obstructions. The answer may depend on the precise pattern we are trying to find, so we formulate only a simple case of this general question.

**Question 8.2.** Which sets  $A \subset \mathbb{N}$  with  $d_\Phi(A) > 0$  for some Følner sequence  $\Phi$  contain  $B_1 \cup B_2 \cup (B_1 + B_2)$  for some infinite sets  $B_1, B_2 \subset \mathbb{N}$ ?

For example, does it suffice that  $1_A - d_\Phi(A)$  is a weak mixing function along  $\Phi$  in the sense of [26] Definition 3.17?

Under the assumption of the Hardy-Littlewood  $k$ -tuplets prime conjecture, Granville [15] shows that the primes contain  $B_1 + B_2$  for infinite sets  $B_1, B_2 \subset \mathbb{N}$ , but an unconditional proof is still unknown (see [26] Question 6.4). Keeping in mind the heuristic that the primes behave pseudo-randomly, we record the following.

**Question 8.3.** Given  $A \subset \mathbb{N}$ , which notions of pseudo-randomness suffice for  $A$  to contain  $B_1 + B_2$  for infinite sets  $B_1, B_2 \subset \mathbb{N}$ ?

More generally, which other notions of largeness on a set  $A$  imply that it contains a sumset  $B + C$ ?

Question 8.4 is inspired by the main result from [25].

**Question 8.4.** Given a finite partition  $\mathbb{N} = A_1 \cup \dots \cup A_r$ , are there infinite sets  $B_1, B_2 \subset \mathbb{N}$  and a color  $A \in \{A_1, \dots, A_r\}$  such that  $(B_1 + B_2) \cup (B_1 B_2) \subset A$ ?

Question 8.4 was also asked by Erdős in [8, Page 58]. An affirmative answer to the significantly weaker question where  $|B_1| = |B_2| = 1$  is given in [25]. The case  $|B_1| = 2$  and  $|B_2| = 1$  also follows from the results in that paper, but already the case  $|B_1| = |B_2| = 2$  seems to be out of reach.

**8.2. Dynamics.** Lemma 6.7 asserts that any nilsystems possesses a continuous ergodic decomposition. This naturally leads to Question 8.5

**Question 8.5.** What other classes of systems always possesses a continuous ergodic decomposition?

A natural class to consider, which extends the class of nilsystems, is distal systems. However, for distal systems the answer to Question 8.5 turns out to be negative, as Furstenberg [10] exhibited a skew-product  $(x, y) \mapsto (x + \alpha, y + h(x))$  which is topologically minimal but not uniquely ergodic. Taking a non-ergodic measure on such system, there can be no continuous ergodic decomposition, since the only continuous invariant functions are constant, contradicting Lemma 6.2

Another natural class to consider, which extends the class of nilsystems, is the class of horocycle flows on homogeneous spaces. Given their algebraic origin, it is possible that a continuous ergodic decomposition always exists.

**8.3. Amenable groups.** The proof given in [26] for the two-fold sumset holds in an arbitrary amenable group. While our proof of Theorem 1.1 does not hold in this generality due to the reliance on the structure theory for cube systems, the new proof of the case  $k = 2$  in Section 3 does, and may lead to the answers of some questions in this direction that have previously been out of reach.

Recall that a countable group  $G$  is *amenable* if there is a sequence  $(\Phi_N)_{N \in \mathbb{N}}$  of finite subsets of  $G$  such that

$$(8.1) \quad \lim_{N \rightarrow \infty} \frac{|\Phi_N g \cap \Phi_N|}{|\Phi_N|} = 1,$$

$$(8.2) \quad \lim_{N \rightarrow \infty} \frac{|\Phi_N \cap g \Phi_N|}{|\Phi_N|} = 1$$

both hold for all  $g \in G$ . Any such sequence is called a *two-sided Følner sequence* on  $G$ . In [26, Theorem 1.3] it was proved that if  $A \subset G$  satisfies

$$(8.3) \quad \limsup_{N \rightarrow \infty} \frac{|A \cap \Phi_N|}{|\Phi_N|} > 0$$

for a two-sided Følner sequence  $(\Phi_N)_{N \in \mathbb{N}}$ , then  $A$  contains the product  $BC = \{bc : b \in B, c \in C\}$  of two infinite sets  $B, C \subset G$ . It is natural to ask whether the same is true for  $k$ -fold products.

**Question 8.6.** Let  $G$  be an amenable group, let  $k \in \mathbb{N}$  and let  $A \subset G$  satisfy (8.3) with respect to some two-sided Følner sequence. Do there always exist infinite sets  $B_1, \dots, B_k \subset G$  such that  $B_1 \cdots B_k \subset A$ ?

In non-commutative groups we are interested in two-sided versions of Question 8.6 and which of the assumptions (8.1), (8.2) are necessary for various types of product sets to be found in all positive density sets.



**Question 8.7.** Let  $G$  be an amenable group and let  $A \subset G$  satisfy (8.3) with respect to some two-sided Følner sequence. Does  $A$  necessarily contain  $BC \cup CB$  for infinite sets  $B, C \subset G$ ?

**Question 8.8.** Let  $G$  be an amenable group. Which of the assumptions (8.1), (8.2) on a sequence  $(\Phi_N)_{N \in \Phi}$  of finite subsets of  $G$  guarantee every  $A \subset G$  satisfying (8.3) contains a product set  $BC$  for infinite sets  $B, C \subset G$ ?

**8.4. Ultrafilters.** Ultrafilters on the natural numbers were a key tool in [26] and so we enquire here about a strengthening of Theorem 3.1 in those terms. One can use ultrafilters on  $\mathbb{N}$  to characterize when a set  $A \subset \mathbb{N}$  contains  $B+C$  with  $B, C \subset \mathbb{N}$  infinite in terms of the associative operation

$$\mathfrak{p} + \mathfrak{q} = \{A \subset \mathbb{N} : \{n \in \mathbb{N} : A - n \in \mathfrak{q}\} \in \mathfrak{p}\}$$

on the set of all ultrafilters. Indeed, the set  $A \subset \mathbb{N}$  contains a sumset if and only if one can find non-principal ultrafilters  $\mathfrak{p}$  and  $\mathfrak{q}$  with  $A$  belonging to both  $\mathfrak{p} + \mathfrak{q}$  and  $\mathfrak{q} + \mathfrak{p}$ . We remark that, using ultrafilter limits, the notion of an Erdős cube arises naturally from this characterization.

**Question 8.9.** Is it true that for any  $A \subset \mathbb{N}$  with positive upper Banach density there exist non-principal ultrafilters  $\mathfrak{p}$  and  $\mathfrak{q}$  with  $\mathfrak{p} + \mathfrak{q} = \mathfrak{q} + \mathfrak{p}$  containing  $A$ ?

One could generalize Question 8.9 to ask about a strengthening of Theorem 1.1 in terms of ultrafilters, and even further to generalizations of the questions in Section 8.3.

#### ACKNOWLEDGMENTS

We thank Dimitrios Charamaras, Andreas Mountakis, Song Shao, Terence Tao and Xiangdong Ye for helpful comments on a preliminary version of this paper, and we are grateful to the anonymous referees for their detailed reports.

#### REFERENCES

- [1] Vitaly Bergelson, *Sets of recurrence of  $\mathbf{Z}^m$ -actions and properties of sets of differences in  $\mathbf{Z}^m$* , J. London Math. Soc. (2) **31** (1985), no. 2, 295–304, DOI 10.1112/jlms/s2-31.2.295. MR809951
- [2] Vitaly Bergelson, Bernard Host, and Bryna Kra, *Multiple recurrence and nilsequences*, Invent. Math. **160** (2005), no. 2, 261–303, DOI 10.1007/s00222-004-0428-6. With an appendix by Imre Ruzsa. MR2138068
- [3] V. Bergelson and A. Leibman, *Polynomial extensions of van der Waerden’s and Szemerédi’s theorems*, J. Amer. Math. Soc. **9** (1996), no. 3, 725–753, DOI 10.1090/S0894-0347-96-00194-4. MR1325795
- [4] Fangzhou Cai and Song Shao, *Topological characteristic factors along cubes of minimal systems*, Discrete Contin. Dyn. Syst. **39** (2019), no. 9, 5301–5317, DOI 10.3934/dcds.2019216. MR3986331
- [5] Mauro Di Nasso, Isaac Goldbring, Renling Jin, Steven Leth, Martino Lupini, and Karl Mahlborg, *On a sumset conjecture of Erdős*, Canad. J. Math. **67** (2015), no. 4, 795–809, DOI 10.4153/CJM-2014-016-0. MR3361013
- [6] Fabien Durand, Alexander Frank, and Alejandro Maass, *Eigenvalues of minimal Cantor systems*, J. Eur. Math. Soc. (JEMS) **21** (2019), no. 3, 727–775, DOI 10.4171/JEMS/849. MR3908764
- [7] P. Erdős and R. L. Graham, *Old and new problems and results in combinatorial number theory*, Monographies de L’Enseignement Mathématique [Monographs of L’Enseignement Mathématique], vol. 28, Université de Genève, L’Enseignement Mathématique, Geneva, 1980. MR592420



- [8] Paul Erdős, *Problems and results on combinatorial number theory. III*, Number theory day (Proc. Conf., Rockefeller Univ., New York, 1976), Lecture Notes in Math., Vol. 626, Springer, Berlin, 1977, pp. 43–72. MR0472752
- [9] Manfred Einsiedler and Thomas Ward, *Ergodic theory with a view towards number theory*, Graduate Texts in Mathematics, vol. 259, Springer-Verlag London, Ltd., London, 2011, DOI 10.1007/978-0-85729-021-2. MR2723325
- [10] H. Furstenberg, *Strict ergodicity and transformation of the torus*, Amer. J. Math. **83** (1961), 573–601, DOI 10.2307/2372899. MR133429
- [11] H. Furstenberg, *The structure of distal flows*, Amer. J. Math. **85** (1963), 477–515, DOI 10.2307/2373137. MR157368
- [12] Harry Furstenberg, *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions*, J. Analyse Math. **31** (1977), 204–256, DOI 10.1007/BF02813304. MR498471
- [13] H. Furstenberg, *Recurrence in ergodic theory and combinatorial number theory*, Princeton University Press, Princeton, N.J., 1981. M. B. Porter Lectures. MR603625
- [14] H. Furstenberg and Y. Katznelson, *An ergodic Szemerédi theorem for commuting transformations*, J. Analyse Math. **34** (1978), 275–291 (1979), DOI 10.1007/BF02790016. MR531279
- [15] Andrew Granville, *A note on sums of primes*, Canad. Math. Bull. **33** (1990), no. 4, 452–454, DOI 10.4153/CMB-1990-073-7. MR1091350
- [16] Ben Green and Terence Tao, *The primes contain arbitrarily long arithmetic progressions*, Ann. of Math. (2) **167** (2008), no. 2, 481–547, DOI 10.4007/annals.2008.167.481. MR2415379
- [17] Bernard Host, *A short proof of a conjecture of Erdős proved by Moreira, Richter and Robertson*, Discrete Anal., posted on 2019, Paper No. 19, 10, DOI 10.19086/da. MR4042162
- [18] Bernard Host and Bryna Kra, *Nonconventional ergodic averages and nilmanifolds*, Ann. of Math. (2) **161** (2005), no. 1, 397–488, DOI 10.4007/annals.2005.161.397. MR2150389
- [19] Bernard Host and Bryna Kra, *Uniformity seminorms on  $\ell^\infty$  and applications*, J. Anal. Math. **108** (2009), 219–276, DOI 10.1007/s11854-009-0024-1. MR2544760
- [20] Bernard Host and Bryna Kra, *Nilpotent structures in ergodic theory*, Mathematical Surveys and Monographs, vol. 236, American Mathematical Society, Providence, RI, 2018, DOI 10.1090/surv/236. MR3839640
- [21] Bernard Host, Bryna Kra, and Alejandro Maass, *Nilsequences and a structure theorem for topological dynamical systems*, Adv. Math. **224** (2010), no. 1, 103–129, DOI 10.1016/j.aim.2009.11.009. MR2600993
- [22] Ehud Lehrer, *Topological mixing and uniquely ergodic systems*, Israel J. Math. **57** (1987), no. 2, 239–255, DOI 10.1007/BF02772176. MR890422
- [23] A. Leibman, *Orbits on a nilmanifold under the action of a polynomial sequence of translations*, Ergodic Theory Dynam. Systems **27** (2007), no. 4, 1239–1252, DOI 10.1017/S0143385706000940. MR2342974
- [24] Elon Lindenstrauss, *Pointwise theorems for amenable groups*, Invent. Math. **146** (2001), no. 2, 259–295, DOI 10.1007/s002220100162. MR1865397
- [25] Joel Moreira, *Monochromatic sums and products in  $\mathbb{N}$* , Ann. of Math. (2) **185** (2017), no. 3, 1069–1090, DOI 10.4007/annals.2017.185.3.10. MR3664819
- [26] Joel Moreira, Florian K. Richter, and Donald Robertson, *A proof of a sumset conjecture of Erdős*, Ann. of Math. (2) **189** (2019), no. 2, 605–652, DOI 10.4007/annals.2019.189.2.4. MR3919363
- [27] Melvyn B. Nathanson, *Sumsets contained in infinite sets of integers*, J. Combin. Theory Ser. A **28** (1980), no. 2, 150–155, DOI 10.1016/0097-3165(80)90081-3. MR563552
- [28] E. Szemerédi, *On sets of integers containing no  $k$  elements in arithmetic progression*, Acta Arith. **27** (1975), 199–245, DOI 10.4064/aa-27-1-199-245. MR369312
- [29] Siming Tu and Xiangdong Ye, *Dynamical parallelepipeds in minimal systems*, J. Dynam. Differential Equations **25** (2013), no. 3, 765–776, DOI 10.1007/s10884-013-9313-6. MR3095274

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

*Email address:* `kra@math.northwestern.edu`

MATHEMATICS INSTITUTE, ZEEMAN BUILDING, UNIVERSITY OF WARWICK, COVENTRY, CV4 7AL, UK

*Email address:* `joel.moreira@warwick.ac.uk`

INSTITUTE OF MATHEMATICS, ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE (EPFL), STATION 8, 1015 LAUSANNE, SWITZERLAND

*Email address:* `f.richter@epfl.ch`

DEPARTMENT OF MATHEMATICS, ALAN TURING BUILDING, UNIVERSITY OF MANCHESTER, MANCHESTER, M13 9PL, UK

*Email address:* `donald.robertson@manchester.ac.uk`