Pointwise Ergodic Theorems for Non-Conventional Polynomial Averages

Ben Krause

July 7, 2023

Ergodic Theory

Let X be a set, and let $T : X \to X$ be a map. If $x \in X$ is a point, the basic question that ergodic theory seeks to study is to what extent the *orbit* of x under iterates of T

$$x, Tx, T^2x, \ldots, T^nx, \ldots$$

"equidistribute" in X. We will shortly impose enough structure on (X, T) for this question to be meaningfully posed, but first let's discuss some examples.

Example One

Let $X := \mathbb{T} \cong [0,1]$, and let $T : X \to X$ be the point transformation given by $x \mapsto x - \alpha \mod 1$ for $0 \leq \alpha \leq 1$. When α is irrational, one can integrate continuous $f \in C(\mathbb{T})$ by approximating

$$\sup_{0 \leq x \leq 1} \left| \frac{1}{N} \sum_{n=1}^{N} f(T^n x) - \int_{\mathbb{T}} f(t)\ dt \right| \to 0$$

i.e.

$$\sup_{0 \leq x \leq 1} \left| \frac{1}{N} \sum_{n=1}^{N} f(x - n\alpha \mod 1) - \int_{\mathbb{T}} f(t)\ dt \right| \to 0$$

as $N \to \infty$.

Example Two

Let $X := [0,1]$, and let $T : X \to X$ be the map,

$$Tx = 2x \mod 1;$$

then for almost every $x \in X$

$$\left| \frac{1}{N} \sum_{n \leq N} \mathbf{1}_{[0,1/2)}(T^n x) - 1/2 \right| \to 0 \text{ as } N \to \infty.$$

Example Two Continued

If we represent x in its binary expansion,

$$x = \frac{\epsilon_1(x)}{2} + \frac{\epsilon_2(x)}{2^2} + \cdots + \frac{\epsilon_n(x)}{2^n} + \ldots,$$

then

$$T^n x \in [0, 1/2) \iff \epsilon_{n+1}(x) = 0,$$

and the conclusion is that for almost every x

$$\frac{|\{j \leq N : \epsilon_j(x) = 1\}|}{N}, \frac{|\{j \leq N : \epsilon_j(x) = 0\}|}{N} \to \frac{1}{2},$$

i.e. almost every x is *normal*.

The Set-Up

To make our discussion of equidistribution rigorous, we impose measure-theoretic structure on our sets and transformations:

Throughout (X, μ, T) will be a measure preserving system: (X, μ) is a probability space, and $T : X \to X$ is an invertible measure-preserving transformation.

For functions $f : X \to \mathbb{C}$, we define

$$T^n f(x) := f(T^n x), \quad n \in \mathbb{Z}.$$

Our previous two examples fit into this frame work.

Ergodic Theory, Take Two

Broadly, ergodic theory is the study of the dynamics of functions defined on X under the action of measure-preserving transformations.

The classical objects of study are the Cesàro averages,

$$C_N f := \frac{1}{N} \sum_{n \leq N} T^n f;$$

the question is to what extent if any $C_N f$ converge. For instance, our first example involved the phenomenon of uniform convergence, while the second example concerned almost everywhere convergence.

The Hilbert-space setting is most familiar, so we will focus on the case where $f \in L^2(X)$ is square integrable.

Two Types of Convergence

There are two types of convergence that we consider

- **Norm convergence**, in which one views $C_N f \in L^2(X)$ as a sequence of vectors, and seeks to understand convergence from this vectorial standpoint;
- **Pointwise convergence**, in which one seeks to evaluate the limit of the $C_N f$ pointwise.

Classical Ergodic Theorems

The two classical ergodic theorems, due to Von Neumann and Birkhoff, dictate that $C_N f$ converge both in norm and pointwise for $f \in L^2(X)$. The bulk of this talk will concern the pointwise regime; norm convergence is better understood.

Polynomial Averages, Norm Convergence I

Classical Hilbert space techniques can be used to address the issue of norm convergence of the $\{C_N f\}$. These methods actually generalize to the study of ergodic averages along polynomial orbits:

$$\frac{1}{N}\sum_{n\leq N} T^{P(n)}f, \quad P \in \mathbb{Z}[\cdot].$$

On the other hand, once the setting switches from the study of averages of a single function to the multi-linear setting, the difficulty of the problems increases dramatically.

Multiple Ergodic Theorems

The departure point for the modern theory of multiple ergodic averages is Furstenberg's ergodic-theoretic proof of Szemerédi's theorem.

Szemerédi's Theorem

Suppose that $E \subset \mathbb{Z}$ satisfies

$$d^*(E) := \limsup_{|I| \to \infty} \frac{|E \cap I|}{|I|} > 0, \qquad I \text{ an interval.}$$

Then for any $k \geq 2$, there exist infinitely many progressions,

$$\{x, x-n, x-2n, \ldots, x-kn\} \subset E.$$

The $k = 2$ case is due to Roth.

Szemerédi via Ergodic Theory

Using his *Furstenberg Correspondence Principle*, Furstenberg deduced Szemerédi's theorem from a dynamical systems result, slightly weaker than the below formulation due to Host-Kra and Leibman.

Szemerédi's Theorem, Ergodic Theoretic Formulation

For any non-trivial $f \in L^\infty(X)$, and any $k \geq 1$, the bounded vectors

$$\frac{1}{N}\sum_{n \leq N} T^n f \cdot \dots \cdot T^{kn} f \in L^\infty(X) \subset L^2(X)$$

converge in norm to a non-trivial function, when viewed as vectors in $L^2(X)$.

Non-Linear Multiple Ergodic Averages

Almost twenty years after Furstenberg's ergodic theoretic proof of Szemerédi's theorem, Furstenberg-Weiss established the following norm convergence result for the non-linear ergodic averages below.

Theorem (Furstenberg-Weiss, 1996)

Suppose $f, g \in L^\infty(X)$ are bounded. Then the sequence of vectors

$$B_N(f,g) := \frac{1}{N} \sum_{n \leq N} T^n f \cdot T^{n^2} g \in L^\infty(X) \subset L^2(X)$$

converges in the L^2 norm.

Ergodic Theory meets Additive Combinatorics

Analogous to the situation of Szemerédi's theorem, the *Furstenberg Correspondence Principle* leads to the following Corollary.

Non-Linear-Roth-Type Corollary

Suppose that $E \subset \mathbb{Z}$ satisfies

$$d^*(E) := \limsup_{|I| \to \infty} \frac{|E \cap I|}{|I|} > 0, \qquad I \text{ an interval.}$$

Then there exist infinitely many progressions,

$$\{x, x - n, x - n^2\} \subset E.$$

The Bigger Picture

The result of Furstenberg-Weiss admits dramatic extensions, all the way to M. Walsh's celebrated norm convergence result concerning the multiple ergodic averages:

$$A_{N,\vec{P}}(f_1, \ldots, f_m) := \frac{1}{N} \sum_{n \leq N} T_1^{P_1(n)} f_1 \cdot \ldots T_m^{P_m(n)} f_m,$$

where $P_i \in \mathbb{Z}[\cdot]$, and $T_i : X \to X$ generate a *nilpotent group* of transformations.

Pointwise Convergence, Pre-COVID

It took breakthrough work of Bourgain in late 80s and early 90s to resolve the issue of pointwise convergence of ergodic averages along polynomial orbits. The state of the art result for pointwise convergence of bilinear ergodic averages was due to Bourgain and Michael Lacey.

Pointwise Convergence, Bilinear Setting

Suppose that $k \in \mathbb{Z}$. Then whenever $f \in L^p(X)$ and $g \in L^q(X)$ for $1 < p, q \leq \infty$ so that

$$f \cdot g \in L^r(X), \;\; r > 2/3$$

the bilinear ergodic averages

$$\frac{1}{N} \sum_{n \leq N} T^n f \cdot T^{kn} g$$

converge almost everywhere.

The Program

The long-term goal is to establish the analogue of Walsh's theorem in the pointwise setting. Suppose that $\{T_1, \ldots, T_m\}$ are commuting measure-preserving transformations on X.

Goal

Prove that for every $P_1, \ldots, P_m \in \mathbb{Z}[\cdot]$, and each $f_1, \ldots, f_m \in L^\infty(X)$,

$$\frac{1}{N} \sum_{n \leq N} T_1^{P_1(n)} f_1 \cdot \cdots \cdot T_m^{P_m(n)} f_m$$

converge pointwise almost everywhere.

Pointwise Convergence, Bilinear Setting

Joint with M. Mirek and T. Tao, we established the following special case in the bilinear setting.

Theorem

Suppose $P \in \mathbb{Z}[\cdot]$ has degree at least 2, and that $f \in L^p(X),\ g \in L^q(X)$ where $p, q > 1$ and

$$f \cdot g \in L^r(X),\ r > 1 - c_d, \quad c_d > 0.$$

Then the following bilinear ergodic averages converge pointwise almost everywhere

$$B_N(f,g) := \frac{1}{N} \sum_{n \leq N} T^n f \cdot T^{P(n)} g.$$

In what follows, we will restrict to $f, g \in L^2(X)$, and $P(n) = n^2$.

A Combinatorial Consequence

To put this result in perspective, it is helpful to present the following combinatorial consequence.

Generic Behavior

Let us suppose that $E \subset \{1, \ldots, N\}$ were a "random" set, with density δ, $|E| = \delta N$. Then from the "vantage point" of any $x \in E$, we expect E to behave randomly with respect to sampling along the parabola:

$$\mathbb{P}(x - n, x - n^2 \in E) = \mathbb{P}(x - n \in E) \cdot \mathbb{P}(x - n^2 \in E) = \delta^2,$$

which sums to

$$\frac{|\{n \leq P : x - n, x - n^2 \in E\}|}{P} = \delta^2$$

for each $P^2 \ll N$.

Arithmetic Obstruction

There is no reason for a general set of density δ to be "random:" for instance one might take

$$E := \lceil \delta^{-1} \rceil \cdot \mathbb{Z}.$$

On the other hand, once we rule out this arithmetic obstruction, we will see that from certain vantage points, E does indeed behave randomly with respect to sampling along the parabola $\{(n, n^2) : n \geq 1\}$.

NLR-Behavior is Local

Suppose that $E \subset \mathbb{Z}$ has positive upper density, $d^*(E) = \delta > 0$. The non-Linear-Roth Corollary of Furstenberg-Weiss says that there exist infinitely many progressions,

$$\{x, x - n, x - n^2\} \subset E.$$

Our "visibilty result" localizes this phenomenon.

NLR-Type Corollary, Refined Formulation

For any $0 < \epsilon < \delta$, there exists some $Q = Q(\delta, \epsilon)$ so that for each $R \geq \Lambda$ sufficiently large there exists some vantage point $x_R \in E$ so that

$$\inf_{\Lambda \leq N \leq R} \frac{|\{n \leq N : x_R - Qn, x_R - Q^2 n^2 \in E\}|}{N} \geq \delta^2 - \epsilon.$$

Special Cases

Before discussing our theorem, there are two special cases that we need to understand first: the cases where one of the two functions is identically constant.

$$B_N(f, \mathbf{1}_X) := C_N f := \frac{1}{N} \sum_{n \le N} T^n f,$$

and

$$B_N(\mathbf{1}_X, g) := D_N g := \frac{1}{N} \sum_{n \le N} T^{n^2} g.$$

The study of the $\{C_N f\}$ gives rise to the pointwise ergodic theorem; the $\{D_N g\}$ were the subject of Bourgain's work in the late 80s-early 90s.

Pointwise Convergence, One Function

Theorem (Birkhoff's Pointwise Ergodic Theorem 1931)

For any MPS (X, μ, T)*, and any* $f \in L^2(X)$*,* $\{C_N f\}$ *converge pointwise,* μ *a.e.*

Theorem (Bourgain, 1988-1990)

For any MPS (X, μ, T) *and any* $f \in L^2(X)$*,* $\{D_N f\}$ *converge pointwise,* μ *a.e.*

On Pointwise Convergence

Pointwise convergence of a sequence of numbers is a qualitative phenomenon: a sequence $\{a_n\}$ converges if for every $t > 0$, there are finitely many "times," $n_0 < n_1 < \cdots < n_K$, so that the sequence a_n "jumps by t:"

$$|a_{n_i} - a_{n_{i-1}}| > t \tag{1}$$

for each $1 \leq i \leq K$, but there are no subsequences of length $K + 2$ so that (1) holds for each $1 \leq i \leq K + 1$.

We define $\mathcal{N}_t(\{a_n\})$ to be this largest K.

The Strategy

The statement that $f_n(x)$ converges almost everywhere is equivalent to the almost-everywhere convergence of

$$x \mapsto \mathcal{N}_t(f_n(x))$$

for each $t > 0$.

Bourgain's Approach

Prove a norm estimate on

$$f \mapsto \mathcal{N}_t(C_N f(x)), \quad f \mapsto \mathcal{N}_t(D_N f(x)).$$

Quantifying Convergence

This technique appears in martingale theory. In particular, the natural estimate is the following:

Goal

Establish the *quantitative* estimate

$$\|t \cdot \mathcal{N}_t(C_N f(x) : N)^{1/r}\|_{L^2(X)} \leq A_r \cdot \|f\|_{L^2(X)}, \quad A_r < \infty$$

for some $r < \infty$, and similarly for D_N.

This says that for each $\|f\|_{L^2(X)} = 1$,

$$\mathcal{N}_t(C_N f(x)) \quad \text{“} \leq \text{''} \quad A_r^r \cdot t^{-r}$$

on average.

Effective Estimates

Theorem (Bourgain, 1990)

For each $r > 2$, and each $t > 0$, there exists an absolute constant A_r, so that

$$\|t \cdot \mathcal{N}_t(C_N f(x) : N)^{1/r}\|_{L^2(X)} \leq A_r \cdot \|f\|_{L^2(X)}.$$

Theorem (K. 2014)

For each $r > 2$, and each $t > 0$, there exists an absolute constant A_r, so that

$$\|t \cdot \mathcal{N}_t(D_N f(x) : N)^{1/r}\|_{L^2(X)} \leq A_r \cdot \|f\|_{L^2(X)}.$$

Below, all estimates involving $t \cdot \mathcal{N}_t^{1/r}$ will be *uniform* in $t > 0$.

Transference and Discrete Harmonic Analysis

We want to prove the estimate

$$\|t \cdot \mathcal{N}_t(C_N f : N)^{1/r}\|_{L^2(X)} \leq A_r \cdot \|f\|_{L^2(X)},$$

but we have no information about X.

Calderón: "Study $\mathbb{Z}$-actions on $\mathbb{Z}$"

By considering sequences of the form,

$$n \mapsto T^n f(x), \quad x \in X \text{ fixed},$$

matters reduce to a single concrete setting: it suffices to prove the sequence-space estimate

$$\left\| t \cdot \mathcal{N}_t \left(\frac{1}{N} \sum_{n \leq N} f(x-n) : N \right)^{1/r} \right\|_{\ell^2(\mathbb{Z})} \leq A_r \cdot \|f\|_{\ell^2(\mathbb{Z})},$$

And similarly for the squares.

Recap

Our problem in pointwise convergence reduces to proving a *discrete* harmonic-analytic estimate, i.e. a harmonic analytic estimate on $\mathbb{Z}$.

- We began by trying to prove a convergence statement in the dynamical systems setting;
- We recast convergence, and reduced matters to showing that $\mathcal{N}_t(C_N f) < \infty$ almost everywhere for each $t > 0$;
- We planned to prove convergence of $\mathcal{N}_t$ by proving it satisfied a norm estimate on $L^2(X)$;
- We transferred the norm estimate to the sequence-space setting of $\ell^2(\mathbb{Z})$.

The Linear Averages

In our goal to estimate

$$\| t \cdot \mathcal{N}_t \left(\frac{1}{N} \sum_{n \leq N} f(x-n) : N \right)^{1/r} \|_{\ell^2(\mathbb{Z})} \leq A_r \cdot \|f\|_{\ell^2(\mathbb{Z})},$$

we can borrow heavily from the Euclidean theory, which combines martingale methods and Fourier analysis to establish the equivalent estimate:

$$\| t \cdot \mathcal{N}_t \left(\frac{1}{N} \int_0^N f(x-t) \, dt : N \right)^{1/r} \|_{L^2(\mathbb{R})} \leq A_r \cdot \|f\|_{L^2(\mathbb{R})}.$$

Linearity is Key

This analogy relies crucially on the fact that the orbits over which we average have similar traces (below, we use the Minkowski sum):

$$\{n \geq 1\} + [0,1) = \{t \geq 1\} \subset \mathbb{R}.$$

The Quadratic Averages, Geometric Issues

Regarding the averages along the squares, our goal is to estimate

$$\Big\| t \cdot \mathcal{N}_t \left(\frac{1}{N} \sum_{n \leq N} f(x - n^2) : N \right)^{1/r} \Big\|_{\ell^2(\mathbb{Z})} \leq A_r \cdot \|f\|_{\ell^2(\mathbb{Z})}.$$

A geometric issue presents: namely, the sequence $\mathcal{O} := \{n^2 : n \geq 1\} \subset \mathbb{Z}$ satisfies

$$|\mathcal{O} \cap I| \leq |I|^{1/2}, \quad I \subset \mathbb{Z} \text{ an interval}$$

$$\big|\big(\mathcal{O} + [0,1)\big) \cap J\big| \leq 100 \cdot |J|^{1/2}, \quad |J| \geq 1, \; J \subset \mathbb{R} \text{ an interval},$$

and thus acts like a "1/2-dimensional" set, from the perspective of density. Compare this to the continuous situation, where the trace is full dimensional:

$$\{t^2 : t \geq 1\} = [1, \infty) \subset \mathbb{R}.$$

A Non-Example

Suppose that we were interested in proving

$$\left\| t \cdot \mathcal{N}_t \left(\frac{1}{N} \int_0^N F(x - y^2) \, dy : N \right)^{1/r} \right\|_{L^2(\mathbb{R})} \leq A_r \cdot \|F\|_{L^2(\mathbb{R})}.$$

By changing variables, $s = y^2$, and using convexity, matters reduce to the linear situation

$$\left\| t \cdot \mathcal{N}_t \left(\frac{1}{N} \int_0^N F(x - y) \, dy : N \right)^{1/r} \right\|_{L^2(\mathbb{R})} \leq A_r \cdot \|F\|_{L^2(\mathbb{R})}.$$

This argument fails in the integers, due to the presence of a *smallest scale*.

The Quadratic Averages, Fourier Analytic Issues

Geometric issues that we have seen necessitate a different type of analysis. Accordingly, we try a Fourier-analytic approach, where the Fourier transform is given by

$$\hat{f}(\beta) = \sum_n f(n) e^{-2\pi i \beta n}.$$

In particular, we may express

$$\frac{1}{N} \sum_{n \leq N} f(x - n^2) = K_N * f(x) = \int \hat{f}(\beta) \cdot \widehat{K_N}(\beta) \cdot e^{2\pi i \beta x} \, d\beta$$

where

$$K_N(x) := \frac{1}{N} \sum_{n \leq N} \delta_{n^2}(x) \quad \Rightarrow \quad \widehat{K_N}(\beta) = \frac{1}{N} \sum_{n \leq N} e^{-2\pi i \beta n^2}.$$

Fourier Analysis Meets Number Theory

First, let's note that the relevant multipliers in the linear case,

$$\frac{1}{N}\sum_{n\leq N} e^{-2\pi i\beta n}$$

could be completely understood in terms of $\|\beta\|$, the distance from β to $0 \in \mathbb{T}$.

On the other hand,

$$\widehat{K_N}(\beta) = \frac{1}{N}\sum_{n\leq N} e^{-2\pi i\beta n^2}$$

is a *Gauss sum*, and requires ideas from analytic number theory to analyze: the *circle method* of Hardy and Littlewood, developed in the study of Waring's problem.

The Circle Method

Probabilistic Intuition

Each multiplier $\widehat{K_N}(\beta)$ is an average of N mean-zero pairwise independent random variables; from probability theory, we might expect a "generic" power savings,

$$|\widehat{K_N}(\beta)| = |\frac{1}{N}\sum_{n \leq N} e^{-2\pi i \beta n^2}| \text{ “}\lesssim\text{” } N^{-\epsilon}$$

for some $\epsilon > 0$.

There is an obstruction to this type of argument: potential correlation of the phases

$$\beta \mapsto \beta n^2 \mod 1.$$

For instance:

$$\widehat{K_N}(1/3) = \frac{1}{N}\sum_{n \leq N} e^{-2\pi i n^2/3} = \frac{1}{3}(1 + 2 \cdot e^{-2\pi i/3}) + O(1/N).$$

Number-Theoretic Consequences

After elementary analytic number theoretic methods, we are able to make the following distinction:

- $\widehat{K_N}(\beta)$ is large and interesting whenever β is "*N*-close" to a rational number with an "*N*-small" denominator, i.e. β lives in a so-called "*N*-major arc;" and
- $\widehat{K_N}(\beta)$ and is "*N*-negligible" otherwise, when β lives in the complementary "*N*-minor arc," and the probabilistic intuition applies.

Back in Physical Space

By translating the information from the circle method back to physical space, matters reduce to understanding the behavior of K_N when testing against functions like

$$\chi_Q(n) = C_Q(n) \cdot \frac{1}{Q^{100}} \mathbf{1}_{[1,Q^{100}]}(n),$$

where

$$C_Q(n) := \sum_{(A,Q)=1} e^{-2\pi i A/Q \cdot n}.$$

The relevant scaling is $Q \leq N^{0.001}$.

Multi-Frequency Harmonic Analysis

Roughly speaking, Bourgain's challenge was to estimate

$$f \mapsto t \cdot \mathcal{N}_t \left(K_N * \sum_{i=1}^{Q} f_{Q_i} : N \geq Q^{1000} \right)^{1/r}$$

as *efficiently* as possible, where we define

$$f_{Q_i}(x) := \sum_n f(x-n)\chi_{Q_i}(n) = \frac{1}{Q_i^{100}} \sum_{n \leq Q_i^{100}} f(x-n) C_{Q_i}(n),$$

and each $Q_i \approx Q$; the triangle inequality is the enemy.

Single Scale Estimate

With

$$f_Q(x) = \frac{1}{Q^{100}} \sum_{n \le Q^{100}} f(x-n) C_Q(n)$$

$$C_Q(n) := \sum_{(A,Q)=1} e^{-2\pi i A/Q \cdot n}$$

elementary number theory yields the following estimate.

For each $N \geq Q^{1000}$ the following estimate holds for $Q_i \approx Q$:

$$\|K_N * \sum_{i=1}^{Q} f_{Q_i}\|_2 \leq 100 \cdot Q^{-1/2} \cdot \|f\|_2.$$

The Key Estimate

With

$$f_Q(x) = \frac{1}{Q^{100}} \sum_{n \leq Q^{100}} f(x-n) C_Q(n)$$

$$C_Q(n) := \sum_{(A,Q)=1} e^{-2\pi i A/Q \cdot n}$$

as above, the following estimate holds.

Quantifying Destructive Interference

For each $r > 2$, the following estimate holds for $Q_i \approx Q$:

$$\| t \cdot \mathcal{N}_t \Big(K_N * \sum_{i=1}^{Q} f_{Q_i} : N \geq Q^{1000} \Big)^{1/r} \|_2 \\ \leq A_r \cdot \log Q \cdot Q^{-1/2} \cdot \|f\|_2.$$

Summarizing the Linear Theory

In seeking to prove pointwise convergence of the linear averages, we have a rough scheme:

- Transfer the problem to a quantitative estimate involving convolution (averaging) operators on the integers;
- Use the Fourier transform to extract the analytic heart of the averaging operators;
- Estimate $\mathcal{N}_t$.

This scheme is most straightforward on ℓ^2, where Plancherel's theorem allows us to quantify heuristics like "$=$."

Back to the Bilinear Setting

Our approach to proving pointwise convergence of the averages

$$B_N(f,g) := \frac{1}{N}\sum_{n\leq N} T^n f \cdot T^{n^2} g$$

goes through a similar mechanism.
If we let

$$M_N(f,g)(x) := \frac{1}{N}\sum_{n\leq N} f(x-n)\cdot g(x-n^2)$$

denote the pertaining discrete bilinear operators, then we are after estimates of the form

$$\left\| t\cdot \mathcal{N}_t(M_N(f,g) : N)^{1/r} \right\|_{\ell^1} \leq A_r \cdot \|f\|_{\ell^2}\cdot\|g\|_{\ell^2}.$$

Harmonic Analysis Meets Additive Combinatorics

We aren't able to appeal directly to Plancherel's theorem to approximate M_N. Our departure part was recent *quantitative* refinements of the Non-Linear-Roth-Type Corollary due to S. Peluse and S. Prendiville:

Key Inverse Theorem

Suppose that $|f|, |g| \leq \mathbf{1}_I$ for some interval, I, with $|I| \approx N^2$. If

$$\|M_N(f,g)\|_{\ell^1} \geq \log^{-O(1)} N \cdot N^2,$$

then there exists some $q \leq \log^{O(1)} N$ so that

$$\|\frac{1}{M} \sum_{n \leq M} f(x-qn)\|_{\ell^1} \geq \log^{-O(1)} N \cdot N^2$$

for some M with $N \cdot \log^{-O(1)} N \leq M \leq N$.

Harmonic Analysis Meets Additive Combinatorics, II

The Peluse-Prendiville Inverse Theorem can be re-formulated as follows, after a little work:

Key Inverse Theorem, Fourier Formulation

Suppose that $|f|, |g| \leq \mathbf{1}_I$ for some interval, I, with $|I| \approx N^2$. If

$$\|M_N(f,g)\|_{\ell^1} \geq \log^{-O(1)} N \cdot N^2,$$

then there exists some $q \leq \log^{O(1)} N$ so that $\hat{f}$ is generically large on

$$\mathbb{Z}/q\mathbb{Z} + O(\log^{O(1)} N \cdot N^{-1})$$

and $\hat{g}$ is generically large on

$$\mathbb{Z}/q\mathbb{Z} + O(\log^{O(1)} N \cdot N^{-2}).$$

Techniques Needed

Using the above inverse theorem as our departure point, our approximation argument lived at the interface of additive combinatorics and Fourier analysis. The principle tools used were

- Ionescu-Wainger Multiplier Theory ("Discrete Littlewood-Paley Theory");
- Hahn Banach separation theorem;
- Vinogradov Mean Value Theorem.

The Upshot

After making these approximations, we needed to understand super-positions of

$$M_N(f_{Q^1}, g_{Q^2}),$$

where

$$\mathrm{lcm}(Q^1, Q^2) \leq N^{0.001},$$

and we recall

$$f_Q(x) = \frac{1}{Q^{100}} \sum_{n \leq Q^{100}} f(x-n) \cdot C_Q(n).$$

Contrast to Bourgain's situation, where the issue was understanding

$$K_N * \sum f_Q.$$

Single Scale Estimate

With

$$f_Q(x) = \frac{1}{Q^{100}} \sum_{n \leq Q^{100}} f(x-n) C_Q(n)$$

$$C_Q(n) := \sum_{(A,Q)=1} e^{-2\pi i A/Q \cdot n}$$

the Peluse-Prendiville theory (eventually) leads to the following estimate.

For each $N \geq Q^{1000}$ there exists some $C, c > 0$ so that following estimate holds for $\operatorname{lcm}(Q_i^1, Q_i^2) \approx Q$:

$$\| \sum_{i=1}^{Q} M_N(f_{Q_i^1}, g_{Q_i^2}) \|_{\ell^1} \leq C \cdot Q^{-c} \cdot \|f\|_{\ell^2} \cdot \|g\|_{\ell^2}.$$

Multi-Frequency Bilinear Harmonic Analysis

Roughly speaking, our challenge became to efficiently estimate

$$\mathcal{N}_t\Big(\sum_{i=1}^{Q} M_N(f_{Q_i^1}, g_{Q_i^2}) : N \geq Q^{1000}\Big),$$

where $\mathrm{lcm}(Q_i^1, Q_i^2) \approx Q$.

Quantifying Destructive Interference, Bilinear Setting

For each $r > 2$, there exists some $c_0 > 0$ so that the following estimate holds for $\mathrm{lcm}(Q_i^1, Q_i^2) \approx Q$ as above:

$$\|t \cdot \mathcal{N}_t\Big(\sum_{i=1}^{Q} M_N(f_{Q_i^1}, g_{Q_i^2}) : N \geq Q^{1000}\Big)^{1/r}\|_{\ell^1}$$
$$\leq A_r \cdot Q^{-c_0} \cdot \|f\|_{\ell^2} \cdot \|g\|_{\ell^2}.$$

Techniques Required

The key additive combinatorial input used was a Non-linear Roth theorem in the *cyclic subgroup setting*.

NLR Theorem, Cyclic Subgroup Setting

Suppose that $\delta > 0$, and that $p \geq \delta^{-O(1)}$ is a prime that is sufficiently large depending on δ. Then for any $Q = p^j, \ j \geq 1$, if $A \subset \mathbb{Z}/Q\mathbb{Z}$ has $|A| \geq \delta \cdot Q$, then A contains a non-linear arithmetic progression, $x, x-n, x-n^2$.

Further Ideas

- Martingale methods;
- Banach space geometry;
- Oscillatory integral techniques over the p-adics.

Our argument by necessity combines physical-space techniques from additive combinatorics with Fourier-space techniques from harmonic analysis.

Moving Forwards

I'm working on upgrading our convergence result to higher degrees of multi-linearity in progress with M. Mirek, S. Peluse, and J. Wright. Using more refined work of Peluse, the following result is expected.

For every $P_1, \dots, P_m \in \mathbb{Z}[\cdot]$ with distinct degrees and each $f_1, \dots, f_m \in L^\infty(X)$,

$$\frac{1}{N} \sum_{n \leq N} T_1^{P_1(n)} f_1 \cdot \dots \cdot T_m^{P_m(n)} f_m$$

converge pointwise almost everywhere. As of mid 2023, we have the result when $T_i = T$ and P_i are polynomials of distinct degrees.

Same Degrees

Regarding polynomials with the same degrees, the model problem is as follows.

Problem

Prove that for $f_1, f_2 \in L^\infty(X)$

$$\frac{1}{N}\sum_{n\leq N} T^{n^2} f_1 \cdot T^{2n^2} f_2$$

converge *pointwise* almost everywhere.

Solving this problem will require additional input from time frequency analysis.

Same Degrees II

Another major open question in pointwise ergodic about which Bourgain thought deeply concerns the following trilinear averages.

Problem

Prove that for $f_1, f_2, f_3 \in L^\infty(X)$

$$\frac{1}{N} \sum_{n \leq N} T^n f_1 \cdot T^{2n} f_2 \cdot T^{3n} f_3$$

converge *pointwise* almost everywhere.

This problem will require higher-order Fourier analysis, and is closely linked to the *trilinear Hilbert transform* from time-frequency analysis.

Thank you!