Séminaire BOURBAKI Novembre 2022
74e année, 2022–2023, no 1199, p. 87 à 120
doi: 10.24033/ast.1208

POINTWISE ERGODIC THEORY: EXAMPLES AND ENTROPY
[*after* Jean Bourgain]

by **Ben Krause**

Overview

Pointwise ergodic theory, the motivation for discrete harmonic analysis, has at its roots the classical theorem of Birkhoff (1931), which can be described as follows:

> For every ergodic —that is, "sufficiently randomizing"— measure-preserving transformation, τ, of a probability space, (X, μ), and any integrable function $f \in L^1(X, \mu)$, μ-almost surely, one can recover the mean of f by considering the Cesáro sums
>
> $$\frac{1}{N} \sum_{n \leqslant N} f(\tau^n x) \to \int_X f \, d\mu \;\; \mu - \text{a.e.}$$

Informally, this theorem says that one can recover the *spatial mean* of f,

$$\int_X f \, d\mu,$$

by considering the *temporal means*

$$\{ \frac{1}{N} \sum_{n \leqslant N} f(\tau^n x) \},$$

formed by "sampling" the function f at the "times" $\{\tau^n x\}$ and taking the appropriate average.[1]

A classical question in pointwise ergodic theory concerned the almost-everywhere existence of limiting behavior of averages

$$\frac{1}{N} \sum_{n=1}^{N} \tau^{a_n} f \tag{1}$$

[1] Even in the case when τ is not ergodic, the temporal means $\{ \frac{1}{N} \sum_{n \leqslant N} \tau^n f(x) \}$ still converge μ-almost everywhere.

where $\{a_n\}$ is "sparse"; as is custom, here and throughout we use $\tau^k f$ to denote the function

$$x \mapsto f(\tau^k x).$$

When the lower density of the sequence $\{a_n\}$ is bounded away from zero

$$\liminf \frac{|\{n : a_n \leqslant N\}|}{N} > 0,$$

detecting convergence becomes more straightforward, and the classical question concerned the existence of sequences $\{a_n\}$ with zero density,

$$\lim \frac{|\{n : a_n \leqslant N\}|}{N} = 0,$$

for which the averages (1) converged almost everywhere. In Bellow and Losert (1984), such a sequence was constructed; it consisted of taking long blocks of natural numbers, followed by much longer gaps, followed by slightly longer blocks, followed by even longer gaps, etc. In particular, this sequence had an upper Banach density of 1

$$d^*(\{a_n\}) := \limsup_{|I| \to \infty \text{ an interval}} \frac{|\{a_n\} \cap I|}{|I|} = 1.$$

The question remained, however, whether or not there existed upper Banach density-zero sequences, $\{a_n\}$ with $d^*(\{a_n\}) = 0$, for which the almost-everywhere convergence of the averages (1) could be proved. In particular, the classical question, explicitly posed first by Furstenberg, see also Bellow (1982), was whether or not the averages along the squares

$$\frac{1}{N} \sum_{n=1}^{N} \tau^{n^2} f$$

converged pointwise almost everywhere, initially for $f \in L^2(X)$. In breakthrough work, Bourgain (1988b,c, 1989b) answered this question affirmatively, and proved the almost everywhere convergence of (1) for any polynomial sequence,

$$\{a_n = P(n)\}, \quad P \in \mathbb{Z}[\cdot],$$

and any $f \in L^p(X)$, $p > 1$, for any σ-finite measure space X; this result was later proven to be sharp (Buczolich and Mauldin, 2007; Lavictoire, 2011).

Theorem 0.1. *Suppose that (X, μ) is a σ-finite measure space, $\tau : X \to X$ is a measure-preserving transformation, and $P \in \mathbb{Z}[\cdot]$ is a polynomial with integer coefficients. Then for each $1 < p < \infty$*

$$\frac{1}{N} \sum_{n=1}^{N} \tau^{P(n)} f$$

converges μ-a.e.

Although the issue of pointwise convergence is qualitative, Bourgain's insight was to quantify the rate at which convergence occurred – and then to use an abstract transference argument first due to Calderón (1968) to deduce these quantitative estimates from a single "universal" measure preserving system. By considering sequences of the form

$$\mathbb{Z} \ni n \mapsto \tau^n f(x), \quad x \in X \text{ fixed}$$

and using the measure-preserving nature of τ, Bourgain was able to reduce matters to proving estimates in the case of the integers with counting measure and the shift $(\mathbb{Z}, |\cdot|, \tau : x \mapsto x-1)$.

In particular, Bourgain was after quantitative estimates on the oscillation of the averaging operators

$$\frac{1}{N} \sum_{n=1}^{N} f(x - P(n)), \tag{2}$$

applied first to $\ell^2(\mathbb{Z})$-functions. A natural perspective on (2) is as a convolution of f and

$$K_N(x) := \frac{1}{N} \sum_{n=1}^{N} \delta_{P(n)}(x)$$

where δ_m denotes the point-mass at $m \in \mathbb{Z}$; as this problem is $\ell^2(\mathbb{Z})$-based, the Fourier transform method is naturally employed, and the key to the analysis is an understanding of the exponential sums

$$\frac{1}{N} \sum_{n \leqslant N} e^{-2\pi i \beta \cdot P(n)},$$

which is accomplished via the *circle method* from analytic number theory; the interplay between the "soft" analytic issue of pointwise convergence and "hard" analytic estimates on the integers/Euclidean space via analytic-number-theoretic means is characteristic of the fields of pointwise ergodic theory and discrete harmonic analysis.

I first came to understand Bourgain's work by reading Thouvenot (1990), which I think explains Theorem 0.1 beautifully; the goal of these notes is to complement Thouvenot (1990) by trying to explain the motivation behind Bourgain's argument.

Accordingly, for the sake of clarity, we will shift our focus slightly from proving Theorem 0.1, and will instead focus on the related maximal estimate, in the representative case of $L^2(X)$.

Theorem 0.2. *Suppose that (X, μ) is a σ-finite measure space, $\tau : X \to X$ is a measure-preserving transformation, and $P \in \mathbb{Z}[\cdot]$ is a polynomial with integer coefficients. Then there exists an absolute constant $\mathbf{C}$, independent of (X, μ, τ), so that*

$$\| \sup_N \Big| \frac{1}{N} \sum_{n=1}^{N} \tau^{P(n)} f \Big| \|_{L^2(X)} \leqslant \mathbf{C} \cdot \|f\|_{L^2(X)}.$$

By Calderón's transference principle, Theorem 0.2 follows from the analoguous estimate of the integers: if we define

$$\mathscr{M} f(x) := \sup_N \Big| \frac{1}{N} \sum_{n=1}^N f(x - P(n)) \Big|, \tag{3}$$

then our focus turns to establishing the following estimate

Theorem 0.3. *For any $P \in \mathbb{Z}[\cdot]$, the following norm inequality holds: there exists an absolute constant $\mathbf{C}$ so that*

$$\|\mathscr{M} f\|_{\ell^2(\mathbb{Z})} \leqslant \mathbf{C} \cdot \|f\|_{\ell^2(\mathbb{Z})}.$$

Below, following the lead of Thouvenot (1990), we will restrict to the case where

$$P(n) = n^d,$$

as this eliminates some number-theoretic technicality while still capturing the essence of the problem.

Notation. — Here and throughout we abbreviate the complex exponential $e(t) := e^{2\pi i t}$, so that we may express the Fourier transform in Euclidean space, and on the integers, respectively as

$$\hat{f}(\xi) = \int_{\mathbb{R}} f(x) \cdot e(-\xi x)\, dx, \quad g^{\vee}(x) = \int_{\mathbb{R}} g(\xi) \cdot e(\xi x)\, d\xi$$
$$\hat{f}(\beta) = \sum_n f(n) \cdot e(-\beta n), \quad g^{\vee}(n) = \int_{\mathbb{T}} g(\beta) \cdot e(\beta n)\, d\beta.$$

We will let

$$\phi_k(t) := 2^{-k} \cdot \phi(2^{-k} \cdot t)$$

denote the usual L^1-normalized dyadic dilations, and for frequencies θ, we let

$$\mathrm{Mod}_\theta g(x) := e(\theta x) \cdot g(x) \tag{4}$$

so that

$$\widehat{\mathrm{Mod}_\theta} g(\beta) = \hat{g}(\beta - \theta),$$

and recall the Hardy–Littlewood Maximal operator

$$M_{\mathrm{HL}} f(x) := \sup_{r>0} \frac{1}{2r} \int_{-r}^{r} |f(x-t)|\, dt \quad \text{or} \quad := \sup_{N \geqslant 0} \frac{1}{2N+1} \sum_{n=-N}^{N} |f(x-n)|;$$

although we use the same notation to refer to both continuous and discrete maximal operator, it will be clear from context which formulation we use.

We will let $[N] := \{1,\dots,N\}$, and abbreviate $\sum_{n\leqslant N} := \sum_{n=1}^{N}$. We will use the symbol $\mathbf{c}$ to denote suitably small constants, which remain bounded away from zero, and $\mathbf{C}$ to denote suitably large constants, which remain bounded above. If we need these constants to depend on parameters, we use subscripts, thus $\mathbf{c}_d$ is a constant that is small depending on d. We use $X = O(Y)$ to denote the statement that $|X| \leqslant \mathbf{C}\cdot Y$, and analogously define $X = O_d(Y)$.

Finally, we will use the heuristic notation

$$f \quad \text{“}=\text{”} \quad g$$

to denote moral equivalence: up to tolerable errors, f and g exhibit the same type of behavior.

1. Discrete Complications

Before beginning our discussion of Theorem 0.3, let us explain why we might expect this to be a challenging problem.

For problems with a “linear” flavor, the discrete theory essentially mirrors the continuous theory

$$\sup_r \frac{1}{r}\int_0^r |f(x-t)|\,dt \quad \text{“}=\text{”} \quad \sup_N \frac{1}{N}\sum_{n=1}^{N} |f(x-n)|$$

as can be seen by experimenting with functions of the form $F(\lfloor x\rfloor)$ and using dilation invariance of the real-variable maximal function to reduce attention to real variable functions that are constant on unit scales.

The problems become dramatically more complicated once linearity is destroyed. In this case, we consider the simple example of the Hardy–Littlewood maximal function along the curve $t \mapsto t^d$. The continuous maximal function

$$Mf := M_d f := \sup_r \Big|\frac{1}{r}\int_0^r f(x-t^d)\,dt\Big| = \sup_r \Big|\frac{1}{r}\int_0^{r^d} f(x-t)\,\frac{1}{dt^{1-1/d}}\,dt\Big|, \qquad (5)$$

is just a weighted version of M_{HL} via the pointwise majorization

$$\begin{aligned}\frac{1}{r}\int_0^{r^d} |f(x-t)|\frac{1}{dt^{1-1/d}}\,dt &\leqslant \sum_{j=1}^{\infty} 2^{-j/d}\cdot\Big(\frac{2^{j/d}}{r}\int_{2^{-j}\cdot r^d}^{2^{1-j}\cdot r^d} |f(x-t)|\frac{1}{dt^{1-1/d}}\,dt\Big)\\ &\leqslant \mathbf{C}/d\cdot\sum_{j=1}^{\infty} 2^{-j/d}\cdot\Big(\frac{2^j}{r^d}\int_{2^{-j}\cdot r^d}^{2^{1-j}\cdot r^d} |f(x-t)|\,dt\Big) \leqslant \mathbf{C}/d\cdot\sum_{j=1}^{\infty} 2^{-j/d}\cdot M_{\mathrm{HL}}f(x)\\ &\leqslant \mathbf{C}\cdot M_{\mathrm{HL}}f(x). \qquad (6)\end{aligned}$$