Bergelson’s polynomial averages

Bogdan Grechuk • 22 Jun 2026

The 2024 preprint of Kosz, Mirek, Peluse, Wan and Wright, now accepted by the *Annals of Mathematics*, gives a major new answer to a question of Bergelson on pointwise convergence of multilinear polynomial ergodic averages (Kosz et al. 2024). In the distinct-degree case, they prove pointwise almost everywhere convergence for an arbitrary number of commuting transformations, and in fact obtain a stronger quantitative theorem: convergence holds for functions in the natural Hölder range $L^{p_1} \times \cdots \times L^{p_d}$.

Let $(X, \mathcal{B}, \mu)$ be a probability space, so $\mu(X) = 1$, and let $T : X \to X$ be a measure-preserving transformation, meaning that T is measurable and

$$\mu(T^{-1}E) = \mu(E) \qquad \text{for every } E \in \mathcal{B}.$$

If T is invertible, then T^n is defined for every $n \in \mathbb{Z}$, with $T^0 = \mathrm{id}$, $T^n = T \circ \cdots \circ T$ for $n > 0$, and $T^n = (T^{-1})^{-n}$ for $n < 0$. For $1 \leq p < \infty$, $L^p(X)$ denotes the space of measurable functions $g : X \to \mathbb{C}$ with

$$\|g\|_{L^p(X)} := \left(\int_X |g|^p \, d\mu \right)^{1/p} < \infty,$$

and $L^\infty(X)$ denotes the space of essentially bounded functions. A sequence of functions F_N converges in $L^2(X)$ if there is $F \in L^2(X)$ such that $\|F_N - F\|_{L^2(X)} \to 0$. It converges pointwise almost everywhere if, outside a set of measure zero, the numerical sequence $F_N(x)$ converges.

Consider the multilinear polynomial ergodic averages

$$\frac{1}{N} \sum_{n=1}^{N} \prod_{j=1}^{d} g_j(T_j^{P_j(n)} x), \tag{1}$$

where $P_1, \ldots, P_d \in \mathbb{Z}[n]$ are polynomials with integer coefficients, $T_1, \ldots, T_d : X \to X$ are invertible measure-preserving transformations, and $g_1, \ldots, g_d$ are measurable functions on X. The assumption $P_j \in \mathbb{Z}[n]$ guarantees that $P_j(n) \in \mathbb{Z}$ for every $n \in \mathbb{Z}$, so the iterates $T_j^{P_j(n)}$ are well defined. The transformations T_i, T_j are said to commute if $T_i T_j = T_j T_i$, or equivalently $T_i(T_j x) = T_j(T_i x)$ for every $x \in X$.

The average (1) is called “multilinear” because it depends linearly on each function g_j separately, and it is called “polynomial” because the orbit of x is sampled at the polynomial times $P_j(n)$. When $d = 1$ and $P_1(n) = n$, this is the classical Birkhoff average. For $d \geq 2$, it is a non-conventional average: instead of following one observable along one orbit, it compares several observables along several polynomially parametrized orbits at the same time. These averages are central in ergodic Ramsey theory. For example, if $g_j = \mathbf{1}_A$, then integration of (1) involves quantities of the form

$$\mu\left(\bigcap_{j=1}^{d} T_j^{-P_j(n)} A\right),$$

and, after adjoining the extra factor $\mathbf{1}_A(x)$, such expressions measure polynomial multiple recurrence of a set A. In finite cyclic models they count polynomial configurations in dense subsets of integers. Thus convergence of (1) says that these recurrence statistics stabilize as $N \to \infty$.

In 1996, Bergelson (Bergelson 1996) asked whether (1) converges, first in the $L^2(X)$ norm and then pointwise almost everywhere, in the case when all g_j are bounded and all T_j are pairwise commuting. This fundamental question shaped much of the modern research on multiple ergodic averages.

We first discuss norm convergence. In 2005, Leibman (Leibman 2005) proved the L^2-convergence of (1) in the single-transformation case $T_1 = \cdots = T_d = T$. In 2008, Tao (Tao 2008) resolved the commuting-transformation case when all the polynomials P_j are linear. In 2011, Chu, Frantzikinakis and Host (Chu et al. 2011) resolved the case when the polynomials P_j have distinct degrees. Finally, in 2012 Walsh (Walsh 2012) answered Bergelson’s norm-convergence question in full. In fact, Walsh proved L^2-convergence for a substantially more general class of averages, in which the transformations need not commute but are required to generate a nilpotent group.

Invertible measure-preserving transformations form a group under composition. In an abstract group G, the commutator of two elements $a, b \in G$ is

$$[a, b] = a^{-1}b^{-1}ab.$$

The lower central series of G is the decreasing sequence of subgroups

$$G = G_1 \supseteq G_2 \supseteq G_3 \supseteq \cdots,$$

where G_{i+1} is generated by all commutators $[a, b]$ with $a \in G$ and $b \in G_i$. The group G is called nilpotent if this sequence eventually reaches the trivial subgroup $\{e\}$, where e is the identity element. Equivalently, there is some $s \in \mathbb{N}$ such that $G_{s+1} = \{e\}$. Abelian groups are exactly the nilpotent groups of step at most one.

Theorem 1 (Walsh) *Let G be a nilpotent group of measure-preserving transformations of a probability space $(X, \mathcal{B}, \mu)$. Let $T_1, \ldots, T_r \in G$. Then, for every $g_1, \ldots, g_d \in L^\infty(X)$ and every collection of integer-valued polynomials $P_{i,j}$, the averages*

$$\frac{1}{N} \sum_{n=1}^{N} \prod_{j=1}^{d} g_j\big(T_1^{P_{1,j}(n)} \cdots T_r^{P_{r,j}(n)} x\big)$$

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

The average (1) is the special case of Theorem 1 obtained by taking $r = d$, $P_{j,j} = P_j$, and $P_{i,j} = 0$ whenever $i \neq j$. Bergelson's original commuting-transformation setting corresponds to the special case in which the group generated by $T_1, \ldots, T_d$ is abelian. As noted in (Bergelson and Leibman 2002), the nilpotency assumption in Walsh's theorem cannot simply be removed: convergence can fail even for $d = 2$, bounded functions, and the linear choice $P_1(n) = P_2(n) = n$.

Walsh's theorem proves that the L^2-limit exists, but in this generality it does not identify the limit explicitly. Recently, Frantzikinakis and Kuca (Frantzikinakis and Kuca 2025) studied the limit of (1) for commuting transformations and bounded functions. Under additional hypotheses on the polynomial family, they identify characteristic factors and obtain concrete limit formulas. For instance, in the linearly independent case, the rational Kronecker factor is characteristic; in particular, for totally ergodic transformations the limit is the product of the integrals of the functions.

Pointwise convergence is much more delicate. Norm convergence is a global statement in $L^2(X)$, while pointwise convergence asks for a limiting law along almost every individual orbit. The case $d = 1$ and $P_1(n) = n$ is Birkhoff's ergodic theorem (Birkhoff 1931). The case $d = 1$ and arbitrary $P_1 \in \mathbb{Z}[n]$ is Bourgain's polynomial ergodic theorem (Bourgain 1989). In 1990, Bourgain (Bourgain 1990) also resolved the bilinear single-transformation linear case $d = 2$, $T_1 = T_2 = T$, $P_1(n) = an$, and $P_2(n) = bn$. Equivalently, he established pointwise almost everywhere convergence of

$$A_N^{P_1,P_2}(g_1, g_2)(x) := \frac{1}{N} \sum_{n=1}^{N} g_1\big(T^{P_1(n)} x\big) g_2\big(T^{P_2(n)} x\big) \tag{2}$$

when g_1, g_2 are bounded and P_1, P_2 are linear polynomials. In 2022, Krause, Mirek and Tao (Krause et al. 2022) treated a much harder bilinear case in which one polynomial is linear and the other is nonlinear.

Theorem 2 (Krause–Mirek–Tao) *Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system, let $P \in \mathbb{Z}[n]$ be a polynomial of degree at least 2, and let $g_1 \in L^{p_1}(X)$, $g_2 \in L^{p_2}(X)$ for some $1 < p_1, p_2 < \infty$ satisfying*

$$\frac{1}{p_1} + \frac{1}{p_2} \leq 1.$$

Then the averages

$$A_N^{n,P(n)}(g_1, g_2)(x) = \frac{1}{N} \sum_{n=1}^{N} g_1(T^n x) g_2(T^{P(n)} x)$$

converge pointwise almost everywhere as $N \to \infty$.

Theorem 2 is very general: it holds even on σ-finite measure spaces, not just probability spaces. Nevertheless, it was new even for probability spaces and bounded functions. The special case $P(n) = n^2$ answers Problem 11 from Frantzikinakis' open problems survey (Frantzikinakis 2016). The endpoint cases are genuinely different: the counterexamples discussed in Theorem **??** show that one cannot expect Theorem 2 to extend in general to $(p_1, p_2) = (1, \infty)$ or $(\infty, 1)$.

On a probability space, the case $g_1 \equiv 1$ of Theorem 2 reduces to Bourgain's theorem (Bourgain 1989) on pointwise convergence along polynomial iterates. The proof of Theorem 2 combines Bourgain's harmonic-analytic ideas with inverse theorems from additive combinatorics and, at large scales, harmonic analysis on the adelic integers.

Kosz, Mirek, Peluse, Wan and Wright (Kosz et al. 2024) have now proved a far-reaching multilinear extension in the distinct-degree case. Their theorem applies to any number of functions and transformations, and it goes beyond the bounded setting by allowing L^{p_i} functions in the natural Hölder range.

Theorem 3 (Kosz–Mirek–Peluse–Wan–Wright) *Let $d \in \mathbb{Z}_+$, let $(X, \mathcal{B}, \mu)$ be a probability space, and let $T_1, \ldots, T_d$ be invertible pairwise-commuting measure-preserving transformations of X. Suppose that $P_1, \ldots, P_d \in \mathbb{Z}[n]$ have distinct degrees. Let $g_i \in L^{p_i}(X)$ for $i = 1, \ldots, d$, where $1 < p_i < \infty$, and assume that*

$$\frac{1}{p} := \sum_{i=1}^{d} \frac{1}{p_i} \leq 1.$$

Then the averages

$$\frac{1}{N} \sum_{n=1}^{N} \prod_{i=1}^{d} g_i\big(T_i^{P_i(n)} x\big)$$

converge pointwise almost everywhere as $N \to \infty$. Moreover, the averages converge in $L^p(X)$, and the associated maximal and lacunary variational estimates hold in the same range of exponents.

The bounded case in Bergelson's question follows immediately from Theorem 3: on a probability space, every bounded function belongs to $L^q(X)$ for every finite q, and one may choose q large enough so that $d/q \leq 1$. Thus Bergelson's pointwise convergence problem is now solved for arbitrary multilinearity in the important case when the polynomial iterates have distinct degrees.

This leaves two natural frontiers. The first is to remove the distinct-degree assumption, since repeated degrees are precisely where new arithmetic resonances appear. The second is to move from commuting transformations toward the nilpotent setting suggested by Walsh's norm-convergence theorem and by the Furstenberg–Bergelson–Leibman conjecture. The result of Kosz, Mirek, Peluse, Wan and Wright does not close these problems, but it changes the landscape: for the first time, pointwise convergence of general multilinear polynomial averages is known beyond the previously accessible bilinear and single-transformation regimes.

References

Bergelson, Vitaly. 1996. "Ergodic Ramsey Theory—an Update." In *Ergodic Theory of* $\mathbf{Z}^d$ *Actions (Warwick, 1993–1994)*, vol. 228. London Math. Soc. Lecture Note Ser. Cambridge Univ. Press, Cambridge. https://doi.org/10.1017/CBO9780511662812.002.

Bergelson, V., and A. Leibman. 2002. "A Nilpotent Roth Theorem." *Invent. Math.* 147 (2): 429–70. https://doi.org/10.1007/s002220100179.

Birkhoff, George D. 1931. "Proof of the Ergodic Theorem." *Proc. Natl Acad. Sci. USA* 17 (12): 656–60.

Bourgain, J. 1990. "Double Recurrence and Almost Sure Convergence." *J. Reine Angew. Math.* 404: 140–61. https://doi.org/10.1515/crll.1990.404.140.

Bourgain, Jean. 1989. "Pointwise Ergodic Theorems for Arithmetic Sets." *Inst. Hautes Études Sci. Publ. Math.*, no. 69: 5–45. http://www.numdam.org/item?id=PMIHES_1989__69__5_0.

Chu, Qing, Nikos Frantzikinakis, and Bernard Host. 2011. "Ergodic Averages of Commuting Transformations with Distinct Degree Polynomial Iterates." *Proc. Lond. Math. Soc. (3)* 102 (5): 801–42. https://doi.org/10.1112/plms/pdq037.

Frantzikinakis, Nikos. 2016. "Some Open Problems on Multiple Ergodic Averages." *Bulletin of the Hellenic Mathematical Society* 60: 41–90.

Frantzikinakis, Nikos, and Borys Kuca. 2025. "Joint Ergodicity for Commuting Transformations and Applications to Polynomial Sequences." *Invent. Math.* 239 (2): 621–706. https://doi.org/10.1007/s00222-024-01313-w.

Kosz, Dariusz, Mariusz Mirek, Sarah Peluse, Renhui Wan, and James Wright. 2024. "The Multilinear Circle Method and a Question of Bergelson." *arXiv Preprint arXiv:2411.09478*.

Krause, Ben, Mariusz Mirek, and Terence Tao. 2022. "Pointwise Ergodic Theorems for Non-Conventional Bilinear Polynomial Averages." *Ann. Of Math. (2)* 195 (3): 997–1109. https://doi.org/10.4007/annals.2022.195.3.4.

Leibman, A. 2005. "Convergence of Multiple Ergodic Averages Along Polynomials of Several Variables." *Israel J. Math.* 146: 303–15. https://doi.org/10.1007/BF02773538.

Tao, Terence. 2008. "Norm Convergence of Multiple Ergodic Averages for Commuting Transformations." *Ergodic Theory Dynam. Systems* 28 (2): 657–88. https://doi.org/10.1017/S0143385708000011.

Walsh, Miguel N. 2012. "Norm Convergence of Nilpotent Ergodic Averages." *Ann. Of Math. (2)* 175 (3): 1667–88. https://doi.org/10.4007/annals.2012.175.3.15.