> **This is a recreation of the original chat by Codex using the data export supplied by the user.**

Quantitative 2-largeness

Conversation ID: 6aa3e14a-7728-83e9-98fe-ec4409df5960

363 recorded user and assistant messages • Dates shown in Asia/Shanghai (UTC+08:00), with source timestamps.

Source: ChatGPT data export

Source: supplied ChatGPT data export dated 2026-10-25. Conversation URL: https://chatgpt.com/c/6aa3e14a-7728-83e9-98fe-ec4409df5960
This PDF reproduces all 363 saved user questions and assistant text responses: 355 on the saved active conversation path and 8 saved alternate message versions in a separate appendix.
Messages retain their original wording. Mathematics and Markdown are typeset; citation chips are reconstructed from exported reference metadata.
Message dates use the original creation timestamps, displayed in Asia/Shanghai (UTC+08:00); UTC source timestamps are also shown.
The exact source message text is embedded in this PDF as a UTF-8 transcript and JSON record. Original uploaded files and available cited file assets are also embedded; use a PDF reader with an attachments panel.
This is a record of the conversation, including any errors and later corrections in the original answers.
Message date range: 2026-09-11 19:08:57 (Asia/Shanghai, UTC+08:00) to 2026-10-12 19:20:22 (Asia/Shanghai, UTC+08:00)

Typesetting note: 19 formulas are shown in unchanged source form because the original notation contains unsupported commands, ambiguous custom macros, or malformed syntax.

Main conversation

1. User

2026-09-11 19:08:57 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-11T11:08:57.896000Z

Does the contents of this draft of a paper with Ryan relate to littlewoods conjecture in diophantine approximation in any way?

> **Attachments recorded in source:**
>
> - Pasted text(20260911-110840).txt

---

2. Assistant

2026-09-11 19:09:11 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-11T11:09:11.600414Z

I'll compare the draft's quantitative approximation theorem with Littlewood's denominator-sensitive bound, then check whether the nilsystem framework provides a direct connection or implication.

---

3. Assistant

2026-09-11 19:12:14 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-11T11:12:14.004893Z

The crucial gap is denominator size: the draft controls approximation errors through progression length k, while leaving the denominator s unbounded.

---

4. Assistant

2026-09-11 19:18:16 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-11T11:18:16.635631Z

Yes. There is a concrete connection through your generalized-polynomial approximation theorem: it gives a quantitative statement about Littlewood's exact expression, but only modulo the integers. There is also a useful Ramsey-theoretic reformulation of Littlewood's conjecture that fits naturally with the paper.

The distinction between approximation **to an integer** and approximation **to zero in** $\mathbb{R}$ is essential here. In fact, a Littlewood conclusion along every 2-large set would be false—even along every large set of natural density one.

Here are the connections and limitations in detail.

1. Your finite theorem gives a "modular Littlewood" corollary

Write

$$F_{\alpha,\beta}(n) := n\|n\alpha\|_{\mathbb{R}/\mathbb{Z}}\|n\beta\|_{\mathbb{R}/\mathbb{Z}}.$$

Littlewood's conjecture asserts that

$$\liminf_{n\to\infty} F_{\alpha,\beta}(n) = 0 \qquad \text{for every } \alpha,\beta \in \mathbb{R}.$$

Thus, the desired smallness is smallness of a nonnegative real number, not its distance to the nearest integer. arXiv

Taking your finite generalized-polynomial approximation theorem as stated, one obtains the following consequence.

Corollary. There are absolute constants $C, k_0 > 0$ such that, for every $k \geq k_0$, every $(2,k)$-large set $S \subseteq \mathbb{N}$, and every $\alpha,\beta \in \mathbb{R}$, there exists $s \in S$ satisfying

$$\boxed{\operatorname{dist}\big(s\|s\alpha\|_{\mathbb{R}/\mathbb{Z}}\|s\beta\|_{\mathbb{R}/\mathbb{Z}}, \mathbb{Z}\big) \leq Ck^{-3}.} \tag{1}$$

Moreover, for fixed α,β, there is one two-coloring such that the difference of every monochromatic k-term progression satisfies (1), simultaneously for every $k \geq k_0$. This follows from the corresponding stronger assertion in your finite theorem. [Pasted text.txt, L189-L203]

Proof. Use the draft's nearest-integer convention and put

$$a_n = \langle \alpha n \rangle, \qquad b_n = \langle \beta n \rangle.$$

Consider the signed expression

$$E(n) := n(\alpha n - a_n)(\beta n - b_n).$$

Its absolute value is exactly Littlewood's expression:

$$|E(n)| = F_{\alpha,\beta}(n).$$

Expanding gives

$$E(n) = \alpha\beta n^3 - \alpha n^2\langle \beta n\rangle - \beta n^2 \langle \alpha n\rangle + n\langle \alpha n\rangle\langle \beta n\rangle.$$

The last summand is an integer. Consequently,

$$E(n) \equiv \alpha\beta n^3 - \alpha n^2\langle \beta n\rangle - \beta n^2\langle \alpha n\rangle \pmod{\mathbb{Z}}. \tag{2}$$

The three expressions on the right are **special generalized polynomials of weight** 3 in precisely the sense used in the draft:

$$\mathsf{L}(\alpha\beta n^3), \qquad \mathsf{L}(\alpha n^2, \beta n), \qquad \mathsf{L}(\beta n^2, \alpha n).$$

Your assigned-weight convention applies regardless of special choices of the coefficients. [Pasted text.txt, L171-L187]

Applying the finite theorem simultaneously to these three expressions gives

$$\|E(s)\|_{\mathbb{R}/\mathbb{Z}} \leq 3C(3,3)k^{-3}.$$

Since distance to $\mathbb{Z}$ is unchanged by replacing a real number with its absolute value,

$$\|E(s)\|_{\mathbb{R}/\mathbb{Z}} = \operatorname{dist}(F_{\alpha,\beta}(s), \mathbb{Z}),$$

which proves the claim.

So there is a direct algebraic connection—not just the broad observation that both topics concern Diophantine approximation. **Your weight-three theorem controls Littlewood's exact quantity modulo $\mathbb{Z}$.**

2. Why this does not yet give Littlewood's conjecture

There are two related obstacles.

The outer distance to $\mathbb{Z}$

The conclusion

$$\operatorname{dist}(F_{\alpha,\beta}(s),\mathbb{Z}) \longrightarrow 0$$

allows $F_{\alpha,\beta}(s)$ to approach 1, 2, or a sequence of integers tending to infinity. It does not force

$$F_{\alpha,\beta}(s) \longrightarrow 0.$$

The expansion (2) therefore identifies exactly where generalized-polynomial recurrence falls short: it discards an integer-valued contribution, whereas Littlewood's conjecture depends on the size of the entire real-valued expression.

The missing control on the denominator

Applying your theorem just to the two linear polynomials gives

$$\|s\alpha\|_{\mathbb{R}/\mathbb{Z}} \le Ck^{-1}, \qquad \|s\beta\|_{\mathbb{R}/\mathbb{Z}} \le Ck^{-1},$$

and hence

$$F_{\alpha,\beta}(s) \le C^2\frac{s}{k^2}. \tag{3}$$

There is no reason for the right-hand side to tend to zero without additional information about s.

Your introduction already states the relevant limitation explicitly: the theorem bounds the approximation error but gives no bound on the size of the allowed difference s. [Pasted text.txt, L206-L214]

The progression length k and the denominator s are different parameters. The powers k^{-d_i}, impressive as they are, cannot simply be read as powers s^{-d_i}.

3. A stronger warning: Littlewood approximation along arbitrary 2-large sets is false

This is an elementary observation, but I think it is particularly useful for interpreting your theorem.

Proposition. For every pair of irrational numbers α,β and every $\varepsilon>0$, the set

$$S_\varepsilon := \{n\in\mathbb{N} : F_{\alpha,\beta}(n)\ge\varepsilon\}$$

has natural density one and is large. In particular, it is 2-large, but

$$\inf_{s\in S_\varepsilon} s\|s\alpha\|_{\mathbb{R}/\mathbb{Z}}\|s\beta\|_{\mathbb{R}/\mathbb{Z}} \ge \varepsilon. \tag{4}$$

Here is a proof, including why density one suffices for largeness.

The good Littlewood times have density zero

Let

$$L_\varepsilon := \{n : F_{\alpha,\beta}(n) < \varepsilon\}.$$

Fix an integer $H \geq 1$. Because α, β are irrational, we can choose $\delta > 0$ such that

$$2\delta < \min_{1\leq h\leq H} \min\big(\|h\alpha\|_{\mathbb{R}/\mathbb{Z}}, \|h\beta\|_{\mathbb{R}/\mathbb{Z}}\big).$$

For either $\theta = \alpha$ or $\theta = \beta$, two distinct elements of

$$A_{\theta,\delta} := \{n : \|n\theta\|_{\mathbb{R}/\mathbb{Z}} < \delta\}$$

cannot differ by an integer between 1 and H. Otherwise the triangle inequality would contradict the choice of δ. Thus

$$\overline{d}(A_{\theta,\delta}) \leq \frac{1}{H+1}.$$

For $n > \varepsilon/\delta^2$, membership in L_ε implies

$$\|n\alpha\|_{\mathbb{R}/\mathbb{Z}}\|n\beta\|_{\mathbb{R}/\mathbb{Z}} < \delta^2,$$

so at least one of the two factors is less than δ. Therefore

$$\overline{d}(L_\varepsilon) \leq \frac{2}{H+1}.$$

Since H is arbitrary, $d(L_\varepsilon) = 0$, proving $d(S_\varepsilon) = 1$.

Every density-one set is large

Suppose $d(S) = 1$. For every fixed M, each set

$$\{q \in \mathbb{N} : jq \notin S\}, \qquad 1 \leq j \leq M,$$

has density zero. Their finite union has density zero, so there is a q with

$$q\{1, \dots, M\} \subseteq S. \tag{5}$$

Now fix r, k, and take $M = W(r,k)$. Given any r-coloring c of $\mathbb{N}$, apply the finite van der Waerden theorem to the coloring

$$j \longmapsto c(qj), \qquad 1 \leq j \leq M.$$

The resulting monochromatic progression in $\mathbb{N}$ has common difference qd for some $1 \leq d \leq M$, and (5) guarantees $qd \in S$. Hence S is large.

This proves the proposition.

Consequently, the statement

$$S \text{ is 2-large} \quad \Longrightarrow \quad \inf_{s\in S} F_{\alpha,\beta}(s) = 0$$

is false, even with “ 2-large” replaced by “large and of density one.”

This does not contradict Littlewood’s conjecture: it merely removes all its good approximation times, which form a density-zero set. It does show that the distinction between fixed nilpotent return conditions and Littlewood’s shrinking targets is substantive.

4. There is an exact Ramsey-theoretic reformulation—and your localization theorem applies to it

For fixed α, β, retain

$$L_\varepsilon = \{n : F_{\alpha,\beta}(n) < \varepsilon\}.$$

The following equivalence puts Littlewood's conjecture directly into the language of your paper:

$$\boxed{\begin{aligned}&\liminf_{n\to\infty} F_{\alpha,\beta}(n) = 0\\ &\qquad\Longleftrightarrow L_\varepsilon \text{ is large for every } \varepsilon > 0\\ &\qquad\Longleftrightarrow L_\varepsilon \text{ is 2-large for every } \varepsilon > 0.\end{aligned}} \tag{6}$$

The forward implication is stronger than mere nonemptiness, but has a short proof.

For every positive integer j,

$$\|jq\alpha\| \le j\|q\alpha\|, \qquad \|jq\beta\| \le j\|q\beta\|,$$

so

$$F_{\alpha,\beta}(jq) \le j^3 F_{\alpha,\beta}(q). \tag{7}$$

Assuming Littlewood's conclusion, given M and ε, choose q with

$$F_{\alpha,\beta}(q) < \frac{\varepsilon}{M^3}.$$

Then (7) gives

$$q\{1, \dots, M\} \subseteq L_\varepsilon.$$

The same van der Waerden argument used above proves that L_ε is large.

Conversely, nonemptiness of every L_ε already gives the required liminf. If a zero occurs at a fixed denominator, its multiples also give zeros; otherwise denominators realizing values tending to zero must tend to infinity.

Thus, **proving 2-largeness of these particular shrinking-target sets would prove Littlewood's conjecture.** But (6) is an exact reformulation, not yet a reduction in difficulty: the draft assumes 2-largeness rather than establishing it for these sets.

A conditional consequence of your nil-Bohr localization theorem

Suppose a particular pair (α, β) satisfies Littlewood's conclusion. Then every L_ε is large. Your localization corollary says that

$$S \text{ is } 2r\text{-large}, \quad B \text{ is finite-step nil-Bohr}_0 \quad\Longrightarrow\quad S \cap B \text{ is } r\text{-large}.$$

Applying this for every r gives

$$\boxed{L_\varepsilon \cap B \text{ is large}} \tag{8}$$

for every $\varepsilon > 0$ and every finite-step nil-Bohr $_0$ set B. [Pasted text.txt, L1513-L1529]

In words: **once Littlewood approximation is available for a pair, your localization theorem lets one impose arbitrary fixed finite-step nilpotent return conditions while retaining largeness of the permitted differences.**

For example, using the normalized generalized-polynomial neighborhoods discussed in the draft, (8) says that

$$\left\{ s : \begin{array}{l} s\|s\alpha\|\|s\beta\| < \varepsilon, \\ \|P_i(s)\| < \eta \quad (1 \le i \le m) \end{array} \right\}$$

is large whenever (α, β) satisfies Littlewood and the P_i belong to your normalized generalized-polynomial class. The draft identifies the simultaneous generalized-polynomial approximation set as a nil-Bohr $_0$ set. [Pasted text.txt, L1558-L1579]

This is a clean connection to your localization result, although I would not claim novelty for this conditional consequence without a more targeted literature check.

5. The homogeneous-dynamics connection is real, but the ambient systems differ

There is also a conceptual connection through homogeneous dynamics. Your structural theorem concerns polynomial orbits on a fixed compact nilmanifold and forces monochromatic orbits to have only smooth and bounded-period rational factors. [Pasted text.txt, L830-L846]

For Littlewood's conjecture, the standard dynamical model is instead

$$X_3 = \mathrm{SL}_3(\mathbb{R}) / \mathrm{SL}_3(\mathbb{Z}),$$

with the lattice

$$\Lambda_{\alpha,\beta} = \begin{pmatrix} 1 & 0 & 0 \\ \alpha & 1 & 0 \\ \beta & 0 & 1 \end{pmatrix} \mathbb{Z}^3$$

and the diagonal semigroup

$$A^+ = \left\{ \mathrm{diag}(e^{-t-u}, e^t, e^u) : t, u \ge 0 \right\}.$$

Littlewood's conclusion for (α, β) is equivalent to $A^+\Lambda_{\alpha,\beta}$ not being relatively compact. This is stated and proved in Proposition 11.1 of Einsiedler–Katok–Lindenstrauss, *Invariant measures and the set of exceptions to Littlewood's conjecture*. arXiv

The desired event is the appearance of arbitrarily short nonzero lattice vectors—excursions into the cusp of a noncompact space—not recurrence to a neighborhood in a fixed compact nilmanifold. The diagonal action therefore does not fall under your nilsystem theorem merely because unipotent matrices appear in the description of the initial lattice.

My assessment for the paper

The most concrete addition would be the weight-three modular corollary (1), accompanied by a clear warning that it does not control the real size of Littlewood's expression. Its proof is short, uses exactly your special generalized-polynomial class, and retains the coefficient-uniform k^{-3} rate.

The density-one obstruction in Section 3 above is also useful: it explains why replacing modular approximation by actual Littlewood approximation along every 2-large set is not a possible strengthening.

So I would describe the relationship as **a quantitative Ramsey-theoretic modular analogue of Littlewood approximation, together with a nil-Bohr localization principle for pairs already satisfying Littlewood**—not, on the basis of the present statements alone, as a proof of a new case of the classical conjecture.

---

5. User

2026-09-11 19:34:26 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-11T11:34:26.632000Z

My colleagues codex generated this shorter alternative to the previous paper. However, it is a little too short for me. Can you add more details to this so that it is easier to understand? For example, if it ended up becoming 10 pages, it would still be a nice shorter alternative to the original. Please give the final result as a standalone tex file.

> **Attachments recorded in source:**
>
> - Pasted text(20260911-113317).txt

---

6. Assistant

2026-09-11 19:34:43 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-11T11:34:43.345346Z

I'll preserve the shorter draft's elementary structure while expanding compressed steps, especially compactness, rounding, and coefficient estimates, then compile the standalone LaTeX for readability and length.

---

7. Assistant

2026-09-11 19:37:46 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-11T11:37:46.170190Z

I'm also unpacking the shell-coloring and first-exit arguments, and making parameter order explicit so one fixed coloring works uniformly in coefficients and progression length.

---

8. Assistant

2026-09-11 19:47:05 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-11T11:47:05.959230Z

I expanded the shorter draft into a **10-page exposition**, retaining its elementary approach through weighted matrices, the shell coloring, and polynomial rounding. [Pasted text.txt, L351-L355]

The added detail covers the construction of canonical lattice representatives, continuity and uniform oscillation of the shell function, the sampling argument at color boundaries, and the Newton-interpolation step that produces integer-valued polynomial corrections. The first-exit argument and the clearing of factorial denominators are also written out more fully, with explicit parameter dependencies showing why the coloring is fixed independently of the progression length and why the constants are coefficient-uniform.

I also included worked examples illustrating an integer-valued correction with noninteger coefficients, leading-coefficient extraction in the scalar case, and the recovery of a nested bracket from a 3×3 matrix.

The LaTeX file requires no separate bibliography or auxiliary source files and compiles with pdfLaTeX without warnings.

Download the standalone LaTeX file

View the compiled PDF — 10 pages

---

9. User

2026-09-12 11:08:49 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-12T03:08:49.070165Z

I don't understand the sentence after equation 2.3. what is the coefficient extraction? Can you give me some examples here in chat?

---

10. Assistant

2026-09-12 11:10:38 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-12T03:10:38.125653Z

Here **"coefficient extraction" means taking the coefficient of a specified power of** t. There is no rounding involved. The notation $[t^w]$ is an operator applied to the polynomial immediately following it. That is what the sentence after (2.3) is distinguishing from the nearest-integer brackets used elsewhere in the paper. [elementary_two_color_obstruction_expanded.tex, L15-L20]

1. Ordinary polynomial examples

If

$$f(t) = c_0 + c_1 t + \cdots + c_d t^d,$$

then

$$\boxed{[t^w] f(t) = c_w,}$$

with the convention that this is 0 when the t^w term is absent.

For example, take

$$f(t) = 7 + 3t - \frac{5}{2}t^2 + \sqrt{2}\,t^4.$$

Then

$$\begin{aligned}
[t^0] f(t) &= 7, \\
[t^1] f(t) &= 3, \\
[t^2] f(t) &= -\frac{5}{2}, \\
[t^3] f(t) &= 0, \\
[t^4] f(t) &= \sqrt{2}, \\
[t^5] f(t) &= 0.
\end{aligned}$$

In particular, the result need not be an integer: we are **reading off a coefficient**, not rounding it.

The polynomial need not already be expanded. For example,

$$[t^2](2+3t)^3 = 54,$$

because

$$(2+3t)^3 = 8 + 36t + 54t^2 + 27t^3.$$

2. What this means in equation (2.3)

The equation is

$$\tau(A)_{ij} = [t^{w_{ij}}]A_{ij}(t).$$

It says:

> In position (i,j), take the coefficient of the power prescribed by that position's weight.

The crucial point is that **different matrix entries may have different prescribed weights**. Also, we take the coefficient at the *prescribed weight*, not necessarily the highest power that actually occurs. As the draft says, an entry whose degree is smaller than its weight contributes zero.
[elementary_two_color_obstruction_expanded.tex, L15-L20]

For example, take a 3×3 block with adjacent weights

$$w_1 = 1, \qquad w_2 = 2.$$

The three off-diagonal positions therefore have weights

$$w_{12} = 1, \qquad w_{23} = 2, \qquad w_{13} = 3.$$

Consider

$$A(t) = \begin{pmatrix} 1 & 5+2t & 11+7t+3t^2+4t^3 \\ 0 & 1 & 6+9t+8t^2 \\ 0 & 0 & 1 \end{pmatrix}.$$

We extract

$$\begin{aligned} \tau(A)_{12} &= [t^1](5+2t) = 2, \\ \tau(A)_{23} &= [t^2](6+9t+8t^2) = 8, \\ \tau(A)_{13} &= [t^3](11+7t+3t^2+4t^3) = 4. \end{aligned}$$

Thus

$$\boxed{\tau(A) = \begin{pmatrix} 1 & 2 & 4 \\ 0 & 1 & 8 \\ 0 & 0 & 1 \end{pmatrix}.}$$

The diagonal entries are kept equal to 1; the extraction rule applies to the off-diagonal entries.

Now suppose we replace the top-right entry by

$$A_{13}(t) = 11 + 7t + 3t^2.$$

Its actual leading coefficient is 3, but its assigned weight is still 3. Consequently,

$$\tau(A)_{13} = [t^3](11+7t+3t^2) = 0,$$

not 3. This is why "weighted leading coefficient" needs a little care.

3. Why it removes the starting point of a progression

Here is the simplest example of why we use this operation.

Suppose we evaluate the monomial $x \mapsto \alpha x^2$ along a progression $a + dt$. Then

$$\alpha(a + dt)^2 = \alpha a^2 + 2\alpha a d\, t + \alpha d^2 t^2.$$

Taking the coefficient of t^2 gives

$$\boxed{[t^2]\big(\alpha(a + dt)^2\big) = \alpha d^2.}$$

The starting point a has disappeared, and we have recovered the original monomial evaluated at the common difference d.

More generally,

$$[t^w]\big(\alpha(a + dt)^w\big) = \alpha d^w.$$

With the normalization used later in the paper, this becomes

$$[t^w]\left(\alpha\left(\frac{a + dt}{L}\right)^w\right) = \alpha\left(\frac{d}{L}\right)^w.$$

Multiplying by L^w then recovers αd^w. This is precisely the scalar mechanism described in the draft's later remark on leading-coefficient extraction. [elementary_two_color_obstruction_expanded.tex, L52-L65]

4. Why the matrix weights are additive

A small example also explains the identity

$$\tau(AB) = \tau(A)\tau(B)$$

that follows (2.3). The draft proves this using the additive weights and the matrix multiplication formula. [elementary_two_color_obstruction_expanded.tex, L5-L9] [elementary_two_color_obstruction_expanded.tex, L23-L35]

Take adjacent weights $w_1 = w_2 = 1$, so the top-right entry has weight 2, and let

$$A(t) = \begin{pmatrix} 1 & 2t + 5 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad B(t) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 3t + 7 \\ 0 & 0 & 1 \end{pmatrix}.$$

Their product is

$$A(t)B(t) = \begin{pmatrix} 1 & 2t + 5 & (2t + 5)(3t + 7) \\ 0 & 1 & 3t + 7 \\ 0 & 0 & 1 \end{pmatrix}.$$

The top-right entry is

$$(2t + 5)(3t + 7) = 6t^2 + 29t + 35,$$

so extracting the prescribed coefficients gives

$$\tau(AB) = \begin{pmatrix} 1 & 2 & 6 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{pmatrix}.$$

On the other hand,

$$\tau(A)\tau(B) = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 6 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{pmatrix}.$$

The cross term $2 \cdot 3 = 6$ survives because its weight is $1 + 1 = 2$.

A clearer replacement for the sentence after (2.3) would therefore be:

> For an ordinary polynomial $f(t) = \sum_{r=0}^{d} c_r t^r$, the notation $[t^w]f(t)$ means the coefficient c_w, interpreted as zero if that power is absent. Thus $\tau(A)$ is obtained by replacing each off-diagonal entry $A_{ij}(t)$ by its coefficient of $t^{w_{ij}}$, while keeping the diagonal entries equal to one. These brackets do not denote nearest-integer rounding.

---

11. User

2026-09-14 17:33:04 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-14T09:33:04.809000Z

in ryan alweiss' paper about new obstacles to multiple recurrence, or ben green's note on this paper, the finite sets they construct that are used to build the global set, do they naturall satisfy the quantitative recurrence condition of the paper in this chat? I am not asking if they can be modified to do so, as I believe that they can. I am asking if as they are already presented, they satisfy this stronger quantitative condition.

---

12. Assistant

2026-09-14 17:33:20 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-14T09:33:20.183879Z

I'll compare their finite constructions directly with your k^{-d} bounds, keeping the sets and parameter choices exactly as presented rather than considering modifications.

---

13. Assistant

2026-09-14 17:36:22 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-14T09:36:22.583977Z

Green's proof already gives a uniform $1/m$ finite-stage return bound. I'm checking whether coefficient-uniformity alone upgrades this to the required higher-weight powers without altering the set.

---

14. Assistant

2026-09-14 17:44:50 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-14T09:44:50.251177Z

Yes—for the finite integer truncations in Green's presentation, they already satisfy the quantitative approximation condition from your draft, without changing the construction or passing to later stages. In fact, one can deduce a stronger statement: for a fixed number of special generalized polynomials of bounded weight, those finite truncations give approximation at **every fixed inverse power of the stage index**.

This stronger estimate is **a consequence of his argument, not a statement that he explicitly formulates**. The key is the uniformity in the polynomial sequence in his Proposition 3.3. arXiv

I will state precisely which finite sets I am discussing and explain the deduction. For Ryan's original presentation, there is a bookkeeping distinction that I address at the end.

1. The precise statement for Green's existing finite sets

Fix the sequences (α_k), (N_k), and the set S constructed in Sections 2–3 of Green's note. Write

$$F_k := S \cap [1, N_k].$$

These are the natural finite truncations at the cutoffs already chosen in his construction. We do **not** enlarge N_k after specifying the approximation problem. arXiv

Here is the conclusion:

> **Proposition.** Fix positive integers r, D. There are constants C, k_0 such that, for every $k \geq k_0$ and every family of special generalized polynomials
>
> $$P_1, \dots, P_r$$
>
> of respective assigned weights $1 \leq d_i \leq D$, there is an $s \in F_k$ satisfying
>
> $$\boxed{\|P_i(s)\| \leq Ck^{-d_i} \qquad (1 \leq i \leq r).} \tag{1}$$
>
> The constants are independent of all the real coefficients.

The threshold also depends on the fixed enumeration of nilmanifolds used in Green's construction. Once that enumeration is fixed, it is a function of r, D.

This is the **approximation conclusion** of your finite theorem. I am not asserting that F_k satisfies its ** $(2, k)$-largeness hypothesis**. Your theorem distinguishes these two things explicitly. [Pasted text.txt, L189-L203]

Moreover, the following stronger version holds:

> For every fixed r, D, A, there are C, k_0 such that every $k \geq k_0$ and every such family admit $s \in F_k$ with
>
> $$\boxed{\max_{1 \leq i \leq r} \|P_i(s)\| \leq Ck^{-A}.} \tag{2}$$

In particular, this uses the **same stage** k, not F_{k^D} or some other reindexed family.

2. What Green's proof actually gives at stage k

The first step is to extract a slightly stronger conclusion from the proof already written in his note.

Fix one of his nilmanifolds $X = G/\Gamma$, of dimension $h > 0$. Choose the auxiliary tolerance

$$\eta = \min\left(\frac{\delta}{20}, \frac{1}{20h}\right), \qquad \delta = \frac{1}{200}.$$

For

$$k \geq \max\left(r(X), \left(\frac{10}{\delta\eta}\right)^{h+1}\right),$$

the argument surrounding equations (3.3)–(3.9) produces $s \in F_k$ with

$$d_X(g^s\Gamma, \Gamma) \leq \frac{1}{k}.$$

Notice that the actual return error is ** $1/k$**, not merely η. The fixed tolerance η is used to ensure membership in S. arXiv

The same verification applies to any normalized polynomial sequence $p(n)$ adapted to the lower central series: Proposition 3.3 is already uniform over all these sequences. Thus the deduction is

$$\boxed{\forall p,\ p(0) = 1_G, \quad \exists s \in F_k : d_X(p(s)\Gamma, \Gamma) \leq \frac{1}{k},} \tag{3}$$

for all sufficiently large k, with threshold depending only on X. arXiv

This is the useful extra uniformity: the same finite set works for every choice of the coefficients of p, even coefficients that themselves depend on k.

3. Passing from (3) to special generalized polynomials

I first explain why (3) implies the following coefficient-uniform estimate:

$$\boxed{\forall P_1, \dots, P_r, \quad \exists s \in F_k : \max_i \|P_i(s)\| \leq \frac{C_0(r, D)}{k}.} \tag{4}$$

One can use the matrix realization from your draft, with a small adjustment to put everything in the **ordinary lower-central-series filtration** used by Green.

For a special expression

$$P(t) = a_1 t^{w_1}\big[a_2 t^{w_2}[\cdots[a_\ell t^{w_\ell}]\cdots]\big], \qquad d = w_1 + \cdots + w_\ell \leq D,$$

define

$$b_0 = 1, \qquad b_j = 1 + w_1 + \cdots + w_j.$$

Work in $\mathrm{UT}_{D+1}(\mathbb{R})$, and put the successive entries

$$-a_1 t^{w_1}, \dots, -a_{\ell-1} t^{w_{\ell-1}}, +a_\ell t^{w_\ell}$$

in the positions

$$(b_0, b_1), \dots, (b_{\ell-1}, b_\ell),$$

with all other off-diagonal entries zero.

The entry containing t^{w_j} is now ** w_j places above the diagonal**. Therefore its degree is permitted by the ordinary lower central series. This is the all-adjacent-weights-equal-to-one case of the coordinate criterion in your draft. [Pasted text.txt, L1053-L1065]

For example, the expression

$$P(t) = at[bt^2]$$

is represented by

$$g_P(t) = \begin{pmatrix} 1 & -at & 0 & 0 \\ 0 & 1 & 0 & bt^2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Here the linear entry has matrix gap 1, and the quadratic entry has gap 2.

Apply (3) to this matrix sequence. Being $O(k^{-1})$-close to the identity coset means that some integer upper-unitriangular matrix γ satisfies

$$g_P(s)\gamma = E, \qquad |E_{uv}| \le C_0/k \quad (u < v).$$

For sufficiently large k, these errors are less than $1/4$. In the example, the relevant equations are

$$bs^2 + \gamma_{24} = E_{24},$$

and

$$\gamma_{14} - as\,\gamma_{24} = E_{14}.$$

The first forces

$$\gamma_{24} = -[bs^2],$$

and substituting into the second gives

$$as[bs^2] + \gamma_{14} = E_{14}.$$

Consequently,

$$\|P(s)\| \le C_0/k.$$

The same downward recursion reads any nested expression; it is exactly the nearest-integer recovery mechanism in your shorter draft. [Pasted text.txt, L325-L349]

For r expressions, use the fixed product nilmanifold

$$\left(\mathrm{UT}_{D+1}(\mathbb{R})/\,\mathrm{UT}_{D+1}(\mathbb{Z})\right)^r.$$

Its dimension, metric, and position in Green's enumeration are independent of the coefficients. This proves (4).

4. Why the k^{-1} estimate already yields the higher powers

This is the main additional observation. It is elementary, but the coefficient-uniformity is essential.

The quadratic improvement

Suppose the same finite set allows us to find s satisfying

$$\|P(s)\| \le \frac{C_0}{k}, \qquad \|qP(s)\| \le \frac{C_0}{k},$$

where

$$q = \left\lfloor \frac{k}{4C_0} \right\rfloor.$$

Write

$$P(s) = a + u, \qquad a \in \mathbb{Z}, \qquad |u| \le C_0/k.$$

Then

$$|qu| \le \frac{1}{4}.$$

There is consequently no wrap around the circle:

$$\|qP(s)\| = \|qa + qu\| = |qu|.$$

Thus

$$|u| \le \frac{C_0}{kq} \ll_{C_0} k^{-2}.$$

So testing the two expressions

$$P, \qquad qP$$

at accuracy $O(k^{-1})$ forces P itself to accuracy $O(k^{-2})$.

Crucially, multiplying a special generalized polynomial by an integer only changes its **outermost coefficient**. It does not change its assigned weight or introduce additional brackets.

The general argument

For each P_i, test the expressions

$$P_i,\ qP_i,\ q^2P_i, \dots, q^{d_i-1}P_i. \tag{5}$$

There are at most rD expressions, all of weight at most D.

Let

$$C_0 = C_0(rD, D)$$

be the constant from (4), enlarged to be at least 1, and choose

$$q = \left\lfloor \frac{k}{4C_0} \right\rfloor.$$

Uniformity in the coefficients lets us apply (4) to the entire family (5), obtaining one $s \in F_k$ such that

$$\|q^j P_i(s)\| \le C_0/k \qquad (0 \le j < d_i). \tag{6}$$

To verify the consequence carefully, put $x = P_i(s)$ and write

$$q^j x = a_j + u_j, \qquad a_j \in \mathbb{Z}, \quad |u_j| \le C_0/k.$$

Then

$$u_{j+1} - qu_j \in \mathbb{Z}.$$

But, for $k \ge 8C_0$,

$$|u_{j+1} - qu_j| \le (q+1)\frac{C_0}{k} \le \frac{1}{4} + \frac{1}{8} < 1.$$

Hence this integer is zero:

$$u_{j+1} = qu_j.$$

It follows that

$$q^{d_i-1}|u_0| = |u_{d_i-1}| \le C_0/k.$$

Since $q \ge k/(8C_0)$,

$$\|P_i(s)\| \le \frac{C_0}{kq^{d_i-1}} \le C_0(8C_0)^{d_i-1}k^{-d_i}.$$

This proves (1).

For (2), simply test

$$P_i,\ qP_i, \dots, q^{A-1}P_i$$

for every i. The same argument gives $O(k^{-A})$.

Nothing about the finite sets changed. We applied an existing uniform property of F_k to a larger, but still fixed-size, family of test expressions. In particular, the number of tests does not grow with k; only their coefficients do.

5. Why this is not just a modification disguised as a proof

The distinction you asked for matters here.

A different argument would be to replace the construction's tolerance $1/k$ by $1/k^D$, enlarge N_k, and choose the next α_{k+1} smaller. That would indeed modify the construction.

The argument above does **none** of that. For a fixed r, D, A, the product nilmanifold needed for the amplified tests is fixed. Once its index is below k, Green's original choice of N_k already covers every normalized polynomial sequence on it—including the sequences with coefficients $q^j a$ that we use.

Thus the answer is affirmative for **every choice of Green's parameters satisfying the requirements of Sections 2–3**, not merely for a specially accelerated implementation.

There is an important normalization caveat: the bounds concern the stage index k, **not** the cutoff N_k, the cardinality $|F_k|$, or the chosen denominator s. No estimate such as

$$\|P_i(s)\| \ll N_k^{-d_i}$$

is being asserted.

6. What I would say about Ryan's original presentation

Ryan's S_m in Definition 2.2 are **finite-dimensional subsets of** $\mathbb{T}^m$, not finite sets of integers. The finite integer cutoffs enter later through the compactness bounds and rapidly shrinking parameters in Theorems 2.8–2.10. That presentation does not specify a stage sequence N_k with Green's precise $1/k$-uniformity. arXiv

Therefore I would distinguish the conclusions as follows:

For Green's natural truncations $S \cap [1, N_k]$, the quantitative condition is already true, with no modifications; the proof above establishes it. For an exact numerical assertion about Ryan's original finite integer cutoffs, I would first need those cutoffs and their indexing fixed more precisely. I would not infer failure from the absence of an explicitly stated rate, but neither would I attribute the exact stage- k assertion to his displayed statements without that extra bookkeeping.

One final observation relevant to your own draft: the same integer-multiplier argument also strengthens the **finite approximation conclusion** of your theorem to arbitrary fixed inverse powers, after allowing the constants to depend on the desired power. It does **not** automatically preserve your stronger conclusion that **one fixed two-coloring works for every progression length**, because the multipliers q used above depend on k. That distinction is worth retaining when comparing the two results. [Pasted text.txt, L189-L203]

---

15. User

2026-09-14 18:03:32 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-14T10:03:32.846000Z

See if you can answer any other question from their paper. If you can, give the final answer as an extra detailed stand alone tex file.

---

16. Assistant

2026-09-14 18:03:47 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-14T10:03:47.253721Z

I'll review the remaining questions in Alweiss's paper and Green's note, test promising ones carefully, and write any complete solution as a detailed standalone LaTeX file.

---

17. Assistant

2026-09-14 18:06:50 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-14T10:06:50.093319Z

Green explicitly leaves measurable recurrence of his set unresolved, so I'm testing that first—including parameter dependence—while also examining Alweiss's two-color question.

---

18. Assistant

2026-09-14 18:15:52 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-14T10:15:52.539376Z

The estimates now give a negative answer for Alweiss's original construction: an explicit two-coloring blocks every 76-term progression. I'm writing the proof and separating Green's half-turn variant.

---

19. Assistant

2026-09-14 18:28:36 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-14T10:28:36.719650Z

Yes. I found a negative answer to Alweiss's Question 3.7 for the original construction in his paper: those sets are not 2-large. The argument applies to the sets actually defined in Section 2, without modifying the allowed differences or imposing additional growth conditions on the sequence (α_j). Question 3.7 is the question asking whether these sets are 2-large. arXiv

Download the standalone LaTeX file

Read the compiled PDF — 10 pages

The explicit conclusion

For the parameters displayed in Ryan's paper,

$$\delta_1 = 10^{-2}, \qquad \delta_2 = 10^{-10},$$

the proof gives a particularly concrete answer. These are already parameter choices presented in the original construction. arXiv

Define

$$F(n) = \sum_{j=1}^{\infty} \bigl(\cos(2\pi n \alpha_j) - 1\bigr).$$

Then the two-coloring

$$\boxed{c(n) = \lfloor 2F(n) \rfloor \pmod 2}$$

has **no monochromatic 76-term arithmetic progression whose common difference belongs to Ryan's original set**.

More generally, the proof works whenever

$$0 < \delta_1 \le \frac{1}{100}, \qquad 0 < \delta_2 \le \delta_1^3,$$

with forbidden progression length

$$K = 3 \left\lfloor \frac{1}{4\delta_1} + \frac{1}{2} \right\rfloor + 1.$$

The main idea

For an allowed difference d, the defining conditions make

$$|F(n+d) - F(n)| < \frac{1}{2}.$$

Consequently, a monochromatic progression cannot move between different half-unit intervals of the same color: **all its F-values must remain in one interval of length $1/2$.**

But the distinguished coordinate makes approximately a quarter-turn after 25 steps. Writing $y_r = F(A + rd)$, this forces

$$\max(|y_0 - 2y_{25} + y_{50}|, |y_{25} - 2y_{50} + y_{75}|) \ge \sqrt{2} - 3 \cdot 10^{-7} > 1.$$

Being trapped in a single half-unit interval forces both expressions to have absolute value less than 1, giving the contradiction. The two overlapping second differences handle every possible initial phase.

Combined with Alweiss's nil-Bohr recurrence theorem, this also gives a set meeting every finite-step nil-Bohr neighborhood that is **not even $(2, 76)$-large**. The nil-Bohr recurrence is the input from his paper; the two-color obstruction is the argument developed here. arXiv

The scope distinction matters: this answers the question for Ryan's original small-step construction. Green's version places the distinguished coordinate near $1/2$, rather than near $-\delta_1$, so the proof does not automatically cover that version. Nor does it settle Green's question about measurable recurrence. The file explains both limitations explicitly. arXiv

20. User

2026-09-15 10:35:24 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T02:35:24.629000Z

can the original result about 2-largeness implying nilrecurrence be improved as follows: For delta in (0,1) let us define D subseteq mathbb{N} to be delta-density large if for any subset A of mathbb{N} with natural upper density delta, there exist arbitrarily long arithmetic progressions in A with common difference from D. I conjecture that if D is delta-density large for some delta in (0,1), the D is good for nilrecurrence.

---

21. Assistant

2026-09-15 10:35:40 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T02:35:40.352011Z

I'll test whether failure of nilrecurrence yields progression obstructions of density arbitrarily close to one; if so, that would establish the conjecture for every $\delta < 1$.

---

22. Assistant

2026-09-15 10:38:43 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T02:38:43.211877Z

The shell argument does extend: a bad long progression must visit every cyclic shell color, so omitting one yields density at least $1 - 1/r$. Thinning reaches exactly δ.

---

23. Assistant

2026-09-15 10:43:31 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T02:43:31.051109Z

Yes. Your conjecture follows from a strengthening of the shell argument in the shorter draft. The useful strengthening is:

> For any prescribed number r of colors, one can arrange that every sufficiently long progression whose difference fails the desired nilpotent approximation must contain **all r colors**.

This is stronger than merely producing an r-coloring with no monochromatic bad progression. It lets us remove just one color, leaving a set of upper density at least $1 - 1/r$, while still excluding all sufficiently long bad progressions.

In fact, **the quantitative generalized-polynomial approximation theorem also survives under your fixed-density hypothesis**. I will prove both statements.

1. Precise statements

Write

$$\overline{d}(A) = \limsup_{N \to \infty} \frac{|A \cap [1, N]|}{N}.$$

I use your definition with **exactly** $\overline{d}(A) = \delta$, not merely $\overline{d}(A) \geq \delta$.

Theorem: fixed-density largeness implies nilrecurrence

Let $0 < \delta < 1$, and suppose that $D \subseteq \mathbb{N}$ is δ-density large. Then D is a set of multiple pointwise recurrence for all nilsystems.

Explicitly, for every nilsystem (X, T), every $x \in X$, every $\ell \geq 1$, and every $\varepsilon > 0$, there exists $d \in D$ such that

$$\boxed{\max_{1 \leq j \leq \ell} d_X(T^{jd}x, x) < \varepsilon.} \tag{1}$$

The main contrapositive statement is stronger:

High-density obstruction

For every finite-step nil-Bohr $_0$ set $B \subseteq \mathbb{N}$ and every $0 < \delta < 1$, there exist a set $A \subseteq \mathbb{N}$ and an integer K such that

$$\overline{d}(A) = \delta$$

and

$$\boxed{a, a+d, \ldots, a+(K-1)d \in A \quad \Longrightarrow \quad d \in B.} \tag{2}$$

Thus failure of nilrecurrence can be witnessed by a set of upper density **arbitrarily close to one**, not just by a color class of upper density at least $1/2$.

The proof below develops this from the shell construction in your shorter draft, whose stated two-color conclusion is the case in which a progression omits one of two colors. [Pasted text.txt, L59-L74]

2. The key strengthening: progressions that omit a color

Use the draft's special generalized polynomials and their assigned weights:

$$P(n) = a_1 n^{w_1} \left[a_2 n^{w_2} [\cdots [a_\ell n^{w_\ell}] \cdots] \right], \qquad \deg_{\mathrm{wt}} P = w_1 + \cdots + w_\ell,$$

where the brackets denote nearest-integer rounding. [Pasted text.txt, L50-L57]

Here is the stronger lemma.

Lemma: omitted-color obstruction

Fix integers $r \geq 2$, $m \geq 1$, and $d_* \geq 1$. There are constants

$$C = C(r, m, d_*), \qquad k_0 = k_0(r, m, d_*)$$

with the following property.

For every family of special generalized polynomials

$$P_1, \ldots, P_m$$

of assigned weights $1 \leq d_i \leq d_*$, there is a coloring

$$c : \mathbb{N} \longrightarrow \{0, \ldots, r-1\}$$

such that, whenever $k \geq k_0$ and

$$a, a+d, \ldots, a+(k-1)d$$

does not contain all r colors, one has

$$\boxed{\|P_i(d)\| \le Ck^{-d_i} \qquad (1 \le i \le m).} \tag{3}$$

The coloring is fixed independently of k, and the constants are independent of the real coefficients.

For $r = 2$, omitting a color is exactly being monochromatic, so this recovers the shorter draft's theorem. For $r > 2$, the progression may use $r - 1$ different colors and still satisfy the approximation conclusion.

I first explain why this lemma proves your conjecture, and then give the details of its proof.

3. From omitted colors to a high-density test set

Choosing a complement of sufficiently high upper density

Choose r so that

$$1 - \frac{1}{r} \ge \delta.$$

Apply the omitted-color lemma to a prescribed family $P_1, \dots, P_m$. Let

$$C_j = \{n : c(n) = j\}, \qquad A_j = \mathbb{N} \setminus C_j.$$

There is some j such that

$$\overline{d}(A_j) \ge 1 - \frac{1}{r} \ge \delta. \tag{4}$$

Here is the argument, taking care not to assume that the color classes have natural densities. For every N, at least one color class satisfies

$$|C_j \cap [1, N]| \le N/r.$$

Since there are only finitely many colors, the same j satisfies this inequality along an unbounded sequence of N's. Along that sequence,

$$\frac{|A_j \cap [1, N]|}{N} \ge 1 - \frac{1}{r},$$

which proves (4).

Every progression in A_j omits color j, so every sufficiently long such progression satisfies (3).

Thinning to exactly the prescribed upper density

The distinction between $\overline{d}(A) \ge \delta$ and $\overline{d}(A) = \delta$ causes no problem.

Suppose $A = \{a_1 < a_2 < \cdots\}$ has upper density $\alpha \ge \delta > 0$, and put

$$\lambda = \frac{\delta}{\alpha}.$$

Define

$$A' = \{a_j : \lfloor \lambda j \rfloor > \lfloor \lambda (j-1) \rfloor\}.$$

Then, for every N,

$$|A' \cap [1, N]| = \lfloor \lambda \, |A \cap [1, N]| \rfloor.$$

Consequently,

$$\overline{d}(A') = \lambda\alpha = \delta. \tag{5}$$

Since $A' \subseteq A$, all restrictions on progressions in A remain true in A'.

We have therefore obtained the following useful quantitative conclusion:

> For each prescribed family $P_1, \dots, P_m$, there is one set A' of upper density exactly δ such that every k-term progression in A', for every $k \geq k_0$, has difference d satisfying
>
> $$\|P_i(d)\| \leq Ck^{-d_i}.$$

This set is chosen once; it does not change with the desired progression length.

Applying your density hypothesis

Because D is δ-density large, this particular A' contains arbitrarily long progressions with differences in D. Hence, for every sufficiently large k, some $d \in D$ satisfies

$$\|P_i(d)\| \leq Ck^{-d_i} \qquad (1 \leq i \leq m).$$

In particular,

$$\forall \varepsilon > 0 \quad \exists d \in D : \quad \max_i \|P_i(d)\| < \varepsilon. \tag{6}$$

Passing to nilrecurrence

Huang–Shao–Ye's Theorem B and Theorem 4.2.11 identify the family of nil-Bohr $_0$ sets with the family generated by simultaneous neighborhoods of zero for special generalized polynomials. In particular, every finite-step nil-Bohr $_0$ set B contains a set

$$\{n \in \mathbb{N} : \|P_i(n)\| < \varepsilon_i \text{ for all } i\} \tag{7}$$

for a finite family of such polynomials. arXiv

Equation (6) therefore shows that D meets every finite-step nil-Bohr $_0$ set. This is pointwise nilrecurrence.

For multiple pointwise recurrence, apply this to

$$\left(X^\ell,\ T \times T^2 \times \cdots \times T^\ell\right)$$

at $(x, \dots, x)$. This gives (1).

Similarly, choosing K sufficiently large that

$$CK^{-d_i} < \varepsilon_i$$

in the construction of A' proves the high-density obstruction (2).

It remains to justify the omitted-color lemma.

4. Proof of the omitted-color lemma

The necessary change in the shell argument is small, but it is important to check that the rest of the proof uses only **absence of a color**, not monochromaticity.

4.1. The same weighted matrices and shell function

For the prescribed weight patterns, use the product G of weighted upper-unitriangular groups from the draft, with integer lattice Γ. Write w_{ij} for the weight of a matrix entry. These weights are positive and satisfy

$$w_{ik} + w_{kj} = w_{ij}.$$

Let $\mathcal{H}$ be the class of polynomial matrices whose (i, j) entry has degree at most w_{ij}. Define

$$F(A) = \min_{\gamma \in \Gamma} \sum_{i<j} (A\gamma)_{ij}^2, \tag{8}$$

summing over all blocks.

The properties of F established in the shorter draft are exactly what we need: it is continuous, invariant under right multiplication by Γ, and nonconstant on every nonconstant polynomial matrix curve. [Pasted text.txt, L109-L132]

For completeness, the last property follows because, on a compact parameter interval, only finitely many γ's can minimize (8). For each such γ,

$$t \longmapsto \sum_{i<j} (Q(t)\gamma)_{ij}^2$$

is a polynomial. If it were constant, every coordinate of $Q(t)\gamma$, hence $Q(t)$ itself, would be constant. A finite collection of nonconstant polynomials cannot attain one fixed value at every point of an interval.

Fix $\rho = 1/4$, and consider polynomial curves $Q : [0, 1] \to G$ with coordinate degrees at most d_*, coordinate bounds at most 1, and

$$\|Q(1) - Q(0)\|_\infty = \rho.$$

This is a compact family in coefficient space. Thus there is an $\eta > 0$ such that

$$\operatorname{osc}_{[0,1]} F(Q(t)) \geq \eta \tag{9}$$

for every curve in the family. This is the compactness argument already used in the draft. [Pasted text.txt, L134-L145]

4.2. Replace alternating shells by r cyclic shell colors

Choose

$$0 < h < \frac{\eta}{10r}$$

and define

$$\boxed{\kappa_r(A) = \left\lfloor \frac{F(A)}{h} \right\rfloor \pmod r.} \tag{10}$$

The shell bands now cycle through

$$0, 1, \ldots, r-1, 0, 1, \ldots$$

instead of alternating between two colors.

By (9), the range of $F(Q(t))$ is an interval of length at least η. It therefore contains points lying strictly inside shell bands of **every** color, with a uniform positive margin from their boundaries.

Uniform continuity over the compact family supplies T_0 such that, for every $T \geq T_0$,

$$Q(j/T), \qquad 0 \leq j \leq \lfloor T \rfloor, \tag{11}$$

contain all r colors.

More explicitly, choose a midpoint of a shell band of each color lying in the range of $F \circ Q$. Each midpoint is distance $h/2$ from its band's boundary. For a sufficiently fine mesh, some sampled value of $F \circ Q$ lies within $h/4$ of that midpoint, and therefore has the required color. The mesh threshold is uniform because the curve family is compact.

This is the essential strengthening: a normalized polynomial excursion forces every color to appear.

4.3. Choose the same interpolation and rounding parameters

Choose $M > d_* + 1$ sufficiently large that any polynomial matrix Q with

$$|Q_{ij}(n)| \leq \tfrac{1}{2} \qquad (0 \leq n < M)$$

satisfies

$$\|Q(t) - Q(0)\|_\infty < \rho \qquad (0 \leq t \leq T_0). \tag{12}$$

This follows by interpolation at well-spaced integer points: the derivative on the initial interval is $O_{d_*}(1/M)$.

Choose a fine grid in the canonical coordinate box, with cell side length

$$\xi < 2^{-d_*-2}.$$

Let R be the number of cells, choose a finite van der Waerden number $W = W(R, M)$, and put

$$L = d_*! \operatorname{lcm}(1, \ldots, W). \tag{13}$$

These parameters depend on the weight patterns and r, but not on the polynomial coefficients or progression length.

The polynomial rounding lemma in the draft is unchanged: if the first M canonical representatives of $A(t) \in \mathcal{H}$ lie in one grid cell, then

$$A(t)C(t) = Q(t), \tag{14}$$

where $C, Q \in \mathcal{H}$, $C(n) \in \Gamma$ for all integers n, and $Q(n)$ is the canonical representative at the initial samples. Moreover,

$$w_{ij}! \, [t^{w_{ij}}] C_{ij}(t) \in \mathbb{Z}. \tag{15}$$

None of these assertions involves the shell colors. [Pasted text.txt, L190-L218]

The mechanism is worth recalling. The reduced coordinates in one grid cell have sufficiently small $(w_{ij}+1)$-st differences. Those differences are integers, so they vanish. The reduced coordinates therefore interpolate to ordinary polynomials, and their lattice corrections are integer-valued polynomials.

4.4. An orbit omitting one color cannot escape

Suppose

$$p(0), \ldots, p(k-1), \qquad p \in \mathcal{H},$$

omit some color e under κ_r.

Apply the auxiliary grid coloring to the first W samples. It gives an M-term subprogression with parameters $b, q \le W$, whose canonical representatives lie in one cell. Polynomial rounding gives

$$p(b+qt)C(t) = Q(t). \tag{16}$$

Set

$$J = \left\lfloor \frac{k-1-b}{q} \right\rfloor.$$

Since $C(j) \in \Gamma$,

$$\kappa_r(Q(j)) = \kappa_r(p(b+qj)), \qquad 0 \le j \le J.$$

Thus all these samples of Q also omit color e.

We claim

$$\|Q(t) - Q(0)\|_\infty < \rho \qquad (0 \le t \le J). \tag{17}$$

Otherwise, let $T \le J$ be the first time the displacement equals ρ. By (12), $T > T_0$. Because $Q(0)$ is a canonical representative, the rescaled curve

$$u \longmapsto Q(Tu), \qquad 0 \le u \le 1,$$

has coordinates bounded by $3/4$, and its endpoint displacement is ρ. It belongs to the compact family used above.

Therefore (11) says that

$$Q(j), \qquad 0 \le j \le \lfloor T \rfloor,$$

contain **all** r colors. This contradicts their omission of e, proving (17).

This is precisely the first-exit argument from the shorter draft, with “the samples contain both colors” replaced by “the samples contain all r colors.” No other use of monochromaticity is needed. [Pasted text.txt, L240-L252]

4.5. Extract the weighted leading coefficients

By (17), all coordinates of Q are bounded by $3/4$ on $[0, J]$. Polynomial interpolation gives

$$\bigl|[t^{w_{ij}}]Q_{ij}(t)\bigr| \le C_{d_*} J^{-w_{ij}}. \tag{18}$$

Let

$$\tau(A)_{ij} = [t^{w_{ij}}]A_{ij}(t), \qquad (\delta_u B)_{ij} = u^{w_{ij}} B_{ij}.$$

The additive weights give

$$\tau(AB) = \tau(A)\tau(B), \qquad \tau(A(b+qt)) = \delta_q \tau(A).$$

Applying these identities to (16), and then dilating by L/q, gives

$$\delta_L \tau(p)\, \gamma = E, \tag{19}$$

where

$$\gamma = \delta_{L/q}\tau(C) \in \Gamma.$$

The integrality follows from (15) and the fact that L/q is a multiple of d_*!.

Moreover,

$$|E_{ij}| \le C_{d_*} \left(\frac{L}{qJ}\right)^{w_{ij}}.$$

Since

$$qJ \ge k-1-b-q \ge k-1-2W,$$

we obtain, for sufficiently large k,

$$|E_{ij}| \le Ck^{-w_{ij}}. \tag{20}$$

These are the same leading-coefficient and denominator-clearing calculations as in the original proof; their hypothesis can now be weakened from monochromaticity to omission of one color. [Pasted text.txt, L253-L284]

4.6. Recover the special generalized polynomials

For

$$P(t) = a_1 t^{w_1}[a_2 t^{w_2}[\cdots[a_\ell t^{w_\ell}]\cdots]],$$

use the band matrix $g(t)$ with first superdiagonal

$$-a_1 t^{w_1}, \dots, -a_{\ell-1} t^{w_{\ell-1}}, +a_\ell t^{w_\ell},$$

and all higher entries zero. For the family $P_1, \dots, P_m$, take the product of these matrices.

Define the integer coloring

$$c(n) = \kappa_r(g(n/L)). \tag{21}$$

If a progression $a, a+d, \dots, a+(k-1)d$ omits a color, apply (19) to

$$p(t) = g((a+dt)/L).$$

Weighted homogeneity gives

$$\delta_L \tau(p) = g(d).$$

Hence

$$g(d)\gamma = E, \qquad |E_{ij}| \le Ck^{-w_{ij}}. \tag{22}$$

For large enough k, all these errors are less than $1/4$. The last-column equations then recover the nested nearest integers, starting from the bottom row. In particular, the top-row equation becomes

$$P_i(d) + z_i = E_{\mathrm{top},i}, \qquad z_i \in \mathbb{Z},$$

where the relevant top coordinate has weight d_i. Thus

$$\|P_i(d)\| \le Ck^{-d_i}.$$

This is the draft's nested-rounding argument without any change. [Pasted text.txt, L325-L349]

Finally, for fixed m, d_*, there are only finitely many possible tuples of positive-integer weight patterns. Maximizing the constants over those patterns makes them depend only on r, m, d_*.

This proves the omitted-color lemma, and therefore your conjecture. □

5. The quantitative version for your paper

Define D to be ** (δ, k)-density large** if every $A \subseteq \mathbb{N}$ with $\overline{d}(A) = \delta$ contains a k-term progression with difference in D.

The argument proves:

> **Quantitative fixed-density approximation.**
> For every $0 < \delta < 1$ and every $m, d_* \geq 1$, there are constants C, k_0 such that, whenever $k \geq k_0$, D is (δ, k)-density large, and $P_1, \ldots, P_m$ are special generalized polynomials of weights $d_i \leq d_*$, there exists $d \in D$ satisfying
>
> $$\boxed{\|P_i(d)\| \leq Ck^{-d_i} \qquad (1 \leq i \leq m).}$$

> The constants depend only on δ, m, d_*, not on the coefficients.

As in your existing theorem, this controls the approximation error but gives no bound on the size of d.

There are two useful qualifications.

First, for $0 < \delta \leq 1/2$, your qualitative conjecture already follows from the two-color theorem: every binary coloring has a color class of upper density at least $1/2$, which can be thinned to upper density δ. **The genuinely additional range is $1/2 < \delta < 1$.** The omitted-color argument handles precisely that difficulty.

Second, the endpoint $\delta = 1$ must be excluded. The set $D = \{1\}$ is 1-density large: a set of upper density one contains arbitrarily long intervals, since absence of a k-term interval would force upper density at most $1 - 1/k$. But $\{1\}$ is certainly not recurrent for an irrational circle rotation.

So the natural upgrade to the paper is an omitted-color version of the shell theorem. It contains the original two-color theorem and yields your fixed-density nilrecurrence conjecture, with the same weight-dependent quantitative exponents.

---

24. User

2026-09-15 10:49:42 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T02:49:42.778000Z

Please create a complete extra detailed standalone tex file with the statement and proof of this result.

---

25. Assistant

2026-09-15 10:50:00 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T02:50:00.109844Z

I'll prepare a self-contained standalone LaTeX proof, detailing the omitted-color argument, parameter choices, density reduction, nil-Bohr input, and the resulting fixed-density nilrecurrence theorem.

26. Assistant

2026-09-15 10:53:03 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T02:53:03.249751Z

I've checked the nil-Bohr characterization; I'll also isolate the all-colors strengthening and carefully address exact upper density and rounding-boundary issues.

27. Assistant

2026-09-15 11:02:06 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T03:02:06.126085Z

The full draft is now written, including coefficient extraction, nested rounding, and denominator clearance; I'm doing a final check of notation, cross-references, and layout.

28. Assistant

2026-09-15 11:14:34 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T03:14:34.115301Z

Download the standalone LaTeX file

Read the compiled PDF — 22 pages

The document proves your conjecture with **natural upper density exactly** δ and includes the stronger quantitative statement: for every prescribed family of special generalized polynomials $P_1, \dots, P_m$, there is one set $A \subseteq \mathbb{N}$ with $\overline{d}(A) = \delta$ such that

$$\{a, a+d, \dots, a+(k-1)d\} \subseteq A \quad \Longrightarrow \quad \|P_i(d)\|_{\mathbb{R}/\mathbb{Z}} \leq Ck^{-d_i},$$

simultaneously for every sufficiently large k. The constants are independent of the real coefficients.

I have written out the **entire omitted-color argument**, expanding the weighted-matrix, polynomial-rounding, and shell construction from the shorter draft rather than invoking its two-color conclusion as a black box. This includes worked coefficient-extraction examples, the sampling argument at color boundaries, the first-exit argument, exact factorial-denominator clearance, and the nested-rounding calculation. [Pasted text.txt, L351-L355]

The two results quoted without proof are finite van der Waerden and the Huang–Shao–Ye characterization of nil-Bohr neighborhoods. The latter is stated explicitly, with the precise references to Theorem B and Theorem 4.2.11. arXiv

The document also proves that the resulting nilrecurrence times are **unbounded**, gives the high-density obstruction for every finite-step nil-Bohr neighborhood, and explains why the endpoint $\delta = 1$ fails. The source includes its bibliography, requires no auxiliary files, and compiles with pdfLaTeX without warnings.

29. User

2026-09-15 11:17:12 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T03:17:12.296000Z

can you now show that the Alweiss example is not delta-density large for any delta in (0,1)?

30. Assistant

2026-09-15 11:17:28 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T03:17:28.956613Z

I'll test whether the existing Alweiss obstruction can exclude long allowed-difference progressions inside witness sets of density arbitrarily close to one, without modifying his construction.

31. Assistant

2026-09-15 11:20:31 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T03:20:31.996847Z

The existing two-color obstruction already handles $0 < \delta \leq \frac{1}{2}$, but higher densities require a genuinely new argument; nilrecurrence itself and simply adding stripe colors do not suffice.

32. Assistant

2026-09-15 11:29:36 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T03:29:36.819605Z

The cosine obstruction does extend further, producing witnesses through density $14/15$. I'm checking whether the same fixed construction can be amplified to every density below one.

33. Assistant

2026-09-15 11:40:53 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T03:40:53.823076Z

I have a proof for every $0 < \delta \leq 24/25$, with the original Alweiss example unchanged, but I have not established the remaining range $24/25 < \delta < 1$. The argument gives a considerably stronger obstruction than the earlier two-coloring: one can construct a **25-coloring such that every 200-term progression with an allowed difference contains all 25 colors**.

There is also a concrete obstruction to simply extending the earlier scalar-observable argument to arbitrarily many colors. I will explain that after the proof, because it is important not to mistake this partial result for the full assertion.

Throughout, I mean the original small-step construction with the displayed parameters

$$a = \delta_1 = \frac{1}{100}, \qquad b = \delta_2 = 10^{-10},$$

not a construction in which these parameters are chosen anew depending on the desired density. These are the parameters explicitly offered in Alweiss's paper. arXiv

1. The stronger statement that can be proved

Let $S \subseteq \mathbb{N}$ be that original Alweiss set.

Theorem. There is a coloring

$$c : \mathbb{N} \longrightarrow \{0, 1, \ldots, 24\}$$

such that every progression

$$n, n+d, \dots, n+199d, \qquad d \in S,$$

contains all 25 colors.

Consequently, there is a set $B \subseteq \mathbb{N}$ with

$$\overline{d}(B) \geq \frac{24}{25}$$

containing **no 200-term arithmetic progression with difference in** S. In particular,

$$\boxed{S \text{ is not } \delta\text{-density large for every } 0 < \delta \leq \frac{24}{25}.}$$

The change from our previous proof is to replace the cosine observable by a smoothed version of the triangular function $t \mapsto \|t\|$. This improves the ratio between the oscillation forced by the distinguished coordinate and the maximum one-step displacement.

The defining conditions we use

As in the earlier note, write the construction using an absolutely summable sequence (α_j). For each $d \in S$, there is a distinguished index $i = i(d)$, and we can choose real representatives σ_j of $d\alpha_j \bmod 1$, such that

$$\sigma_i = -a + \eta, \qquad |\eta| \leq 2b, \tag{1}$$

and

$$|\sigma_j| \leq 2b \quad (j \neq i), \qquad \sum_{j \neq i} |\sigma_j| \leq a + 2b. \tag{2}$$

These are precisely the relaxed inequalities used in the preceding two-color proof; Alweiss's original defining conditions imply them. [alweiss_original_set_not_two_large.tex, L13-L25] [alweiss_original_set_not_two_large.tex, L52-L56]

Two immediate consequences are

$$\sum_{j} |\sigma_j| \leq 2a + 4b = 0.0200000004 \tag{3}$$

and

$$\sum_{j \neq i} \sigma_j^2 \leq 2b \sum_{j \neq i} |\sigma_j| \leq 2b(a + 2b). \tag{4}$$

The first controls consecutive values of our observable. The second controls the combined curvature contribution from every coordinate other than the distinguished one.

2. A smoothed triangular observable

Put

$$\rho(t) = \|t\|_{\mathbb{R}/\mathbb{Z}}, \qquad \varepsilon = 10^{-4},$$

and define the periodic function

$$\boxed{\psi(t) = \frac{1}{2\varepsilon}\int_{-\varepsilon}^{\varepsilon} \rho(t-u)\,du - \frac{\varepsilon}{2}.} \tag{5}$$

The subtraction makes $\psi(0) = 0$. We need three elementary properties.

First, ψ is 1-Lipschitz, because ρ is 1-Lipschitz and averaging preserves that bound. In particular, periodicity and $\psi(0) = 0$ give

$$|\psi(t)| \le \|t\|. \tag{6}$$

Second, the triangle inequality gives

$$\rho(t) - |u| \le \rho(t-u) \le \rho(t) + |u|.$$

Averaging, and noting that the average of $|u|$ on $[-\varepsilon, \varepsilon]$ is $\varepsilon/2$, yields

$$\boxed{\rho(t) - \varepsilon \le \psi(t) \le \rho(t).} \tag{7}$$

Third, ψ is continuously differentiable, with

$$\psi'(t) = \frac{\rho(t+\varepsilon) - \rho(t-\varepsilon)}{2\varepsilon}.$$

Since ρ is 1-Lipschitz,

$$|\psi'(t) - \psi'(u)| \le \frac{|t-u|}{\varepsilon}.$$

Consequently,

$$\boxed{|\psi(t) - 2\psi(t+v) + \psi(t+2v)| \le \frac{v^2}{\varepsilon} \qquad (t, v \in \mathbb{R}).} \tag{8}$$

For example, this follows by expressing the second difference as an integral of a difference of first derivatives.

Now define

$$\boxed{H(n) = \sum_{j=1}^{\infty} \psi(n\alpha_j).} \tag{9}$$

This series converges absolutely: by (6),

$$|\psi(n\alpha_j)| \le \|n\alpha_j\| \le n|\alpha_j|,$$

and $\sum_j |\alpha_j| < \infty$.

Fix an allowed progression and write

$$y_t = H(n + td), \qquad 0 \le t \le 199.$$

We will prove that consecutive y_t's move by less than 0.0204, while the whole sequence has oscillation greater than 0.48985.

Small consecutive increments

Periodicity and the Lipschitz bound give

$$\begin{aligned}|y_{t+1}-y_t| &\le \sum_j |\psi((n+td)\alpha_j+\sigma_j)-\psi((n+td)\alpha_j)| \\ &\le \sum_j |\sigma_j| \\ &\le 0.0200000004.\end{aligned}$$

Thus, with

$$h=\frac{51}{2500}=0.0204,$$

we have

$$\boxed{|y_{t+1}-y_t|<h.} \tag{10}$$

3. A uniformly large excursion within 200 steps

The distinguished coordinate has step very close to $-1/100$. During the first 100 steps, it therefore comes close to the bottom of the triangular function. Fifty steps later, it comes close to the top; another fifty steps later, it comes close to the bottom again.

The other coordinates cannot cancel this excursion, because their combined second differences are small by (4) and (8).

Here are the details.

Let

$$\theta = n\alpha_i \bmod 1.$$

The points

$$\theta-\frac{r}{100}, \qquad 0\le r\le 99,$$

form a translate of the $1/100$-grid on the circle. Therefore some $r\in\{0,\dots,99\}$ satisfies

$$\left\|\theta-\frac{r}{100}\right\| \le \frac{1}{200}.$$

Using $\sigma_i=-1/100+\eta$, choose a representative u of $\theta+r\sigma_i$ with

$$|u|\le \frac{1}{200}+99|\eta| \le 0.005+198b. \tag{11}$$

Set $q=50$. Since

$$q\sigma_i=-\frac{1}{2}+50\eta, \qquad 2q\sigma_i=-1+100\eta,$$

the distinguished coordinate's contribution to

$$2y_{r+50}-y_r-y_{r+100} \tag{12}$$

is

$$2\psi\left(u-\frac{1}{2}+50\eta\right)-\psi(u)-\psi(u+100\eta).$$

By the Lipschitz property of ρ and (7),

$$\psi\left(u-\frac{1}{2}+50\eta\right)\ge\frac{1}{2}-|u|-\varepsilon-100b.$$

Also,

$$\psi(u)\le|u|,\qquad \psi(u+100\eta)\le|u|+200b.$$

Hence the distinguished contribution is at least

$$\begin{aligned}1-4|u|-2\varepsilon-400b&\ge1-4(0.005+198b)-2\varepsilon-400b\\&=0.9798-1192b.\end{aligned}\tag{13}$$

For the remaining coordinates, (8) and (4) give the uniform bound

$$\begin{aligned}&\left|\sum_{j\neq i}\left(2\psi((n+(r+50)d)\alpha_j)-\psi((n+rd)\alpha_j)-\psi((n+(r+100)d)\alpha_j)\right)\right|\\&\le\frac{50^2}{\varepsilon}\sum_{j\neq i}\sigma_j^2\\&\le\frac{2500}{10^{-4}}\,2b(a+2b)\\&<0.000051.\end{aligned}\tag{14}$$

Combining (13) and (14), with $b=10^{-10}$, gives

$$\boxed{2y_{r+50}-y_r-y_{r+100}>0.9797.}\tag{15}$$

All three indices belong to $\{0,\dots,199\}$. If

$$R=\max_{0\le t\le199}y_t-\min_{0\le t\le199}y_t,$$

then the expression in (15) is at most $2R$. Therefore

$$\boxed{R>0.48985>0.4896=24h.}\tag{16}$$

This estimate is independent of the starting point n, the difference $d\in S$, and the distinguished index $i(d)$.

4. Every allowed progression contains all 25 colors

Define

$$\boxed{c(n)=\left\lfloor\frac{H(n)}{h}\right\rfloor\pmod{25},\qquad h=\frac{51}{2500}.}\tag{17}$$

For the progression under consideration, put

$$v_t=\left\lfloor\frac{y_t}{h}\right\rfloor.$$

By (10),

$$|v_{t+1}-v_t|\le1.\tag{18}$$

By (16),

$$\max_t v_t-\min_t v_t\ge24.\tag{19}$$

An integer-valued sequence satisfying (18) must attain every integer between its minimum and maximum: it cannot skip an integer when moving from one extreme to the other. Thus (19) forces the v_t's to contain at least 25 consecutive integers.

Reducing these integers modulo 25 proves that the progression contains all 25 colors.

Passing to upper density exactly δ

For each color e, let

$$B_e = \{n \in \mathbb{N} : c(n) \neq e\}.$$

No B_e contains a 200-term progression with difference in S.

For every N, at least one color occurs at most $N/25$ times in $[1, N]$. The same color e has this property along an unbounded sequence of N's, so

$$\overline{d}(B_e) \geq \frac{24}{25}.$$

Finally, fix any $0 < \delta \leq 24/25$. Write

$$B_e = \{b_1 < b_2 < \cdots\}, \qquad \alpha = \overline{d}(B_e), \qquad \lambda = \frac{\delta}{\alpha}.$$

Retain b_j exactly when

$$\lfloor \lambda j \rfloor > \lfloor \lambda (j-1) \rfloor,$$

and call the resulting subset A. Then

$$|A \cap [1, N]| = \lfloor \lambda |B_e \cap [1, N]| \rfloor,$$

so

$$\overline{d}(A) = \lambda \alpha = \delta.$$

Since $A \subseteq B_e$, it contains no allowed 200-term progression. This proves the theorem. □

5. Why this does not yet extend to every $\delta < 1$

The omitted-color argument in the nilrecurrence proof works for an arbitrary number of colors. For the Alweiss obstruction above, however, **one cannot simply replace 25 by an arbitrary** r: the available oscillation is bounded, while the stripe width must remain larger than the one-step displacement.

There is an even more concrete limitation on this approach.

Consider the group

$$E = \ell^1(\mathbb{N}; \mathbb{R}/\mathbb{Z})$$

used in Alweiss's construction, and put

$$M = \frac{a}{b} = 10^8.$$

The finite-support direction

$$s = (-a, \underbrace{b, \ldots, b}_{M \text{ coordinates}}, 0, \ldots) \tag{20}$$

satisfies the original group-level defining conditions: its distinguished coordinate is exactly $-a$, every other nonzero coordinate is b, and their total norm is exactly a. arXiv

Choose the starting phases in those M tail coordinates to be

$$0, \frac{1}{M}, \frac{2}{M}, \dots, \frac{M-1}{M}.$$

After 100 steps, the distinguished coordinate returns to itself, because

$$100a = 1.$$

Each tail coordinate advances by

$$100b = \frac{1}{M},$$

which simply permutes the collection of tail phases.

Consequently, for **every** periodic function φ with $\varphi(0) = 0$, the symmetric additive observable

$$J_\varphi(x) = \sum_j \varphi(x_j)$$

satisfies, along this orbit,

$$\boxed{J_\varphi(x + (t+100)s) = J_\varphi(x + ts) \qquad (t \in \mathbb{Z}).} \tag{21}$$

This includes both our original cosine observable and the smoothed triangular observable above. Therefore a coloring depending only on such an observable sees at most 100 colors on this orbit, regardless of the progression length. It cannot yield a uniform "every allowed progression visits all r colors" statement for arbitrary r.

This is an obstruction to that proof method, not a counterexample to your proposed conclusion. The finite-support direction and starting point in (20)–(21) need not occur exactly on the particular irrational integer orbit defining S. A proof using additional features of that integer orbit, or a substantially different observable, could still succeed.

Also, the theorem we just proved about fixed-density largeness implying nilrecurrence does not resolve this: Alweiss's example already meets every finite-step nil-Bohr neighborhood. arXiv

Thus the conclusion established here, for the **same fixed original example**, is

$$\boxed{\forall\, 0 < \delta \le \frac{24}{25}, \quad \exists A \subseteq \mathbb{N} :\ \overline{d}(A) = \delta \text{ and } A \text{ has no 200-term AP with difference in } S.}$$

The range $24/25 < \delta < 1$ remains unproved in this argument; neither the earlier two-color obstruction nor the new nilrecurrence theorem supplies the missing amplification.

---

34. User

2026-09-15 12:06:59 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T04:06:59.822000Z

This is great progress! It would be amazing if it turned out that the Alweiss example is not delta-density large for some delta in (0,1). Please keep investigating this.

35. Assistant

2026-09-15 12:07:16 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T04:07:16.640421Z

The previous argument already establishes failure for some δ, including $1/2$. I'll verify it against Alweiss's definitions and investigate whether the obstruction extends to densities nearer 1.

36. Assistant

2026-09-15 12:10:19 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T04:10:19.593458Z

Alweiss's defining inequalities do confirm the coordinate bounds behind the $24/25$ argument. I'm now testing whether its fixed rational scales create a genuine barrier to further amplification.

37. Assistant

2026-09-15 12:19:22 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T04:19:22.584888Z

I found a modest improvement: shifted stripes raise the cutoff to about 0.960767, with natural-density witnesses. The additive-observable stripe method itself appears capped precisely at 0.99.

38. Assistant

2026-09-15 12:27:52 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T04:27:52.794899Z

Failure for at least one positive density is already established—in fact, it follows directly from Alweiss's original finite-color obstruction. The stronger question we have been investigating is whether the **same fixed example fails δ-density largeness for every $\delta < 1$.** Alweiss constructs a finite coloring with no three-term progression having an allowed difference; one of its color classes has positive upper density and therefore already witnesses failure at some positive density. arXiv

I have obtained two further results. There is a **small improvement beyond** $24/25$, and the witnessing sets can be chosen to have **uniform density exactly** δ, rather than merely upper density δ. I also found a sharper limitation of the current observable: above density $49/50$, a broad class of sets defined using that observable necessarily contain arbitrarily long allowed progressions.

I have not proved failure for every $\delta < 1$. Here are the new results and their proofs.

1. A stronger density obstruction for the unchanged example

Let S be Alweiss's original example with

$$a = \delta_1 = \frac{1}{100}, \qquad b = \delta_2 = 10^{-10}.$$

These are the parameter values displayed in his construction. arXiv

Define

$$\boxed{\delta_* = \frac{489899}{509900} = 0.960774661698\ldots.}$$

Theorem

For every

$$0 < \delta \le \delta_*,$$

there exists $A \subseteq \mathbb{N}$ such that

$$\boxed{\lim_{N\to\infty} \sup_{M\ge 0} \left| \frac{|A \cap \{M+1, \ldots, M+N\}|}{N} - \delta \right| = 0,} \tag{1}$$

and A contains **no 200-term arithmetic progression with common difference in** S.

In particular, S is not δ-density large for any $0 < \delta \le \delta_*$.

The numerical gain over $24/25 = 0.96$ is modest. The stronger density conclusion is more substantial: every sufficiently long interval has approximately the prescribed proportion of points of A, uniformly in the position of the interval. Thus the witnesses do not depend on favorable subsequences of initial intervals.

The proof has two independent ingredients: a deterministic obstruction along allowed progressions, and an equidistribution statement for the infinite sum defining the obstruction.

2. The deterministic obstruction, with slightly better parameters

Write

$$\rho(t) = \|t\|_{\mathbb{R}/\mathbb{Z}}, \qquad \varepsilon = \frac{1}{20000},$$

and define

$$\psi(t) = \frac{1}{2\varepsilon} \int_{-\varepsilon}^{\varepsilon} \rho(t-u)\, du - \frac{\varepsilon}{2}. \tag{2}$$

This is the same type of smoothed triangular function as before, with a slightly different smoothing scale. It satisfies

$$\psi(0) = 0, \qquad \mathrm{Lip}(\psi) \le 1, \qquad \rho(t) - \varepsilon \le \psi(t) \le \rho(t). \tag{3}$$

Its derivative is $1/\varepsilon$-Lipschitz, so

$$|\psi(x) - 2\psi(x+v) + \psi(x+2v)| \le \frac{v^2}{\varepsilon}. \tag{4}$$

Let (α_j) be the sequence in the original construction, and put

$$H(n) = \sum_{j=1}^{\infty} \psi(n\alpha_j). \tag{5}$$

The series converges absolutely because

$$|\psi(n\alpha_j)| \le |n|\,|\alpha_j|, \qquad \sum_j |\alpha_j| < \infty.$$

What an allowed difference does

For $d \in S$, Alweiss's definition supplies a distinguished index i and representatives σ_j of $d\alpha_j \bmod 1$ such that

$$\sigma_i = -a + \eta, \qquad |\eta| \le 2b,$$

and

$$|\sigma_j| \le 2b \quad (j \ne i), \qquad \sum_{j \ne i} |\sigma_j| \le a + 2b. \tag{6}$$

These are consequences of the original defining inequalities, not additional restrictions on S. arXiv

Consequently,

$$\sum_{j} |\sigma_j| \le 2a + 4b \tag{7}$$

and

$$\sum_{j \ne i} \sigma_j^2 \le 2b(a + 2b). \tag{8}$$

Fix a progression and write

$$y_t = H(n + td), \qquad 0 \le t \le 199.$$

The Lipschitz bound immediately gives

$$\boxed{|y_{t+1} - y_t| \le 2a + 4b = 0.0200000004.} \tag{9}$$

We also have the following lower bound on its oscillation:

$$\boxed{\max_{0 \le t \le 199} y_t - \min_{0 \le t \le 199} y_t \ge 0.489899940399.} \tag{10}$$

Here is the calculation, so that the small numerical improvement can be checked.

The 100 points

$$n\alpha_i - ra, \qquad 0 \le r \le 99,$$

form a translated $1/100$-grid on the circle. Choose r for which one of these points is within $a/2$ of zero. Taking the small representative u of $n\alpha_i + r\sigma_i$, we obtain

$$|u| \le \frac{a}{2} + 198b. \tag{11}$$

Set $q = 50$. The distinguished coordinate's contribution to

$$2y_{r+q} - y_r - y_{r+2q}$$

is

$$2\psi\left(u - \frac{1}{2} + 50\eta\right) - \psi(u) - \psi(u + 100\eta).$$

Using (3), this is at least

$$1 - 4|u| - 2\varepsilon - 400b. \tag{12}$$

By (4) and (8), the absolute value of the combined contribution of all other coordinates is at most

$$\frac{q^2}{\varepsilon}\sum_{j\neq i}\sigma_j^2 \le \frac{2q^2b(a+2b)}{\varepsilon}. \tag{13}$$

All three indices $r, r+50, r+100$ lie between 0 and 199. Since the expression $2y_{r+50} - y_r - y_{r+100}$ is at most twice the full oscillation, (11)–(13) imply

$$\begin{aligned}\operatorname{osc}_{0\le t\le 199} y_t &\ge \frac{49}{100} - \varepsilon - 596b - \frac{2500b(a+2b)}{\varepsilon}\\ &= 0.489899940399,\end{aligned}$$

which proves (10).

Use intervals instead of an integral number of colors

The previous argument divided the real line into equal bands and colored them cyclically. We can do slightly better by allowing the retained and deleted intervals to have different lengths.

Set

$$h = \frac{20001}{10^6} = 0.020001, \qquad w = \frac{489899}{10^6} = 0.489899,$$

and

$$p = h + w = \frac{5099}{10000} = 0.5099.$$

Define

$$\boxed{A_* = \{n \in \mathbb{N} : H(n) \bmod p \in [h, p)\}.} \tag{14}$$

Thus we delete, periodically, intervals of length h, and retain the intervening intervals of length w.

Suppose an allowed 200-term progression were entirely contained in A_*. Every y_t would then belong to

$$\bigcup_{m\in\mathbb{Z}} [mp + h, (m+1)p).$$

Consecutive y_t's cannot move between different components of this union: doing so requires a jump of at least h, whereas (9) gives a strictly smaller jump. Therefore all y_t's must belong to one retained interval of length w.

But (10) says their oscillation is strictly greater than w. This is a contradiction.

We have proved, independently of any density calculation, that

$$\boxed{A_* \text{ contains no allowed 200-term progression.}} \tag{15}$$

The remaining issue is to calculate the density of this particular set—not merely find a favorable translate of the bands.

3. An infinite-sum equidistribution lemma

The useful additional fact is that

$$H(n) \bmod p$$

is **well-distributed**, for every $p > 0$.

Alweiss's construction uses a summable sequence whose positive integer orbit is dense in

$$E = \ell^1(\mathbb{N}; \mathbb{R}/\mathbb{Z}).$$

In particular, every finite coordinate projection is dense in its torus, giving the rational independence needed below. This is the dense-orbit property stated in Section 2.3 of his paper. arXiv

The following lemma isolates the general mechanism.

Lemma

Suppose that

$$\sum_{j=1}^{\infty} |\alpha_j| < \infty$$

and that

$$1, \alpha_1, \dots, \alpha_m$$

are linearly independent over $\mathbb{Q}$ for every m.

Let $\psi : \mathbb{R}/\mathbb{Z} \to \mathbb{R}$ be nonconstant and Lipschitz, with $\psi(0) = 0$, and define

$$H(n) = \sum_{j=1}^{\infty} \psi(n\alpha_j).$$

Then, for every $p > 0$, $H(n) \bmod p$ is well-distributed. Explicitly, for every interval $I \subseteq \mathbb{R}/p\mathbb{Z}$,

$$\lim_{N\to\infty} \sup_{M\geq 0} \left| \frac{1}{N} \#\{M < n \leq M + N : H(n) \bmod p \in I\} - \frac{|I|}{p} \right| = 0. \tag{16}$$

The proof uses uniform correlation limits followed by van der Corput.

Step 1: the correlations have uniform limits

Fix $\lambda \neq 0$, and put

$$u(n) = e^{i\lambda H(n)}.$$

For a fixed integer h, the tail of the difference $H(n+h) - H(n)$ is uniformly small in n:

$$\left| \sum_{j>m} \big(\psi((n+h)\alpha_j) - \psi(n\alpha_j)\big) \right| \leq \mathrm{Lip}(\psi)\, |h| \sum_{j>m} |\alpha_j|. \tag{17}$$

This is the crucial point. The tail of $H(n)$ itself need not be uniformly small in n, but the tail of each **fixed difference** is.

Define

$$K_\lambda(v) = \int_0^1 e^{i\lambda(\psi(x+v)-\psi(x))}\, dx.$$

Finite-dimensional torus equidistribution gives, uniformly in M,

$$\frac{1}{N}\sum_{n=M+1}^{M+N}\exp\left(i\lambda\sum_{j=1}^{m}\bigl(\psi((n+h)\alpha_j)-\psi(n\alpha_j)\bigr)\right) \longrightarrow \prod_{j=1}^{m}K_\lambda(h\alpha_j). \tag{18}$$

Uniformity in the starting point follows directly by approximating the continuous torus function by trigonometric polynomials: every nonconstant Fourier term has a geometric-sum bound independent of M.

Using (17) and then letting $m\to\infty$, we obtain

$$\lim_{N\to\infty}\sup_{M\ge 0}\left|\frac{1}{N}\sum_{n=M+1}^{M+N}u(n+h)\overline{u(n)}-C_\lambda(h)\right|=0, \tag{19}$$

where

$$C_\lambda(h)=\prod_{j=1}^{\infty}K_\lambda(h\alpha_j).$$

Step 2: the absolute correlations have Cesàro mean zero

We have $|K_\lambda(v)|\le 1$. Moreover,

$$q_\lambda:=\int_0^1|K_\lambda(v)|\,dv<1. \tag{20}$$

To justify the strict inequality, let $f(x)=e^{i\lambda\psi(x)}$. If $|K_\lambda(v)|=1$ for every v, equality in the triangle inequality would force every translate of f to be a scalar multiple of f. Thus f, up to a constant factor, would be a character of the circle.

But f has a continuous periodic real phase $\lambda\psi$, so such a character must be constant. That would force ψ to be constant, a contradiction.

For every fixed m,

$$|C_\lambda(h)|\le\prod_{j=1}^{m}|K_\lambda(h\alpha_j)|.$$

A second application of finite-dimensional torus equidistribution therefore yields

$$\limsup_{H\to\infty}\frac{1}{H}\sum_{h=1}^{H}|C_\lambda(h)|\le q_\lambda^m.$$

Letting $m\to\infty$, we get

$$\boxed{\lim_{H\to\infty}\frac{1}{H}\sum_{h=1}^{H}|C_\lambda(h)|=0.} \tag{21}$$

Step 3: apply van der Corput uniformly in the interval

The van der Corput inequality, together with the uniform limits (19), gives

$$\limsup_{N\to\infty}\sup_{M\ge 0}\left|\frac{1}{N}\sum_{n=M+1}^{M+N}u(n)\right|^2\le\frac{1}{H}+\frac{2}{H}\sum_{h=1}^{H-1}|C_\lambda(h)|.$$

Letting $H \to \infty$ and using (21), we obtain

$$\lim_{N\to\infty} \sup_{M\ge 0} \left| \frac{1}{N} \sum_{n=M+1}^{M+N} e^{i\lambda H(n)} \right| = 0. \tag{22}$$

Take $\lambda = 2\pi k/p$, for every nonzero integer k. The uniform Weyl criterion proves (16). □

This lemma also verifies that the **25 color classes from the previous argument each have uniform density** $1/25$. The earlier pigeonhole argument was sufficient, but it was not the whole story.

4. Finish the density calculation and obtain every smaller density

Applying the lemma to (14),

$$d_{\text{uniform}}(A_*) = \frac{p-h}{p} = \frac{w}{p} = \boxed{\frac{489899}{509900}}. \tag{23}$$

Together with (15), this proves the theorem at $\delta = \delta_*$.

For an arbitrary $0 < \delta \le \delta_*$, enumerate

$$A_* = \{a_1 < a_2 < \cdots\}, \qquad \theta = \frac{\delta}{\delta_*},$$

and retain a_j exactly when

$$\lfloor \theta j \rfloor > \lfloor \theta (j-1) \rfloor.$$

Call the resulting subset A.

Writing

$$B(N) = |A_* \cap [1, N]|,$$

we have

$$|A \cap [1, N]| = \lfloor \theta B(N) \rfloor.$$

Consequently, for every M, N,

$$\begin{aligned} |A \cap \{M+1, \dots, M+N\}| &= \lfloor \theta B(M+N) \rfloor - \lfloor \theta B(M) \rfloor \\ &= \theta\, |A_* \cap \{M+1, \dots, M+N\}| + O(1), \end{aligned}$$

with an absolute error bound independent of M, N.

Thus A has uniform density $\theta\delta_* = \delta$. Since $A \subseteq A_*$, it still contains no allowed 200-term progression.

This proves the theorem completely.

5. A sharper obstruction to pushing this particular observable to density one

The investigation also gives a useful negative result about the method. It is stronger than the observation that a certain group-level orbit is periodic.

Proposition: a limitation of value-defined test sets

Keep the same function H. Let $p > 0$, and let

$$J \subseteq \mathbb{R}/p\mathbb{Z}$$

be a finite union of open intervals with

$$\frac{|J|}{p} > \frac{49}{50}.$$

Then

$$A_J = \{n \in \mathbb{N} : H(n) \bmod p \in J\}$$

contains arbitrarily long arithmetic progressions with common difference in S.

By the preceding lemma, A_J also has uniform density $|J|/p$.

Thus, merely changing the period or the shape of the retained intervals in the range of this same H cannot produce witnesses at densities above $49/50$.

Here is the proof.

A group-level orbit on which H takes only 50 values

Put

$$m_0 = \frac{a}{b} = 10^8$$

and consider the finite-support direction

$$s = (-a, \underbrace{b, \dots, b}_{m_0 \text{ coordinates}}, 0, \dots) \in E. \tag{24}$$

This belongs to Alweiss's original S_∞: its distinguished coordinate is $-a$, its other nonzero coordinates are b, and their total norm is a.

Choose a finite-support starting point x whose distinguished coordinate is

$$x_1 = \frac{a}{2},$$

and whose next m_0 coordinates are

$$0, \frac{1}{m_0}, \dots, \frac{m_0 - 1}{m_0}.$$

The smoothed triangular function satisfies

$$\psi(t) + \psi(t + 1/2) = \frac{1}{2} - \varepsilon. \tag{25}$$

Because m_0 is even, we can pair the tail coordinates antipodally. Therefore

$$\sum_{j=0}^{m_0-1} \psi\left(\frac{j}{m_0} + tb\right) = m_0\left(\frac{1}{4} - \frac{\varepsilon}{2}\right) \qquad (t \in \mathbb{R}). \tag{26}$$

The entire tail contribution is **constant**, not merely periodic.

It follows that

$$H(x+ts) = C + \psi\left(\frac{a}{2} - ta\right). \tag{27}$$

At integer t, the argument of ψ ranges over the 100 odd multiples of $1/200$ modulo one. Since ψ is even, there are only **50 distinct values** in (27).

A set of measure greater than $49/50$ contains a translate of any 50-point set

Let these values, before adding C, be $v_1, \dots, v_{50}$ modulo p. Normalizing Lebesgue measure on $\mathbb{R}/p\mathbb{Z}$, the union bound gives

$$\mu\left(\bigcap_{\ell=1}^{50}(J - v_\ell)\right) \geq 1 - 50(1 - \mu(J)) > 0.$$

Hence some common translate of all 50 values lies in J.

We can realize the required additive constant in H by adding finitely many starting coordinates with **zero step**. Indeed, ψ is continuous and its range is the interval

$$[0, 1/2 - \varepsilon],$$

so finite sums of such contributions realize any desired nonnegative constant, and therefore any constant modulo p.

We have consequently constructed finite-support x, s, with $s \in S_\infty$, such that

$$H(x+ts) \bmod p \in J \qquad \text{for every integer } t. \tag{28}$$

Transfer any finite portion of this orbit to the original integers

The function

$$H_E(z) = \sum_j \psi(z_j)$$

is 1-Lipschitz on E. Also, the original map

$$P(n) = (n\alpha_1, n\alpha_2, \dots)$$

has dense positive orbit in E, as stated in Alweiss's construction. arXiv

Fix a desired length K. Since the finitely many values in (28) lie in the open set J, they have a positive margin from its complement.

Choose positive integers n, d such that

$$P(n) \approx x, \qquad P(d) \approx s,$$

with errors sufficiently small in terms of that margin and K. The approximation to s can also be chosen small enough that

$$P(d) \in S_\infty,$$

so $d \in S$.

For $0 \leq t < K$,

$$\|P(n+td)-(x+ts)\|_{\ell^1} \le \|P(n)-x\|_{\ell^1} + t\|P(d)-s\|_{\ell^1}.$$

Lipschitz continuity then preserves membership in J, giving

$$n, n+d, \dots, n+(K-1)d \in A_J.$$

Since K was arbitrary, the proposition follows. □

6. Where this leaves the main question

For the **same original Alweiss example**, the established conclusion is now

$$\boxed{\forall\, 0 < \delta \le \frac{489899}{509900}, \quad \exists A \subseteq \mathbb{N} \text{ of uniform density } \delta \text{ with no allowed 200-term AP.}}$$

So the existential assertion in your latest message is firmly settled, and the obstruction remains valid under a much stronger density requirement.

The unresolved question in this investigation is whether failure continues throughout

$$\frac{489899}{509900} < \delta < 1.$$

The proposition above does **not** show that S is δ-density large for $\delta > 49/50$: it concerns only sets obtained by imposing interval conditions on this particular observable H, not arbitrary subsets of $\mathbb{N}$. What it does show is that reaching density arbitrarily close to one requires more than finer bands or a longer progression in the present construction.

---

39. User

2026-09-15 12:33:19 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T04:33:19.270000Z

This is great progress! Please keep investigating further.

---

40. Assistant

2026-09-15 12:33:34 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T04:33:34.965843Z

I'll focus on the remaining high-density range, testing alternative or multiple observables and amplification arguments to overcome the scalar construction's current limitation.

---

41. Assistant

2026-09-15 12:36:38 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T04:36:38.294675Z

A new lead is the unused lower bound on tail mass; meanwhile, periodic examples force me beyond symmetric coordinate-sum observables toward coordinate labels or rotation roots.

42. Assistant

2026-09-15 12:45:42 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T04:45:42.972358Z

The two-scale construction works: wider sparse gaps constrain long progressions, then thinner gaps raise the obstruction density to 0.979 without modifying Alweiss's set.

---

43. Assistant

2026-09-15 13:03:50 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T05:03:50.616336Z

I found a substantially stronger obstruction for the unchanged Alweiss example: it is not δ-density large for every

$$\boxed{0 < \delta \le \frac{79297129}{80400000} = 0.986282699004975\ldots.}$$

The witnesses again have **uniform density exactly** δ.

This is not just a better numerical estimate for the previous argument. The new ingredient is a **two-scale construction**: sparse, wider gaps first constrain the whole progression, and this allows much thinner gaps to provide the actual obstruction.

I have written the complete argument, including the density calculation and two limitations of the method, in a standalone document:

Download the detailed LaTeX file

Read the compiled PDF — 13 pages

The full assertion for every $\delta < 1$ remains unproved. Here is the new result and its mechanism.

1. Precise statement

Throughout, S is the original Alweiss example with the displayed parameters

$$a = \delta_1 = \frac{1}{100}, \qquad b = \delta_2 = 10^{-10}.$$

Neither the parameters nor the sequence (α_j) defining the integer set is changed. These are the parameter choices given in Section 2.1 of his paper. arXiv

Theorem

Put

$$K = 1{,}100{,}200, \qquad \delta_{\mathrm{new}} = \frac{79297129}{80400000}.$$

There is an explicitly defined set $B \subseteq \mathbb{N}$ of uniform density δ_{new} containing no K-term arithmetic progression whose common difference belongs to S.

Moreover, for every $0 < \delta \le \delta_{\mathrm{new}}$, there is a subset $A \subseteq B$ satisfying

$$\lim_{N\to\infty}\sup_{M\ge 0}\left|\frac{|A\cap\{M+1,\dots,M+N\}|}{N}-\delta\right|=0.$$

Thus S fails δ-density largeness throughout this interval.

The tradeoff is the longer forbidden progression: the previous construction excluded 200-term progressions, whereas this one excludes $1{,}100{,}200$-term progressions. For disproving density largeness, either is sufficient.

2. The new idea: constrain the tail before trying to overpower it

Along an allowed progression, we write our observable as

$$y_t = u_t + z_t,$$

where u_t is the contribution of the distinguished coordinate and z_t is the sum of all the other contributions.

Previously, we bounded the possible change of z_t directly. The improvement is to exploit its **very small second differences**.

Here is the elementary estimate behind the argument.

Discrete drift lemma

Suppose a real sequence $z_0,\dots,z_{K-1}$ has oscillation at most M and

$$|z_{t+2}-2z_{t+1}+z_t|\le B.$$

Then, whenever $t+q\le K-1$,

$$\boxed{|z_{t+1}-z_t|\le \frac{M}{q}+\frac{B(q-1)}{2}.} \tag{1}$$

Indeed, successive summation gives

$$z_{t+q}-z_t = q(z_{t+1}-z_t)+\sum_{r=0}^{q-2}(q-1-r)(z_{t+r+2}-2z_{t+r+1}+z_{t+r}),$$

and division by q proves the estimate.

So a sequence with tiny curvature cannot maintain a substantial drift while staying in a bounded interval for a long time.

The coarse gaps enforce the bounded interval. The fine gaps are introduced only after this drift estimate becomes available.

3. Replace the triangle by an asymmetric ramp

Define a continuous, one-periodic function by

$$\chi(t)=\begin{cases} t, & 0\le t\le 9/10,\\ 9(1-t), & 9/10\le t\le 1.\end{cases}$$

It rises slowly for nine tenths of a period and falls rapidly during the remaining tenth.

Smooth it at scale

$$\varepsilon = \frac{1}{50}$$

and normalize at zero:

$$\boxed{\psi(t) = \frac{1}{2\varepsilon}\int_{-\varepsilon}^{\varepsilon} \chi(t-u)\,du - \frac{1}{20}.} \tag{2}$$

The subtracted constant is exactly the averaged value at zero, so $\psi(0) = 0$.

The relevant properties are

$$\operatorname{Lip}(\psi) \le 9, \qquad \operatorname{osc}\psi \le 0.9, \tag{3}$$

and, importantly,

$$\boxed{\psi(t) = t - \frac{1}{20} \qquad (0.02 \le t \le 0.88).} \tag{4}$$

Also,

$$\operatorname{Lip}(\psi') \le 250,$$

which gives

$$|\psi(t) - 2\psi(t+v) + \psi(t+2v)| \le 250v^2. \tag{5}$$

Now put

$$H(n) = \sum_{j=1}^{\infty} \psi(n\alpha_j). \tag{6}$$

This converges absolutely because

$$|\psi(n\alpha_j)| \le 9|n|\,|\alpha_j|,$$

and the original sequence is summable. The summability and dense finite-dimensional projections come from the construction in Section 2.3. arXiv

The advantage over the symmetric triangle is the long interval on which ψ is **exactly linear**. Once the tail drift is controlled, the distinguished coordinate forces a long, nearly uniform decrease of H.

4. The explicit high-density set

We impose two periodic restrictions on the same real-valued observable:

$$\boxed{B = \left\{ n \in \mathbb{N} : \begin{array}{l} H(n) \bmod 300 \in [0.181, 300), \\ H(n) \bmod 0.804 \in [0.01055, 0.804) \end{array} \right\}.} \tag{7}$$

The first condition deletes intervals of length 0.181, spaced 300 units apart. Its density cost is only

$$\frac{0.181}{300} \approx 0.0006033.$$

The second deletes intervals of length 0.01055, spaced 0.804 units apart. Its gaps are much narrower.

Here is why an allowed long progression cannot stay in both retained sets.

First: coarse confinement

For an allowed difference d, choose representatives σ_j of $d\alpha_j \bmod 1$ and a distinguished index i such that

$$\sigma_i = -a + \eta, \qquad |\eta| \le 2b,$$

and

$$|\sigma_j| \le 2b \quad (j \ne i), \qquad \sum_{j \ne i} |\sigma_j| \le a + 2b.$$

These are relaxed versions of the original strict defining inequalities. arXiv

They give

$$\sum_j |\sigma_j| \le 2a + 4b, \qquad \sum_{j \ne i} \sigma_j^2 \le 2b(a + 2b). \tag{8}$$

Suppose, toward a contradiction, that

$$n, n + d, \dots, n + (K-1)d \in B, \qquad d \in S.$$

Write

$$y_t = H(n + td), \quad u_t = \psi(n\alpha_i + t\sigma_i), \quad z_t = y_t - u_t.$$

From (3) and (8),

$$|y_{t+1} - y_t| \le 9(2a + 4b) = 0.1800000036 < 0.181. \tag{9}$$

Thus consecutive y_t's cannot jump across a coarse deleted interval. Since every y_t satisfies the coarse condition, **all of them lie in one retained interval of length** 299.819.

Consequently,

$$\operatorname{osc} z_t \le 299.819 + \operatorname{osc} u_t \le 300.719. \tag{10}$$

Second: the tail now has a much smaller drift

Using (5) and (8),

$$|\Delta^2 z_t| \le 250 \sum_{j \ne i} \sigma_j^2 \le 250 \cdot 2b(a + 2b) = 5.0000001 \times 10^{-10}. \tag{11}$$

Apply the drift lemma with

$$q = 1{,}100{,}000.$$

For every $0 \le t \le 199$, there are q subsequent steps available, since $K = q + 200$. Equations (10)–(11) therefore give

$$\begin{aligned} |z_{t+1} - z_t| &\le \frac{300.719}{1{,}100{,}000} + \frac{(5.0000001 \times 10^{-10})(1{,}099{,}999)}{2} \\ &= 0.000548380664590904\ldots \\ &< 0.000549. \end{aligned} \tag{12}$$

This is the essential improvement: after coarse confinement, the entire tail changes by less than 0.000549 per step during the portion we need.

Third: find 85 consecutive samples on the linear part

Let $\theta = n\alpha_i \bmod 1$. The 100 points

$$\theta - \frac{r}{100}, \qquad 0 \le r \le 99,$$

form a translated $1/100$-grid. Some r therefore gives a representative in

$$[0.865, 0.875).$$

For the actual step $\sigma_i = -1/100 + \eta$, the distinguished phases at

$$r, r+1, \dots, r+84$$

all lie between

$$0.025 - 366b \quad \text{and} \quad 0.875 + 366b.$$

This interval is strictly inside $[0.02, 0.88]$, so (4) applies throughout.

Hence

$$u_{r+j+1} - u_{r+j} = \sigma_i \qquad (0 \le j < 84).$$

Combining this with (12),

$$-a - 2b - 0.000549 \le y_{r+j+1} - y_{r+j} \le -a + 2b + 0.000549 < 0. \tag{13}$$

In particular, every step has magnitude at most

$$0.0105490002 < 0.01055.$$

So these 85 samples cannot jump across a fine deleted interval either. They must all lie in one fine retained interval of length

$$0.804 - 0.01055 = 0.79345. \tag{14}$$

But their total decrease, by (13), is at least

$$\begin{aligned} y_r - y_{r+84} &\ge 84(a - 2b - 0.000549) \\ &= 0.7938839832 \\ &> 0.79345. \end{aligned} \tag{15}$$

This contradicts (14).

Therefore the set B in (7) contains no allowed K-term progression.

5. Why its density is exactly $0.986282699\ldots$

The infinite-sum equidistribution lemma from the previous investigation applies to this new ψ as well:

> If φ is a nonconstant periodic Lipschitz function with $\varphi(0) = 0$, and (α_j) is summable with all finite tuples $1, \alpha_1, \dots, \alpha_m$ rationally independent, then
>
> $$\sum_j \varphi(n\alpha_j) \bmod p$$
>
> is well-distributed for every $p > 0$, uniformly over translated intervals of integers.

The standalone file reproves this carefully. The crucial point is that, although the tail of the infinite sum need not be uniformly small in n, the tail of every **fixed difference** is:

$$\left|\sum_{j>m}\bigl(\varphi((n+h)\alpha_j)-\varphi(n\alpha_j)\bigr)\right| \le \mathrm{Lip}(\varphi)|h|\sum_{j>m}|\alpha_j|.$$

This gives uniform correlation limits; finite-dimensional equidistribution makes their absolute Cesàro averages tend to zero; uniform van der Corput then proves the assertion.

For our two periodic restrictions, a common period is

$$L_0 = 20100 = 67\cdot 300 = 25000\cdot 0.804.$$

Thus B has uniform density equal to the proportion of the common period retained by both conditions.

Even without calculating overlaps, the union bound gives

$$d(B) \ge 1 - \frac{0.181}{300} - \frac{0.01055}{0.804} = 0.986274776119403\ldots.$$

So the clean conclusion for every $\delta \le 0.986$ already follows.

For the exact value, the two families of deleted intervals have total lengths

$$67(0.181) = 12.127, \qquad 25000(0.01055) = 263.75,$$

and their overlap has length

$$0.15925.$$

Therefore

$$\begin{aligned} d(B) &= 1 - \frac{12.127 + 263.75 - 0.15925}{20100} \\ &= \boxed{\frac{79297129}{80400000}}. \end{aligned} \tag{16}$$

The overlap computation is written out in the file using the common spacing 0.012; all these numbers are exact rational calculations.

Finally, the usual enumeration-based thinning preserves uniform density. If

$$B = \{b_1 < b_2 < \cdots\}, \qquad \vartheta = \frac{\delta}{d(B)},$$

retain b_j exactly when

$$\lfloor \vartheta j \rfloor > \lfloor \vartheta (j-1) \rfloor.$$

The resulting subset has uniform density exactly δ, and cannot acquire an allowed progression that was absent from B.

6. A new limitation: this particular potential cannot reach every density

I also checked whether one could simply keep refining periodic restrictions on the new H. There is a definite obstruction.

Proposition

For this particular asymmetric potential, let $p > 0$ and let $J \subseteq \mathbb{R}/p\mathbb{Z}$ be a finite union of open intervals with

$$\frac{|J|}{p} > \frac{90}{91}.$$

Then

$$A_J = \{n : H(n) \bmod p \in J\}$$

contains arbitrarily long arithmetic progressions with common difference in S.

This concerns **only sets defined through this particular** H, not arbitrary subsets of the integers.

The mechanism is a finite-value orbit. Put

$$m = \frac{a}{b} = 10^8, \qquad s = (-a, \underbrace{b, \ldots, b}_{m \text{ coordinates}}, 0, \ldots) \in S_\infty.$$

Give the distinguished starting coordinate phase zero, and give the tail coordinates the phases

$$0, \frac{1}{m}, \ldots, \frac{m-1}{m}.$$

Because m is divisible by 10, at every generic point precisely $9m/10$ terms of

$$\sum_{j=0}^{m-1} \chi(t + j/m)$$

have slope 1, and $m/10$ have slope -9. The sum is therefore constant. Smoothing preserves this cancellation.

Consequently the whole observable along this orbit is

$$H_E(x + ts) = C + \psi(-t/100).$$

At integer t, this takes only **91 distinct real values**. More explicitly, those values before adding C form

$$\left\{ \frac{k}{100} : -3 \le k \le 83 \right\} \cup \left\{ -\frac{11}{400}, \frac{331}{400}, \frac{299}{400}, \frac{21}{400} \right\}.$$

A set of normalized measure greater than $90/91$ contains a translate of any prescribed 91-point set: if V is that set, then

$$\mu\left(\bigcap_{v \in V} (J - v) \right) \ge 1 - 91(1 - \mu(J)) > 0.$$

The common translate can be realized by adding finitely many zero-step coordinates to the starting point.

Finally, openness of J, openness of S_∞, and density of the original integer orbit in E transfer any finite portion of this exact orbit to an arithmetic progression in the original integers. The ambient density is part of Alweiss's Section 2.3 construction. arXiv

Thus this potential has a genuine limitation at

$$\frac{90}{91} = 0.989010989\ldots.$$

The two-scale result gets close to that threshold, but changing the retained intervals alone cannot take it to density one.

7. What this changes about the remaining problem

The established interval has expanded from roughly 96.08% to ** 98.63%**, with uniform-density witnesses and no change to the original allowed differences.

More importantly, the new argument shows that the crude tail bound was not an intrinsic obstruction: **one can first force a long orbit to remain bounded, then use its tiny tail curvature to obtain a much sharper local estimate.** That is the useful new mechanism.

The remaining assertion is still

$$\forall \delta \in \left(\frac{79297129}{80400000}, 1\right), \quad S \text{ is not } \delta\text{-density large.}$$

The finite-value proposition does not refute this assertion. It shows that resolving the entire remaining range requires a witness that breaks the cancellation behind the finite-value orbit, rather than merely imposing more elaborate interval conditions on the same asymmetric sum.

---

44. User

2026-09-15 15:07:52 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T07:07:52.633000Z

great progress! please keep working on this.

---

45. Assistant

2026-09-15 15:08:09 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T07:08:09.538535Z

I'll test whether the two-scale mechanism can approach density one while seeking a route beyond the fixed-observable obstruction, carefully rechecking Alweiss's original hypotheses.

---

46. Assistant

2026-09-15 15:11:13 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T07:11:13.351113Z

A promising new route is dividing the orbit generator by a fixed integer: this preserves S while changing the rational motions responsible for the short-period obstruction.

---

47. Assistant

2026-09-15 15:20:18 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T07:20:18.019261Z

The divided-frequency route introduces rational offsets needing extra control, so I'm also testing higher-order differences as a separate way to sharpen the tail estimate.

48. Assistant

2026-09-15 15:47:20 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T07:47:20.964467Z

A gated two-observable construction now yields uniform-density witnesses of at least 0.9990506513, excluding 10^{18}-term allowed progressions. I'm checking starting-point uniformity and writing the estimates carefully.

49. Assistant

2026-09-15 15:53:18 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T07:53:18.459597Z

The unchanged Alweiss example is not 0.999-density large. I obtained a new construction proving the stronger bound

$$\boxed{0 < \delta \le \delta_* := \frac{9075692815599}{9084317000000} = 0.999050651314677\ldots.}$$

The witnesses have **uniform density exactly** δ. This improves the previous threshold of approximately 0.98628, and uses a different mechanism rather than just sharper numerical constants. The previous result and its limitations are recorded in the earlier note. [alweiss_two_scale_density_obstruction.tex, L114-L125]

Download the complete standalone LaTeX file

Read the compiled PDF — 13 pages

The full assertion for **every** $\delta < 1$ is still not established. Here is the new result and the argument.

1. Precise statement

Let S be Alweiss's original example, with the fixed parameters

$$a = \delta_1 = \frac{1}{100}, \qquad b = \delta_2 = 10^{-10}.$$

These are the parameters displayed in his paper; we do not change them or the sequence (α_j) defining the integer set. arXiv

Theorem

For every $0 < \delta \le \delta_*$, there is a set $A \subseteq \mathbb{N}$ such that

$$\lim_{N\to\infty} \sup_{M\ge 0} \left| \frac{|A \cap \{M+1, \ldots, M+N\}|}{N} - \delta \right| = 0$$

and A contains no arithmetic progression of length

$$\boxed{K = 10^{18}}$$

whose common difference belongs to S.

Thus S is not δ-density large throughout this range.

The progression length is enormous, but it is an **explicit, verified bound**, not an unspecified compactness constant. Neither this length nor the density threshold is claimed to be optimal.

2. The new idea: impose the expensive restriction only occasionally

Previously, we imposed coarse and fine restrictions everywhere on a single observable. The fine restriction accounted for most of the density loss.

The new construction uses two observables. A narrow band for the first acts as a **gate**: only while that gate is active do we impose the fine restriction on the second.

The crucial properties are:

- Every sufficiently long allowed progression must enter the gate.
- After entering, it stays in the gate along a suitable 200-term subprogression.
- On that subprogression, the second observable cannot avoid its fine forbidden band.

The density cost of the conditional restriction is a **product**, not a sum. Numerically, the fine band occupies about 1.6621% of its circle, but it is imposed only when the gate—occupying about 4.5907% of its circle—is active. Their joint cost is therefore only about

$$0.016621 \times 0.045907 = 0.000763015.$$

There are also two inexpensive coarse restrictions. Together, all three deletions cost less than 0.00095.

Two additional ingredients make this possible: using the second observable at a **root scale**, and obtaining a much better derivative estimate for the tail of the first observable.

3. The explicit construction

We retain the smoothed asymmetric ramp from the previous argument. Define the continuous one-periodic function

$$\chi(t) = \begin{cases} t, & 0 \le t \le 0.9, \\ 9(1-t), & 0.9 \le t \le 1, \end{cases}$$

and put

$$\psi(t) = \frac{1}{0.04} \int_{-0.02}^{0.02} \chi(t-u)\, du - 0.05.$$

Its useful properties are

$$\psi(0) = 0, \qquad \mathrm{Lip}(\psi) \le 9, \qquad \operatorname{osc} \psi \le 0.9,$$

$$\psi(t) = t - 0.05 \quad (0.02 \le t \le 0.88), \qquad \mathrm{Lip}(\psi') \le 250.$$

These properties are also proved in the previous note and are reproved in the new file. [alweiss_two_scale_density_obstruction.tex, L12-L48]

Now define

$$\boxed{F(n) = \sum_{j \ge 1} \big(\cos(2\pi n \alpha_j) - 1\big), \qquad G(n) = \sum_{j \ge 1} \psi(n\alpha_j/100).} \tag{1}$$

Both series converge absolutely for each n.

Notice the different scales: F uses $n\alpha_j$, whereas G uses $n\alpha_j/100$.

Delete the following sets:

$$E_0 = \{n : F(n) \bmod 1000 \in [0, 0.126)\},$$

$$E_1 = \{n : G(n) \bmod 3000 \in [0, 0.181)\},$$

and

$$E_f = \left\{ n : \begin{array}{l} F(n) \bmod 1.93 \in [0, 0.0886), \\ G(n) \bmod 0.706035 \in [0, 0.011735) \end{array} \right\}.$$

Set

$$\boxed{A_* = \mathbb{N} \setminus (E_0 \cup E_1 \cup E_f).} \tag{2}$$

Thus E_0, E_1 are coarse deletions imposed everywhere. In E_f, the fine G-band is deleted **only when the F-gate is active**.

I will show that A_* has uniform density at least δ_* and contains no allowed K-term progression.

4. A sharper tail estimate keeps the gate active

For an allowed difference d, the original defining conditions supply a distinguished index i and representatives σ_j of $d\alpha_j \bmod 1$ satisfying

$$\sigma_i = -a + \eta, \qquad |\eta| < 2b,$$

$$|\sigma_j| < 2b \quad (j \neq i), \qquad \sum_{j \neq i} |\sigma_j| < a + 2b.$$

These follow directly from Definition 2.1 of the original construction. arXiv

In particular,

$$\sum_j |\sigma_j| \leq 2a + 4b, \qquad \sum_{j \neq i} \sigma_j^2 \leq 2b(a + 2b). \tag{3}$$

Suppose an allowed K-term progression lies entirely in A_*. Write

$$y_t = F(n + td) = u_t + z_t,$$

where

$$u_t = \cos\bigl(2\pi(n\alpha_i + t\sigma_i)\bigr) - 1$$

is the distinguished contribution.

Coarse confinement

By (3),

$$|y_{t+1} - y_t| \leq 2\pi(2a + 4b) < 0.126.$$

Since every y_t avoids the coarse deleted bands of width 0.126, the sequence cannot move between different retained components. All its values lie in one interval of length $1000 - 0.126$.

After subtracting u_t, and extending the tail continuously between integer times, this gives

$$\operatorname{osc}_{[0,K-1]} z < 1002. \tag{4}$$

The new analytic estimate

The tail z is a sum of cosines whose frequencies have absolute value at most

$$\omega = 4\pi b.$$

Its global Lipschitz constant is at most

$$D_0 = 2\pi(a + 2b).$$

The following finite-window estimate is useful here.

> **Finite-window slope lemma.** Suppose
>
> $$f(t) = \sum_j c_j\big(\cos(\theta_j + \nu_j t) - \cos\theta_j\big), \quad |\nu_j| \le \omega, \quad \sum_j |c_j \nu_j| \le D.$$

> If f has oscillation at most M on
>
> $$[t_0 - R_N, t_0 + R_N], \qquad R_N = \frac{(2N-1)\pi}{2\omega},$$

> then
>
> $$\boxed{|f'(t_0)| \le \frac{\omega M}{2} + \frac{5D}{N}.} \tag{5}$$

The file gives an elementary proof using the differentiation identity

$$f'(t) = \frac{4\omega}{\pi^2} \sum_{k \ge 0} \frac{(-1)^k}{(2k+1)^2} \left(f\left(t + \frac{(2k+1)\pi}{2\omega}\right) - f\left(t - \frac{(2k+1)\pi}{2\omega}\right)\right).$$

This follows from the Fourier series of the triangle wave. Summation by parts bounds the tail by $5D/N$.

Take $N = 50{,}000{,}000$. The required radius is less than 1.25×10^{17}, so the windows fit comfortably inside our hypothetical progression when centered near $K/2$. Equations (4)–(5) give

$$\boxed{|z'(t)| < e_0 := 6.4 \times 10^{-7}} \tag{6}$$

throughout the portion of the progression used below.

This is much stronger than what the second-difference estimate alone supplied.

Entering and remaining in the gate

Among 100 consecutive samples near the middle of the progression, the distinguished cosine makes almost a complete turn. The estimates give

$$\operatorname{osc}(y_t) > 1.998, \qquad |y_{t+1} - y_t| < 0.063.$$

A real-valued walk with steps smaller than 0.063 and this much oscillation must meet one of the intervals

$$[1.93m + 0.0128,\ 1.93m + 0.0758].$$

Choose an index r where this happens.

These intervals lie inside the gate bands

$$[1.93m,\ 1.93m + 0.0886),$$

with margin 0.0128 on both sides.

For $0 \le q \le 199$,

$$100q\sigma_i = -q + 100q\eta.$$

The integer term disappears in the cosine. Combining this with (6),

$$\begin{aligned} |y_{r+100q} - y_r| &\le 100q(e_0 + 4\pi b) \\ &\le 19900(e_0 + 4\pi b) \\ &< 0.0128. \end{aligned}$$

Therefore

$$\boxed{F(n + (r + 100q)d) \bmod 1.93 \in (0, 0.0886) \quad (0 \le q \le 199).} \tag{7}$$

The gate stays active on all 200 selected points.

5. The root scale now forces a contradiction

Consider

$$Y_q = G(n + (r + 100q)d).$$

The j-th phase in this expression is

$$\frac{(n + rd)\alpha_j}{100} + q\, d\alpha_j.$$

This is why the factor $1/100$ in G matters: along the index-spacing- 100 subprogression, the increment becomes exactly $d\alpha_j$, so the original allowed-step estimates apply again.

We do **not** assume that $100d \in S$.

For a long extension of this subprogression, the coarse G-restriction gives

$$|Y_{q+1} - Y_q| \le 9(2a + 4b) < 0.181.$$

Consequently, all these Y_q's lie in one retained coarse interval of length $3000 - 0.181$.

Write $Y_q = U_q + Z_q$, separating the distinguished ramp coordinate. Then

$$\operatorname{osc} Z_q \le 3000.719,$$

and

$$|\Delta^2 Z_q| \le 250 \sum_{j \ne i} \sigma_j^2 \le 5.0000001 \times 10^{-10}.$$

The discrete drift estimate from the previous argument says that

$$|\Delta Z_q| \le \frac{M}{Q} + \frac{B(Q-1)}{2}.$$

Using $Q = 3{,}500{,}000$, we obtain

$$\boxed{|\Delta Z_q| < e_1 := 0.001734 \qquad (0 \le q \le 199).} \tag{8}$$

Among these first 200 samples, there are 85 consecutive ones for which the distinguished phase stays inside the linear interval $[0.02, 0.88]$ of ψ. On this portion,

$$\Delta U_q = \sigma_i = -a + \eta.$$

Thus the corresponding Y_q's decrease strictly. Each step has magnitude at most

$$a + 2b + e_1 = 0.0117340002 < 0.011735. \tag{9}$$

Their total decrease over the 84 steps is at least

$$84(a - 2b - e_1) = 0.6943439832. \tag{10}$$

But the gate is active throughout these samples by (7). Membership in A_* therefore forces every one of them to avoid the fine G-bands of width 0.011735.

By (9), the sequence cannot jump across such a forbidden band. It must stay in one retained interval, whose length is

$$0.706035 - 0.011735 = 0.6943.$$

This contradicts (10).

Hence ** A_* contains no allowed 10^{18}-term progression**.

6. The density calculation: why the product is legitimate

It would not be enough merely to assert that the two observables are independent: they are built from the same sequence (α_j). The file proves the required **joint well-distribution**.

Write

$$\beta_j = \alpha_j/100,$$

and consider the two functions

$$\varphi_0(x) = \cos(200\pi x) - 1, \qquad \varphi_1(x) = \psi(x).$$

Then

$$F(n) = \sum_j \varphi_0(n\beta_j), \qquad G(n) = \sum_j \varphi_1(n\beta_j).$$

No nonzero real linear combination of these functions is constant. Indeed, on the linear interval of ψ, the derivative of such a combination would be

$$-200\pi c_0 \sin(200\pi x) + c_1,$$

which cannot vanish identically unless $c_0 = c_1 = 0$.

The infinite-sum equidistribution argument consequently gives joint well-distribution of (F, G) modulo any two prescribed periods, uniformly over translated intervals of integers.

The proof handles the infinite tails through **fixed differences**, rather than incorrectly assuming that the original tails are uniformly small:

$$\sup_{n}\left|\sum_{j>m}\bigl(\varphi((n+h)\beta_j)-\varphi(n\beta_j)\bigr)\right|\le \mathrm{Lip}(\varphi)|h|\sum_{j>m}|\beta_j|\longrightarrow 0.$$

This yields uniform correlation limits. Finite-dimensional torus equidistribution and van der Corput then prove the assertion. The complete argument is included in the file.

All conditions in (2) have common periods in their respective variables. Thus A_* has a uniform density γ, and the union bound gives

$$\begin{aligned}1-\gamma &\le \frac{0.126}{1000}+\frac{0.181}{3000}+\frac{0.0886}{1.93}\,\frac{0.011735}{0.706035}\\ &=\frac{8624184401}{9084317000000}\\ &=0.000949348685322\ldots.\end{aligned}$$

Therefore

$$\boxed{\gamma\ge\delta_*.}$$

This is a lower bound for the explicit set's density, not a claim that the union-bound expression is its exact density.

Finally, enumeration-based thinning gives every smaller density exactly. If

$$A_*=\{a_1<a_2<\cdots\},\qquad \theta=\delta/\gamma,$$

retain a_j precisely when

$$\lfloor\theta j\rfloor>\lfloor\theta(j-1)\rfloor.$$

Counts on every integer interval are θ times the corresponding counts for A_*, up to a uniformly bounded error. The resulting subset has uniform density δ, and it cannot acquire a progression absent from A_*.

7. What this changes about the remaining problem

The important gain is that we have escaped the previous **single-observable interval restriction**. The fine obstruction now costs only a product of two small proportions, and the root scale restores the original allowed step on the subprogression where that obstruction is needed.

This proves, in particular, that a set occupying **exactly** 99.9% **of every sufficiently long interval**, uniformly in the interval's location, can still avoid all sufficiently long APs with differences in the unchanged Alweiss example.

The remaining interval is

$$\frac{9075692815599}{9084317000000}<\delta<1.$$

An iteration of the gate is a natural possibility, but the present argument does **not** automatically supply the longer persistence estimates such an iteration would require. The new conclusion rests on the explicit proof above, not on an assumed density-amplification principle.

50. User

2026-09-15 16:31:01 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T08:31:01.224000Z

how about this, can you modify the example of alweiss to obtain an example of a set of multiple recurrence for nilrotations that fails to be delta-density large for every delta in (0,1), and also fails to be a set of double topological recurrence?

51. Assistant

2026-09-15 16:31:19 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T08:31:19.577912Z

I'll investigate whether a diagonal modification of Alweiss's construction can preserve nilrecurrence and the double-topological-recurrence obstruction while producing density obstructions arbitrarily close to one.

52. Assistant

2026-09-15 16:34:23 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T08:34:23.239550Z

Two promising routes emerged: an irrational distinguished step with shrinking tail errors, or a diagonal selection across parameter versions. The latter requires proving simultaneous nilrecurrence rather than assuming intersection stability.

53. Assistant

2026-09-15 16:43:26 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T08:43:26.870770Z

The diagonal route now appears to achieve all three goals: it preserves a fixed three-term coloring obstruction, while each density-level obstruction leaves only finitely many exceptional differences.

54. Assistant

2026-09-15 17:01:36 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T09:01:36.277669Z

Yes. A diagonal modification of Alweiss's construction gives all three properties simultaneously. The two changes are to use a **fixed irrational distinguished step** and to make the **tail tolerances tend to zero at successive stages**. The nilrecurrence argument can accommodate both changes, while one fixed finite coloring still obstructs three-term progressions.

Download the complete standalone LaTeX file

Read the compiled PDF — 14 pages

The document includes the construction, the parameter dependencies, and proofs of all three assertions. Here is the result and its mechanism.

1. Precise statement

Theorem

There exists a single infinite set $D \subseteq \mathbb{N}$ with the following properties.

1. **Multiple pointwise recurrence for all nilsystems.** For every nilsystem (X, T), every $x \in X$, every $\ell \geq 1$, and every neighborhood U of x, there are arbitrarily large $d \in D$ such that

$$x, T^d x, T^{2d} x, \ldots, T^{\ell d} x \in U.$$

2. **Failure of double topological recurrence.** There is one finite coloring of $\mathbb{Z}$ containing no monochromatic progression

$$n, n+d, n+2d, \qquad d \in D.$$

Consequently, there is a compact minimal system (Y, S) and a nonempty clopen set $V \subseteq Y$ such that

$$V \cap S^{-d} V \cap S^{-2d} V = \varnothing \qquad (d \in D).$$

3. **Failure of δ-density largeness for every $\delta < 1$.** For every $0 < \delta < 1$, there are $A_\delta \subseteq \mathbb{N}$ and $K_\delta \geq 3$ such that

$$\overline{d}(A_\delta) = \delta$$

and A_δ contains no K_δ-term arithmetic progression whose common difference belongs to D.

The important quantifier is

$$\boxed{\exists D \, \forall \delta \in (0,1),}$$

not a separate choice of D for each density.

This retains Alweiss's basic construction: one coordinate has a definite displacement, while the other coordinates are individually tiny but have a controlled total mass. His original construction and complex-exponential coloring appear in Definition 2.1 and Section 2.2. arXiv

2. The two modifications

Fix

$$\boxed{a = \frac{\sqrt{2}}{1000}.}$$

For stage $j \geq 1$, set

$$L_j = 10^{12} 2^j, \qquad b_j = \frac{a}{L_j}, \qquad \varepsilon_j = \frac{b_j}{100}. \tag{1}$$

Thus $a / b_j = L_j$ remains an integer, but

$$b_j, \varepsilon_j \longrightarrow 0.$$

We construct a summable sequence of positive frequencies (α_r), together with finite sets

$$D_j \subseteq (N_{j-1}, N_j], \qquad D = \bigcup_{j \geq 1} D_j.$$

For every $d \in D_j$, there is a distinguished coordinate $i = i(d)$ and centered real representatives σ_r of $d\alpha_r \bmod 1$ satisfying

$$\boxed{\begin{aligned} \sigma_i &= -a + \eta, & |\eta| &< \varepsilon_j, \\ |\sigma_r| &\le b_j + \varepsilon_j & (r &\neq i), \\ a - \varepsilon_j < \sum_{r \neq i} |\sigma_r| &< a + \varepsilon_j. \end{aligned}} \tag{2}$$

In particular,

$$\sum_r |\sigma_r| < 2a + 2\varepsilon_j \tag{3}$$

and

$$\boxed{\sum_{r \neq i} \sigma_r^2 \le (b_j + \varepsilon_j)(a + \varepsilon_j) \longrightarrow 0.} \tag{4}$$

The distinguished displacement stays bounded away from zero, preserving the three-term obstruction. But the combined curvature contribution of all the other coordinates tends to zero, which is what makes the density witnesses approach density one.

3. Why nilrecurrence survives the shrinking tolerances

This is the delicate construction step. **We do not take an arbitrary intersection of nilrecurrent sets.** Instead, every finite stage is built to satisfy a uniform collection of nilrecurrence tests.

The input is Green's compactness argument: for a fixed finite collection of nilmanifolds, fixed rationally independent torus frequencies, and any prescribed accuracy, one finite orbit prefix works uniformly over all normalized polynomial sequences. Green states this at accuracy $1/m$; his proof works unchanged at any prescribed positive accuracy. arXiv

The finite-dimensional observation

Suppose $H \le \mathbb{R}^m$ has codimension at most k. In a sufficiently large prescribed coordinate block, we can find a vector

$$v = -ae_i + b_j \sum_{r \in J} e_r, \qquad |J| = L_j, \tag{5}$$

arbitrarily close, modulo $\mathbb{Z}^m$, to H.

Here is the elementary reason. Choose a coordinate set C, with $|C| \le k$, such that

$$H + \mathbb{R}^C = \mathbb{R}^m.$$

Write

$$b_j e_r = h_r + x_r, \qquad h_r \in H, \quad x_r \in \mathbb{R}^C.$$

Among sufficiently many indices, pigeonhole on the torus $\mathbb{R}^C/\mathbb{Z}^C$ gives $L_j + 1$ of the x_r's lying in one tiny box. Designate one index i, and call the others J. Then

$$
\begin{aligned}
-ae_i + b_j \sum_{r\in J} e_r &= \sum_{r\in J} (b_j e_r - b_j e_i) \\
&= \underbrace{\sum_{r\in J} (h_r - h_i)}_{\in H} + \underbrace{\sum_{r\in J} (x_r - x_i)}_{\text{small modulo } \mathbb{Z}^m}.
\end{aligned} \tag{6}
$$

The block can be chosen before the subspace H is known: its required size depends only on k, L_j, and the requested error. The file gives an explicit sufficient size.

Apply this to nilmanifold orbit closures

Enumerate standard nilmanifold presentations

$$X_1, X_2, \dots.$$

At stage j, use a new coordinate block large enough for the codimensions associated with $X_1, \dots, X_j$. Choose the new α_r's very small, with

$$1, \alpha_1, \dots, \alpha_m$$

rationally independent.

For a normalized polynomial orbit $p(n)\Gamma$, the uniform finite-orbit argument supplies a subnilmanifold through the identity inside

$$\mathbb{T}^m \times G/\Gamma$$

whose presenting subgroup has dimension at least m. Its directions with nilmanifold coordinate equal to the identity form a subspace

$$H \le \mathbb{R}^m, \qquad \operatorname{codim} H \le \dim G.$$

Thus (6) gives a target of the form

$$(v, \Gamma),$$

up to an arbitrarily small torus error. A sufficiently long finite orbit prefix approximates this target.

Choosing one cutoff N_j uniformly gives a finite set D_j such that

$$
\boxed{\begin{array}{c}\text{for every } r \le j \text{ and every normalized polynomial sequence } p \text{ on } X_r, \\ \exists d \in D_j : \qquad d_{X_r}(p(d)\Gamma_r, \Gamma_r) \le \dfrac{1}{2j}.\end{array}} \tag{7}
$$

Uniformity over the orbit—not just a countable selection of rotations—is essential here. It comes from bounded polynomial representatives, continuity of finite evaluations, and a finite subcover, as in Green's Lemma 3.1 and Proposition 3.3. arXiv

Two safeguards in the recursion

The new block's frequencies are chosen smaller than

$$\frac{a}{10N_{j-1}}.$$

Therefore an integer $n \le N_{j-1}$ cannot approximate a template whose distinguished coordinate is $-a$ in that block. This ensures that the witnesses in (7) are genuinely new:

$$d > N_{j-1}.$$

After choosing N_j, all future frequencies are constrained so that

$$N_j \sum_{r>m_j} \alpha_r < \frac{\varepsilon_j}{8}. \tag{8}$$

Consequently, future stages do not spoil the template approximations already established at stage j.

It follows from (7) that D is pointwise recurrent for every nilsystem, with unbounded return times. Arbitrary basepoints are handled by conjugating the translating element. General presentations reduce to connected, simply connected ones by the usual nilsystem embedding reduction. arXiv

Finally, applying pointwise recurrence to

$$(X^\ell, T \times T^2 \times \cdots \times T^\ell)$$

gives multiple pointwise recurrence.

4. Why one fixed coloring still obstructs double recurrence

Define

$$\Phi(n) = \sum_{r\geq 1} \big(e(n\alpha_r) - 1\big), \qquad e(t) = e^{2\pi i t}. \tag{9}$$

This converges absolutely for every integer n, since

$$|e(n\alpha_r) - 1| \leq 2\pi |n| \alpha_r.$$

We use Alweiss's complex second-difference mechanism, but verify the estimates for the modified parameters directly. arXiv

For $d \in D_j$,

$$\begin{aligned} W &= \Phi(n) - 2\Phi(n+d) + \Phi(n+2d) \\ &= \sum_r e(n\alpha_r)(1 - e(\sigma_r))^2. \end{aligned} \tag{10}$$

The distinguished summand has magnitude

$$4\sin^2(\pi\sigma_i).$$

Because $a/2 < |\sigma_i| < 2a < 1/2$, this is at least $4a^2$.

The sum of the other terms has magnitude at most

$$4\pi^2 \sum_{r\neq i} \sigma_r^2 \leq 4\pi^2 (b_j + \varepsilon_j)(a + \varepsilon_j) < a^2.$$

An upper bound on the distinguished term also gives

$$\boxed{3a^2 < |W| < (16\pi^2 + 1)a^2 < \frac{1}{4}.} \tag{11}$$

Now partition $\mathbb{C}/\mathbb{Z}[i]$ into finitely many squares of Euclidean diameter at most $a^2/100$, and color n according to the square containing

$$\Phi(n) \bmod \mathbb{Z}[i].$$

If $n, n+d, n+2d$ were monochromatic, then W would be within $a^2/50$ of a Gaussian integer. Since $|W| < 1/4$, that Gaussian integer must be zero. This contradicts $|W| > 3a^2$.

Thus **the same coloring works for every stage**. Passing to a minimal subsystem of the coloring's shift orbit closure gives the claimed dynamical failure of double topological recurrence.

5. The new density argument

Put

$$F(n) = \Re\Phi(n) = \sum_r \big(\cos(2\pi n\alpha_r) - 1\big). \tag{12}$$

Fix an integer $P \geq 2$ and $0 < h < 1/4$. Consider the periodic retained set

$$\mathcal{R}_{P,h} = \{y \in \mathbb{R} : y \bmod P \in [1, P),\ y \bmod 1 \in [h, 1)\}. \tag{13}$$

Its relative measure is exactly

$$\boxed{\beta = (1 - 1/P)(1 - h).} \tag{14}$$

There are two kinds of deleted intervals: gaps of width 1, separated by distance P, and gaps of width h, separated by distance 1. Both density costs can be made arbitrarily small.

For each real phase θ, let

$$B_\theta = \{n : F(n) + \theta \in \mathcal{R}_{P,h}\}.$$

The key conclusion is:

> For every P, h, there are J, K_0 such that **every** B_θ contains no K_0-term progression with difference in
>
> $$\bigcup_{j \geq J} D_j.$$

Here is the proof.

Coarse confinement bounds the tail

Suppose an allowed progression lies in B_θ, and write

$$y_t = F(n + td) + \theta = u_t + z_t,$$

where

$$u_t = \cos(2\pi(n\alpha_i + t\sigma_i)) - 1$$

is the distinguished coordinate's contribution.

By (3),

$$|y_{t+1} - y_t| \leq 2\pi(2a + 2\varepsilon_j) < 1. \tag{15}$$

The progression cannot jump across a coarse deleted gap of width 1. Therefore all its y_t's lie in one coarse retained interval, giving

$$\operatorname{osc} y_t \leq P - 1.$$

Since $\operatorname{osc} u_t \leq 2$,

$$\boxed{\operatorname{osc} z_t \le P + 1.} \tag{16}$$

Meanwhile,

$$\boxed{|\Delta^2 z_t| \le B_j := 4\pi^2(b_j + \varepsilon_j)(a + \varepsilon_j) \longrightarrow 0.} \tag{17}$$

Bounded range and small curvature make the tail nearly constant

The elementary drift estimate is

$$\boxed{|\Delta z_t| \le \frac{M}{Q} + \frac{B(Q-1)}{2}} \tag{18}$$

whenever the sequence has oscillation at most M, second differences at most B, and Q subsequent steps are available.

It follows from the exact identity

$$z_{t+Q} - z_t = Q\Delta z_t + \sum_{r=0}^{Q-2} (Q-1-r)\Delta^2 z_{t+r}.$$

Choose $L \ge 2$ so that

$$\{0, a, \dots, (L-1)a\}$$

is $h/(16\pi)$-dense on the circle. This is possible because ** a is irrational**.

Next choose

$$Q \ge \frac{16L(P+1)}{h}, \qquad K_0 = Q + L. \tag{19}$$

Finally choose J so large that, for every $j \ge J$,

$$\frac{B_j(Q-1)}{2} \le \frac{h}{16L}, \qquad 2\pi L\varepsilon_j \le \frac{h}{8}. \tag{20}$$

Applying (18) with (16)–(20) gives

$$|\Delta z_t| \le \frac{h}{8L}$$

during the first L samples. Hence

$$|z_t - z_0| \le h/8 \qquad (0 \le t < L). \tag{21}$$

The irrational cosine must enter a fine forbidden band

Because $\sigma_i = -a + \eta$, with $|\eta| < \varepsilon_j$, the actual values y_t differ by at most $h/4$ from

$$z_0 + \cos(2\pi(n\alpha_i - ta)) - 1, \qquad 0 \le t < L. \tag{22}$$

For every real c, the range of

$$c + \cos(2\pi x) - 1$$

is $[c-2, c]$, which contains a point of $\mathbb{Z} + h/2$. By our choice of L, the sampled values in (22) include one within $h/8$ of such a point.

The corresponding actual y_t is therefore within $3h/8$ of an integer plus $h/2$, so

$$y_t \bmod 1 \in [h/8, 7h/8] \subset (0, h).$$

This lies in a fine deleted band, a contradiction.

That proves the obstruction for all sufficiently late stages, uniformly in the translating phase θ.

6. Handle the finite initial stages and obtain exactly the requested density

Fix $0 < \delta < 1$. Choose P, h such that

$$\beta = (1 - 1/P)(1 - h) > \delta.$$

The preceding argument gives J, K_0. The exceptional set

$$E_0 = \bigcup_{j<J} D_j$$

is finite.

Choose a prime $q > \max E_0$, sufficiently large that

$$\beta(1 - 1/q) > \delta,$$

and define

$$A_\theta = \{n : q \nmid n,\ F(n) + \theta \in \mathcal{R}_{P,h}\}. \tag{23}$$

For any $d \in E_0$, we have $\gcd(d, q) = 1$, so every q-term progression with difference d contains a multiple of q. Thus A_θ excludes such progressions.

For the later differences, the preceding argument excludes length K_0. Consequently, **every** A_θ contains no D -progression of length

$$K = \max\{q, K_0, 3\}. \tag{24}$$

Phase averaging supplies high upper density

No equidistribution theorem for the infinite sum is needed. For every N,

$$\frac{1}{P}\int_0^P \frac{|A_\theta \cap [1, N]|}{N}\, d\theta = \beta\left(1 - \frac{\lfloor N/q \rfloor}{N}\right).$$

Reverse Fatou's inequality gives

$$\frac{1}{P}\int_0^P \overline{d}(A_\theta)\, d\theta \geq \beta(1 - 1/q) > \delta.$$

Therefore some **single fixed phase** θ satisfies

$$\alpha := \overline{d}(A_\theta) > \delta.$$

Finally, enumerate $A_\theta = \{a_1 < a_2 < \cdots\}$, set $\lambda = \delta/\alpha$, and retain a_j exactly when

$$\lfloor \lambda j \rfloor > \lfloor \lambda(j-1) \rfloor.$$

The resulting set A_δ satisfies

$$|A_\delta \cap [1,N]| = \lfloor \lambda |A_\theta \cap [1,N]| \rfloor ,$$

and hence

$$\overline{d}(A_\delta) = \lambda\alpha = \delta.$$

As a subset of A_θ, it still avoids all K-term progressions with differences in D.

This proves the all-densities assertion.

What has been resolved

Allowing the modification does give the complete example you requested. The construction is fixed independently of density, remains multiply pointwise recurrent for every finite-step nilsystem, and retains one uniform obstruction to double topological recurrence.

The argument for densities approaching one is now qualitative rather than a sequence of numerical improvements: a smaller density deficit merely requires a later stage, where the tail curvature is sufficiently small.

It does **not** settle the earlier question for the unchanged Alweiss example. The fixed irrational distinguished step and the stage-dependent shrinking errors are precisely the additional features that make this construction work.

---

55. User

2026-09-15 17:09:42 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T09:09:42.576000Z

can you construct a set of measurable recurrence that fails to be delta density regular for some delta in (0,1)?

---

56. Assistant

2026-09-15 17:10:01 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T09:10:01.573468Z

I'll interpret " δ-density regular" as your earlier fixed-density largeness condition, and check whether classical single-versus-multiple recurrence constructions already provide an explicit example.

---

57. User

2026-09-15 17:11:33 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T09:11:33.197000Z

can you construct a set of measurable recurrence that fails to be delta density regular for every delta in (0,1)?

58. Assistant

2026-09-15 17:11:51 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T09:11:51.225378Z

I'll investigate whether one fixed measurably recurrent set can fail your density-largeness condition simultaneously for every $\delta \in (0,1)$.

59. Assistant

2026-09-15 17:14:54 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T09:14:54.727700Z

A quadratic example works: $D = \{n : \{\sqrt{2}\, n^2\} \in [1/4, 3/4]\}$ is measurably recurrent yet fails density largeness for every $\delta < 1$. The tradeoff is losing pointwise nilrecurrence.

60. Assistant

2026-09-15 17:23:59 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T09:23:59.862459Z

For each δ, the obstruction can be witnessed by a quadratic-phase set of uniform density exactly δ; I'll make the finite-difference argument and progression-length bound explicit.

61. Assistant

2026-09-15 17:29:55 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-15T09:29:55.411555Z

Yes. A single classical quadratic set already works, and the witnessing sets can even have uniform density exactly δ. No modification of Alweiss's construction is needed for the properties in this question.

Take

$$\boxed{D = \left\{ n \in \mathbb{N} : \{\sqrt{2}\, n^2\} \in [1/4, 3/4] \right\}.}$$

This is the $k = 2$ case of the construction of Frantzikinakis–Lesigne–Wierdl. Their Theorem A establishes that it is measurably recurrent but not doubly recurrent. The **failure of your density condition for every** $\delta < 1$ will follow from the direct argument below; I am not attributing that particular formulation to their paper. arxiv.org

Download the complete standalone LaTeX proof

Read the compiled PDF — 9 pages

Here " δ-density regular" means your earlier " δ-density large" condition: every set of natural upper density exactly δ contains arbitrarily long arithmetic progressions with differences in D.

1. Precise theorem

In fact, any irrational α works.

Theorem

Let

$$D_\alpha = \left\{ n \in \mathbb{N} : \|\alpha n^2\|_{\mathbb{R}/\mathbb{Z}} \geq \frac{1}{4} \right\}, \qquad \alpha \notin \mathbb{Q}.$$

Then D_α is a set of measurable recurrence. Moreover, for every $0 < \delta < 1$, there are integers $M_\delta \geq 1$ and $K_\delta \geq 3$ such that

$$\boxed{A_\delta = \left\{ n \in \mathbb{N} : \left\| \frac{\alpha n^2}{M_\delta} \right\|_{\mathbb{R}/\mathbb{Z}} \geq \frac{1-\delta}{2} \right\}} \tag{1}$$

has **uniform density exactly** δ and contains no K_δ-term arithmetic progression whose difference belongs to D_α.

Uniform density here means

$$\lim_{N\to\infty} \sup_{L \geq 0} \left| \frac{|A_\delta \cap \{L+1, \ldots, L+N\}|}{N} - \delta \right| = 0. \tag{2}$$

Thus

$$\boxed{\exists D \text{ measurably recurrent} \quad \forall \delta \in (0,1), \quad D \text{ is not } \delta\text{-density regular.}}$$

The allowed difference set is fixed once and for all. Only the witness A_δ and its forbidden progression length depend on δ.

2. Why D_α is measurably recurrent

There is actually a stronger averaging statement. Let $(X, \mathcal{B}, \mu, T)$ be a probability-preserving system, let $Uf = f \circ T$, and let P denote the orthogonal projection onto the T-invariant functions. Then

$$\boxed{\frac{1}{|D_\alpha \cap [1,N]|} \sum_{d \in D_\alpha \cap [1,N]} U^d f \longrightarrow Pf \quad \text{in } L^2(\mu).} \tag{3}$$

This is a special case of the weighted averaging result in Proposition 3.2 of Frantzikinakis–Lesigne–Wierdl. Here is a proof of the case we need. arXiv

For a nonzero integer m, consider

$$v_n = e(m\alpha n^2) U^n f, \qquad e(t) = e^{2\pi i t}.$$

Its lag- h correlation is

$$\langle v_{n+h}, v_n \rangle = e\big(m\alpha(2nh + h^2)\big) \langle U^h f, f \rangle.$$

For each $h \geq 1$, its average tends to zero, because $2mh\alpha$ is irrational. Hilbert-space van der Corput therefore gives

$$\frac{1}{N} \sum_{n=1}^{N} e(m\alpha n^2) U^n f \longrightarrow 0 \qquad (m \neq 0). \tag{4}$$

Trigonometric approximation, followed by approximation of the interval indicator, now yields

$$\frac{1}{N}\sum_{n=1}^{N}\mathbf{1}_{[1/4,3/4]}(\alpha n^2)U^n f-\frac{1}{2}\frac{1}{N}\sum_{n=1}^{N}U^n f\longrightarrow 0. \tag{5}$$

For the indicator approximation, the norm of the error is bounded by $\|f\|_2$ times the scalar average of the absolute approximation error; quadratic equidistribution identifies its limit with the corresponding integral.

The ordinary mean ergodic theorem and

$$\frac{|D_\alpha\cap[1,N]|}{N}\longrightarrow\frac{1}{2}$$

prove (3).

Taking $f=\mathbf{1}_B$ and then its inner product with $\mathbf{1}_B$, we obtain

$$\boxed{\lim_{N\to\infty}\frac{1}{|D_\alpha\cap[1,N]|}\sum_{d\in D_\alpha\cap[1,N]}\mu(B\cap T^{-d}B)=\|P\mathbf{1}_B\|_2^2\geq\mu(B)^2.} \tag{6}$$

In particular, whenever $\mu(B)>0$, there are arbitrarily large $d\in D_\alpha$ with positive intersection.

So the example is not merely measurably recurrent: it is **nicely measurably recurrent**.

3. The key to the density obstruction: clear all small rational denominators

The tempting unscaled witness

$$\{n:\|\alpha n^2\|\geq(1-\delta)/2\}$$

does not automatically work. Along an arithmetic progression, a quadratic phase can resemble a rational periodic orbit that misses a short interval.

The integer M_δ in (1) removes precisely this obstruction.

We need the following elementary lemma.

Lemma: uniform finite hitting

For every $0<\varepsilon<1$, there are integers $M\geq 1$ and $K\geq 3$ such that

$$\boxed{\|M\beta\|\geq\frac{1}{4}\quad\Longrightarrow\quad\min_{0\leq j<K}\|\beta j^2+\lambda j+c\|<\frac{\varepsilon}{2}} \tag{7}$$

for **every** $\beta,\lambda,c\in\mathbb{R}$.

Thus one finite orbit length works uniformly in all three coefficients.

An explicit choice of M is

$$H=\left\lceil\frac{32}{\varepsilon^2}\right\rceil,\qquad Q=\left\lceil\frac{128H}{\varepsilon^2}\right\rceil,\qquad\boxed{M=\operatorname{lcm}(1,\dots,Q).} \tag{8}$$

The proof below gives the existence of K; no numerical bound on K is needed for the conclusion.

Step 1: an infinite avoiding orbit has a bounded rational denominator

Suppose

$$\|\beta j^2 + \lambda j + c\| \geq \frac{\varepsilon}{2} \qquad (j \geq 0). \tag{9}$$

If β is irrational, quadratic equidistribution contradicts avoidance of the open arc

$$(-\varepsilon/2, \varepsilon/2) \pmod 1.$$

Therefore $\beta = p/q \pmod 1$, in lowest terms.

We claim that $q \leq Q$.

Consider the tent function

$$\tau_\varepsilon(x) = \max\left(1 - \frac{2\|x\|}{\varepsilon}, 0\right).$$

It vanishes at every point outside the forbidden arc. Its absolutely convergent Fourier series has coefficients

$$c_0 = \frac{\varepsilon}{2}, \qquad c_m = \frac{2}{\varepsilon}\left(\frac{\sin(\pi m \varepsilon/2)}{\pi m}\right)^2 \quad (m \neq 0). \tag{10}$$

In particular,

$$c_m \geq 0, \qquad \sum_{m \in \mathbb{Z}} c_m = 1, \qquad \sum_{|m| > H} c_m \leq \frac{4}{\pi^2 \varepsilon H} < \frac{\varepsilon}{8}. \tag{11}$$

For the rational quadratic phase, the exponential averages

$$L_m = \lim_{N \to \infty} \frac{1}{N} \sum_{j=0}^{N-1} e\left(m\left(\frac{p}{q} j^2 + \lambda j + c\right)\right)$$

exist and satisfy

$$\boxed{|L_m| \leq \sqrt{\frac{\gcd(2m, q)}{q}} \leq \sqrt{\frac{2|m|}{q}}.} \tag{12}$$

To check that this estimate allows arbitrary λ, split the sum into residue classes modulo q. If $m\lambda q \notin \mathbb{Z}$, the average on each class is zero by a geometric sum. Otherwise the sequence is q-periodic. Squaring the absolute value of its period average and taking the difference of the two summation indices gives an inner sum

$$\sum_{r=0}^{q-1} e(2mptr/q),$$

which vanishes unless $q \mid 2mt$. There are exactly $\gcd(2m, q)$ such residues t, proving (12).

Now suppose $q > Q$. For $0 < |m| \leq H$, our choice of Q gives

$$|L_m| \leq \sqrt{\frac{2H}{q}} < \frac{\varepsilon}{8}.$$

Averaging the Fourier series of the tent function, and using (11), therefore gives

$$\begin{aligned}\lim_{N\to\infty}\frac{1}{N}\sum_{j=0}^{N-1}\tau_\varepsilon(\beta j^2+\lambda j+c) &= \frac{\varepsilon}{2}+\sum_{m\neq 0}c_m L_m\\ &\geq \frac{\varepsilon}{2}-\frac{\varepsilon}{8}-\frac{\varepsilon}{8}\\ &= \frac{\varepsilon}{4}>0.\end{aligned}\tag{13}$$

But every summand is zero by (9). This contradiction proves

$$q\leq Q,\qquad \text{and consequently}\qquad M\beta\in\mathbb{Z}.\tag{14}$$

Step 2: compactness makes the hitting time uniform

For each K, let $\mathcal{C}_K\subseteq\mathbb{T}^3$ consist of coefficient triples satisfying

$$\|M\beta\|\geq\frac{1}{4}$$

and

$$\|\beta j^2+\lambda j+c\|\geq\frac{\varepsilon}{2}\qquad(0\leq j<K).$$

These are nested compact sets.

If every $\mathcal{C}_K$ were nonempty, their intersection would contain an infinite avoiding orbit satisfying (9). But (14) would then contradict $\|M\beta\|\geq 1/4$.

Thus some $\mathcal{C}_K$ is empty, proving (7). □

4. Construct the witness at every prescribed density

Fix $0<\delta<1$, set

$$\varepsilon=1-\delta,$$

and choose $M=M_\delta$ and $K=K_\delta$ from the lemma.

Define

$$A_\delta=\left\{n\in\mathbb{N}:\left\|\frac{\alpha n^2}{M}\right\|\geq\frac{\varepsilon}{2}\right\}.$$

Because α/M is irrational, the sequence

$$\frac{\alpha n^2}{M}\pmod 1$$

is well-distributed, including uniformly over translated intervals of integers. The retained arc

$$[\varepsilon/2,1-\varepsilon/2]$$

has length $1-\varepsilon=\delta$, so A_δ has uniform density exactly δ.

The uniformity follows directly from van der Corput: for a fixed irrational quadratic coefficient, the lag- h correlation is a geometric sum with irrational slope $2h\alpha/M$. Its magnitude is independent of the linear and constant coefficients, which absorb any translation of the integer interval.

Now suppose

$$a, a+d, \dots, a+(K-1)d \in A_\delta.$$

The tested phases along this progression are

$$\frac{\alpha}{M}(a+jd)^2 = \underbrace{\frac{\alpha d^2}{M}}_{\beta} j^2 + \underbrace{\frac{2\alpha a d}{M}}_{\lambda} j + \underbrace{\frac{\alpha a^2}{M}}_{c}. \tag{15}$$

All K phases avoid the open arc of length ε. The lemma forces

$$\|M\beta\| < \frac{1}{4}.$$

But

$$M\beta = \alpha d^2,$$

so

$$\|\alpha d^2\| < \frac{1}{4}.$$

This means $d \notin D_\alpha$.

Therefore

$$\boxed{A_\delta \text{ has no } K_\delta\text{-term arithmetic progression with difference in } D_\alpha.}$$

This proves the all-densities assertion, with the **same fixed measurably recurrent set** throughout.

5. Relation to the previous nilrecurrence construction

There is an important scope distinction: **this example does not retain multiple recurrence for nilrotations**. That was an additional requirement in your previous question.

Indeed, consider the minimal skew shift

$$R(x,y) = (x+\alpha, y+x+\alpha/2) \quad \text{on } \mathbb{T}^2.$$

Its iterates are

$$R^n(x,y) = (x+n\alpha, y+nx+\alpha n^2/2).$$

Let

$$V = \{(x,y) : \|y\| < 1/32\}.$$

If a point and its dth and $2d$th iterates all belonged to V, choose second-coordinate representatives $y_0, y_1, y_2 \in (-1/32, 1/32)$. Then

$$y_0 - 2y_1 + y_2 \equiv \alpha d^2 \pmod 1,$$

giving

$$\|\alpha d^2\| \le |y_0| + 2|y_1| + |y_2| < \frac{1}{8}.$$

This is impossible for $d \in D_\alpha$. Hence

$$V \cap R^{-d}V \cap R^{-2d}V = \varnothing \qquad (d \in D_\alpha).$$

The file includes an explicit two-step nilrotation presentation of this skew shift.

Thus the present question has a complete affirmative answer, even with nice measurable recurrence and uniform-density witnesses. The stronger simultaneous requirement of measurable recurrence **and** multiple nilrecurrence is not established by this quadratic example.

---

62. User

2026-09-16 14:28:20 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-16T06:28:20.427000Z

it is clear that delta density large for some delta in (0,1/2] implies 2 large. Does 2 large imply delta density large for some absolute delta in (0,1)?

---

63. Assistant

2026-09-16 14:29:14 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-16T06:29:14.921359Z

I'll investigate whether two-color largeness forces any absolute density threshold, paying particular attention to possible separations between color-based, topological, and density-based recurrence.

---

64. Assistant

2026-09-16 14:32:18 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-16T06:32:18.550107Z

The quadratic example is inapplicable because it is not 2-large, and pair-recurrence counterexamples such as Kříž's do not address arbitrarily long progressions.

---

65. Assistant

2026-09-16 14:41:26 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-16T06:41:26.241812Z

I have not found the converse; the unresolved point is whether 2-largeness yields a density threshold uniform over all progression lengths.

---

66. Assistant

2026-09-16 14:47:21 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-16T06:47:21.169724Z

I do not currently have a proof or a counterexample for an absolute $\delta_0 < 1$. I also did not find a published result resolving this precise implication. I would treat it as an additional conjecture, rather than as a consequence of the nilrecurrence theorem.

There is a useful finite-length result:

$$\boxed{D \text{ is } 2\text{-large} \quad \Longrightarrow \quad D \text{ is } (1-1/k,k)\text{-density large for every } k \geq 2.} \tag{1}$$

In fact, for each fixed D, k, the density threshold can be made **strictly smaller** than $1 - 1/k$. The difficulty is obtaining a threshold bounded away from 1 **uniformly in the progression length**.

I will prove this and explain precisely why the earlier examples and the usual color-versus-density comparisons do not settle your question.

1. The quantifier that needs to be controlled

Your question is whether

$$\boxed{\exists \delta_0 < 1 \quad \forall D \subseteq \mathbb{N}: \quad D \text{ 2-large} \Longrightarrow D \text{ } \delta_0\text{-density large.}} \tag{2}$$

Here the same δ_0 must work both for **every 2-large set** and for **all progression lengths**.

As you observed, if $\delta \leq 1/2$, then δ-density largeness implies 2-largeness: one cell of a binary coloring has upper density at least $1/2$, and we can thin it to upper density exactly δ.

The reverse argument does not work. Given a high-density set A, 2-largeness applied to

$$\mathbb{N} = A \sqcup A^c$$

allows all the arbitrarily long monochromatic progressions to occur in A^c. Small density does not, by itself, exclude arbitrarily long progressions in that cell.

This is the general distinction between chromatic recurrence and density recurrence. The literature explicitly distinguishes the two; Robertson's discussion of the 2-large conjecture highlights the difficulty of forcing the permitted progressions into a positive-density cell. arXiv

2. A finite-length consequence, including the endpoint

Here is a precise lemma that goes slightly beyond the elementary union-bound argument.

Proposition

Suppose $D \subseteq \mathbb{N}$ is $(2, k)$-large, where $k \geq 2$. Then there is a number

$$\eta = \eta(D, k) > 0$$

such that every $A \subseteq \mathbb{N}$ with

$$\overline{d}(A) > 1 - \frac{1}{k} - \eta$$

contains a k-term arithmetic progression with common difference in D.

In particular, every set of upper density **at least**

$$1 - \frac{1}{k}$$

contains such a progression.

Proof

By compactness, there is an integer N such that every binary coloring of $[N]$ has a monochromatic k-term progression with common difference in D.

Let $\mathcal{P}$ be the finite collection of all such progressions contained in $[N]$, and write

$$m = |\mathcal{P}|.$$

Suppose A contains no k-term progression with difference in D. Put

$$b(n) = \mathbf{1}_{\mathbb{N}\setminus A}(n).$$

For $t \geq 0$ and $P \in \mathcal{P}$, define

$$Y_P(t) = \sum_{s \in P} b(t+s).$$

Since $t + P$ cannot lie entirely in A,

$$Y_P(t) \geq 1 \qquad (P \in \mathcal{P}). \tag{3}$$

Now apply the finite binary-coloring property to the block $t + [N]$, coloring according to membership in A. Its monochromatic progression cannot lie in A, so some $P \in \mathcal{P}$ satisfies

$$Y_P(t) = k.$$

Combining this with (3),

$$\boxed{\sum_{P \in \mathcal{P}} Y_P(t) \geq m + k - 1 \qquad (t \geq 0).} \tag{4}$$

Average over $0 \leq t < H$. For each fixed s,

$$\frac{1}{H}\sum_{t=0}^{H-1} b(t+s) = 1 - \frac{|A \cap [1, H]|}{H} + O_N(H^{-1}).$$

Choose $H \to \infty$ along a sequence realizing $\overline{d}(A) = \alpha$. Because each P has k elements, (4) yields

$$mk(1-\alpha) \geq m + k - 1.$$

Consequently,

$$\boxed{\overline{d}(A) \leq 1 - \frac{1}{k} - \frac{k-1}{km}.} \tag{5}$$

Take

$$\eta(D, k) = \frac{k-1}{km} > 0.$$

This proves the proposition. □

The result handles natural **upper** density directly, so no existence of an ordinary density is assumed.

For comparison, without any 2-largeness hypothesis, a fixed $d \geq 1$ already gives a k-term progression in every set with upper density **strictly greater** than $1 - 1/k$. The proposition includes the endpoint and gives a further D-dependent improvement. It does **not**, however, keep the threshold away from 1 as $k \to \infty$.

3. An exact extremal formulation of your question

A useful way to isolate the remaining issue is to define

$$a_k(D) = \sup\left\{\overline{d}(A) : A \subseteq \mathbb{N} \text{ contains no } k\text{-term AP with difference in } D\right\}. \tag{6}$$

These numbers are nondecreasing in k: avoiding a shorter progression is a stronger restriction than avoiding a longer one.

The proposition proves

$$\boxed{D \text{ 2-large} \quad \Longrightarrow \quad a_k(D) < 1 - \frac{1}{k} \qquad (k \geq 2).} \tag{7}$$

Your question is equivalent to asking whether

$$\boxed{\sup_{\substack{D \subseteq \mathbb{N} \\ D \text{ 2-large}}} \sup_{k \geq 2} a_k(D) < 1.} \tag{8}$$

Indeed, if the supremum in (8) is $c < 1$, choose any $\delta_0 \in (c, 1)$. No set of upper density δ_0 can avoid a prescribed length, so every 2-large D is δ_0-density large.

Conversely, suppose (2) holds. An avoiding set of upper density greater than δ_0 could be thinned to upper density exactly δ_0, contradicting (2). Thus

$$a_k(D) \leq \delta_0$$

for every 2-large D and every k, proving (8).

This formulation makes the limitation of (7) particularly clear: **strict inequalities at every individual length do not prevent the supremum over lengths from being 1.**

It also clarifies what a negative answer requires. One would need

$$D_j \text{ 2-large}, \qquad A_j \text{ avoiding a } k_j\text{-term } D_j\text{-AP}, \qquad \overline{d}(A_j) \longrightarrow 1. \tag{9}$$

The sets D_j may depend on j. Constructing one fixed 2-large set failing every density would be stronger than necessary.

4. What the related results do—and do not—supply

The quadratic example is not a counterexample

Our previous measurably recurrent set

$$R_\alpha = \{n : \|\alpha n^2\| \geq 1/4\}$$

fails every δ-density condition, but it is **not 2-large**. This follows directly from your draft's two-color obstruction, applied to the special generalized polynomial $P(n) = \alpha n^2$. [Pasted text.txt, L61-L74]

Likewise, the modified Alweiss construction was designed to preserve nilrecurrence; we did not establish 2-largeness for it. Nilrecurrence cannot replace the hypothesis in your present question. Your manuscript itself distinguishes its nilrecurrence conclusion from full topological multiple recurrence. [Pasted text.txt, L286-L292]

Kříž's separation does not automatically answer it

Kříž-type examples separate **single topological recurrence** from measurable recurrence. Griesmer's exposition and extension concerns precisely that separation. It does not assert that the constructed sets are 2-large. arXiv

Thus one cannot conclude "no" merely by citing topological-but-not-measurable recurrence.

The 2-large-is-large conjecture is related, but not the same statement

The Brown–Graham–Landman conjecture asks whether

$$2\text{-large} \implies r\text{-large for every } r.$$

That is the conjecture stated in your manuscript. [Pasted text.txt, L93-L104]

“Large” corresponds to multiple topological recurrence, equivalently to finding the prescribed progressions in every syndetic set. This is the content of the standard topological recurrence equivalences. arxiv.org

An affirmative answer to that conjecture would therefore still need an additional argument to give your fixed-density conclusion. Conversely, a conclusion at a density such as 0.99 would not immediately yield arbitrary-color largeness: an r-coloring only guarantees a cell of upper density at least $1/r$.

There is a genuine positive result for a stronger covering hypothesis on A

If

$$A \cup (A - h) = \mathbb{N} \tag{10}$$

for some fixed $h \geq 1$, then every 2-large D gives arbitrarily long D-progressions in A.

To see this, color by A and A^c. A monochromatic progression in A is already suitable. If it lies in A^c, condition (10) says that translating the entire progression by h places it in A, without changing its difference.

For $h = 1$, this says that every 2-large set works for every set meeting each pair of consecutive integers. It is the $r = 2$ case of Host–Kra–Maass’s implication from r-largeness to r-syndetic largeness. arXiv

But high upper density does not ensure a two-translate covering property such as (10).

Conclusion

The definite implication obtained here is

$$\boxed{D \text{ 2-large} \implies D \text{ } (1 - 1/k, k)\text{-density large for every } k \geq 2,}$$

with a strict improvement at each fixed D, k.

Replacing these length-dependent thresholds by one absolute $\delta_0 < 1$ is the unresolved step in my investigation. The sources checked supply chromatic, syndetic, and nilpotent recurrence implications, but none supplies the uniform bound (8); and none of our preceding counterexamples has the required 2-largeness.

---

67. User

2026-09-16 17:18:23 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-16T09:18:23.422000Z

ryan thinks that our paper on this topic should just focus on the less general qualitative result rather than the quantitative result. He thinks that not many people will be interested in the quantitative generalization, so it makes the paper more technical and harder to read, which will decrease the chances that it gets accepted into a top journal. Do you agree or disagree? Why?

68. Assistant

2026-09-16 17:18:44 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-16T09:18:44.016748Z

I'll compare what the quantitative statement adds mathematically with what it costs the reader, especially in the shorter proof.

69. Assistant

2026-09-16 17:21:47 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-16T09:21:47.949171Z

The shorter proof changes that tradeoff: the quantitative rates emerge directly from the bounded polynomial lift, so deleting them may simplify the statement more than the argument.

70. Assistant

2026-09-16 17:24:39 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-16T09:24:39.926485Z

I largely agree with Ryan about what should lead the paper. I do not think that automatically means deleting the quantitative result. My recommendation is:

> **Make the qualitative nilrecurrence theorem the main result, choose the simplest rigorous proof, and retain the quantitative estimate only if it can be presented as a short consequence without making that proof harder to understand.**

For the **shorter elementary argument you shared**, I lean toward keeping the quantitative conclusion in a subordinate position. For the **longer factorization-based manuscript**, I am much more sympathetic to Ryan's proposal to simplify substantially.

The important distinction is between **the mathematical value of a stronger theorem** and **whether organizing the paper around it makes a stronger submission**.

1. Why I agree with Ryan about the main emphasis

Your qualitative theorem has a clear conceptual message:

$$\boxed{\text{2-large} \quad \Longrightarrow \quad \text{multiple pointwise recurrence for every nilsystem.}}$$

It connects a binary Ramsey property to recurrence throughout the finite-step nilpotent hierarchy. It directly addresses Conjecture 3.6 in Ryan's *New Obstacles to Multiple Recurrence*, and your introduction explains how it answers the question of Host–Kra–Maass with two colors. That gives the reader a reason to care before learning anything about weighted matrices or quantitative factorization. arXiv [Pasted text.txt, L128-L154]

In contrast, appreciating the full quantitative theorem requires understanding several additional distinctions: $(2, k)$-largeness rather than 2-largeness, special generalized polynomials rather than arbitrary normalized generalized polynomials, assigned weights rather than actual degrees after simplification, and uniformity in coefficients. All of these are mathematically legitimate, but they raise the entry cost. The longer draft introduces precisely this additional framework before stating the quantitative theorem. [Pasted text.txt, L171-L203]

My judgment is that **the qualitative result currently has the stronger immediate claim to broad mathematical interest**. The quantitative result has to explain what its additional precision enables.

That is relevant to your journal goal. For example, *Forum of Mathematics, Pi* explicitly emphasizes interest to a broad cross-section of mathematicians, while JEMS states that it selects papers for outstanding quality and interest. Those policies do not decide your case, but they support thinking about conceptual significance and accessibility, rather than simply maximizing the strength of the displayed theorem. Cambridge University Press EMS Press

I would therefore not advertise the paper as two equally weighted projects—one on nilrecurrence and another on quantitative approximation—unless the second has an independently compelling application.

2. Why the quantitative theorem is valuable, but may not justify substantial extra machinery

I would not go as far as "not many people will be interested" as a confident prediction. The theorem has genuine features worth preserving.

The conclusion

$$\|P_i(s)\| \le C(m,D)k^{-d_i}$$

is not merely the qualitative theorem with an unnecessary constant written down. It works under a hypothesis at **one finite progression length**, its constants are **independent of the real coefficients**, and the stronger formulation gives **one coloring that works simultaneously for every sufficiently large length**. These are substantive uniformity statements. [Pasted text.txt, L189-L203]

For instance, a reader studying finite Ramsey obstructions could use the theorem without knowing that their allowed difference set is 2-large. That is a different use from invoking the qualitative nilrecurrence theorem.

Nevertheless, there are important limitations to how I would promote it.

First, the parameter being controlled is **progression length**, not the size of the approximating integer:

$$\|P_i(s)\| \ll k^{-d_i}$$

does not give a bound on s. Your draft correctly says this explicitly. It should therefore not be presented as though it supplies an effective bound for how far one must search for a nilrecurrence time. [Pasted text.txt, L206-L214]

Second, the draft leaves the dependence of $C(m,D)$ and $k_0(m,D)$ as a further question and does not claim optimality of the exponents. That is entirely acceptable mathematics, but it means the paper is not currently offering a sharp quantitative theory in the strongest usual sense of that phrase. [Pasted text.txt, L1624-L1635]

Third, **I do not see in the supplied drafts a major application that specifically needs the powers k^{-d_i}, rather than sufficiently small error**. The displayed proof of qualitative nilrecurrence uses the structural theorem and a slow-motion contradiction, while the localization consequence uses the qualitative nilrecurrence theorem itself. Neither deduction needs the special-generalized-polynomial rate. [Pasted text.txt, L1372-L1430] [Pasted text.txt, L1513-L1529]

That is probably the strongest version of Ryan's concern:

> **Why should every reader pay a substantial technical cost for a refinement whose main additional uses are not demonstrated in the paper?**

I think that is a fair question. “The statement is stronger” is not, by itself, a sufficient answer.

3. But the shorter proof changes the calculation substantially

Here is where I would push back against removing the quantitative result automatically.

In the shorter elementary draft, the rates appear to be a natural output of the argument, not the reason the argument is technical.

The proof already constructs a polynomial matrix $Q(t)$, shows that its coordinates remain bounded on a long interval $[0, J]$, and extracts their weighted leading coefficients. The relevant estimate is

$$|[t^w]Q_{ij}(t)| \leq CJ^{-w}.$$

The shorter draft then clears denominators and reads off the nested brackets. These steps yield the stated rates directly. [Pasted text.txt, L240-L280] [Pasted text.txt, L310-L349]

The underlying scalar fact is elementary. If

$$q(t) = c_0 + c_1 t + \cdots + c_w t^w$$

is bounded on $[0, J]$, then rescaling $t = Ju$ and interpolating on $[0, 1]$ gives

$$|c_w| \ll_w J^{-w}.$$

For the qualitative conclusion, one uses the fact that this coefficient tends to zero. For the quantitative conclusion, one retains the displayed power.

At that point, suppressing the rate does not remove the main mathematical work.

You still need the polynomial lift, the boundedness argument, the elimination of the progression’s starting point, and the arithmetic treatment of the lattice corrections. The short draft itself identifies these as the mechanisms making the proof work. [Pasted text.txt, L351-L355]

So I would ask a very concrete question when comparing versions:

> **Which definitions, lemmas, and arguments actually disappear when we remove the quantitative theorem?**

If the answer is “a substantial section, several new parameters, and a more complicated proof of the main lemma,” Ryan’s case is strong.

If the answer is “mostly the J^{-w} notation and a final theorem statement,” then I would keep the result. Removing useful information is not an expository improvement when the reader has already paid almost the entire cost of obtaining it.

This is a judgment about the architecture of the supplied proof, not a substitute for checking every step of it.

4. There are really three choices here, not two

I think your discussion will be more productive if you separate three layers:

1. **The qualitative nilrecurrence theorem.**
2. **The finite coefficient-uniform approximation theorem.**
3. **The general structural theorem for monochromatic polynomial orbits on arbitrary fixed filtered nilmanifolds.**

The longer manuscript proves the third and uses it to derive the first two. Its infrastructure includes rational-subgroup complexity, robust colorings, rational conjugation, and intrinsic-to-ambient metric comparisons. Your introduction describes these dependencies explicitly. [Pasted text.txt, L234-L284]

The shorter manuscript instead proves the special-generalized-polynomial obstruction directly through elementary matrix algebra, finite differences, interpolation, compactness, and the shell coloring. It still obtains the rates. [Pasted text.txt, L46-L74]

Thus you do **not** have to choose between:

> a broad quantitative paper with substantial nilmanifold machinery,

and

> a short paper that contains no quantitative statement.

There is a third possibility:

> **a qualitative-first paper using the elementary proof, with the finite estimate recorded as a concise additional consequence.**

That is my preferred option based on the drafts here.

There is also an important qualification in the other direction: in the long manuscript, not all the quantitative machinery is present solely for the quantitative corollary. It also supports the qualitative proof. Simply deleting the finite-approximation section would not automatically produce a much simpler nilrecurrence proof. The simplification has to happen at the level of the argument, not just the theorem list.

5. What I would actually submit

I would organize the paper so that a reader can understand its main contribution, and then follow a complete proof of it, without first committing to the quantitative generalization.

Title and abstract. I would use a title centered on 2-largeness and nilrecurrence. The abstract should emphasize the qualitative theorem, the question it resolves, and the proof's main idea. At most, I would end with one sentence saying that the argument also yields coefficient-uniform finite approximation bounds. I would not put the full weighted formula in the abstract unless you decide that the quantitative theorem is independently central.

Introduction. I would explain the qualitative problem first. Then give a proof overview that highlights the actual insight: after passing to a short subprogression, the lattice corrections become polynomial; the shell coloring prevents a nonconstant polynomial excursion; leading coefficients remove the starting point. That is a coherent explanation of the elementary proof without immediately making the reader track all the constants. [Pasted text.txt, L351-L353]

Main proof. I would use the weakest formulation of each lemma that makes the argument transparent—but not artificially conceal a simple estimate that is already being proved. A displayed coefficient bound can be easier to understand than a vague "sufficiently small" statement with several implicit dependencies.

Quantitative conclusion. I would put it in a short final subsection, or an appendix if that genuinely separates optional material. State precisely the coefficient-uniformity and the fixed-coloring feature, give the deduction, and stop. I would not add a substantial discussion of arbitrary generalized-polynomial syntax, expression complexity, or further quantitative problems unless those are needed for the paper's main purpose.

An appendix is not a magic solution, however. If proving its theorem forces the main text to carry much more general machinery, then the technical cost has not really been isolated. In that situation I would omit the generalization from this submission.

Nor would I automatically split it into a second paper. A refinement that is a short consequence of the same argument may not justify a separate publication. Its appropriate home may simply be a brief theorem at the end of the main paper.

6. What I can—and cannot—say about acceptance prospects

I think Ryan's proposed emphasis is likely to improve the paper's presentation. I cannot reliably predict whether removing the quantitative theorem itself will improve its acceptance probability.

My expectation is that the important questions will be:

Is the qualitative theorem significant enough for the journal? Is the proof genuinely insightful? Is that insight presented in a form that other mathematicians can understand and use?

The answer is not determined by whether the strongest statement contains an explicit power of k.

I would also resist the converse temptation: adding a technically stronger theorem is not a reliable way to make a paper qualify for a more selective journal. If the qualitative result is judged insufficiently consequential, a refinement without a demonstrated application may not change that judgment. Conversely, a short optional quantitative consequence need not detract from an otherwise focused paper.

The strongest case for retaining the quantitative theorem would be either **a substantial application that genuinely uses it**, or **the fact that it comes almost for free from a particularly elegant proof**. In your shorter draft, I think the second is the more convincing argument.

My verdict

On the paper's identity, I agree with Ryan: this should primarily be a paper about 2-largeness implying nilrecurrence, not a paper about quantitative generalized-polynomial approximation.

On deleting the quantitative result altogether, I lean against deletion in the elementary version, because the estimate is already built into the proof. I would demote it rather than let it dictate the exposition.

The proposal I would put to Ryan is:

> "Let us make the qualitative theorem the sole headline and optimize the proof for clarity. We will keep the finite estimate only to the extent that it follows from that proof with negligible additional machinery."

That preserves his strongest editorial point without discarding mathematical information unnecessarily. **The goal should be the clearest paper that conveys the important insight—not the weakest theorem, and not the longest list of stronger ones.**

---

71. User

2026-09-16 17:28:27 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-16T09:28:27.752000Z

can you find any noteworthy applications of the quantitative result?

72. Assistant

2026-09-16 17:28:45 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-16T09:28:45.420020Z

I'll focus on applications that genuinely exploit the finite-length rates or coefficient-uniformity, separating them from consequences already obtainable from qualitative nilrecurrence.

73. Assistant

2026-09-16 17:31:49 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-16T09:31:49.684255Z

Two promising directions emerge: exact homogeneity of rounded expressions on blocks proportional to k, and quantitative localization forcing differences into shrinking generalized-polynomial neighborhoods.

74. Assistant

2026-09-16 17:44:05 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-16T09:44:05.733631Z

Yes. I found two applications that I think are genuinely worth considering for the paper: a finite extremal-transfer theorem, and a strong finite form of multiple nilrecurrence in which the number of simultaneous returns grows polynomially with the Ramsey parameter. The first has a very short proof; the second uses coefficient-uniformity in a more substantial way.

These are deductions from your stated quantitative theorem, not claims that I have located an independent major open problem that it resolves. I checked related primary sources, but I have not established the publication novelty of every corollary below.

The main lesson is that **the useful feature is not just the displayed exponent k^{-d_i}. It is that a finite Ramsey hypothesis controls a shrinking target, uniformly in coefficients that may themselves depend on k.** That permits applications unavailable from the qualitative theorem without losing control of the progression length.

1. A finite Szemerédi theorem with Diophantine constraints at a factor-two cost

This is the application I would most readily include.

Write

$$r_k(N) = \max\{|A| : A \subseteq [N] \text{ contains no } k\text{-term arithmetic progression}\}.$$

Fix special generalized polynomials $P_1, \dots, P_m$ of assigned weights $d_i \le D$. Your theorem supplies constants C, k_0, depending only on m, D, and one binary coloring such that every monochromatic k-term progression, $k \ge k_0$, has difference d satisfying

$$\|P_i(d)\| \le Ck^{-d_i}.$$

The fact that this is a statement about **every monochromatic progression for a particular coloring**, rather than merely about 2-large sets, is crucial. [Pasted text.txt, L61-L74]

Define

$$B_k = \{d \in \mathbb{N} : \|P_i(d)\| \le Ck^{-d_i} \text{ for all } i\},$$

and

$$r_k(N; B_k) = \max\left\{|A| : \begin{array}{l} A \subseteq [N], \\ A \text{ contains no } k\text{-term AP with difference in } B_k \end{array}\right\}.$$

Corollary: extremal transfer

For every $k \ge k_0$ and every N,

$$\boxed{r_k(N) \le r_k(N; B_k) \le 2r_k(N).} \tag{1}$$

Thus **restricting the common difference to the simultaneous shrinking-target set increases the extremal number by at most a factor of two**.

Proof

The lower bound is immediate: a set containing no k-term progression certainly contains no such progression with difference in B_k.

For the upper bound, let $c : \mathbb{N} \to \{0, 1\}$ be your obstruction coloring. Suppose $A \subseteq [N]$ contains no k-term progression with difference in B_k. Partition

$$A = A_0 \sqcup A_1, \qquad A_j = A \cap c^{-1}(j).$$

Each A_j contains no k-term progression at all. Otherwise that progression would be monochromatic for c, forcing its difference into B_k, a contradiction. Hence

$$|A| = |A_0| + |A_1| \le 2r_k(N). \qquad \square$$

A concrete formulation

There are absolute constants C, k_0 such that, for all real numbers $\alpha, \beta, \gamma, \theta$, every $A \subseteq [N]$ satisfying

$$|A| > 2r_k(N), \qquad k \ge k_0,$$

contains a progression

$$a, a + d, \ldots, a + (k - 1)d$$

for which

$$\boxed{\|\alpha d\| \le \frac{C}{k}, \qquad \|\beta d^2\| \le \frac{C}{k^2}, \qquad \|\gamma d[\theta d]\| \le \frac{C}{k^2}.} \tag{2}$$

Here $[\,\cdot\,]$ denotes nearest-integer rounding.

More generally, if $N_{\mathrm{Sz}}(k, \eta)$ is any valid finite Szemerédi threshold at density η, then

$$\boxed{N \ge N_{\mathrm{Sz}}(k, \eta/2)} \tag{3}$$

suffices to find the constrained progression in every $A \subseteq [N]$ of density at least η, once $k \ge k_0(m, D)$.

The density loss is exactly a factor of two. It does not grow with the number of constraints, their degrees, or their coefficients. The complexity still enters C and the lower threshold k_0.

Why this actually uses the quantitative theorem

The qualitative result gives the following weaker finite conclusion: for fixed polynomials and fixed tolerances, some binary coloring forces the desired difference condition on progressions of some length

$$L = L(P_1, \dots, P_m, \varepsilon_1, \dots, \varepsilon_m).$$

The same argument then gives a bound involving

$$2r_{\max(k,L)}(N),$$

rather than $2r_k(N)$.

Your quantitative result allows the tolerances to shrink with k **while keeping the extremal function at the same progression length** k. That is the substantive parameter improvement.

I would describe (1) as a **finite Szemerédi transfer principle for shrinking generalized-polynomial neighborhoods**. It is simple enough to explain in an introduction and short enough not to burden the paper.

2. The same transfer works in sparse ambient sets, including the primes

The preceding proof never used anything special about $[N]$.

For any finite set $\Omega \subseteq \mathbb{N}$, define

$$r_k(\Omega) = \max\{|A| : A \subseteq \Omega \text{ contains no } k\text{-term AP}\}.$$

Then precisely the same argument proves

$$\boxed{r_k(\Omega; B_k) \le 2r_k(\Omega).} \tag{4}$$

This means that **any finite dense-subset theorem for arithmetic progressions in an ambient set transfers to your constrained differences at a factor-two density cost**.

For example, Green–Tao's relative-density theorem says that every subset of the primes of positive relative upper density contains arbitrarily long arithmetic progressions. arXiv Combining it with your coloring gives:

> If $\mathcal{A}$ is a subset of the primes of positive relative upper density, then, for every sufficiently large k, it contains a k-term progression
>
> $$p, p + d, \dots, p + (k-1)d$$
>
> satisfying
>
> $$\|P_i(d)\| \le Ck^{-d_i} \qquad (1 \le i \le m).$$

The finite statement is stronger and more relevant to your question: **whatever finite threshold one has for k-term progressions in relative density $\eta/2$ in the primes also works at relative density η with these difference constraints**, without introducing an additional coefficient-dependent threshold.

There is an important editorial qualification. The *bare existence* of prime progressions with arbitrarily accurate fixed Diophantine constraints can already be deduced using qualitative results and a sufficiently long auxiliary progression. I would therefore **not** advertise this as an independent breakthrough concerning the primes. The noteworthy feature is the finite parameter preservation in (4).

This is also not a result about d itself being prime. The prime variables are the terms of the progression.

3. Finite Ramsey localization without increasing the progression length

There is a similarly clean application on the coloring side.

Corollary: quantitative localization

With B_k as above, for every $r \geq 1$,

$$\boxed{S \text{ is } (2r,k)\text{-large} \quad \Longrightarrow \quad S \cap B_k \text{ is } (r,k)\text{-large},} \tag{5}$$

provided $k \geq k_0$.

Proof

Given an r-coloring c_1, combine it with your binary obstruction coloring c_0. A monochromatic k-term progression for

$$n \longmapsto (c_0(n), c_1(n))$$

has difference in S by the $(2r,k)$-largeness hypothesis, and in B_k by the choice of c_0. It is monochromatic for c_1. □

Your long manuscript already records qualitative nil-Bohr localization. The additional point here is that the neighborhood is shrinking with k, and **the input and output progression lengths are identical**. [Pasted text.txt, L1513-L1537]

In terms of van der Waerden numbers, define $W(r,k)$ using r colors and progression length k. Then every r-coloring of

$$[W(2r,k)]$$

contains a monochromatic k-term progression satisfying all the inequalities

$$\|P_i(d)\| \leq Ck^{-d_i}. \tag{6}$$

So the extra Diophantine conditions require **one additional binary coordinate in the coloring**, not a number of colors growing as the target shrinks.

This contrasts with the earlier dimension-dependent Bohr-localization statement of Host–Kra–Maass, whose Proposition 7.7 uses a factor 2^d for a d-dimensional Bohr neighborhood. The dimension-free qualitative improvement belongs to your existing story; equation (5) supplies its finite shrinking-target version. arXiv

I would put (1) and (5) together in one short subsection. They are two manifestations of the same useful principle, rather than material for two long application sections.

4. A more substantial application: polynomially many simultaneous returns from a finite hypothesis

Here is the consequence that I find most interesting dynamically.

Theorem: growing families of shrinking-target returns

Fix positive integers m, D, u, v. There are constants C, k_0, depending only on these integers, such that the following holds.

If $k \geq k_0$, S is $(2,k)$-large, and $P_1, \dots, P_m$ are special generalized polynomials of weights at most D, then some $d \in S$ satisfies

$$\boxed{\max_{\substack{1\leq i\leq m\\ 1\leq t\leq k^u}} \|P_i(td)\| \leq Ck^{-v}.} \tag{7}$$

Thus a finite Ramsey hypothesis at length k supplies ** k^u consecutive multiples of one allowed difference**, all meeting simultaneous targets of radius $O(k^{-v})$. The powers u, v can be prescribed in advance.

The dependence on u, v is hidden in the constants and threshold; this is not a uniform assertion as those exponents tend to infinity.

There are two ingredients in the proof. Both are worth spelling out, because simply applying your theorem to the mk^u expressions $P_i(tn)$ would not prove (7): the number of tests would grow with k, destroying the required parameter control.

4.1. Coefficient-uniformity amplifies the approximation rate

First, your theorem implies:

> For every fixed positive integer R, finitely many special generalized polynomials of bounded weight can be approximated simultaneously at rate $O(k^{-R})$ under the $(2,k)$-largeness hypothesis.

This uses the integer-multiplier observation we discussed earlier, but here it has a concrete purpose.

Fix a family $F_1, \dots, F_s$. Apply the quantitative theorem simultaneously to

$$F_i,\ qF_i,\ q^2F_i, \dots, q^{R-1}F_i \qquad (1 \leq i \leq s), \tag{8}$$

where

$$q = \left\lfloor \frac{k}{4C_0} \right\rfloor$$

and $C_0 \geq 1$ is a uniform constant for sR tests of the relevant maximum weight.

This is legitimate because multiplying a special generalized polynomial by an integer only changes its outermost real coefficient. The assigned weight does not change. The constants in your theorem are independent of those coefficients. [Pasted text.txt, L182-L203]

For sufficiently large k, we obtain one $d \in S$ such that

$$\|q^j F_i(d)\| \leq C_0/k \qquad (0 \leq j < R).$$

Write

$$q^j F_i(d) = a_j + e_j, \qquad a_j \in \mathbb{Z}, \quad |e_j| \leq C_0/k.$$

Then

$$e_{j+1} - qe_j \in \mathbb{Z},$$

but

$$|e_{j+1} - qe_j| \leq (q+1)C_0/k < 1.$$

Hence

$$e_{j+1} = qe_j.$$

It follows that

$$\begin{aligned}\|F_i(d)\| &\le \frac{C_0}{kq^{R-1}} \\ &\le C_0(8C_0)^{R-1}k^{-R}.\end{aligned} \tag{9}$$

The number of expressions in (8) is fixed. Only their coefficients depend on k.

A qualification: this argument does not automatically produce one coloring working for all k, because q depends on k. It proves the amplified finite statement, and for each fixed k it supplies a suitable binary coloring. Your original fixed-coloring conclusion remains a distinct strength.

4.2. Approximate the inner expressions so that rounding becomes homogeneous

For an ordinary monomial,

$$P(td) = t^{\deg P}P(d),$$

so sufficiently accurate approximation at d immediately controls many multiples.

For a nested expression, this identity is not automatic. The remedy is to approximate all its nested suffixes.

For example, let

$$P(n) = \alpha n[\beta n].$$

If

$$\|\beta d\| \le \xi, \qquad \|P(d)\| \le \xi,$$

and $t\xi < 1/2$, then

$$[\beta td] = t[\beta d].$$

Consequently,

$$P(td) = t^2P(d), \qquad \|P(td)\| \le t^2\xi. \tag{10}$$

The same reasoning works recursively. For

$$P(n) = a_1n^{w_1}[a_2n^{w_2}[\cdots[a_\ell n^{w_\ell}]\cdots]],$$

let

$$Q_j(n) = a_jn^{w_j}[a_{j+1}n^{w_{j+1}}[\cdots[a_\ell n^{w_\ell}]\cdots]]$$

be its suffixes, with

$$Q_\ell(n) = a_\ell n^{w_\ell}.$$

Their weights are

$$e_j = w_j + \cdots + w_\ell \le D.$$

If

$$t^{e_j}\|Q_j(d)\| < 1/2$$

for every suffix, downward induction gives

$$Q_j(td) = t^{e_j} Q_j(d). \tag{11}$$

Indeed, once this holds for Q_{j+1}, the small-error condition gives

$$[Q_{j+1}(td)] = t^{e_{j+1}}[Q_{j+1}(d)],$$

which proves the next identity.

For the original m polynomials, there are at most mD suffixes. Apply (9) to all of them with

$$R = uD + v.$$

Then

$$\|Q_j(d)\| \le C_1 k^{-(uD+v)}.$$

For $1 \le t \le k^u$,

$$t^{e_j}\|Q_j(d)\| \le C_1 k^{-v} < 1/2$$

once k is sufficiently large. Thus all the rounding identities stabilize, and

$$\|P_i(td)\| \le t^{d_i}\|P_i(d)\| \le C_1 k^{-v}.$$

This proves (7). □

Why this is more than a restatement

The output simultaneously has a growing number of return times and a shrinking error. More importantly, **the hypothesis only concerns one finite progression length**.

The qualitative theorem says that a 2-large set supplies any prescribed finite collection of returns. It does not provide the relation

$$\text{input Ramsey length } k \quad \rightsquigarrow \quad k^u \text{ returns at accuracy } k^{-v}.$$

Equation (7) is a precise use of both coefficient-uniformity and the rates, with a fixed number of auxiliary tests.

5. A genuine nilsystem formulation of the preceding application

The previous theorem yields a clean statement without mentioning generalized polynomials.

Corollary: polynomial amplification of finite nilrecurrence

Fix a nilsystem (X, T), a point $x \in X$, a neighborhood U of x, and a positive integer u. There exists k_0 such that

$$\boxed{\begin{aligned} &k \ge k_0, \quad S \text{ is } (2,k)\text{-large} \\ \Longrightarrow \quad &\exists d \in S: \quad T^{td}x \in U \quad \text{for every } 1 \le t \le k^u. \end{aligned}} \tag{12}$$

Proof

The Huang–Shao–Ye characterization supplies finitely many special generalized polynomials P_i and a tolerance $\varepsilon > 0$ such that

$$\{n : \max_i \|P_i(n)\| < \varepsilon\} \subseteq \{n : T^n x \in U\}. \tag{13}$$

This follows from their Theorem B together with Theorem 4.2.11, which identifies the families generated by general and special generalized-polynomial neighborhoods. arXiv

Apply (7), for example with $v = 1$, and take k large enough that $C/k < \varepsilon$. Then (13) holds at every td, $1 \le t \le k^u$. □

This corollary has the same conceptual vocabulary as your headline theorem. It says:

> **For a fixed nilsystem and return neighborhood, finite binary Ramsey richness at length k forces a string of nilreturns of any prescribed polynomial length in k.**

The neighborhood is fixed in (12). I am not asserting a coefficient-uniform metric shrinking-target theorem for arbitrary nilmanifold presentations; obtaining such a formulation would require additional quantitative control of the representation in (13).

Also, only d is required to belong to S. The multiples $2d, 3d, \dots$ need not belong to S.

For your paper, I find (12) a more compelling dynamical application than simply adding another example of simultaneous fractional-part approximation.

6. The stronger recurrence condition also inherits the factor-two transfer

There is a useful way to combine the applications.

For fixed u, v, define

$$B_k^{u,v} = \left\{ d : \max_{\substack{1 \le i \le m \\ 1 \le t \le k^u}} \|P_i(td)\| \le Ck^{-v} \right\}.$$

The proof of (7), applied at the coloring level, gives a binary coloring for each fixed k whose monochromatic k-term progressions have differences in $B_k^{u,v}$.

Consequently,

$$\boxed{r_k(\Omega; B_k^{u,v}) \le 2r_k(\Omega)} \tag{14}$$

for every finite ambient set Ω, and

$$\boxed{S \text{ is } (2r,k)\text{-large} \Longrightarrow S \cap B_k^{u,v} \text{ is } (r,k)\text{-large.}} \tag{15}$$

Thus the factor-two cost can enforce not merely a few small fractional parts at d, but **polynomially many simultaneous conditions at its multiples**.

The constants depend on the prescribed powers and polynomial complexity, but not on the coefficients. The coloring in this amplified formulation may depend on k.

7. What I would—and would not—claim about these applications

I would not promote the theorem as improving the best unrestricted bounds for small fractional parts of ordinary polynomials. Those results measure the approximation against the size of the search interval. For example, Maynard proves coefficient-uniform bounds of the form

$$\min_{1\le n\le N}\max_{1\le i\le m}\|f_i(n)\| \ll_{m,D} N^{-c_D/m}$$

for ordinary polynomials of degree at most D with zero constant term. Your theorem, by itself, does not supply an improvement on such denominator bounds. arXiv

Nor would I sell fixed-neighborhood localization alone as an application requiring the rate: your qualitative theorem already gives that. The distinctions worth emphasizing are **same-length finite transfer**, **shrinking targets**, and **a growing number of simultaneous returns**.

These deductions therefore strengthen the case for retaining a short quantitative subsection, but they do not establish that the paper should be reorganized around a long quantitative theory.

My recommendation to you and Ryan would be to consider including the extremal inequality

$$r_k(N;B_k)\le 2r_k(N)$$

and the finite nilrecurrence amplification

$$(2,k)\text{-large} \Longrightarrow k^u \text{ consecutive multiples return to a prescribed nil-neighborhood.}$$

The first is an accessible additive-combinatorial application with an almost immediate proof. The second gives a genuine dynamical payoff for the coefficient-uniform quantitative statement. Together, they provide a stronger justification for keeping the quantitative result than the observation that it is simply a refinement of qualitative nilrecurrence.

---

75. User

2026-09-16 17:56:26 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-16T09:56:26.056000Z

These looks like interesting applications. Please preparing a pdf containing these applications and their proofs so that I can share it with ryan.

---

76. Assistant

2026-09-16 17:56:56 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-16T09:56:56.627279Z

I'll prepare a shareable PDF with detailed proofs, clearly separating deductions from your quantitative theorem from the external results used for nilsystems and primes.

---

77. Assistant

2026-09-16 18:00:00 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-16T10:00:00.294130Z

I'll detail the multiplier-and-rounding amplification, emphasizing that it preserves length k with finitely many tests, while noting the amplified coloring may depend on k.

78. Assistant

2026-09-16 18:08:19 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-16T10:08:19.100464Z

Download the PDF — Applications of the quantitative two-color theorem (16 pages)

Download the standalone LaTeX source

The note states the precise **coefficient-uniform, single-coloring theorem** from your drafts as its input, then gives detailed proofs of the applications rather than reproducing the underlying theorem's proof. [Pasted text.txt, L68-L74] [Pasted text.txt, L189-L203]

The two principal conclusions are the **same-length extremal transfer**

$$r_k(\Omega) \le r_k(\Omega; B_k) \le 2r_k(\Omega)$$

and the **growing family of simultaneous approximations**

$$S \text{ is } (2,k)\text{-large} \quad \Longrightarrow \quad \exists d \in S : \max_{\substack{1\le i\le m\\ 1\le t\le k^u}} \|P_i(td)\| \le Ck^{-v}.$$

It also includes the finite Szemerédi formulation, Ramsey localization, the relative-density application in the primes, and the deduction of polynomial-length strings of returns to a fixed nilsystem neighborhood.

For the growing-return argument, I have written out the integer-multiplier lemma, the exact stabilization of nested rounding, worked examples explaining why the inner expressions must be controlled, and the complete order of parameter choices.

Two important distinctions are explicit throughout: **the amplified coloring may depend on** k, unlike the original coloring; and the nilsystem application concerns a **fixed neighborhood**, not a metric shrinking-target estimate uniform over nilmanifolds. The external results are stated and referenced separately, and the note makes no claim of publication novelty for these corollaries.

---

79. User

2026-09-17 10:43:42 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-17T02:43:42.393000Z

If we use the result that if A is delta density large for some delta in (0,1), then it is a set of multiple recurrence for nilrotations, can we deduce this quantitative bound r(Omega) <= r(Omega,B) <= 2r(Omega), but with 2 replaced by 1+epsilon for arbitrary epsilon > 0?

---

80. Assistant

2026-09-17 10:44:07 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-17T02:44:07.510468Z

I'll check whether the density argument yields a $1+\varepsilon$ factor at the same progression length, paying particular attention to the shrinking-neighborhood issue.

81. Assistant

2026-09-17 10:47:12 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-17T02:47:12.236806Z

Yes: both fixed nil-Bohr neighborhoods and the quantitative shrinking targets admit the $1+\varepsilon$ improvement, without enlarging the target constant; only the required minimum progression length increases.

82. Assistant

2026-09-17 10:48:55 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-17T02:48:55.918400Z

Yes. For a fixed nil-Bohr neighborhood B, your fixed-density theorem gives the factor $1+\varepsilon$ by a short translation-averaging argument. Moreover, the quantitative omitted-color theorem proved in the fixed-density note gives the improvement for the **shrinking sets** B_k from the applications note—even with the original approximation constant unchanged.

The important qualification is that **the minimum progression length may depend on** ε. Writing the progression length explicitly, the fixed-neighborhood conclusion is

$$\boxed{\forall \varepsilon>0\ \exists k_0(B,\varepsilon)\ \forall k\ge k_0\ \forall \text{ finite } \Omega\subseteq\mathbb{N}:\quad r_k(\Omega)\le r_k(\Omega;B)\le (1+\varepsilon)r_k(\Omega).} \tag{1}$$

Here is the complete deduction, followed by the stronger shrinking-target version.

1. The general principle: average translates of a high-density test set

Recall the notation from the applications note:

$$r_k(\Omega)=\max\{|E|: E\subseteq\Omega \text{ contains no } k\text{-term AP}\},$$

and

$$r_k(\Omega;B)=\max\left\{|E|: E\subseteq\Omega \text{ contains no } k\text{-term AP with difference in } B\right\}.$$

These are the extremal quantities in the earlier factor-two comparison.
[quantitative_nilrecurrence_applications.tex, L83-L91]

The following lemma is the key.

Lemma: high-density transfer

Let $B\subseteq\mathbb{N}$, $0<\rho\le 1$, and $K\ge 2$. Suppose there is $H\subseteq\mathbb{N}$ with

$$\overline{d}(H)=\rho$$

such that **every K-term arithmetic progression in H has its common difference in B.**

Then, for every $k\ge K$ and every finite $\Omega\subseteq\mathbb{N}$,

$$\boxed{r_k(\Omega)\le r_k(\Omega;B)\le \frac{1}{\rho}\,r_k(\Omega).} \tag{2}$$

Proof

The lower bound is immediate.

For the upper bound, let $E \subseteq \Omega$ contain no k-term progression with difference in B. For each integer $t \geq 1$, put

$$E_t = E \cap (H - t) = \{x \in E : x + t \in H\}.$$

Every E_t is k-AP-free. Indeed, suppose

$$a, a + d, \ldots, a + (k-1)d \in E_t.$$

Translating by t gives a k-term progression in H. Since $k \geq K$, its first K terms force $d \in B$. But the original progression also lies in E, contradicting the choice of E.

Consequently,

$$|E_t| \leq r_k(\Omega) \qquad (t \geq 1). \tag{3}$$

Now choose $N_j \to \infty$ along which

$$\frac{|H \cap [1, N_j]|}{N_j} \longrightarrow \rho.$$

For each fixed $x \in E$,

$$\frac{1}{N_j} \sum_{t=1}^{N_j} \mathbf{1}_H(t + x) \longrightarrow \rho,$$

because shifting an interval by the fixed integer x changes its count by only $O(x)$.

Therefore

$$\begin{aligned} \frac{1}{N_j} \sum_{t=1}^{N_j} |E_t| &= \sum_{x \in E} \frac{1}{N_j} \sum_{t=1}^{N_j} \mathbf{1}_H(t + x) \\ &\longrightarrow \rho |E|. \end{aligned}$$

Averaging (3) and passing to the limit gives

$$\rho |E| \leq r_k(\Omega).$$

Maximizing over E proves (2). □

The useful point is that **upper density is enough**. Neither natural density nor uniform density is needed. Also, the argument does not require Ω to be an interval or to have any density.

2. Your fixed-density theorem supplies precisely the required test set

Let B be a fixed finite-step nil-Bohr $_0$ set, and put

$$D = \mathbb{N} \setminus B.$$

Since D misses a nil-Bohr neighborhood, it is not pointwise recurrent for all nilsystems. Your fixed-density theorem says that any set which is ρ-density large for some $\rho \in (0, 1)$ is multiply pointwise recurrent for all nilsystems. Thus, by contraposition,

$$\boxed{D \text{ is not } \rho\text{-density large for every } 0 < \rho < 1.} \tag{4}$$

This uses precisely the pointwise conclusion stated in the fixed-density note. [fixed_density_nilrecurrence.tex, L44-L57]

Fix $\varepsilon > 0$, and take

$$\rho = \frac{1}{1+\varepsilon}.$$

By (4), there are $H \subseteq \mathbb{N}$ and an integer $K \geq 2$ such that

$$\overline{d}(H) = \rho$$

and H contains no K-term progression with difference in D.

Equivalently,

$$\text{every } K\text{-term progression in } H \text{ has difference in } B.$$

The high-density transfer lemma now gives

$$r_k(\Omega; B) \leq \rho^{-1} r_k(\Omega) = (1+\varepsilon) r_k(\Omega)$$

for every $k \geq K$ and every finite Ω.

This proves (1).

So, for a fixed nil-Bohr neighborhood, the qualitative fixed-density result alone is sufficient. No quantitative estimate on the nilrecurrence error is needed.

One particularly clean formulation is

$$\boxed{\lim_{k\to\infty} \sup_{\substack{\Omega \subseteq \mathbb{N} \\ 0<|\Omega|<\infty}} \frac{r_k(\Omega; B)}{r_k(\Omega)} = 1.} \tag{5}$$

The convergence is uniform over all nonempty finite ambient sets Ω; its rate may depend on B.

3. For shrinking targets, distinguish the qualitative theorem from its stronger proof

The applications note uses

$$B_k = \{d : \|P_i(d)\| \leq C k^{-d_i} \text{ for every } i\}. \tag{6}$$

Here B_k changes with k. The original two-color theorem supplies one coloring and coefficient-independent constants for precisely these weight-dependent rates. [Pasted text.txt, L68-L74]

The qualitative argument above, applied separately to B_k, gives an auxiliary length

$$K(B_k, \rho).$$

It does **not** itself show that

$$k \geq K(B_k, \rho).$$

Thus one cannot obtain the same-length shrinking-target estimate merely by substituting $B = B_k$ into (1).

However, the quantitative omitted-color theorem from the fixed-density proof supplies exactly the missing uniformity.

It says that, for every prescribed number $r \geq 2$ of colors, there are coefficient-independent constants such that every sufficiently long progression **omitting any one of the r colors** has a difference satisfying the required quantitative approximations. Its coloring is fixed independently of the progression length. [fixed_density_nilrecurrence.tex, L15-L33]

This gives the following stronger result.

Theorem: almost-lossless shrinking-target transfer

Fix

$$\varepsilon > 0, \qquad c > 0, \qquad m, D \in \mathbb{N}.$$

There is

$$k_0 = k_0(\varepsilon, c, m, D)$$

such that, for every family of special generalized polynomials

$$P_1, \ldots, P_m$$

of assigned weights $1 \leq d_i \leq D$, every $k \geq k_0$, and every finite $\Omega \subseteq \mathbb{N}$, the set

$$B_k(c) = \{d \in \mathbb{N} : \|P_i(d)\| \leq ck^{-d_i} \text{ for every } i\}$$

satisfies

$$\boxed{r_k(\Omega) \leq r_k(\Omega; B_k(c)) \leq (1+\varepsilon) r_k(\Omega).} \tag{7}$$

The threshold is independent of all real coefficients.

In particular, ** c can be the original constant C from the applications note**. There is no need to enlarge the target neighborhoods when replacing 2 by $1+\varepsilon$; increasing the lower threshold on k suffices.

Proof, first step: choose enough colors

Choose

$$r = 1 + \left\lceil \frac{1}{\varepsilon} \right\rceil.$$

Then

$$\frac{r}{r-1} \leq 1 + \varepsilon. \tag{8}$$

We want an r-coloring such that every k-term progression omitting a color has its difference in $B_k(c)$.

The omitted-color theorem almost gives this immediately, except that its approximation constant may be larger than the prescribed c. The following small integer-multiplier argument removes that issue.

Second step: keep the target constant c unchanged

Let $C_0 \geq 1$ and k_1 be the constants in the omitted-color theorem for

$$r \text{ colors}, \qquad 2m \text{ expressions}, \qquad \text{maximum weight } D.$$

Choose the fixed positive integer

$$q = \max\left\{1, \left\lceil \frac{C_0}{c} \right\rceil \right\}. \tag{9}$$

Apply that theorem to the $2m$ expressions

$$P_1, \dots, P_m,\ qP_1, \dots, qP_m.$$

Multiplication by q changes only the outermost real coefficient of a special generalized polynomial. It does not change its assigned weight.

We obtain one r-coloring such that every k-term progression omitting a color, for $k \geq k_1$, has a difference d satisfying

$$\|P_i(d)\| \leq C_0 k^{-d_i}, \qquad \|qP_i(d)\| \leq C_0 k^{-d_i}. \tag{10}$$

Increase the threshold so that

$$k \geq 4qC_0.$$

Write

$$P_i(d) = a_i + u_i, \qquad a_i \in \mathbb{Z}, \qquad |u_i| = \|P_i(d)\|.$$

Then

$$|qu_i| \leq qC_0 k^{-d_i} \leq \frac{1}{4}.$$

There is no wrap around the circle, and therefore

$$\|qP_i(d)\| = q|u_i|.$$

Using the second inequality in (10),

$$\|P_i(d)\| = |u_i| \leq \frac{C_0}{q} k^{-d_i} \leq ck^{-d_i}. \tag{11}$$

Thus, above the enlarged threshold, **every progression omitting a color has its difference in the original prescribed set** $B_k(c)$.

Notice that q is fixed independently of k. The resulting coloring still works simultaneously for all sufficiently large k.

Third step: remove the least numerous color

Let $E \subseteq \Omega$ contain no k-term progression with difference in $B_k(c)$.

Among the r color classes, choose one whose intersection with E has cardinality at most $|E|/r$. Remove that class, leaving $E' \subseteq E$ with

$$|E'| \geq \left(1 - \frac{1}{r}\right) |E|. \tag{12}$$

The set E' contains no k-term progression at all. Otherwise that progression would omit the removed color, so (11) would force its difference into $B_k(c)$, contradicting the choice of E.

Hence

$$r_k(\Omega) \geq |E'| \geq \frac{r-1}{r}|E|.$$

Therefore

$$|E| \leq \frac{r}{r-1} r_k(\Omega) \leq (1+\varepsilon) r_k(\Omega).$$

Taking the maximum over E proves (7). □

This is the direct improvement of the quantitative extremal theorem in the PDF. The old proof retained one of two color classes. The new proof retains the complement of one of many color classes—almost the entire set.

4. Why the dependence on progression length matters

We cannot fix a short progression length and then let $\varepsilon \to 0$ without changing anything else.

For example, take the fixed Bohr neighborhood

$$B = 2\mathbb{N}$$

and the ambient set

$$\Omega = [k].$$

The only k-term progression in $[k]$ is the full interval, with common difference 1. Thus

$$r_k([k]) = k-1, \qquad r_k([k]; 2\mathbb{N}) = k.$$

Consequently,

$$\frac{r_k([k]; 2\mathbb{N})}{r_k([k])} = 1 + \frac{1}{k-1}. \tag{13}$$

So the $1+\varepsilon$ assertion cannot hold here unless

$$\varepsilon \geq \frac{1}{k-1}.$$

This explains why the correct statement has an ε-dependent lower threshold on k, and why it does not imply exact equality of the extremal numbers.

The parameters in the two versions should therefore be kept distinct:

For fixed B:

$$k_0 = k_0(B, \varepsilon)$$

comes from the qualitative fixed-density theorem.

For the shrinking generalized-polynomial sets $B_k(c)$:

$$k_0 = k_0(\varepsilon, c, m, D)$$

comes from the quantitative omitted-color theorem and is uniform in the coefficients.

Conclusion

Your proposed improvement is correct, and it substantially strengthens the application.

For fixed nil-Bohr neighborhoods, the qualitative fixed-density theorem already gives

$$r_k(\Omega; B) = (1 + o_{k\to\infty}(1))\, r_k(\Omega),$$

uniformly over finite nonempty Ω.

For the original shrinking-target sets in the applications note, the quantitative strengthening proved along the way gives

$$\boxed{r_k(\Omega) \le r_k(\Omega; B_k) \le (1+\varepsilon) r_k(\Omega)}$$

for every sufficiently large k, **with the same approximation constant in B_k** and coefficient-independent thresholds.

This also sharpens the distinction relevant to your discussion with Ryan: the **fixed-neighborhood almost-lossless transfer is already an application of the qualitative density theorem**; the **uniform same-length shrinking-target version is the additional quantitative payoff**.

---

83. User

2026-09-17 10:56:13 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-17T02:56:13.925000Z

Please give me a 1-2 page pdf file that contains the 1) qualitative statement of delta-density large implies good diophantine approximation for arbitrary finite families of special generalized polynomials, 2) the quantitative shrinking targets version, 3) the (1+epsilon) corollary of 2). Include enough definitions that I could share this pdf with Sarah peluse so that I can ask for her opinion on the quality of the result 3).

---

84. Assistant

2026-09-17 10:56:43 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-17T02:56:43.952971Z

I'll make a two-page note with the necessary definitions, precise parameter dependencies, and a short proof of the $1+\varepsilon$ corollary.

---

85. Assistant

2026-09-17 10:59:50 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-17T02:59:50.212408Z

I'll keep the shrinking-target constant fixed and include the high-density test-set formulation, which preserves progression length; the $1+\varepsilon$ conclusion requires an ε-dependent length threshold.

---

86. Assistant

2026-09-17 11:01:28 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-17T03:01:28.366999Z

Download the two-page PDF

Download the standalone LaTeX source

The note contains the three requested statements, the definitions needed to read them independently, and a **complete short proof of the $1+\varepsilon$ corollary** by translation averaging. It preserves your drafts' nearest-integer convention and assigned-weight definition for special generalized polynomials. [Pasted text.txt, L50-L57] [Pasted text.txt, L182-L187]

It also includes a concrete example involving

$$\|\alpha s\| \le k^{-1}, \qquad \|\beta s^2\| \le k^{-2}, \qquad \|\gamma s\langle \theta s\rangle\| \le k^{-2}.$$

The parameter dependencies are explicit: **the target constant remains fixed**, the minimum progression length may depend on ε, and the extremal comparison is uniform over all real coefficients and finite ambient sets Ω. The note identifies the main approximation statements as results from the work in progress, rather than published results.

---

87. User

2026-09-17 17:55:23 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-17T09:55:23.790000Z

why does the direct special generalized polynomial proof only work with w_i > 0. Does this proof still handle generalized polynomials such as sqrt{2}floor(sqrt{3}floor(n^2sqrt{5}))?

---

88. Assistant

2026-09-17 17:55:40 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-17T09:55:40.291870Z

I'll check exactly where positivity enters the matrix argument, while distinguishing ordinary floors from the draft's nearest-integer brackets in your example.

---

89. Assistant

2026-09-17 17:58:44 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-17T09:58:44.828106Z

The positivity restriction is stronger than necessary: zero outer weights are allowed if the innermost weight is positive. However, literal floors introduce a genuine obstruction absent with nearest-integer brackets.

---

90. Assistant

2026-09-17 18:02:24 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-17T10:02:24.707324Z

The strict positivity requirement is stronger than the elementary proof actually needs. I should correct the sentence in the draft saying that it is "essential." The direct matrix argument extends to nonnegative exponents, provided the expression has positive total weight, with a small adjustment to the final step. [Pasted text.txt, L351-L355]

However, there is a second, genuinely important issue in your example: **nearest-integer rounding and ordinary floor are not interchangeable here.**

For

$$P_{\mathrm{near}}(n) = \sqrt{2}\left\langle \sqrt{3}\left\langle \sqrt{5}\,n^2 \right\rangle \right\rangle,$$

where $\langle x \rangle$ denotes a nearest integer, the answer is **yes**, including a quantitative $O(k^{-2})$ conclusion.

For the literal floor expression

$$P_{\mathrm{floor}}(n) = \sqrt{2}\left\lfloor \sqrt{3}\left\lfloor \sqrt{5}\,n^2 \right\rfloor \right\rfloor,$$

the analogous guarantee along every 2-large set is **false**. In fact, below I construct a set that is δ-density large for every $0 < \delta < 1$, but on which

$$\|P_{\mathrm{floor}}(n)\| > \frac{1}{10}$$

always holds.

I will separate these two issues.

1. What positive weights accomplish—and what the direct proof really requires

The draft defines the special class using

$$P(n) = a_1 n^{w_1} \left\langle a_2 n^{w_2} \left\langle \cdots \left\langle a_\ell n^{w_\ell} \right\rangle \cdots \right\rangle \right\rangle, \qquad w_i \geq 1.$$

This agrees with Huang–Shao–Ye's definition of the *special* subclass. Their larger normalized generalized-polynomial class also permits constant scalar multipliers outside nested rounding operations. Thus your nearest-integer example belongs to the larger class, though not to the special subclass as presently defined. [Pasted text.txt, L50-L57] arXiv

For the weighted matrices, recall that

$$w_{ij} = w_i + \cdots + w_{j-1}, \qquad \tau(A)_{ij} = [t^{w_{ij}}]A_{ij}(t).$$

The proof uses

$$\tau(AB) = \tau(A)\tau(B), \qquad \tau(A(b+qt)) = \delta_q \tau(A).$$

These identities remain valid when some weights are zero. They require additivity of the weights and the corresponding coordinate-degree bounds, not strict positivity. The displayed identities are the ones in equation (3) of the elementary draft. [Pasted text.txt, L77-L107]

What changes at weight zero is the following.

If $w_{ij} = 0$, then

$$[t^0]A_{ij}(t) = A_{ij}(t),$$

because this coordinate is constant. Consequently, constant matrices need not map to the identity under τ. Also, the bound

$$|\tau(Q)_{ij}| \ll J^{-w_{ij}}$$

becomes merely $O(1)$, rather than something tending to zero.

But—and this is the crucial observation—**the bracket-recovery step only needs small errors in the last column of each block.** It does not need every off-diagonal error to tend to zero. That is visible directly in the last-column recursion in the draft. [Pasted text.txt, L325-L345]

If

$$w_1, \dots, w_{\ell-1} \geq 0, \qquad w_\ell > 0,$$

then every relevant last-column weight is positive:

$$w_{i,\ell+1} = w_i + \cdots + w_\ell > 0.$$

Therefore all the errors actually needed to recover the brackets still tend to zero.

So the elementary argument works with zero outer or intermediate exponents. The necessary repair is to require smallness only in the last columns, rather than in every off-diagonal coordinate.

2. Your nearest-integer example works directly with weights $(0, 0, 2)$

For

$$P_{\mathrm{near}}(n) = \sqrt{2}\left\langle \sqrt{3}\left\langle \sqrt{5}\, n^2 \right\rangle \right\rangle,$$

use

$$g(t) = \begin{pmatrix} 1 & -\sqrt{2} & 0 & 0 \\ 0 & 1 & -\sqrt{3} & 0 \\ 0 & 0 & 1 & \sqrt{5}\, t^2 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Give the adjacent entries weights

$$w_1 = 0, \qquad w_2 = 0, \qquad w_3 = 2.$$

Then

$$w_{12} = w_{23} = w_{13} = 0, \qquad w_{34} = w_{24} = w_{14} = 2.$$

The shell, polynomial-rounding, first-exit, and coefficient-extraction arguments still give, for the difference d of a sufficiently long monochromatic k-term progression,

$$g(d)\gamma = E, \qquad \gamma \in \mathrm{UT}_4(\mathbb{Z}),$$

with

$$\boxed{|E_{34}|,\ |E_{24}|,\ |E_{14}| \leq Ck^{-2}.} \tag{1}$$

The weight-zero coordinates need only remain bounded. Indeed,

$$E_{12} = -\sqrt{2} + \gamma_{12},$$

so it is impossible to make $|E_{12}| < 1/4$ by any choice of integer γ_{12}. This demonstrates exactly why the draft's instruction to make **all** off-diagonal entries smaller than $1/4$ must be narrowed. That instruction is stronger than its subsequent recursion uses. [Pasted text.txt, L315-L339]

Now read the last column.

The bottom equation is

$$\sqrt{5}\, d^2 + \gamma_{34} = E_{34}.$$

For sufficiently large k, $|E_{34}| < 1/4$, so

$$\gamma_{34} = -\left\langle \sqrt{5}\, d^2 \right\rangle .$$

The next equation is

$$\gamma_{24} - \sqrt{3}\,\gamma_{34} = E_{24},$$

or

$$\gamma_{24} + \sqrt{3}\left\langle \sqrt{5}\, d^2 \right\rangle = E_{24}.$$

Thus

$$\gamma_{24} = -\left\langle \sqrt{3}\left\langle \sqrt{5}\, d^2 \right\rangle \right\rangle .$$

Finally,

$$\gamma_{14} - \sqrt{2}\,\gamma_{24} = E_{14},$$

which becomes

$$\gamma_{14} + \sqrt{2}\left\langle \sqrt{3}\left\langle \sqrt{5}\, d^2 \right\rangle \right\rangle = E_{14}.$$

Therefore

$$\boxed{\|P_{\mathrm{near}}(d)\| \le |E_{14}| \le Ck^{-2}.} \tag{2}$$

This is the original direct argument, not an appeal to the more elaborate nilmanifold proof.

Why the intervening steps tolerate weight zero

There are only a few things to check.

In the polynomial-rounding lemma, a weight-zero coordinate is constant. Its reduced values in one small grid cell have integer first differences of magnitude less than one, so those differences vanish. The resulting correction is an integer constant.

In denominator clearance,

$$w!\,[t^w]C_{ij}(t) \in \mathbb{Z}$$

also holds for $w = 0$: it simply says that the constant correction is an integer. The dilation does nothing to that coordinate, which is harmless. The original rounding and denominator-clearance steps appear in the draft's equations (10)–(16). [Pasted text.txt, L194-L218] [Pasted text.txt, L253-L280]

Finally, the weighted-homogeneity identity

$$\tau\big(g(x+ut)\big) = g(u)$$

still holds: the weight-zero entries remain the same constants, while the quadratic entry contributes $\sqrt{5}\,u^2$.

The observation that constant matrices no longer disappear **does** obstruct a verbatim extension of the longer factorization proof, which explicitly discards a constant matrix using that property. But the elementary proof does not have that step. [Pasted text.txt, L1173-L1185]

3. A simpler scalar verification for these particular coefficients

There is also a short deduction using only the already-stated positive-weight theorem.

Apply it to the three ordinary monomials

$$\sqrt{5}\,n^2, \qquad \sqrt{15}\,n^2, \qquad \sqrt{30}\,n^2.$$

Suppose the resulting difference d satisfies

$$\begin{aligned} \sqrt{5}\,d^2 &= N + u, \\ \sqrt{15}\,d^2 &= M + v, \qquad N, M, L \in \mathbb{Z}, \quad |u|, |v|, |w| \le \eta. \\ \sqrt{30}\,d^2 &= L + w, \end{aligned} \tag{3}$$

The theorem gives this with $\eta = C_0 k^{-2}$. [Pasted text.txt, L68-L74]

For sufficiently small η,

$$\left\langle \sqrt{5}\,d^2 \right\rangle = N.$$

Furthermore,

$$\sqrt{3}\,N = \sqrt{15}\,d^2 - \sqrt{3}\,u = M + v - \sqrt{3}\,u.$$

If

$$(1+\sqrt{3})\eta < \frac{1}{2},$$

then

$$\left\langle \sqrt{3}\,N \right\rangle = M.$$

Hence

$$P_{\mathrm{near}}(d) = \sqrt{2}\,M = \sqrt{30}\,d^2 - \sqrt{2}\,v = L + w - \sqrt{2}\,v,$$

and therefore

$$\boxed{\|P_{\mathrm{near}}(d)\| \le (1+\sqrt{2})\eta \ll k^{-2}.} \tag{4}$$

Thus, even without extending the matrix statement, **your particular nearest-integer example already has the desired quantitative conclusion**.

4. The appropriate extension of the elementary theorem

The preceding verification gives the following extension of the proof in the draft.

> **Nonnegative-weight extension.**
> Fix finitely many nested monomial expressions
>
> $$P(n) = a_1 n^{w_1} \langle a_2 n^{w_2} \langle \cdots \langle a_\ell n^{w_\ell} \rangle \cdots \rangle \rangle,$$
>
> with $w_i \ge 0$ and $\sum_i w_i > 0$. For fixed weight patterns, the elementary proof gives a coefficient-uniform coloring and bounds
>
> $$\|P(d)\| \ll k^{-\sum_i w_i}$$

> for differences of sufficiently long monochromatic progressions.

Trailing zero weights cause no additional problem: their entire constant tail can first be evaluated and absorbed into the coefficient of the last positive-power term. For example,

$$an^2 \langle b \langle c \rangle \rangle = a' n^2, \qquad a' = a \langle b \langle c \rangle \rangle .$$

There is, however, a **parameter distinction** worth preserving in the paper. When all $w_i > 0$,

$$\ell \le \sum_i w_i,$$

so the total weight automatically bounds the bracket depth and matrix dimension. Once zero weights are permitted, it no longer does.

Accordingly, for the extended class, the proof gives constants depending on the number of expressions, maximum weight, **and maximum depth**. It does not establish degree-only uniformity over expressions containing arbitrarily many zero-weight layers.

So positivity is useful for the clean statement $C = C(m, D)$, but it is not an absolute limitation of the elementary method.

5. Literal floors are a different matter: a counterexample for your exact expression

Now take the expression exactly as you wrote it:

$$F(n) = \sqrt{2} \left\lfloor \sqrt{3} \left\lfloor \sqrt{5}\, n^2 \right\rfloor \right\rfloor .$$

The draft uses nearest integers, not ordinary floors. Its larger generalized-polynomial corollary also explicitly retains that convention and warns that merely requiring an arbitrary floor expression to vanish at zero is insufficient. [Pasted text.txt, L1542-L1556] [Pasted text.txt, L1615-L1622]

For this particular F, we can show something stronger than a failure of the proof.

Proposition

The set

$$\boxed{S = \left\{ n \in \mathbb{N} : \|F(n)\| > \frac{1}{10} \right\}} \tag{5}$$

is large. In fact, it is δ-density large for every $0 < \delta < 1$.

Thus no conclusion of the form

$$\inf_{n \in S} \|F(n)\| = 0$$

can follow from either of these largeness hypotheses.

Step 1: approaching integers from below creates a fixed nonzero offset

Let $\eta = 1/100$. Suppose

$$\sqrt{5}\, n^2 = N + u, \qquad \sqrt{15}\, n^2 = M + v, \qquad \sqrt{30}\, n^2 = L + w, \tag{6}$$

where

$$-\eta < u < 0, \qquad |v|, |w| < \eta, \qquad N, M, L \in \mathbb{Z}.$$

Because $u < 0$,

$$\left\lfloor \sqrt{5}\, n^2 \right\rfloor = N - 1.$$

Therefore

$$\sqrt{3} \left\lfloor \sqrt{5}\, n^2 \right\rfloor = M + v - \sqrt{3}\, u - \sqrt{3}.$$

Our error bounds ensure

$$-2 < v - \sqrt{3}\, u - \sqrt{3} < -1,$$

so

$$\left\lfloor \sqrt{3} \left\lfloor \sqrt{5}\, n^2 \right\rfloor \right\rfloor = M - 2.$$

Consequently,

$$F(n) = \sqrt{2}(M - 2) = L + w - \sqrt{2}\, v - 2\sqrt{2}. \tag{7}$$

It follows that

$$\begin{aligned} \|F(n)\| &\geq \|2\sqrt{2}\| - |w - \sqrt{2} v| \\ &> 3 - 2\sqrt{2} - \frac{1 + \sqrt{2}}{100} \\ &> \frac{1}{10}. \end{aligned} \tag{8}$$

Unlike nearest-integer errors, the floor errors here do not disappear when the input approaches an integer from below. They produce the fixed offset $-2\sqrt{2}$.

Step 2: S contains arbitrarily long dilated initial intervals

I claim that

$$\boxed{\forall R \geq 1 \quad \exists q \geq 1 : \quad q\{1, \ldots, R\} \subseteq S.} \tag{9}$$

The vector sequence

$$n^2(\sqrt{5}, \sqrt{15}, \sqrt{30}) \pmod 1$$

is equidistributed in $\mathbb{T}^3$. To check the relevant hypothesis, every nonzero integer combination of the three radicals is irrational. Indeed, an equality

$$h_1\sqrt{5} + h_2\sqrt{15} + h_3\sqrt{30} \in \mathbb{Q}$$

first forces the rational value to be zero by changing the sign of $\sqrt{5}$, and then forces all three h_i to vanish by the other sign-changing automorphisms of the multiquadratic field. The Weyl criterion reduces the equidistribution assertion to irrational quadratic phases; their exponential averages vanish by one application of van der Corput.

We may therefore choose q so that the signed errors in (6), with $n = q$, satisfy

$$-\frac{\eta}{R^2} < u < 0, \qquad |v|, |w| < \frac{\eta}{R^2}.$$

For $n = tq$, $1 \le t \le R$, the corresponding errors are exactly

$$t^2u, \qquad t^2v, \qquad t^2w.$$

They satisfy the hypotheses of Step 1. Hence $tq \in S$, proving (9).

Property (9) immediately gives largeness: apply finite van der Waerden to the coloring of $\{1, \dots, R\}$ induced by $j \mapsto c(qj)$.

Step 3: property (9) also gives density largeness at every positive density

Fix $0 < \delta < 1$ and a desired progression length k. Choose R sufficiently large that every subset of $[R]$ of density at least $\delta/2$ contains a k-term progression, by finite Szemerédi. arXiv

Choose q with $q[R] \subseteq S$. Partition $\mathbb{N}$ into blocks of length qR, and decompose each block into its q residue-class progressions of length R and common difference q.

If $A \subseteq \mathbb{N}$ has upper density δ, at least one of these length- R progressions must have relative A-density at least $\delta/2$. Otherwise the density of A in every complete block would be less than $\delta/2$, forcing

$$\overline{d}(A) \le \delta/2.$$

Apply the choice of R inside that progression. We obtain a k-term progression in A with common difference

$$qt, \qquad 1 \le t \le R.$$

Since $qt \in S$, this is an allowed progression. As k was arbitrary, S is δ-density large. □

What should change in the draft

I would replace the assertion that positive exponents are essential with something like:

> The positive-exponent convention ensures that total weight bounds the nesting depth and that all off-diagonal coordinates have positive weight. The elementary argument also applies to fixed nonnegative weight patterns of positive total weight, after absorbing constant tails: only the last-column errors are required to tend to zero. For this extension, the constants may additionally depend on the nesting depth. All rounding operations are nearest-integer operations, not ordinary floors.

So your nearest-integer example is covered, and the direct proof is more flexible than the draft currently says. But the literal floor expression exposes a real obstruction, not just a technical gap in the proof.

---

91. User

2026-09-17 18:22:24 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-17T10:22:24.199000Z

I dont understand the example that appears before the proof of Theorem 1.1. If says P(d) = x_1floor(x_2), so what is d?

92. Assistant

2026-09-17 18:23:48 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-17T10:23:48.276545Z

Here d is the positive integer at which P is evaluated—in the proof, it is the common difference of the monochromatic arithmetic progression. The notation x_1, x_2 suppresses their dependence on d. The underlying substitution is

$$\boxed{x_1 = a_1 d^{w_1}, \qquad x_2 = a_2 d^{w_2}.}$$

This is the two-layer case of the substitution $x_i = a_i d^{w_i}$ used in the draft's bracket-recovery argument. [Pasted text.txt, L310-L325]

What the sentence means

Start with the special generalized polynomial

$$P(n) = a_1 n^{w_1} \left\langle a_2 n^{w_2} \right\rangle .$$

After fixing d, evaluating at $n = d$ gives

$$P(d) = a_1 d^{w_1} \left\langle a_2 d^{w_2} \right\rangle = x_1 \langle x_2 \rangle .$$

So ** d has not disappeared mathematically; its dependence has merely been absorbed into x_1 and x_2**. At that point, the example is explaining an algebraic calculation with the two real numbers x_1, x_2.

Here I write $\langle x \rangle$ for the draft's **nearest-integer rounding**, not the ordinary floor $\lfloor x \rfloor$. The draft explicitly uses that convention. [Pasted text.txt, L50-L57]

A concrete example

Suppose

$$P(n) = \sqrt{2}\, n \left\langle \sqrt{3}\, n \right\rangle ,$$

and the progression under consideration is

$$10, 13, 16, 19, \ldots .$$

Its common difference is $d = 3$, so we set

$$x_1 = 3\sqrt{2}, \qquad x_2 = 3\sqrt{3}.$$

Since $3\sqrt{3} \approx 5.196$, its nearest integer is 5. Therefore

$$P(d) = P(3) = (3\sqrt{2})\langle 3\sqrt{3} \rangle = (3\sqrt{2}) \cdot 5 = 15\sqrt{2}.$$

Notice that the starting point 10 does not enter this evaluation: the desired conclusion concerns P evaluated at the **difference**, not at the terms of the progression.

A clearer wording for the example

I should have written:

> Consider $P(n) = a_1 n^{w_1} \langle a_2 n^{w_2} \rangle$. Fix a positive integer d, which will be the common difference of the progression in the proof, and abbreviate $x_1 = a_1 d^{w_1}$ and $x_2 = a_2 d^{w_2}$. Then $P(d) = x_1 \langle x_2 \rangle$.

That makes the role of d explicit before introducing the shorter notation.

---

93. User

2026-09-17 18:26:01 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-17T10:26:01.303000Z

Explain the idea and proof of Lemma 7.1 to me in more detail please

---

94. Assistant

2026-09-17 18:26:15 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-17T10:26:15.368339Z

I'll identify the exact statement of Lemma 7.1, then unpack its proof step by step and explain the intuition and its role in the argument.

---

95. Assistant

2026-09-17 18:29:23 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-17T10:29:23.388508Z

Lemma 7.1, "Decoding a small last column," says that the integer entries in the last column of γ must be the nested nearest integers occurring in $P(d)$. The small entries of E force those integer choices uniquely. Once we identify them, the top-right entry of E is exactly the rounding error of $P(d)$. [elementary_two_color_obstruction_expanded.tex, L632-L659]

The main point is that **we recover the rounded integers exactly, one at a time—not approximately**. Thus no errors accumulate as we move through the brackets.

I will use $\langle x \rangle$ for nearest-integer rounding, to distinguish it from the ordinary floor function.

1. Restating the lemma and identifying all the objects

Start with

$$P(n) = a_1 n^{w_1} \left\langle a_2 n^{w_2} \left\langle \cdots \left\langle a_\ell n^{w_\ell} \right\rangle \cdots \right\rangle \right\rangle .$$

Fix a positive integer d, and abbreviate

$$x_i = a_i d^{w_i}.$$

Thus

$$P(d) = x_1 \left\langle x_2 \left\langle \cdots \left\langle x_\ell \right\rangle \cdots \right\rangle \right\rangle .$$

The matrix encoding these real numbers is

$$g(d) = \begin{pmatrix} 1 & -x_1 & 0 & \cdots & 0 \\ 0 & 1 & -x_2 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & x_\ell \\ 0 & \cdots & 0 & 0 & 1 \end{pmatrix}. \tag{1}$$

There are $\ell+1$ rows and columns. All entries above the first superdiagonal are zero. The superdiagonal entries are negative except for the last one, which is $+x_\ell$. This is exactly the matrix chosen in the draft. [elementary_two_color_obstruction_expanded.tex, L612-L628]

The lemma **assumes** that we have an integer upper-unitriangular matrix γ such that

$$g(d)\gamma = E \tag{2}$$

and

$$|E_{i,\ell+1}| < \frac{1}{4} \qquad (1 \le i \le \ell). \tag{3}$$

Here "integer upper-unitriangular" means that γ has integer entries, ones on the diagonal, and zeros below the diagonal.

Its conclusion is

$$\boxed{P(d) + \gamma_{1,\ell+1} = E_{1,\ell+1}.} \tag{4}$$

Because $\gamma_{1,\ell+1}$ is an integer,

$$\boxed{\|P(d)\| \le |E_{1,\ell+1}|.}$$

The lemma does not construct γ or prove that these small entries exist. The preceding argument does that. This lemma explains how to interpret the resulting matrix identity.

2. The elementary fact behind every step

Everything rests on the following observation:

> If $x \in \mathbb{R}$, $m \in \mathbb{Z}$, and
>
> $$|x+m| < \frac{1}{2},$$
>
> then the unique nearest integer to x is $-m$.

Indeed, x is less than $1/2$ away from the integer $-m$. Any other integer is farther away:

$$|x-z| \ge |z+m| - |x+m| \ge 1 - |x+m| > \frac{1}{2} \qquad (z \ne -m).$$

So an equation

$$x + m = e, \qquad |e| < \frac{1}{4},$$

does not merely tell us that m is *approximately* $-x$. It tells us the exact identity

$$\boxed{m = -\langle x \rangle.} \tag{5}$$

This is why the small residuals in the matrix identity are so useful.

The particular constant $1/4$ is not essential to this decoding step: a strict bound below $1/2$ would suffice. The draft uses $1/4$ to leave a comfortable margin.

3. First understand the case of one bracket

Take

$$P(d) = x_1\langle x_2\rangle.$$

Then

$$g(d) = \begin{pmatrix} 1 & -x_1 & 0 \\ 0 & 1 & x_2 \\ 0 & 0 & 1 \end{pmatrix}.$$

Write

$$\gamma = \begin{pmatrix} 1 & \gamma_{12} & \gamma_{13} \\ 0 & 1 & \gamma_{23} \\ 0 & 0 & 1 \end{pmatrix}, \qquad \gamma_{12}, \gamma_{13}, \gamma_{23} \in \mathbb{Z}.$$

Multiplying gives

$$E = g(d)\gamma = \begin{pmatrix} 1 & \gamma_{12} - x_1 & \gamma_{13} - x_1\gamma_{23} \\ 0 & 1 & x_2 + \gamma_{23} \\ 0 & 0 & 1 \end{pmatrix}. \tag{6}$$

Focus only on the last column.

The bottom equation recovers the inner nearest integer

We have

$$E_{23} = x_2 + \gamma_{23}.$$

Since $|E_{23}| < 1/4$, observation (5) gives

$$\boxed{\gamma_{23} = -\langle x_2\rangle.} \tag{7}$$

This is an exact equality.

Substitute that exact integer into the top equation

The top equation is

$$E_{13} = \gamma_{13} - x_1\gamma_{23}.$$

Substituting (7),

$$\begin{aligned} E_{13} &= \gamma_{13} + x_1\langle x_2\rangle \\ &= \gamma_{13} + P(d). \end{aligned}$$

Therefore

$$\boxed{\|P(d)\| \le |E_{13}|.}$$

That is the whole lemma in the one-bracket case. The example immediately before the proof of Theorem 1.1 is illustrating exactly these two substitutions. [elementary_two_color_obstruction_expanded.tex, L661-L672]

Notice that γ_{12} never enters the argument. We do not need the entry E_{12} to be small.

4. Two brackets: why the same argument repeats

Now take

$$P(d) = x_1 \langle x_2 \langle x_3 \rangle \rangle .$$

The relevant matrix is

$$g(d) = \begin{pmatrix} 1 & -x_1 & 0 & 0 \\ 0 & 1 & -x_2 & 0 \\ 0 & 0 & 1 & x_3 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The last column of $g(d)\gamma = E$ gives

$$\begin{aligned} E_{34} &= x_3 + \gamma_{34}, \\ E_{24} &= \gamma_{24} - x_2\gamma_{34}, \\ E_{14} &= \gamma_{14} - x_1\gamma_{24}. \end{aligned} \tag{8}$$

We read these equations **from the bottom upward**.

First,

$$|E_{34}| < \frac{1}{4} \quad \Longrightarrow \quad \gamma_{34} = -\langle x_3 \rangle.$$

Substituting this into the next equation,

$$E_{24} = \gamma_{24} + x_2\langle x_3 \rangle.$$

Its smallness forces

$$\gamma_{24} = -\langle x_2 \langle x_3 \rangle \rangle .$$

Substitute again:

$$\begin{aligned} E_{14} &= \gamma_{14} + x_1 \langle x_2 \langle x_3 \rangle \rangle \\ &= \gamma_{14} + P(d). \end{aligned}$$

So the last column records, successively,

$$\boxed{\begin{aligned} \gamma_{34} &= -\langle x_3 \rangle, \\ \gamma_{24} &= -\langle x_2 \langle x_3 \rangle \rangle , \\ \gamma_{14} &= -\langle x_1 \langle x_2 \langle x_3 \rangle \rangle \rangle . \end{aligned}} \tag{9}$$

The final equality uses the smallness of E_{14} as well.

Each row adds precisely one more layer of the nested expression.

5. A numerical example

For an illustration, suppose that, after fixing d,

$$x_1 = 2.003, \qquad x_2 = 1.401, \qquad x_3 = 5.02.$$

For example, these values arise with $d = 1$, $w_1 = w_2 = w_3 = 1$, and coefficients equal to the displayed numbers.

The nested calculation is

$$\langle x_3 \rangle = \langle 5.02 \rangle = 5,$$

then

$$x_2\langle x_3 \rangle = 1.401 \cdot 5 = 7.005, \qquad \langle 7.005 \rangle = 7,$$

and finally

$$P(d) = 2.003 \cdot 7 = 14.021.$$

Thus

$$\|P(d)\| = 0.021.$$

The corresponding integer matrix can be chosen as

$$\gamma = \begin{pmatrix} 1 & 0 & 0 & -14 \\ 0 & 1 & 0 & -7 \\ 0 & 0 & 1 & -5 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Then

$$\begin{pmatrix} 1 & -2.003 & 0 & 0 \\ 0 & 1 & -1.401 & 0 \\ 0 & 0 & 1 & 5.02 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & -14 \\ 0 & 1 & 0 & -7 \\ 0 & 0 & 1 & -5 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -2.003 & 0 & 0.021 \\ 0 & 1 & -1.401 & 0.005 \\ 0 & 0 & 1 & 0.02 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{10}$$

The last-column residuals are exactly

$$\begin{aligned} E_{34} &= 5.02 - 5 = 0.02, \\ E_{24} &= 7.005 - 7 = 0.005, \\ E_{14} &= 14.021 - 14 = 0.021. \end{aligned}$$

This example shows the two sides of the argument:

Starting with the nested rounding calculation, we can see which integers should occur in γ. Conversely, **given the small last-column entries in (10), the lemma forces those integers to be** $-5, -7, -14$.

It also illustrates why only the last column matters: $E_{12} = -2.003$ and $E_{23} = -1.401$ are not small, but they do not interfere with decoding.

6. The general proof, with the induction written out

For arbitrary nesting depth, define the suffix values

$$H_\ell = x_\ell, \qquad H_i = x_i\langle H_{i+1} \rangle \quad (i = \ell - 1, \dots, 1). \tag{11}$$

Thus

$$H_1 = P(d).$$

For example,

$$H_{\ell-1} = x_{\ell-1}\langle x_\ell \rangle,$$

and

$$H_{\ell-2} = x_{\ell-2}\,\langle x_{\ell-1}\langle x_\ell\rangle\rangle\,.$$

We prove, working upward through the last column, that

$$\boxed{\gamma_{i,\ell+1} = -\langle H_i\rangle, \qquad E_{i,\ell+1} = H_i - \langle H_i\rangle.} \tag{12}$$

This is the induction carried out in the draft. [elementary_two_color_obstruction_expanded.tex, L641-L658]

Base case: $i = \ell$

Row ℓ of $g(d)$ has only two nonzero entries: its diagonal entry 1 and the final entry x_ℓ. Consequently,

$$E_{\ell,\ell+1} = \gamma_{\ell,\ell+1} + x_\ell \gamma_{\ell+1,\ell+1} = \gamma_{\ell,\ell+1} + x_\ell,$$

because $\gamma_{\ell+1,\ell+1} = 1$.

Since $H_\ell = x_\ell$,

$$E_{\ell,\ell+1} = \gamma_{\ell,\ell+1} + H_\ell.$$

The residual has absolute value below $1/4$, so

$$\gamma_{\ell,\ell+1} = -\langle H_\ell\rangle.$$

Substitution proves both identities in (12) for $i = \ell$.

Inductive step

Suppose we already know

$$\gamma_{i+1,\ell+1} = -\langle H_{i+1}\rangle. \tag{13}$$

For $i < \ell$, row i of $g(d)$ has only the two nonzero entries

$$g(d)_{ii} = 1, \qquad g(d)_{i,i+1} = -x_i.$$

Therefore the matrix product gives

$$E_{i,\ell+1} = \gamma_{i,\ell+1} - x_i \gamma_{i+1,\ell+1}. \tag{14}$$

Use the exact identity (13):

$$\begin{aligned} E_{i,\ell+1} &= \gamma_{i,\ell+1} + x_i\langle H_{i+1}\rangle \\ &= \gamma_{i,\ell+1} + H_i. \end{aligned}$$

Again $\gamma_{i,\ell+1}$ is an integer and the residual is smaller than $1/4$. Hence

$$\gamma_{i,\ell+1} = -\langle H_i\rangle,$$

and therefore

$$E_{i,\ell+1} = H_i - \langle H_i\rangle.$$

This completes the induction.

At the top row, $H_1 = P(d)$, so

$$E_{1,\ell+1} = P(d) - \langle P(d)\rangle.$$

In particular,

$$P(d) + \gamma_{1,\ell+1} = E_{1,\ell+1},$$

as claimed.

Under the lemma's hypotheses, we actually obtain the slightly stronger identity

$$\boxed{\|P(d)\| = |E_{1,\ell+1}|.} \tag{15}$$

The lemma only states the inequality because that is all the subsequent proof needs.

7. Why this is useful, rather than just a complicated way of rounding

The matrix encoding avoids a potentially serious error-amplification problem.

For example, knowing

$$x_2 = m + e, \qquad |e| \ll 1,$$

does not let us harmlessly replace x_2 by m after multiplying by x_1: the resulting error is

$$x_1 e,$$

and x_1 may be arbitrarily large.

The lemma does **not** make such an approximate substitution. Instead, the first small residual proves

$$\gamma_{23} = -\langle x_2 \rangle$$

exactly. The next equation then becomes

$$E_{13} = \gamma_{13} + x_1 \langle x_2 \rangle$$

with no error term at all. Its residual is controlled separately by the matrix estimate.

That is how the argument retains uniformity in potentially enormous coefficients.

The sign pattern in $g(d)$ is chosen to make this work cleanly. The bottom $+x_\ell$ produces

$$\gamma_{\ell,\ell+1} = -\langle x_\ell \rangle.$$

Each subsequent $-x_i$ cancels this negative sign:

$$-x_i\big(-\langle H_{i+1} \rangle\big) = x_i \langle H_{i+1} \rangle.$$

Thus we recover precisely the desired nesting, without alternating signs or needing any identity about rounding negative numbers.

Finally, the earlier part of the proof supplies

$$g(d)\gamma = E, \qquad |E_{ij}| \le K_D \left(\frac{L}{N - 1 - 2W} \right)^{w_{ij}}.$$

The top-right entry has weight

$$w_{1,\ell+1} = w_1 + \cdots + w_\ell = \deg P.$$

Lemma 7.1 translates that matrix estimate into

$$\|P(d)\| \le K_D \left(\frac{L}{N-1-2W}\right)^{\deg P} \ll N^{-\deg P}.$$

This is exactly how it is used in the proof of Theorem 1.1. [elementary_two_color_obstruction_expanded.tex, L674-L700]

In one sentence: read the last column from bottom to top; every small residual identifies one nearest integer exactly, and the final residual is the fractional error of the full nested expression. This is also the mechanism summarized in the shorter draft's bracket-recovery paragraph. [Pasted text.txt, L325-L345]

---

96. User

2026-09-17 21:39:10 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-17T13:39:10.541000Z

why does lemma 7.1 merely say ||P(d)|| <= |E_{1,ell+1}| instead of =?

---

97. Assistant

2026-09-17 21:40:38 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-17T13:40:38.520903Z

You are right: under Lemma 7.1's hypotheses, equality holds. The inequality is simply weaker than what the proof establishes; it was sufficient for the subsequent estimate, but there is no reason not to state the equality.

The proof gives

$$P(d) + \gamma_{1,\ell+1} = E_{1,\ell+1}, \qquad \gamma_{1,\ell+1} \in \mathbb{Z}, \qquad |E_{1,\ell+1}| < \frac{1}{4}.$$

These are precisely the integer correction and small-residual conditions used in the last-column argument. [Pasted text.txt, L323-L339]

Since adding an integer does not change distance to $\mathbb{Z}$,

$$\|P(d)\| = \|P(d) + \gamma_{1,\ell+1}\| = \|E_{1,\ell+1}\|.$$

But $|E_{1,\ell+1}| < 1/4 < 1/2$, so its nearest integer is zero. Therefore

$$\boxed{\|P(d)\| = |E_{1,\ell+1}|.}$$

The distinction is that **the identity modulo integers alone** would give only

$$\|P(d)\| = \|E_{1,\ell+1}\| \le |E_{1,\ell+1}|.$$

The smallness assumption upgrades the last inequality to equality. For example, without smallness, $P(d) = 0.2$ and $\gamma_{1,\ell+1} = 1$ would give $E_{1,\ell+1} = 1.2$, so the inequality would be strict.

I would therefore replace the lemma's conclusion by

$$\boxed{P(d) + \gamma_{1,\ell+1} = E_{1,\ell+1}, \qquad \|P(d)\| = |E_{1,\ell+1}|.}$$

No change to the proof is needed beyond explicitly noting that the residual lies in $(-1/2, 1/2)$.

98. User

2026-09-17 21:57:54 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-17T13:57:54.903000Z

before equation (7.6), why is tau(p) = g(d/L)?

---

99. Assistant

2026-09-17 21:58:57 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-17T13:58:57.155426Z

Because p is obtained from the weighted-homogeneous matrix g by the affine substitution

$$t \longmapsto \frac{a+dt}{L} = \frac{a}{L} + \frac{d}{L}\,t,$$

and τ extracts, from each matrix entry, the coefficient of the **highest power allowed by that entry's weight**.

More precisely, the proof defines

$$p(t) = g\!\left(\frac{a+dt}{L}\right) = g\!\left(\frac{a}{L} + \frac{d}{L}\,t\right).$$

Earlier it records the identity

$$\tau\big(g(x+ut)\big) = g(u).$$

Taking

$$x = \frac{a}{L}, \qquad u = \frac{d}{L}$$

immediately gives

$$\boxed{\tau(p) = g(d/L).}$$

This is exactly the weighted-homogeneity identity used in that passage. [Pasted text.txt, L286-L317]

Here is the entry-by-entry calculation.

1. The nonzero entries of g

For one special generalized polynomial, $g(s)$ has only first-superdiagonal entries:

$$g(s)_{i,i+1} = \begin{cases} -a_i s^{w_i}, & i < \ell, \\ +a_\ell s^{w_\ell}, & i = \ell, \end{cases}$$

and all entries farther above the diagonal are zero.

The entry $(i, i+1)$ has assigned weight w_i. Therefore

$$\tau(p)_{i,i+1} = [t^{w_i}]\, p_{i,i+1}(t).$$

For $i < \ell$,

$$p_{i,i+1}(t) = -a_i \left(\frac{a}{L} + \frac{d}{L}t\right)^{w_i}.$$

By the binomial theorem,

$$\left(\frac{a}{L}+\frac{d}{L}t\right)^{w_i}=\sum_{r=0}^{w_i}\binom{w_i}{r}\left(\frac{a}{L}\right)^{w_i-r}\left(\frac{d}{L}\right)^{r}t^r.$$

The coefficient of t^{w_i} is obtained only when $r=w_i$, and is

$$\left(\frac{d}{L}\right)^{w_i}.$$

Hence

$$\tau(p)_{i,i+1}=-a_i\left(\frac{d}{L}\right)^{w_i}=g(d/L)_{i,i+1}.$$

For the final superdiagonal entry, the same calculation gives

$$\tau(p)_{\ell,\ell+1}=+a_\ell\left(\frac{d}{L}\right)^{w_\ell}=g(d/L)_{\ell,\ell+1}.$$

All higher off-diagonal entries of both matrices are zero. Thus the complete matrices agree:

$$\tau(p)=g(d/L).$$

2. Why the starting point a disappears

The essential scalar identity is

$$[t^w](x+ut)^w=u^w.$$

Every term involving x has degree strictly less than w in t, so coefficient extraction at degree w removes x.

For example,

$$\left(\frac{a+dt}{L}\right)^2=\frac{a^2}{L^2}+\frac{2ad}{L^2}t+\frac{d^2}{L^2}t^2.$$

Thus

$$[t^2]\left(\frac{a+dt}{L}\right)^2=\frac{d^2}{L^2}.$$

The constant a, which is the starting point of the arithmetic progression, vanishes from the leading coefficient. Only the difference d remains.

This is the whole reason τ is introduced: it transforms the polynomial orbit along

$$a,\ a+d,\ a+2d,\dots$$

into an expression depending only on its common difference.

3. A two-layer example

Suppose

$$P(n)=\alpha n^2\langle\beta n\rangle.$$

Then

$$g(s) = \begin{pmatrix} 1 & -\alpha s^2 & 0 \\ 0 & 1 & \beta s \\ 0 & 0 & 1 \end{pmatrix}.$$

Along the progression, put

$$p(t) = g\left(\frac{a+dt}{L}\right).$$

Then

$$p(t) = \begin{pmatrix} 1 & -\alpha\left(\frac{a+dt}{L}\right)^2 & 0 \\ 0 & 1 & \beta\left(\frac{a+dt}{L}\right) \\ 0 & 0 & 1 \end{pmatrix}.$$

The weights of entries $(1,2)$ and $(2,3)$ are 2 and 1, respectively. Therefore

$$\tau(p)_{12} = [t^2]\left[-\alpha\left(\frac{a+dt}{L}\right)^2\right] = -\alpha\frac{d^2}{L^2},$$
$$\tau(p)_{23} = [t]\left[\beta\left(\frac{a+dt}{L}\right)\right] = \beta\frac{d}{L}.$$

Consequently,

$$\tau(p) = \begin{pmatrix} 1 & -\alpha(d/L)^2 & 0 \\ 0 & 1 & \beta(d/L) \\ 0 & 0 & 1 \end{pmatrix} = g(d/L).$$

Then applying the dilation δ_L multiplies a weight- w entry by L^w, so

$$\delta_L\tau(p) = \delta_L g(d/L) = g(d).$$

That is the next identity used in the proof:

$$\boxed{\tau(p) = g(d/L), \qquad \delta_L\tau(p) = g(d).}$$

So the two operations have complementary roles:

$$\text{coefficient extraction } \tau \quad \text{removes } a,$$

while

$$\text{dilation } \delta_L \quad \text{removes the denominator } L.$$

---

100. User

2026-09-17 22:14:31 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-17T14:14:31.345000Z

I think that the only part of the proof that using 2-colorings and 2-largeness instead of sets of density delta and delta-density largeness is Section 6. Check whether or not I am correct. If I am correct, give me the complete tex code for a replacement of section 6 of the original document to change it to delta-density largeness. The other document that was produced with delta-density largeness went a bit too far and got a bit too fancy with the whole intersecting all colors stuff. I don't need the extra stuff.

101. Assistant

2026-09-17 22:14:53 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-17T14:14:53.881053Z

I'm checking the proof line by line, especially whether monochromaticity appears outside Section 6 and whether a density test set can replace it without altering later sections.

---

102. Assistant

2026-09-17 22:17:59 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-17T14:17:59.770630Z

Your intuition is nearly right: most machinery remains unchanged, but binary coloring also appears in the shell construction and final integer passage. I'll use the minimal cyclic-shell modification.

---

103. Assistant

2026-09-17 22:22:24 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-17T14:22:24.521918Z

Your diagnosis is **almost correct at the level of the proof mechanism, but not literally correct as a matter of editing the document**.

The algebraic material in Sections 2, 4, and 5, together with the nested-bracket decoding in Lemma 7.1, does not use two colors or 2-largeness. The only substantive place where **monochromaticity** is used is the first-exit argument in Section 6: the polynomial lift cannot escape because a normalized excursion would force the sampled orbit to exhibit both shell colors. [elementary_two_color_obstruction_expanded.tex, L494-L584]

However, three additional portions require small changes:

1. **Section 3:** it currently defines the binary shell coloring and proves that a sufficiently fine sample receives both colors. [elementary_two_color_obstruction_expanded.tex, L258-L316]
   For the density version, this is replaced by one periodic family of narrow forbidden barriers.
2. **Section 7:** it currently defines the integer coloring

$$c(n) = \kappa(g(n/L))$$

and applies monochromaticity and $(2, N)$-largeness. [elementary_two_color_obstruction_expanded.tex, L674-L712]
For the density version, one instead defines a family A_θ, averages over the phase θ, selects a phase of sufficiently high upper density, and thins it to upper density exactly δ.

3. The theorem statement, introductory proof outline, and final explanatory paragraph need the corresponding terminology changes.

So **Section 6 is the only part requiring a new mathematical argument**, but replacing Section 6 alone would leave the paper logically incomplete.

The simpler density argument

This version does not use the "all colors must occur" construction.

After obtaining the uniform oscillation

$$\operatorname{osc}_{[0,1]} F(Q) \geq \eta$$

for normalized nonconstant curves, choose

$$\sigma = \frac{\eta}{2}, \qquad 0 < h < \min\left\{\frac{\eta}{10}, \frac{(1-\delta_0)\eta}{4}\right\}.$$

For each phase $\theta \in [0, \sigma)$, retain matrices satisfying

$$F(A) + \theta \in \bigcup_{v \in \mathbb{Z}} [v\sigma + h, (v+1)\sigma).$$

Thus one periodically deletes gaps of width h. The retained proportion is

$$1 - \frac{h}{\sigma} > \delta_0.$$

If all the integer samples of the polynomial lift lie in this retained region, consecutive samples cannot cross a deleted gap. Hence they all remain in one retained component, whose length is $\sigma - h$. But at the first exit, the normalized curve has F-oscillation at least η, and sufficiently fine sampling gives sample oscillation at least

$$\eta - \frac{h}{2} > \sigma - h,$$

a contradiction.

For the final integer test set, define

$$A_\theta = \{n : g(n/L) \text{ lies in the retained region of phase } \theta\}.$$

Averaging in θ gives some A_{θ_0} with upper density at least $1 - h/\sigma > \delta_0$. An elementary deterministic thinning produces a subset of upper density exactly δ_0. Every progression in that subset satisfies the same quantitative estimate as in the original two-color proof.

Files

The complete replacement for Section 6 is here:

Download the replacement Section 6 TeX code

Because the necessary edits are not confined literally to Section 6, I also prepared a minimally modified complete version of the original expanded document:

Download the complete fixed-density TeX document

View the compiled 11-page PDF

The full version proves the following direct quantitative statement:

> For every $\delta_0 \in (0,1)$ and every fixed finite collection of exponent patterns, there are coefficient-independent constants C, N_0 such that, for every corresponding family $P_1, \ldots, P_r$, there is one set $A \subseteq \mathbb{N}$ with
>
> $$\overline{d}(A) = \delta_0$$
>
> and every N-term progression in A, $N \geq N_0$, has difference d satisfying

$$\|P_i(d)\| \le CN^{-\deg P_i}.$$

Consequently, every (δ_0, N)-density-large set contains such a d, and the corresponding bad set is not δ_0-density large.

This proof uses only **one periodically repeated family of barriers, phase averaging, and thinning**—not an r-color omitted-color theorem and not an intersection of color conditions.

---

104. User

2026-09-17 22:27:39 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-17T14:27:39.809000Z

on page 5 of the most recent 11 page document you just produced for me, what is the m in I^m?

---

105. Assistant

2026-09-17 22:28:51 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-17T14:28:51.137201Z

Here m is **the total number of off-diagonal matrix coordinates in the product group** G. It is not the number of special generalized polynomials and not the maximum weight D.

For each expression

$$P_s(n) = a_{s,1}n^{w_{s,1}}\left[a_{s,2}n^{w_{s,2}}[\cdots]\right]$$

with nesting depth ℓ_s, the proof uses one block

$$U_{\ell_s+1}(\mathbb{R}).$$

An $(\ell_s+1)\times(\ell_s+1)$ upper-unitriangular matrix has

$$\binom{\ell_s+1}{2} = \frac{\ell_s(\ell_s+1)}{2}$$

off-diagonal entries. Consequently, for a family $P_1, \dots, P_r$,

$$\boxed{m = \sum_{s=1}^{r}\binom{\ell_s+1}{2}.}$$

This is how m is defined earlier in the document. [Pasted text.txt, L79-L84]

The interval

$$I = [-1/2, 1/2)$$

is the centered fundamental interval used for every off-diagonal coordinate of the canonical representative. Therefore

$$I^m = [-1/2, 1/2)^m$$

is simply the coordinate box containing **all off-diagonal coordinates in all blocks**.

When each copy of I is divided into K subintervals of length $1/K$, the product I^m is divided into

$$R = K^m$$

small boxes. The auxiliary R-coloring records which of these boxes contains the canonical representative.

Example

For one two-layer expression

$$P(n) = a_1 n^{w_1}[a_2 n^{w_2}],$$

we use $U_3(\mathbb{R})$:

$$\begin{pmatrix} 1 & * & * \\ 0 & 1 & * \\ 0 & 0 & 1 \end{pmatrix}.$$

There are three off-diagonal coordinates,

$$(1,2), \quad (1,3), \quad (2,3),$$

so

$$m = 3, \qquad I^m = I^3, \qquad R = K^3.$$

Even though the original band matrix $g(t)$ has only first-superdiagonal entries, we must include coordinates such as $(1,3)$, because matrix multiplication and the lattice correction $C(t)$ generally create nonzero entries on the higher superdiagonals.

A clarifying replacement for the sentence on page 5 would be:

> Let $m = \sum_{s=1}^{r} \binom{\ell_s+1}{2}$ be the total number of off-diagonal coordinates in the product group $G = \prod_{s=1}^{r} U_{\ell_s+1}(\mathbb{R})$. Since every canonical representative has these m coordinates in $I = [-1/2, 1/2)$, its coordinate vector belongs to I^m.

---

106. User

2026-09-17 22:34:04 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-17T14:34:04.415000Z

ah, so G is a finite direct product of r groups of the form U_{ell+1}(mathbb{R}), right?

---

107. Assistant

2026-09-17 22:34:46 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-17T14:34:46.452344Z

Exactly. If the finite family is

$$P_s(t) = a_{s,1} t^{w_{s,1}} \left[a_{s,2} t^{w_{s,2}} [\cdots [a_{s,\ell_s} t^{w_{s,\ell_s}}] \cdots] \right], \qquad 1 \le s \le r,$$

then the proof assigns to P_s one block

$$G_s = U_{\ell_s+1}(\mathbb{R}),$$

and takes

$$\boxed{G = \prod_{s=1}^{r} U_{\ell_s+1}(\mathbb{R}).}$$

Thus G is a finite direct product of r upper-unitriangular groups, one for each special generalized polynomial. The depths ℓ_s need not all be equal. [Pasted text.txt, L79-L84]

Correspondingly,

$$\Gamma = \prod_{s=1}^{r} U_{\ell_s+1}(\mathbb{Z})$$

is the product integer lattice, and the total number of off-diagonal coordinates is

$$m = \sum_{s=1}^{r} \binom{\ell_s + 1}{2}.$$

If every polynomial has the same nesting depth ℓ, this simplifies to

$$G = U_{\ell+1}(\mathbb{R})^r, \qquad m = r\binom{\ell+1}{2}.$$

---

108. User

2026-09-19 21:26:20 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-19T13:26:20.927000Z

in the most recent tex file that you produced for me, what is the point of applying delta_{L/q}? Don't we already have that C \in Gamma, so can't we just work with equation 6.8 instead?

---

109. Assistant

2026-09-19 21:26:54 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-19T13:26:54.195544Z

I'm checking (6.8) carefully, because the key distinction is that $C(n) \in \Gamma$ at integer n does not imply $\tau(C) \in \Gamma$.

---

110. Assistant

2026-09-19 21:29:05 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-19T13:29:05.000673Z

The point is that **the polynomial $C(t)$ is lattice-valued at integers, but its weighted leading-coefficient matrix $\tau(C)$ need not lie in the lattice**. Equation (6.8) involves $\tau(C)$, not an evaluation $C(n)$.

More precisely, the rounding lemma gives

$$C(n) \in \Gamma \qquad (n \in \mathbb{Z}),$$

not that C is a single constant element of Γ. Here $C : \mathbb{R} \to G$ is a polynomial matrix-valued map. After applying τ, all we initially know is

$$w_{ij}!\,\tau(C)_{ij} \in \mathbb{Z}.$$

Thus $\tau(C)$ may have factorial denominators. [Pasted text.txt, L194-L218]

There are therefore **two distinct reasons** for applying $\delta_{L/q}$.

1. It turns $\tau(C)$ into a genuine lattice element

Equation (6.8) is

$$\delta_q \tau(p)\, \tau(C) = \tau(Q). \tag{6.8}$$

To use the last-column decoding lemma, we need an identity of the form

$$g(d)\gamma = E \qquad \text{with } \gamma \in \Gamma.$$

But $\tau(C)$ is not necessarily in Γ.

Scalar example

Take an integer-valued polynomial

$$C(t) = \binom{t}{2} = \frac{t^2 - t}{2}.$$

Then

$$C(n) \in \mathbb{Z} \qquad (n \in \mathbb{Z}),$$

but its degree-two leading coefficient is

$$[t^2]C(t) = \frac{1}{2} \notin \mathbb{Z}.$$

Thus, in a weight-two $U_2(\mathbb{R})$-coordinate,

$$C(\mathbb{Z}) \subseteq \Gamma \qquad \text{but} \qquad \tau(C) \notin \Gamma.$$

This is exactly why the source emphasizes integer-valued polynomials rather than polynomials with integer coefficients.

Now set

$$s = \frac{L}{q}.$$

Since

$$L = D!\,\mathrm{lcm}(1, \dots, W)$$

and $1 \le q \le W$, the integer $s = L/q$ is a multiple of $D!$. If an entry has weight $w = w_{ij} \le D$, write

$$\tau(C)_{ij} = \frac{a}{w!}, \qquad a \in \mathbb{Z}.$$

Then

$$\big(\delta_s \tau(C)\big)_{ij} = s^w \tau(C)_{ij} = a\frac{s^w}{w!} \in \mathbb{Z}.$$

Therefore

$$\boxed{\gamma := \delta_{L/q}\tau(C) \in \Gamma.}$$

So the dilation is what converts the rational leading-coefficient correction into an integer matrix suitable for the rounding lemma. This is the “factorial denominator clearance” step in the proof. [Pasted text.txt, L259-L280]

2. It changes the left-hand side from the auxiliary step q to the desired difference d

There is a second, separate issue.

The integer q comes from the short auxiliary subprogression

$$b,\ b+q,\dots,b+(M-1)q$$

found by finite van der Waerden. It is not the original common difference d. Equation (6.8) contains

$$\delta_q\tau(p),$$

because coefficient extraction from $p(b+qt)$ records the auxiliary step q.

For the original progression

$$a,\ a+d,\dots,a+(N-1)d,$$

we defined

$$p(t)=g\left(\frac{a+dt}{L}\right).$$

Weighted coefficient extraction gives

$$\tau(p)=g(d/L).$$

Consequently,

$$\delta_q\tau(p)=\delta_q g(d/L)=g(qd/L).$$

Thus equation (6.8), even if $\tau(C)$ happened to be integral, would naturally concern

$$g(qd/L),$$

not $g(d)$.

Applying $\delta_{L/q}$ fixes this:

$$\delta_{L/q}\big(\delta_q\tau(p)\big)=\delta_L\tau(p)=\delta_L g(d/L)=g(d).$$

Hence equation (6.8) becomes

$$\boxed{g(d)\gamma=E,\qquad \gamma=\delta_{L/q}\tau(C)\in\Gamma,\qquad E=\delta_{L/q}\tau(Q).}$$

This is exactly the form required to decode the nested nearest integers and obtain information about $P(d)$.

3. The scalar model makes both roles visible

Suppose one coordinate has weight w, and write

$$\theta=[t^w]p(t).$$

The scalar version of equation (6.8) is

$$q^w\theta+\tau(C)=\tau(Q),\qquad w!\tau(C)\in\mathbb{Z}. \tag{1}$$

This is not enough to conclude that $q^w\theta$ is close to an integer, because $\tau(C)$ may equal, say, $1/2$.

Multiply (1) by

$$s^w = \left(\frac{L}{q}\right)^w.$$

Then

$$L^w\theta + s^w\tau(C) = s^w\tau(Q),$$

and now

$$s^w\tau(C) \in \mathbb{Z}.$$

Therefore

$$\|L^w\theta\| \le s^w|\tau(Q)|.$$

For

$$p(t) = \alpha\left(\frac{a+dt}{L}\right)^w,$$

we have

$$\theta = \alpha\left(\frac{d}{L}\right)^w,$$

so

$$L^w\theta = \alpha d^w.$$

Thus the conclusion becomes

$$\|\alpha d^w\| \le \left(\frac{L}{q}\right)^w |\tau(Q)|.$$

This shows the two functions of the dilation simultaneously:

$$\boxed{\begin{array}{c}\text{it clears the factorial denominator in } \tau(C),\\ \text{and it converts the scaled coefficient into the expression at } d.\end{array}}$$

So you are right that equation (6.8) contains essentially all the information. But **applying $\delta_{L/q}$ is the necessary normalization that turns (6.8) into usable arithmetic information**. Working directly with (6.8) would leave a generally nonintegral correction and would concern the auxiliary scale qd/L, rather than the original common difference d.

---

111. User

2026-09-20 11:45:29 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-20T03:45:29.997000Z

I want you to make a complete paper, ready to publish in either Journal of the european mathematical society, or forum of math pi, co-authored by Sohail Farhangi (BIMSA affiliation) and Ryan Alweiss (cambridge affiliation) on the topic of the 2-large is large conjecture. The main result is that 2-large sets are good for nilrecurrence. This answers questions of alweiss, host-kra-maass, is significant progress on the original 2-large is large conjecture of brown-graham-landman, and all of this should be discussed in the introduction. However, the paper should not prove the main result using nilmanifold machinery. The main result should be

proven using direct elementary computations with special generalized polynomials, and it should be the strengthened main result of the first attached note about progressions that avoid one coloring in any finite coloring fail to satisfying some shrinking target problem. Huang-Shao-Ye should be cited to show that our special generalized polynomial result proves the nilrecurrence statements, and we should also cite the foundational papers of Bergelson and Leibman (and maybe just Leibman as well) that established the connection between nilrotations and generalized polynomials. The first main result of the paper should just be stated using 2-vdw largeness (we want to use the 2-vdw large language of the paper of alweiss instead of the 2-large language of the past), and then later in the introduction we mention that we introduce a new but natural density analogue called delta-szemeredi largeness (not delta-density large as we have been doing in this chat thus far, I want to change that name as well), and then state the more general result that implies the vdW large result and the szemeredi large result simultaneously. We want to also ask as a question somewhere in the paper (intro, or at the end, i don't know what is best) whether or not delta-szemeredi largeness for some delta in (0,1) implies delta szemeredi largeness for every delta in (0,1). Please give the final result as a complete tex file along with a bibtex file (do not put the references in the main tex file, I want them in a separate bibtex file).

> **Attachments recorded in source:**
>
> - fixed_density_and_coloring_nilrecurrence.tex
> - Pasted text(20260920-033743).txt

---

112. Assistant

2026-09-20 11:46:08 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-20T03:46:08.645770Z

I'll build the manuscript around the elementary special-polynomial proof, verify the historical attributions, and prepare separate LaTeX and BibTeX files with a compiled PDF for review.

---

113. Assistant

2026-09-20 11:49:12 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-20T03:49:12.660872Z

The attached note already contains the common strengthening: progressions omitting one color force approximation, yielding both the two-color and high-density consequences.

---

114. Assistant

2026-09-20 11:58:17 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-20T03:58:17.649937Z

I'll lead with qualitative nilrecurrence, keep pointwise recurrence distinct from open-set recurrence, and use the common theorem to derive the density and $1+\varepsilon$ extremal consequences.

115. Assistant

2026-09-20 12:02:20 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-20T04:02:20.838020Z

Van der Waerden largeness and nilrecurrence

Ryan Alweiss and Sohail Farhangi

Download the complete LaTeX manuscript
Download the separate BibTeX bibliography

Read the compiled PDF — 15 pages
Download the complete package, including compilation instructions

The paper opens with **Theorem 1.1: every 2-vdW-large set is a set of multiple pointwise recurrence for nilrotations**, with arbitrarily large return times. The introduction places this in the Brown–Graham–Landman conjecture and explains the answers to Alweiss's Conjecture 3.6 and Host–Kra–Maass's Question 7.8, while distinguishing pointwise recurrence from ordinary open-set recurrence. arXiv arXiv

The elementary theorem is **Theorem 1.2**: for every prescribed finite family of special generalized polynomials and every palette size r, there is **one coloring, independent of progression length**, such that every sufficiently long progression omitting a color has difference d satisfying

$$\|P_i(d)\|_{\mathbb{R}/\mathbb{Z}} \le Ck^{-d_i}.$$

Its proof is written out completely through weighted matrix algebra, polynomial rounding, the shell construction, the first-exit argument, factorial-denominator clearance, and exact last-column decoding. The Huang–Shao–Ye characterization enters only afterward, in the passage to nilrecurrence.

The manuscript introduces ** δ-Szemerédi-largeness**, proves its nilrecurrence consequence, and ends with the requested question about whether largeness at one positive density implies largeness at every positive density. A short application section also establishes the coefficient-uniform, same-length extremal comparison

$$r_k(\Omega) \le r_k(\Omega; B_k) \le (1+\varepsilon) r_k(\Omega).$$

The bibliography includes the foundational Bergelson–Leibman and Leibman papers and is entirely separate from the main source. The manuscript compiles without LaTeX or BibTeX warnings, and I checked the rendered pages.

Before submission, you and Ryan should approve the final arguments, contact information, and AI-assistance acknowledgement. The files provide a complete journal-style manuscript, rather than a claim of external referee verification or guaranteed acceptance.

---

116. User

2026-09-20 16:22:27 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-20T08:22:27.011000Z

why did you use <x> for the nearest integer to x instead of [x], and why did you use [t^w] for the coefficient extraction of t^w? Were these personal choices, or were they motivated by the notation in some other paper, perhaps the huang-shao-ye paper? If it was motivated by another paper, please give me pinpoint references to where the notation is used, so that I can also check it.

117. Assistant

2026-09-20 16:22:44 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-20T08:22:44.834690Z

I'll verify both conventions against Huang–Shao–Ye and the current manuscript, distinguishing inherited nearest-integer notation from the standard coefficient-extraction notation and my own editorial choices.

---

118. Assistant

2026-09-20 16:26:02 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-20T08:26:02.904104Z

The two choices had **different motivations**.

1. Nearest integer: $\langle x \rangle$ was an editorial choice, not Huang–Shao–Ye's notation

Huang–Shao–Ye actually use **square brackets** for nearest-integer rounding. On printed page 23, §4.1.1, they define

$$\|a\| = \inf\{|a - n| : n \in \mathbb{Z}\}$$

and then

$$[a] = \min\{m \in \mathbb{Z} : |a - m| = \|a\|\}.$$

Thus their tie-breaking rule chooses the smaller of the two nearest integers. On the same page, equation (4.1) defines

$$L(a_1, \dots, a_\ell) = a_1[L(a_2, \dots, a_\ell)].$$

Then printed page 24, Definition 4.1.2, defines the special generalized polynomials SGP_d using this L-notation. arXiv

So the precise pinpoint references are:

- **Huang–Shao–Ye, printed p. 23, §4.1.1:** definition of $[a]$ as the nearest integer;
- **printed p. 23, equation (4.1):** recursive definition of $L(a_1, \dots, a_\ell)$;
- **printed p. 24, Definition 4.1.2:** definition of SGP_d.

Therefore, the sentence in the manuscript saying

> "We use the nearest-integer convention of Huang, Shao, and Ye: $\langle t \rangle$ is a nearest integer..."

is accurate regarding the **operation and tie-breaking convention**, but not regarding the symbol. The manuscript uses $\langle t \rangle$, whereas Huang–Shao–Ye use $[t]$. The manuscript identifies the special class with their Definition 4.1.2. [Pasted text.txt, L171-L187]

I chose $\langle x \rangle$ in the final manuscript primarily to avoid overloading square brackets. The earlier elementary note used both

$$[x] \quad \text{for nearest-integer rounding}$$

and

$$[t^w]f(t) \quad \text{for coefficient extraction.}$$

Those two meanings are distinguishable from context, but putting both in the same technical proof creates unnecessary visual ambiguity. The earlier note indeed used $[x]$ for nearest integers. [Pasted text.txt, L50-L53] It then also used $[t^{w_{ij}}]$ as the coefficient-extraction operator. [Pasted text.txt, L91-L103]

A secondary reason is that many readers initially read $[x]$ as an ordinary floor, especially before they have absorbed the paper's convention. Angle brackets make it more visually obvious that a special rounding operation is intended.

So, to be completely candid:

$$\boxed{\langle x \rangle \text{ was our expository choice, not notation copied from HSY.}}$$

2. Coefficient extraction: $[t^w]f(t)$ is standard notation, but it did not come from Huang–Shao–Ye

The notation

$$[t^w]f(t)$$

means "the coefficient of t^w in $f(t)$." For example,

$$[t^2](3 + 5t - 7t^2 + t^4) = -7.$$

This is standard coefficient-operator notation in enumerative combinatorics and generating-function literature. A particularly precise reference is Donald Knuth's paper

> **"Bracket notation for the 'coefficient of' operator."**

Its opening sentence says that many authors write

$$[z^n]G(z)$$

for the coefficient of z^n in $G(z)$, and it attributes the notation to Goulden and Jackson, page 1 of their book. arXiv

Thus the pinpoint reference is:

- **Donald E. Knuth, "Bracket notation for the 'coefficient of' operator," opening paragraph**, arXiv:math/9402216;
- Knuth there credits **Goulden and Jackson, *Combinatorial Enumeration*, p. 1**, for the notation $[z^n]G(z)$.

I used it because it makes the key identities compact:

$$[t^w]f(b + qt) = q^w[t^w]f(t),$$

and, for weighted matrix entries,

$$\tau(A)_{ij} = [t^{w_{ij}}]A_{ij}(t).$$

It is particularly convenient when proving

$$\tau(AB) = \tau(A)\tau(B).$$

Huang–Shao–Ye do **not** use this coefficient-extraction operator in their Definitions 4.1.1–4.1.2. It belongs to our matrix proof, not to their generalized-polynomial notation.

3. What I recommend for the paper

There are two clean options.

Option A: retain the current notation

Use

$$\langle x\rangle \quad \text{for nearest integer}, \qquad [t^w]f(t) \quad \text{for coefficient extraction}.$$

Then revise the attribution to say:

```
For \(x\in\mathbb R\), let \(\langle x\rangle\) denote the nearest
integer to \(x\), with ties resolved toward the smaller integer.
This is the rounding convention of Huang--Shao--Ye
\cite[\S4.1.1]{HSY}, who denote the same quantity by \([x]\).
For a polynomial \(f(t)\), the notation \([t^w]f(t)\) denotes the
coefficient of \(t^w\) in \(f(t)\).
```

This is probably the clearest option for readers. It avoids overloading and gives exact credit.

Option B: follow Huang–Shao–Ye's notation exactly

Use

$$[x]$$

for nearest integer, but replace coefficient extraction by something such as

$$\operatorname{coeff}_{t^w} f(t)$$

or define

$$\mathrm{LC}_w(f) := \text{the coefficient of } t^w \text{ in } f.$$

For example:

```
\tau(A)_{ij}
=
\operatorname{coeff}_{t^{w_{ij}}} A_{ij}(t).
```

This makes the generalized-polynomial expressions look exactly like those of Huang–Shao–Ye, at the price of slightly longer matrix formulas.

Recommendation

I would retain

$$\boxed{\langle x\rangle \text{ for rounding and } [t^w] \text{ for coefficient extraction}}$$

because the distinction is visually effective throughout a proof in which the two operations occur repeatedly. But I would explicitly say that Huang–Shao–Ye themselves write $[x]$. Without that sentence, the present wording can naturally be read as attributing the angle-bracket notation to them, which would be inaccurate.

119. User

2026-09-20 16:41:36 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-20T08:41:36.131000Z

In theorem 1.2, did you mean to also assert that every color class in this coloring has natural density 1/r? I think you are implicitly using this in theorem 5.2.

120. Assistant

2026-09-20 16:41:54 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-20T08:41:54.519703Z

I'll check Theorems 1.2 and 5.2 carefully, distinguishing whether equal color-class densities are actually proved or whether Theorem 5.2 only needs averaging and thinning.

121. Assistant

2026-09-20 16:44:04 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-20T08:44:04.419333Z

No. **Theorem 1.2 was not intended to assert that the color classes have natural density $1/r$, and that stronger statement is not proved by the construction.** Theorem 5.2 does not require it.

The coloring theorem only says that there is an r-coloring such that every sufficiently long progression omitting at least one color has a difference satisfying the desired estimates. It makes no density assertion about the individual color classes. [fixed_density_and_coloring_nilrecurrence.tex, L41-L60]

What Theorem 5.2 actually uses

Write

$$C_e = \{n \in \mathbb{N} : c(n) = e\}, \qquad 0 \le e < r.$$

For every N,

$$\sum_{e=0}^{r-1} |C_e \cap [1, N]| = N.$$

Therefore, for each N, at least one color $e = e_N$ satisfies

$$|C_e \cap [1, N]| \le \frac{N}{r}.$$

There are only r possible colors, so one fixed color e has this property for infinitely many N. Along that subsequence,

$$\frac{|(\mathbb{N} \setminus C_e) \cap [1, N]|}{N} = 1 - \frac{|C_e \cap [1, N]|}{N} \ge 1 - \frac{1}{r}.$$

It follows that

$$\boxed{\overline{d}(\mathbb{N} \setminus C_e) \ge 1 - \frac{1}{r}.} \tag{1}$$

This is the elementary finite-palette observation used in Theorem 5.2. It is stated and proved in the longer source note. [fixed_density_and_coloring_nilrecurrence.tex, L165-L180]

Now choose r so that

$$1 - \frac{1}{r} \geq \delta.$$

Let

$$A_e = \mathbb{N} \setminus C_e.$$

Every arithmetic progression lying in A_e omits color e. Therefore Theorem 1.2 applies to it. Since A_e has upper density at least δ, the thinning lemma gives a subset

$$A \subseteq A_e, \qquad \overline{d}(A) = \delta.$$

Every progression in A is still contained in A_e, so it still omits color e. That is the complete deduction.

Thus the logic is

$$\begin{aligned} &\text{arbitrary } r\text{-coloring} \\ &\quad \Longrightarrow \text{complement of some color has upper density at least } 1 - 1/r \\ &\quad \Longrightarrow \text{thin that complement to upper density exactly } \delta. \end{aligned}$$

It does **not** use

$$d(C_e) = \frac{1}{r}$$

for any e, much less for every e.

The constructed coloring need not be balanced

In fact, the coloring produced by the proof can be extremely unbalanced.

Take the degenerate family

$$P_1 = \cdots = P_m = 0$$

while retaining their assigned positive weights. Then the associated matrix polynomial $g(n/L)$ is constant. Consequently, its shell color is constant as well. The color classes then have densities

$$1, 0, \ldots, 0,$$

rather than $1/r$.

Theorem 1.2 remains true in this case because every difference automatically satisfies

$$\|P_i(d)\| = 0.$$

So this is not a defect in the theorem; it shows concretely that balanced color classes are not part of its conclusion.

Even outside the degenerate case, the present proof supplies no reason for the color classes to have equal densities.

Why the upper-density argument must be phrased carefully

One should not argue that “the upper densities of the r color classes sum to one.” That is false in general. Nor is it generally true that

$$\overline{d}(\mathbb{N} \setminus C_e) = 1 - \overline{d}(C_e).$$

The correct identity is

$$\overline{d}(\mathbb{N} \setminus C_e) = 1 - \underline{d}(C_e),$$

where $\underline{d}$ denotes lower density. The subsequence argument above directly proves the needed estimate without assuming any density exists. The source note explicitly warns against subtracting upper densities. [fixed_density_and_coloring_nilrecurrence.tex, L182-L199]

Recommended clarification in the paper

I would add the following sentence immediately after Theorem 1.2:

```
No assertion is made about the densities of the individual color
classes.  For the density application we use only the elementary fact
that, for every \(r\)-coloring \(c\), the complement of some color
class has upper density at least \(1-1/r\); see
Section~\ref{sec:density-nil}.
```

I would also turn the finite-palette observation preceding Theorem 5.2 into a formally numbered lemma. That would make it unmistakable that Theorem 5.2 relies on (1), rather than on an unstated equidistribution or balanced-coloring property.

So there is **no gap in the proof of Theorem 5.2**, but your reading identifies a place where the exposition should be made more explicit.

---

122. User

2026-09-20 17:10:10 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-20T09:10:10.412000Z

in lemma 6.1, does the new k_0 really depend on the choice of c? We explicitly stated earlier that it does not depend on the constants of the polynomial.

---

123. Assistant

2026-09-20 17:10:31 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-20T09:10:31.618477Z

I’ll check Lemma 6.1’s parameter dependencies against the earlier coefficient-uniform theorem, especially whether the auxiliary multiplier involving c genuinely forces k_0 to depend on c.

124. Assistant

2026-09-20 17:12:34 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-20T09:12:34.746588Z

Yes. **The new threshold in Lemma 6.1 really must depend on the prescribed target constant c.** This does not contradict the earlier statement that the threshold is independent of the real coefficients of the special generalized polynomials.

The earlier omitted-color theorem gives constants

$$C_0 = C_0(r, 2m, D), \qquad k_1 = k_1(r, 2m, D),$$

uniformly in all the real coefficients of the $2m$ expressions to which it is applied. [fixed_density_and_coloring_nilrecurrence.tex, L41-L59] In particular, after choosing an integer q, applying that theorem to

$$P_1, \dots, P_m,\ qP_1, \dots, qP_m$$

does **not** make k_1 depend on q. This is precisely where coefficient uniformity is being used.

However, an additional step is needed to convert the estimates for P_i and qP_i into an estimate with the smaller prefactor c.

Where the c-dependence enters

Choose

$$q = \max\left\{1, \left\lceil \frac{C_0}{c} \right\rceil \right\}.$$

The omitted-color theorem gives

$$\|P_i(d)\| \le C_0 k^{-d_i}, \qquad \|qP_i(d)\| \le C_0 k^{-d_i}.$$

Write

$$P_i(d) = z_i + u_i, \qquad z_i \in \mathbb{Z}, \qquad |u_i| = \|P_i(d)\|.$$

We want to conclude

$$\|qP_i(d)\| = q|u_i|.$$

That equality is not automatic: multiplication by q can wrap u_i around the circle. It is valid only when $q|u_i| < 1/2$. The first estimate guarantees this if

$$qC_0 k^{-d_i} \le \frac{1}{4}.$$

Because $d_i \ge 1$, it suffices to require

$$k \ge 4qC_0.$$

Then

$$q|u_i| = \|qP_i(d)\| \le C_0 k^{-d_i},$$

and therefore

$$\|P_i(d)\| \le \frac{C_0}{q} k^{-d_i} \le c k^{-d_i}.$$

Thus an explicit valid choice is

$$\boxed{k_0(r,m,D,c) = \max\left\{k_1(r,2m,D),\ \left\lceil 4C_0(r,2m,D)\max\left(1,\left\lceil\frac{C_0(r,2m,D)}{c}\right\rceil\right)\right\rceil\right\}.}$$

So:

- k_1 from the original theorem is independent of q, hence independent of c;
- the **additional no-wrap threshold** $4qC_0$ depends on c.

There is no contradiction.

Why coefficient uniformity alone does not remove this dependence

Knowing

$$\|qP_i(d)\| \ll k^{-d_i}$$

does not by itself imply that $P_i(d)$ is close to an integer. It might instead be close to

$$\frac{1}{q}, \frac{2}{q}, \dots, \frac{q-1}{q} \pmod 1.$$

The simultaneous estimate for $P_i(d)$ says that it is already close to an integer; the lower bound on k ensures that multiplying this small residual by q does not pass another integer.

So the proof is not treating q as a problematic polynomial coefficient. The original theorem is completely uniform in q. Rather, q appears in a subsequent arithmetic decoding step.

In fact, some c-dependence is mathematically necessary

This is not merely an artifact of the proof. Already take one linear expression

$$P(n) = \alpha n$$

with α irrational, and suppose that there were a universal threshold

$$K = K(2,1,1)$$

working for **every** $c > 0$.

Let $W = W(2,K)$ be a finite van der Waerden number. Then every two-coloring of $[W]$ has a monochromatic K-term progression, whose common difference belongs to

$$\{1,\dots,W\}.$$

Since α is irrational,

$$\mu := \min_{1\le d\le W} \|\alpha d\| > 0.$$

Choose

$$0 < c < K\mu.$$

The hypothetical c-uniform statement would give a two-coloring such that every monochromatic K-term progression has difference d satisfying

$$\|\alpha d\| \le \frac{c}{K} < \mu.$$

But finite van der Waerden supplies such a progression inside $[W]$, and its difference $d \le W$ necessarily satisfies

$$\|\alpha d\| \ge \mu,$$

a contradiction.

Therefore:

$$\boxed{\text{the threshold cannot be chosen independently of } c.}$$

I would clarify Lemma 6.1 by adding:

```
The threshold \(k_1\) supplied by Theorem~\ref{thm:omitted}
is independent of \(q\), since \(q\) enters only as a real
coefficient of the expressions \(qP_i\).  The dependence on \(c\)
in the present lemma arises only from the additional no-wrap
condition \(qC_0k^{-d_i}\leq 1/4\).
```

It may also be worth consistently saying that the original constants are independent of the **real coefficients of the input expressions**, rather than saying simply “independent of the constants,” since the latter could be misread as including the desired output prefactor c.

---

125. User

2026-09-21 14:00:53 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-21T06:00:53.757000Z

stylistically, why did you use rho = 1/4 instead of just writing 1/4?

---

126. Assistant

2026-09-21 14:01:50 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-21T06:01:50.516108Z

The notation

$$\rho = \frac{1}{4}$$

was meant to give a **name to the fixed escape radius** used in the first-exit argument. The value $1/4$ is not itself important; its role is.

It appears in three conceptually linked places:

1. The normalized family of polynomial curves is required to satisfy

$$\|Q(1) - Q(0)\|_\infty = \rho.$$

2. The initial interpolation estimate is arranged so that

$$\|Q(t) - Q(0)\|_\infty < \rho \qquad (0 \le t \le T_0).$$

3. If the lift later escapes, T is defined as the first time at which

$$\|Q(T) - Q(0)\|_\infty = \rho.$$

Thus ρ labels one geometric scale throughout the proof: "the amount of motion constituting an escape." The document first fixes $\rho = 1/4$, defines the compact family using that radius, and later chooses M so that motion on the initial interval remains below the same radius. [Pasted text.txt, L134-L168]

There is also a practical reason for choosing $1/4$. Since $Q(0)$ is a canonical representative with coordinates in

$$[-1/2, 1/2),$$

remaining within distance $1/4$ of $Q(0)$ ensures that every coordinate stays bounded by

$$\frac{1}{2} + \frac{1}{4} = \frac{3}{4} < 1.$$

That places the rescaled first-exit curve inside the normalized family whose coordinates are required to be bounded by 1. This is exactly what is used in the first-exit argument. [Pasted text.txt, L245-L255]

Naming the constant also helps distinguish this $1/4$ from another $1/4$ later in the paper: the bound

$$|E_{i,\ell+1}| < \frac{1}{4}$$

used to decode nearest integers. Those two occurrences have completely different meanings:

- $\rho = 1/4$: a geometric escape radius;
- $|E_{i,\ell+1}| < 1/4$: a rounding-error threshold, which could be replaced by any strict bound below $1/2$.

So the notation was not mathematically necessary, but it was intended to prevent two unrelated constants from looking like the same device.

That said, stylistically, I think either choice is defensible. In the present short proof, I would probably retain ρ, but make its role explicit:

```
Fix the escape radius
\[
\rho:=\frac14.
\]
```

Then later one can say "the first time the lift moves distance ρ" rather than repeatedly saying "the first time the lift moves distance $1/4$."

If you prefer a more concrete and less parameter-heavy presentation, replacing every occurrence of ρ by $1/4$ would not alter the argument at all. My slight preference is to keep the symbol because the same constant controls several choices and because it separates the escape scale from the later rounding threshold.

127. User

2026-09-21 16:52:44 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-21T08:52:44.510000Z

can you give me the tex code for a more detailed proof of lemma 3.4? For example, you could write out what is the lagrange basis since it was not mentioned before. I already back earlier in the paper and defined delta_s instead of just delta.

128. Assistant

2026-09-21 16:53:02 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-21T08:53:02.063867Z

I'll first match Lemma 3.4 to the surrounding notation, then give a drop-in expanded proof defining the Lagrange basis explicitly and using your δ_s notation.

129. Assistant

2026-09-21 16:54:38 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-21T08:54:38.967147Z

Below is replacement code for the entire proof of Lemma 3.4. It uses your earlier notation

$$\Delta_s f(t) = f(t+s) - f(t)$$

and writes out the Lagrange basis and both interpolation calculations explicitly.

```
\begin{proof}
We prove the two assertions separately.

\medskip
\noindent
\emph{Proof of \textup{(i)}.}
For \(0\leq j\leq D\), let
\begin{equation}\label{eq:lagrange-basis}
 \lambda_j(u)
 :=
 \prod_{\substack{0\leq m\leq D\\ m\neq j}}
 \frac{u-m}{j-m}.
\end{equation}
These are the Lagrange basis polynomials for the interpolation nodes
\(0,1,\ldots,D\).  Thus
\[
 \lambda_j(m)=
 \begin{cases}
  1,&m=j,\\
  0,&m\neq j,
 \end{cases}
 \qquad 0\leq j,m\leq D.
\]
Consequently, every real polynomial \(g\) of degree at most \(D\)
satisfies the interpolation identity
\begin{equation}\label{eq:lagrange-interpolation}
 g(u)=\sum_{j=0}^D g(j)\lambda_j(u).
```

```
\end{equation}

Since the finitely many polynomials \(\lambda_j'\) are continuous, the
constant
\[
 A_D
 :=
 \frac12\sum_{j=0}^D
 \max_{0\leq u\leq1}|\lambda_j'(u)|
\]
is finite and depends only on \(D\).

Choose an integer \(H\) such that
\begin{equation}\label{eq:H-interpolation-choice}
 H>\max\left\{1,T_0,\frac{A_DT_0}{\rho}\right\},
\end{equation}
and put
\[
 M:=DH+1.
\]
Because \(H\geq2\) and \(D\geq1\), we have \(M\geq D+2\), as required.

Now let \(f\) be a polynomial of degree at most \(D\) such that
\[
 |f(n)|\leq\frac12
 \qquad(0\leq n<M).
\]
In particular, since
\[
 0,H,2H,\ldots,DH=M-1
\]
are among these integer sample points, we have
\[
 |f(jH)|\leq\frac12
 \qquad(0\leq j\leq D).
\]

Define the rescaled polynomial
\[
 \widetilde f(u):=f(Hu).
\]
It also has degree at most \(D\), and
\[
 |\widetilde f(j)|=|f(jH)|\leq\frac12
 \qquad(0\leq j\leq D).
\]
Applying \eqref{eq:lagrange-interpolation} to \(\widetilde f\) gives
\[
 \widetilde f(u)
 =
 \sum_{j=0}^D f(jH)\lambda_j(u).
\]
Differentiating, we obtain
\[
 \widetilde f'(u)
 =
 \sum_{j=0}^D f(jH)\lambda_j'(u).
\]
```

```
Hence, for \(0\leq u\leq1\),
\begin{align*}
 |\widetilde f'(u)|
 &\leq
 \sum_{j=0}^D |f(jH)|\,|\lambda_j'(u)|\\
 &\leq
 \frac12\sum_{j=0}^D
 \max_{0\leq v\leq1}|\lambda_j'(v)|\\
 &=A_D.
\end{align*}

If \(0\leq t\leq T_0\), then \(0\leq t/H\leq1\), by
\eqref{eq:H-interpolation-choice}.  The mean value theorem therefore
gives
\begin{align*}
 |f(t)-f(0)|
 &=
 \left|
 \widetilde f\left(\frac{t}{H}\right)-\widetilde f(0)
 \right|\\
 &\leq
 A_D\frac{t}{H}\\
 &\leq
 \frac{A_DT_0}{H}\\
 &<\rho.
\end{align*}
This proves \eqref{eq:initial-motion}.

\medskip
\noindent
\emph{Proof of \textup{(ii)}.}
Let
\[
 a_w:=[t^w]f(t)
\]
be the coefficient of \(t^w\) in \(f\), and put
\[
 s:=\frac{J}{w}.
\]
Then
\[
 0,s,2s,\ldots,ws=J
\]
all belong to \([0,J]\).

Recall that
\[
 \Delta_s f(t)=f(t+s)-f(t).
\]
Iterating this identity gives
\begin{equation}\label{eq:iterated-finite-difference}
 \Delta_s^w f(0)
 =
 \sum_{j=0}^w
 (-1)^{w-j}\binom{w}{j}f(js).
\end{equation}
Since \(|f(js)|\leq B\) for \(0\leq j\leq w\), it follows that
\begin{align}
```

```
 |\Delta_s^w f(0)|
 &\leq
 \sum_{j=0}^w\binom{w}{j}|f(js)|\notag\\
 &\leq
 B\sum_{j=0}^w\binom{w}{j}\notag\\
 &=2^wB.
 \label{eq:difference-upper-bound}
\end{align}

On the other hand, taking one \(s\)-difference lowers the degree of a
nonconstant polynomial by one. More precisely,
\[
 \Delta_s(t^m)
 =
 (t+s)^m-t^m
 =
 ms\,t^{m-1}
 +\text{a polynomial of degree at most \(m-2\)}.
\]
It follows by iteration that
\[
 \Delta_s^w(t^w)=w!s^w,
\]
whereas
\[
 \Delta_s^w(t^m)=0
 \qquad(0\leq m<w).
\]
Therefore, for every polynomial \(f\) of degree at most \(w\),
\begin{equation}\label{eq:leading-coefficient-difference}
 \Delta_s^w f(0)=w!s^w[t^w]f(t)=w!s^wa_w.
\end{equation}
This identity also covers the case \(\deg f<w\), when both sides are
zero.

Combining \eqref{eq:difference-upper-bound} and
\eqref{eq:leading-coefficient-difference}, we obtain
\[
 w!s^w|a_w|\leq2^wB.
\]
Since \(s=J/w\), this gives
\begin{align*}
 |[t^w]f(t)|
 =|a_w|
 &\leq
 \frac{2^wB}{w!s^w}\\
 &=
 \frac{2^wB}{w!}
 \left(\frac{w}{J}\right)^w\\
 &=
 \frac{2^ww^w}{w!}\,B J^{-w}\\
 &=b_wBJ^{-w}.
\end{align*}
This proves \eqref{eq:coefficient-bound}.
\end{proof}
```

A minor benefit of this version is that it makes clear that the two parts use **different interpolation mechanisms**:

- Part (i) uses Lagrange interpolation at the widely spaced integer points

$$0, H, \dots, DH$$

to control motion near 0.

- Part (ii) uses the exact w-fold finite-difference identity at the equally spaced points

$$0, \frac{J}{w}, \dots, J$$

to isolate the leading coefficient.

---

130. User

2026-09-21 17:04:53 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-21T09:04:53.692000Z

this is way too detailed. I wanted a proof that was at best twice as long as the original.

---

131. Assistant

2026-09-21 17:05:21 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-21T09:05:21.780187Z

I'll replace it with a compact proof, adding only the missing Lagrange-basis details and retaining your Δ_s notation.

---

132. Assistant

2026-09-21 17:06:16 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-21T09:06:16.034013Z

You are right; the previous version expanded every elementary step unnecessarily. Here is a replacement proof only modestly longer than the original. It writes out the Lagrange basis and uses your previously defined operator Δ_s, while keeping the proof compact. The level of detail now matches the surrounding interpolation arguments in the draft. [Pasted text(20260920-033743).txt, L14-L32]

```
\begin{proof}
For \textup{(i)}, let
\[
 \lambda_j(u)
 =
 \prod_{\substack{0\leq m\leq D\\m\neq j}}
 \frac{u-m}{j-m},
 \qquad 0\leq j\leq D,
\]
be the Lagrange basis for the nodes \(0,\ldots,D\), and choose
\[
 B_D>\frac12\sum_{j=0}^D\sup_{0\leq u\leq1}|\lambda_j'(u)|.
\]
Choose an integer
\[
 H>\max\{1,T_0,B_DT_0/\rho\},
```

```
 \qquad M=DH+1.
\]
If \(\deg f\leq D\) and \(|f(n)|\leq1/2\) for \(0\leq n<M\), then
Lagrange interpolation applied to \(\widetilde f(u)=f(Hu)\) gives
\[
 \widetilde f(u)=\sum_{j=0}^D f(jH)\lambda_j(u),
 \qquad
 \sup_{0\leq u\leq1}|\widetilde f'(u)|\leq B_D.
\]
Hence, for \(0\leq t\leq T_0\),
\[
 |f(t)-f(0)|
 =
 \left|\widetilde f(t/H)-\widetilde f(0)\right|
 \leq \frac{B_DT_0}{H}<\rho.
\]

For \textup{(ii)}, put \(s=J/w\). By the formula for an iterated
finite difference,
\[
 \Delta_s^w f(0)
 =
 \sum_{j=0}^w(-1)^{w-j}\binom{w}{j}f(js),
\]
and therefore
\[
 |\Delta_s^w f(0)|\leq 2^wB.
\]
On the other hand, \(\Delta_s^w\) annihilates every polynomial of
degree less than \(w\), while
\[
 \Delta_s^w(t^w)=w!s^w.
\]
Consequently,
\[
 \Delta_s^w f(0)=w!s^w[t^w]f(t),
\]
and hence
\[
 |[t^w]f(t)|
 \leq \frac{2^wB}{w!s^w}
 =\frac{2^ww^w}{w!}\,BJ^{-w}
 =b_wBJ^{-w}.
\]
\end{proof}
```

133. User

2026-09-21 17:21:42 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-21T09:21:42.474000Z

is the lean library for combinatorics and dynamics sufficiently well developed that you or codex could produce a lean certificate for this paper?

134. Assistant

2026-09-21 17:22:09 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-21T09:22:09.855102Z

I'll separate the elementary matrix/combinatorial core from the nilsystem deduction, then compare each dependency with current Lean/mathlib infrastructure and identify what would require new formalization.

---

135. Assistant

2026-09-21 17:25:18 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-21T09:25:18.726128Z

The elementary special-generalized-polynomial theorem looks realistically formalizable; the likely bottlenecks are van der Waerden/Szemerédi infrastructure and especially the Huang–Shao–Ye nilsystem realization step.

---

136. Assistant

2026-09-21 17:27:28 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-21T09:27:28.776258Z

Assessment

Not yet for a fully self-contained Lean certificate of the entire paper. However:

- **A kernel-checked formalization of the elementary special-generalized-polynomial theorem looks feasible with current mathlib**, although it would be a substantial bespoke project rather than a routine translation.
- **A full certificate of the headline nilrecurrence theorem would presently require formalizing the Huang–Shao–Ye bridge—and probably a significant amount of nilsystem theory—from scratch.** The existing dynamics library is not sufficiently developed for that part.
- Codex could materially help write the elementary formalization, but I would not trust an unattended, one-shot formalization. It would need continual compilation, theorem decomposition, and mathematical supervision.

The distinction matters because the paper contains two logically different layers. Its new elementary core concerns finite colorings, special generalized polynomials, weighted unitriangular matrices, finite differences, compactness, and interpolation. [Pasted text.txt, L46-L74] The headline conclusion is then obtained by invoking the established characterization of nil-Bohr neighborhoods and pointwise nilrecurrence through Huang–Shao–Ye and Host–Kra–Maass. [Pasted text.txt, L1280-L1290] The full theorem package includes both the nilrecurrence statement and the coefficient-uniform finite approximation result. [Pasted text.txt, L63-L78]

What mathlib already provides

1. Ramsey-theoretic infrastructure is reasonably strong

Mathlib contains a formal proof of Hales–Jewett and deduces a general form of van der Waerden's theorem: every finite coloring of a commutative monoid contains a monochromatic homothetic copy of any prescribed finite set. Lean Prover Community

The actual theorem is already available as

```
Combinatorics.exists_mono_homothetic_copy
```

and Hales–Jewett itself is formalized in a genuinely finitary form over finite types. Lean Prover Community

There is one inconvenience: the documentation still lists a **finitary interval version of van der Waerden's theorem** as a TODO. Lean Prover Community Your proof needs a number $W = W(R, M)$ such that every R-coloring of $\{0, \dots, W-1\}$ contains an M-term progression. That statement could be extracted directly from the finitary Hales–Jewett theorem, or derived from the existing infinitary van der Waerden theorem by a compactness argument, but the desired wrapper theorem is apparently not already packaged.

So this part is not a serious conceptual obstacle, but it would require some library work.

2. The basic algebra and analysis are present

Mathlib has an extensive univariate-polynomial API, including coefficients, degrees, evaluation, products, finite supports, and polynomial algebra. Lean Prover Community It also has the general matrix, finite-dimensional linear algebra, real topology, compactness, continuity, and integer arithmetic needed for the proof.

Consequently, the following ingredients are all formalizable without developing new foundational mathematics:

$$\Delta_s^w f(0) = w! s^w [t^w] f,$$

Lagrange interpolation,

boundedness of coefficients on a compact interval,

the extreme-value theorem,

uniform continuity on compact sets,

and finite combinatorial pigeonhole arguments.

Most of those facts would still need lemmas tailored to your exact use, but they sit on mature foundations.

3. There is some dynamics and abstract nilpotent-group theory

Mathlib formalizes measure-preserving and conservative systems, including several versions of Poincaré recurrence. Lean Prover Community It also has a substantial API for **abstract nilpotent groups**, their upper and lower central series, nilpotency classes, subgroups, quotients, and products. Lean Prover Community

But those two resources do not combine into anything close to the theory needed here. Abstract nilpotent groups are not the same as nilpotent Lie groups with lattices and compact homogeneous quotients.

A useful indicator of the current level of the dynamics library is that the circle-dynamics documentation still lists

> proving that irrational rotation orbits are dense

as a future task. Lean Prover Community

That is far below the Huang–Shao–Ye theory of generalized polynomials, nil-Bohr families, and nilsystem return-time sets.

What would need to be developed for the elementary theorem

The elementary proof is unusually well suited to formalization because it avoids Green–Tao factorization, Mal'cev coordinates, quantitative equidistribution, and general nilmanifold geometry. Nevertheless, most of its specialized objects are not already in mathlib.

A serious Lean development would need custom modules for the following.

Special generalized polynomials

One would define an inductive syntax recording the fixed nesting pattern, for example

```
inductive SpecialGP
  | monomial : ℝ → ℕ → SpecialGP
  | nest : ℝ → ℕ → SpecialGP → SpecialGP
```

together with:

- evaluation on $\mathbb{N}$ or $\mathbb{R}$;
- assigned weight;
- the specified nearest-integer tie convention;
- suffix expressions used in last-column decoding.

The theorem should be stated with the exponent patterns fixed, so coefficient uniformity is visible in the Lean quantifiers.

Weighted upper-unitriangular matrices

One would need a convenient representation of

$$U_{\ell+1}(\mathbb{R}), \qquad U_{\ell+1}(\mathbb{Z}),$$

their finite products, the entry weights

$$w_{ij} = w_i + \cdots + w_{j-1},$$

the class $\mathcal{H}$ of polynomial matrices satisfying

$$\deg A_{ij} \le w_{ij},$$

and the operations

$$(\delta_u A)_{ij} = u^{w_{ij}} A_{ij}, \qquad \tau(A)_{ij} = [t^{w_{ij}}] A_{ij}(t).$$

The central exact identities

$$\tau(AB) = \tau(A)\tau(B), \qquad \tau(A(b+qt)) = \delta_q \tau(A)$$

should be quite tractable once the definitions are chosen well. Your paper's weighted-matrix component is finite-dimensional and explicit. [Pasted text.txt, L77-L107]

Canonical reduction and the shell function

This is probably the most technically demanding part of the elementary formalization.

One must construct the canonical right-lattice representative recursively, prove uniqueness, and define

$$F(A) = \min_{\gamma \in \Gamma} \sum_{i<j} (A\gamma)_{ij}^2.$$

Then one must prove:

- the minimum is attained;
- only finitely many lattice elements can minimize on a fixed compact set;
- F is continuous;
- if $Q(t)$ is a nonconstant polynomial matrix, then $F(Q(t))$ is not constant on an interval;
- the normalized family of curves is compact;
- the oscillation of $F \circ Q$ has a uniform positive lower bound.

None of these arguments is beyond mathlib's foundations, but encoding the locally finite minimum over the integer lattice and proving continuity cleanly would require care. The paper's shell construction and compactness argument occupy a genuinely nontrivial part of the formalization. [Pasted text.txt, L109-L159]

Polynomial rounding

The polynomial-rounding lemma requires formalizing:

- coordinatewise finite differences;
- the "small integer is zero" step;
- Newton interpolation;
- integer-valued polynomials in the binomial basis;
- the factorial-denominator conclusion

$$w! \, [t^w] C(t) \in \mathbb{Z}.$$

This is algebraically elementary but likely to create a fair amount of Lean code because one must manage casts between $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$.

First exit and decoding

The remaining argument needs:

- a first-exit time for a continuous polynomial curve;
- rescaling that curve into the compact normalized family;
- the sampling contradiction;
- the coefficient estimate;
- denominator clearance under $\delta_{L/q}$;
- exact last-column recovery of the nested nearest integers.

These are all finite computations and compactness arguments. The last-column lemma should be among the easiest parts once the matrix multiplication formulas have been established. [Pasted text.txt, L222-L349]

The real obstacle: the Huang–Shao–Ye bridge

A full proof of

$$2\text{-vdW-large} \Longrightarrow \text{multiple pointwise nilrecurrence}$$

needs a formal version of something like:

> Every finite-step nil-Bohr neighborhood contains a simultaneous small-value set for finitely many special generalized polynomials.

That is the Huang–Shao–Ye input used by the paper. As far as I could determine from the current mathlib documentation, there is no developed API for:

- nilmanifolds G/Γ;
- lattices in nilpotent Lie groups;
- nilsystems and nilrotations;
- nil-Bohr families;
- generalized polynomials in the Bergelson–Leibman sense;
- the Huang–Shao–Ye equivalence of the nil-Bohr and generalized-polynomial families.

Mathlib's abstract nilpotent-group theory does not provide those objects, and the present state of circle dynamics indicates how much topological-dynamical infrastructure would still be needed. Lean Prover Community Lean Prover Community

There are therefore three possible meanings of "Lean certificate":

1. **Certificate for the new elementary theorem:** realistic.
2. **Certificate for nilrecurrence conditional on a stated HSY theorem:** easy once the elementary theorem is formalized, but it contains an axiom or unformalized assumption.
3. **Fully self-contained certificate for the entire paper:** requires formalizing the relevant part of Huang–Shao–Ye and substantial nilsystem infrastructure.

A Lean file containing

```
axiom hsy_bridge : ...
```

would compile, and the rest of the deduction would be kernel checked, but it would not be a complete certificate. Running `#print axioms` would reveal that the main theorem still depends on `hsy_bridge`.

Could Codex produce it?

Codex could probably produce a large fraction of the elementary formalization with expert supervision. In particular, it should be useful for:

- translating finite matrix calculations;
- finding existing polynomial and compactness lemmas;
- generating repetitive coordinate proofs;
- repairing type coercions;
- building the finite-difference and last-column APIs;
- iterating against Lean compiler errors.

What I would not rely on Codex to do unattended is choose the entire architecture correctly and then emit thousands of lines that compile on the first attempt. The hard parts are not only proving individual facts; they are choosing definitions that make all later facts usable.

A credible project structure would be:

```
VdWLargeness/
  VanDerWaerdenFinite.lean
  SpecialGP/Basic.lean
  WeightedUT/Basic.lean
  WeightedUT/LeadingCoeff.lean
  Shell/CanonicalRep.lean
  Shell/Oscillation.lean
  PolynomialRounding.lean
  OmittedColor.lean
  SzemerediLarge.lean
  NilBohrBridge.lean
  Main.lean
```

The first eight files could yield a genuinely self-contained certificate of the elementary omitted-color and shrinking-target theorems. `NilBohrBridge.lean` is where the present library gap would become explicit.

I would expect the elementary certificate to be a **multi-file development of several thousand lines**, not a lightly annotated transcription of the paper. The full HSY formalization would be a separate major project.

Practical recommendation

The strongest realistic first target is:

> **Formally certify the omitted-color special-generalized-polynomial theorem, including its quantitative k^{-d_i} estimate and the δ-Szemerédi-large corollary.**

That would certify the novel technical contribution of the paper without waiting for a general nilsystem library. The final nilrecurrence deduction could initially be placed in a small file depending explicitly on an `HSYBridge` assumption, making the exact remaining dependency transparent.

After that, the bridge could be replaced incrementally by a formal development of the relevant Huang–Shao–Ye theorem.

One practical limitation in this chat is that the current runtime has neither `lean` nor `lake` installed. I therefore could draft Lean source here, but I could not honestly label it a certificate until it had been compiled by Lean's kernel in an appropriate project environment.

So the bottom line is:

Elementary new theorem: feasible with current mathlib and substantial work; entire nilrecurrence paper: not yet supported without formalizing HSY.

137. User

2026-09-22 16:39:25 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-22T08:39:25.356000Z

can we maybe showed that if f:mathbb{R} to mathbb{R} is a tempered function and S subseteq mathbb{N} is 2-large, then for every epsilon > 0 there is a s in S for which f(s) mod 1 in (-epsilon,epsilon)?

---

138. Assistant

2026-09-22 16:39:46 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-22T08:39:46.613871Z

I'll first pin down the relevant notion of "tempered," then test simple cases and determine whether the polynomial argument extends or a counterexample appears.

---

139. Assistant

2026-09-22 16:42:52 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-22T08:42:52.913464Z

The square-root case already appears to refute the extension; I'm checking two candidate vdW-large witnesses and verifying that $\sqrt{x}$ satisfies the intended tempered hypotheses.

---

140. Assistant

2026-09-22 16:46:18 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-22T08:46:18.697962Z

No—the proposed extension is false, even for $f(x)=\sqrt{x}$, and even if S is vdW-large for every number of colors. In fact, the counterexample can also be chosen to be δ-Szemerédi-large for every $\delta \in (0,1)$.

The obstruction is that a tempered function can stay uniformly away from integers on arbitrarily long intervals of integer inputs. Those intervals already make the bad set sufficiently rich to be vdW-large.

1. The counterexample

Take

$$f(x)=\sqrt{x} \qquad (x \geq 1),$$

and extend it smoothly to $\mathbb{R}$. The extension can even be chosen with $f(0)=0$; nothing below depends on its values below 1.

This is tempered in the usual ergodic-theoretic sense. Namely, a function g is tempered of degree $i \geq 0$ when

$$g^{(i+1)}(x) \longrightarrow 0 \quad \text{monotonically}, \qquad x|g^{(i+1)}(x)| \longrightarrow \infty.$$

Degree-zero tempered functions are also called **Fejér functions**. This definition appears in §1.2 of Koutsogiannis, *Multiple ergodic averages for tempered functions*. arXiv

Here

$$f'(x) = \frac{1}{2\sqrt{x}} \downarrow 0, \qquad xf'(x) = \frac{\sqrt{x}}{2} \longrightarrow \infty.$$

Now define

$$\boxed{S = \left\{ n \in \mathbb{N} : \{\sqrt{n}\} \in \left[\frac{1}{4}, \frac{3}{4}\right] \right\}.} \tag{1}$$

By definition,

$$\boxed{\|f(s)\|_{\mathbb{R}/\mathbb{Z}} \geq \frac{1}{4} \qquad (s \in S).} \tag{2}$$

Thus your desired conclusion fails for $\varepsilon = 1/4$.

What remains is to verify that ** S is genuinely a set of allowed differences with the required largeness property**, not merely a set containing long progressions.

2. The set S is thick, and every thick set is vdW-large

A set is *thick* if it contains intervals of consecutive integers of arbitrarily large length.

For every integer $m \geq 1$,

$$\left[\left(m + \frac{1}{4}\right)^2, \left(m + \frac{3}{4}\right)^2\right] \cap \mathbb{N} \subseteq S.$$

The underlying real interval has length

$$\left(m + \frac{3}{4}\right)^2 - \left(m + \frac{1}{4}\right)^2 = m + \frac{1}{2} \longrightarrow \infty.$$

Hence S is thick.

Here is a complete proof of the relevant combinatorial fact.

Lemma

Every thick set $S \subseteq \mathbb{N}$ is vdW-large.

Proof

First construct positive integers $x_1, x_2, \ldots$ such that

$$\mathrm{FS}(x_j) := \left\{ \sum_{j \in F} x_j : F \subseteq \mathbb{N} \text{ finite and nonempty} \right\} \subseteq S. \tag{3}$$

Choose $x_1 \in S$. Suppose $x_1, \ldots, x_m$ have been chosen and all their nonempty finite sums lie in S. Put

$$M = x_1 + \cdots + x_m.$$

Thickness supplies an integer interval

$$[b, b + M] \subseteq S.$$

Set $x_{m+1} = b$. Every new sum involving x_{m+1} belongs to this interval, so the induction continues.

Now fix an r-coloring $c : \mathbb{N} \to \{1, \dots, r\}$ and a desired progression length $k \geq 2$. Choose m large enough for the finite Hales–Jewett theorem with r colors and an alphabet of size k. That theorem guarantees a monochromatic combinatorial line in every r-coloring of $\{0, \dots, k-1\}^m$. arXiv

Color a word $w = (w_1, \dots, w_m)$ by

$$\tilde{c}(w) = c\left(1 + \sum_{j=1}^{m} w_j x_j\right).$$

A monochromatic combinatorial line has a nonempty set F of variable coordinates, all carrying the same symbol t, and fixed symbols w_j outside F. Its image is

$$1 + \sum_{j \notin F} w_j x_j + t \sum_{j \in F} x_j, \qquad t = 0, \dots, k-1.$$

This is a monochromatic k-term arithmetic progression with common difference

$$d = \sum_{j \in F} x_j \in S$$

by (3). Since r, k were arbitrary, S is vdW-large. □

Combining the lemma with (1)–(2) proves

$$\boxed{S \text{ is vdW-large, but } \inf_{s \in S} \|\sqrt{s}\| \geq \frac{1}{4}.}$$

3. Restricting to superlinear tempered functions does not fix the statement

There is also a counterexample for

$$f(x) = x^{3/2}.$$

It is tempered of degree 1, since

$$f''(x) = \frac{3}{4\sqrt{x}} \downarrow 0, \qquad x f''(x) = \frac{3}{4}\sqrt{x} \longrightarrow \infty.$$

I claim that

$$\boxed{S_{3/2} := \{n \in \mathbb{N} : \|n^{3/2}\| \geq 1/4\}}$$

is thick as well.

For a large integer m, put

$$h_m = \left\lfloor \sqrt{\frac{8m}{3}} \right\rfloor.$$

Consider the consecutive integers

$$n = 4m^2 + h_m + j, \qquad 0 \leq j \leq \lfloor m^{1/4} \rfloor.$$

Writing $h = h_m + j$, Taylor expansion gives

$$(4m^2 + h)^{3/2} = 8m^3 + 3mh + \frac{3h^2}{16m} + O\left(\frac{h^3}{m^3}\right). \tag{4}$$

The first two terms are integers. Moreover, uniformly over the indicated range of j,

$$h = \sqrt{\frac{8m}{3}} + O(m^{1/4}),$$

so

$$\frac{3h^2}{16m} = \frac{1}{2} + O(m^{-1/4}), \qquad \frac{h^3}{m^3} = O(m^{-3/2}).$$

Consequently,

$$n^{3/2} = \text{an integer} + \frac{1}{2} + O(m^{-1/4})$$

uniformly throughout this interval.

For sufficiently large m, the whole interval lies in $S_{3/2}$. Its length tends to infinity. Thus $S_{3/2}$ is thick, hence vdW-large.

So the failure is not confined to the sublinear case.

4. These examples even satisfy every positive-density largeness condition

There is a stronger standard consequence of thickness:

> Every thick set $S \subseteq \mathbb{N}$ is δ-Szemerédi-large for every $0 < \delta < 1$.

To see this, let $A \subseteq \mathbb{N}$ have positive upper density and fix k. The multiple recurrence theorem, together with Furstenberg's correspondence principle, implies that

$$R_k(A) = \{d \in \mathbb{N} : A \text{ contains a } k\text{-term AP of difference } d\}$$

is syndetic. The required syndetic multiple-recurrence statement is recorded, in greater polynomial generality, as Theorem 1.5 of Bergelson–Moreira–Richter, *Single and Multiple Recurrence Along Non-polynomial Sequences*. arXiv

A thick set meets every syndetic set: if every interval of length M meets $R_k(A)$, choose such an interval contained in S. Therefore

$$S \cap R_k(A) \neq \varnothing.$$

This gives the required progression in A.

Thus our counterexamples satisfy substantially more than your hypothesis:

$$\boxed{\begin{gathered} S \text{ is vdW-large and } \delta\text{-Szemerédi-large for every } \delta \in (0,1), \\ \text{but } \|f(s)\| \geq 1/4 \quad \text{for every } s \in S. \end{gathered}}$$

5. Why the special-generalized-polynomial proof does not extend

Your elementary proof uses much more than controlled growth or small higher derivatives. It uses the exact algebraic identities

$$\Delta^{w+1} p = 0$$

for ordinary polynomial coordinates, and

$$\tau\big(g(x+ut)\big) = g(u),$$

which removes the starting point x and produces the expression evaluated at the common difference. These are the polynomial-rounding and weighted-homogeneity steps in the draft. [Pasted text.txt, L202-L213] [Pasted text.txt, L297-L301]

For tempered functions, a Taylor approximation along $a+td$ generally involves derivatives evaluated near a; it does not recover $f(d)$. More decisively, the examples above show that **no alternative proof can establish the proposed assertion for all tempered functions**.

The distinction is therefore structural, not just technical: **tempered growth permits thick sets of inputs on which the fractional part stays away from zero, whereas your normalized generalized-polynomial theorem rules out vdW-largeness of precisely such bad sets.**

---

141. User

2026-09-22 17:31:58 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-22T09:31:58.934000Z

is there a reasonable definition of generalized polynomials on other abelian groups, such as direct sum of infinitely many copes of Z/2Z?

---

142. Assistant

2026-09-22 17:32:17 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-22T09:32:17.803409Z

I'll compare generalized-polynomial notions on abelian groups, focusing on torsion and how they differ from the real-valued bracket polynomials used in your paper.

---

143. Assistant

2026-09-22 17:35:22 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-22T09:35:22.655739Z

The target is crucial: real-valued finite-difference polynomials collapse on torsion groups, while circle-valued ones remain rich. Scalar progressions also behave degenerately, affecting recurrence analogues.

---

144. Assistant

2026-09-22 17:38:04 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-22T09:38:04.106384Z

Yes. For

$$V = \bigoplus_{j=1}^{\infty} \mathbb{Z}/2\mathbb{Z} = \mathbb{F}_2^{(\mathbb{N})},$$

there is a well-developed theory of circle-valued polynomial maps, usually called non-classical polynomials. Their exponentials are called *phase polynomials*. These are the first objects I would consider for extending your recurrence questions to V. The terminology "non-classical" includes the classical polynomials as a subclass. arXiv

However, there is an important distinction:

> **Non-classical polynomials are a natural finite-characteristic analogue of polynomial phases, but they are not obtained simply by replacing $\mathbb{Z}$ by V in the real-valued bracket-polynomial definition.**

I will explain the definition, give explicit examples, and then point out a particularly relevant obstruction to extending your 2-vdW-largeness theorem.

1. Why the direct real-valued definition becomes trivial

Your manuscript's generalized polynomials start with real monomials and close under addition, multiplication, scalar multiplication, and nearest-integer rounding. [Pasted text.txt, L1542-L1556]

On an arbitrary abelian group G, a natural substitute for an ordinary polynomial is a function $p : G \to \mathbb{R}$ satisfying

$$\Delta_{h_1} \cdots \Delta_{h_{d+1}} p(x) = 0 \qquad (x, h_1, \ldots, h_{d+1} \in G),$$

where

$$\Delta_h p(x) = p(x+h) - p(x).$$

But on V, **every such real-valued polynomial is constant**.

Indeed, because $h + h = 0$,

$$\begin{aligned} \Delta_h^2 p(x) &= p(x+2h) - 2p(x+h) + p(x) \\ &= 2p(x) - 2p(x+h) \\ &= -2\Delta_h p(x). \end{aligned}$$

Thus

$$\Delta_h^{d+1} p(x) = (-2)^d \Delta_h p(x).$$

If p has degree at most d, the left-hand side is zero. Since the target is $\mathbb{R}$,

$$\Delta_h p(x) = 0$$

for every x, h, proving that p is constant.

Consequently, starting with these real-valued polynomials and applying finitely many arithmetic and rounding operations produces only constants. With a zero-at-the-origin normalization, it produces only the zero function.

The problem is not the finite-difference definition; it is the torsion-free target $\mathbb{R}$.

2. The useful definition: take values in $\mathbb{T} = \mathbb{R}/\mathbb{Z}$

For an abelian group G, define

$$\mathrm{Poly}_{\le d}(G, \mathbb{T}) = \{P : G \to \mathbb{T} : \Delta_{h_1} \cdots \Delta_{h_{d+1}} P(x) = 0 \text{ for all } x, h_1, \ldots, h_{d+1} \in G\}. \tag{1}$$

All subtraction and equality here take place **in** $\mathbb{T}$.

For finite-dimensional $\mathbb{F}_p$-vector spaces, this is exactly **Definition 1.2** in Tao–Ziegler, *The inverse conjecture for the Gowers norm over finite fields in low characteristic*. The same formula defines the class on the direct sum V; equivalently, every restriction to a finite-dimensional subspace must have degree at most d, with the same d throughout. arXiv

The exponential

$$\phi(x) = e(P(x)) = \exp(2\pi i P(x))$$

then satisfies

$$\partial_{h_1} \cdots \partial_{h_{d+1}} \phi(x) = 1, \qquad \partial_h \phi(x) = \phi(x+h)\overline{\phi(x)}.$$

This is the phase-polynomial formulation; the corresponding definition for actions of arbitrary countable abelian groups appears in **Definition 1.13** of Bergelson–Tao–Ziegler. arXiv

Example A: linear polynomials

Write $x = (x_1, x_2, \ldots)$, with $x_j \in \{0,1\}$ and finite support. For any fixed sequence $a_j \in \{0,1\}$,

$$P(x) = \frac{1}{2} \sum_j a_j x_j \pmod 1$$

is a homomorphism $V \to \mathbb{T}$, hence has degree at most one.

The sum is well-defined even when infinitely many a_j are nonzero, because each input x has finite support.

Example B: a classical quadratic

The function

$$Q(x) = \frac{1}{2} \sum_{j \geq 1} x_{2j-1} x_{2j} \pmod 1 \tag{2}$$

has degree two.

It is the usual $\mathbb{F}_2$-valued quadratic form

$$q(x) = \sum_{j \geq 1} x_{2j-1} x_{2j} \pmod 2$$

embedded into $\mathbb{T}$ by $0 \mapsto 0$, $1 \mapsto 1/2$. Explicitly,

$$Q(x+h) - Q(x) - Q(h) = \frac{1}{2} \sum_j \big(x_{2j-1} h_{2j} + h_{2j-1} x_{2j} \big) \pmod 1.$$

The right-hand side is bilinear, so three additive derivatives vanish.

Example C: a genuinely non-classical quadratic

Even the function

$$P(x) = \frac{x_1}{4} \pmod 1 \tag{3}$$

has degree two, although it looks linear in the coordinate x_1.

Let $e_1 = (1, 0, \ldots)$. Then

$$\Delta_{e_1} P(x) = \frac{1 - 2x_1}{4},$$

and therefore

$$\Delta_{e_1}^2 P(x) = \frac{1}{2} \pmod 1, \qquad \Delta_{e_1}^3 P(x) = 0.$$

Since P depends only on the first coordinate, these computations also establish the required identities for arbitrary directions.

This is precisely the basic example in **Tao–Ziegler, Example 1.6**: the map $0 \mapsto 0,\ 1 \mapsto 1/4$ on $\mathbb{F}_2$ is quadratic rather than linear. arXiv

The denominator matters because addition in the domain is modulo two. Dividing by another power of two raises the polynomial degree.

Example D: a non-classical quadratic involving infinitely many coordinates

Let

$$W(x) = \sum_{j \geq 1} x_j$$

be the Hamming weight, and put

$$\boxed{P(x) = \frac{W(x)}{4} \pmod 1.} \tag{4}$$

This is also quadratic. Indeed,

$$W(x + h) = W(x) + W(h) - 2 \sum_j x_j h_j,$$

so

$$\Delta_h P(x) = \frac{W(h)}{4} - \frac{1}{2} \sum_j x_j h_j \pmod 1.$$

The second term is a linear character of x, so the first derivative is affine-linear and the third derivative vanishes.

More explicitly,

$$\Delta_k \Delta_h P(x) = -\frac{1}{2} \sum_j k_j h_j \pmod 1,$$

which is independent of x.

This gives a genuinely infinite-coordinate example without any convergence issue.

3. There is an explicit normal form

The finite-dimensional classification gives a concrete description of all these functions. In characteristic two, a normalized polynomial $P : V \to \mathbb{T}$ of degree at most d can be written

$$\boxed{P(x) = \sum_{\substack{I \subseteq \mathbb{N} \text{ finite},\, I \neq \varnothing \\ j \geq 0,\, |I|+j \leq d}} \frac{c_{I,j}}{2^{j+1}} \prod_{i \in I} x_i \pmod 1, \qquad c_{I,j} \in \{0,1\}.} \tag{5}$$

The sum is pointwise finite: only subsets $I \subseteq \operatorname{supp}(x)$ contribute. The formula for finite-dimensional spaces is **Tao–Ziegler, Lemma 1.7(iii)**; it is also restated as **Lemma 2.4** in Berger–Sah–Sawhney–Tidor, *Non-classical polynomials and the inverse theorem*. The direct-sum version follows by restricting to the first n coordinates and using uniqueness to make the coefficients compatible. arXiv

For example, degree at most two means

$$P(x) = \frac{1}{2}\sum_i a_i x_i + \frac{1}{4}\sum_i b_i x_i + \frac{1}{2}\sum_{i<j} c_{ij} x_i x_j \pmod 1,$$

with coefficients in $\{0,1\}$.

Thus the relevant weight is

$$\boxed{\text{number of coordinate factors} + \text{extra denominator depth}.}$$

For instance,

$$\frac{x_1 x_2}{2}, \qquad \frac{x_1}{4}$$

both have degree two, whereas

$$\frac{x_1 x_2}{4}, \qquad \frac{x_1}{8}$$

have degree three.

A consequence for Diophantine approximation

A normalized degree- d polynomial on V takes values in

$$\frac{1}{2^d}\mathbb{Z}/\mathbb{Z}. \tag{6}$$

There is a very short direct proof. Apply the repeated-direction identity from Section 1, now with target $\mathbb{T}$, at $x = 0$:

$$0 = \Delta_h^{d+1} P(0) = (-2)^d (P(h) - P(0)).$$

When $P(0) = 0$, this says $2^d P(h) = 0$.

Consequently,

$$\boxed{\|P(s)\| < 2^{-d} \quad \Longrightarrow \quad P(s) = 0.} \tag{7}$$

So, for a bounded-degree family on $\mathbb{F}_2^{(\mathbb{N})}$, sufficiently accurate simultaneous approximation becomes **simultaneous exact vanishing**. This is a major difference from the real-coefficient setting of your paper.

4. How close is this to your bracket-polynomial class?

There are two distinctions to preserve.

First, arbitrary real scalar multiplication is not intrinsically defined on $\mathbb{T}$. Multiplication by an integer is well-defined, but

$$t+\mathbb{Z} \longmapsto \sqrt{2}\,t+\mathbb{Z}$$

is not: changing the representative t by 1 changes the output by $\sqrt{2}$.

One can choose the representative in $[0,1)$ and then multiply it by $\sqrt{2}$, but the resulting function need not be a finite-degree polynomial. For example,

$$x \longmapsto \frac{\sqrt{2}}{2}x_1 \pmod 1$$

has no finite degree, since its repeated e_1-differences never vanish.

Second, **the finite-difference condition is not the definition of a bracket polynomial over** $\mathbb{Z}$. The connection is through higher-order structure, not a literal identity between the two syntactic classes.

A reasonable *broader bracket-style class* on V would be the following: start with canonical real lifts of finitely many normalized circle-valued polynomials, and apply finitely many additions, multiplications, real scalar multiplications, and rounding operations. I would present this as a chosen definition, not claim it is the unique standard meaning of "generalized polynomial on V."

Every such expression has the form

$$F(P_1(x),\dots,P_m(x)), \tag{8}$$

where each P_i has finite image. Hence it also has finite image. With the normalization $F(0,\dots,0)=0$, simultaneous vanishing of the P_i forces the whole expression to vanish.

For recurrence questions, this suggests that **finite families of normalized circle-valued polynomials are the fundamental objects to study first**, before adding a more elaborate bracket syntax.

5. The dynamical connection is already substantial

This is not merely an algebraic definition invented for the example. Bergelson, Tao, and Ziegler study ergodic actions of

$$\mathbb{F}_p^{(\mathbb{N})}$$

and their Gowers–Host–Kra characteristic factors using phase polynomials.

Their paper, *An inverse theorem for the uniformity seminorms associated with the action of* $\mathbb{F}^\omega$, contains the definition in **Definition 1.13** and the principal characteristic-factor results in **Theorems 1.19–1.20**. The latter include bounded-degree phase-polynomial control in low characteristic, which is the relevant setting for $p=2$. arXiv

For a framework encompassing general abelian groups and nilpotent structure, there are also **nilspace polynomials**: functions of the form

$$G \xrightarrow{\varphi} X \xrightarrow{F} \mathbb{C},$$

where φ preserves the appropriate cube structure and X is a compact nilspace. Candela–González-Sánchez–Szegedy discuss this explicitly in the introduction to *Free nilspaces, double-coset nilspaces, and Gowers norms*, immediately before Example 1.1. This is a broader analogue of structured nilsequences, rather than an exact replacement for your nearest-integer syntax. arXiv

6. An important warning for extending your largeness theorem

Even with this good polynomial definition, **your notion of vdW-largeness cannot simply be copied to V.**

Since $2s = 0$, a scalar arithmetic progression is

$$x,\ x+s,\ x+2s,\ x+3s, \ldots = x,\ x+s,\ x,\ x+s, \ldots.$$

If distinct terms are required, there are no nondegenerate progressions of length three. If repeated terms are allowed, arbitrarily long progressions are just two-point configurations.

There is a concrete counterexample to the latter interpretation.

Let $e_1, e_2, \ldots$ be the standard basis, and set

$$S = \{e_i + e_j : i < j\}. \tag{9}$$

For every finite coloring of V, two distinct basis vectors e_i, e_j receive the same color. Taking

$$x = e_i, \qquad s = e_i + e_j \in S,$$

the progression

$$x,\ x+s,\ x+2s, \ldots = e_i,\ e_j,\ e_i, \ldots$$

is monochromatic at every length.

Thus S satisfies the naive repeated-term vdW-largeness condition for **every finite number of colors**.

But the quadratic polynomial from (4) satisfies

$$P(e_i + e_j) = \frac{2}{4} = \frac{1}{2} \pmod 1.$$

Therefore

$$\boxed{\|P(s)\| = \frac{1}{2} \qquad (s \in S).} \tag{10}$$

So this naive largeness condition does not force approximation to zero even for one normalized quadratic polynomial.

For a meaningful analogue of your project, the combinatorial configurations should therefore involve **independent directions**, such as affine cubes

$$x + \left\{ \sum_{j=1}^{k} \omega_j h_j : \omega_j \in \{0,1\} \right\},$$

or affine subspaces of increasing dimension, rather than arbitrarily long scalar progressions.

My recommendation is to begin with normalized non-classical polynomials $V \to \mathbb{T}$, using the finite-difference definition (1). They have an explicit algebra, nontrivial examples of every degree, and an established dynamical theory. The real conceptual task is then to choose the right higher-dimensional replacement for vdW-largeness; the polynomial side is already available.

145. User

2026-09-22 17:45:32 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-22T09:45:32.169000Z

going back to the paper you produced for us, why is mathcal{K} defined with |Q(1)-Q(0)| = rho instead of |Q(1)-Q(0)| <= rho?

---

146. Assistant

2026-09-22 17:46:54 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-22T09:46:54.355202Z

Because **we need a fixed positive amount of motion, not merely an upper bound on the motion**. The condition

$$\|Q(1) - Q(0)\|_\infty = \rho > 0$$

excludes both constant curves and curves that are arbitrarily close to constant. This is what lets compactness give a **uniform positive lower bound on the oscillation of** $F \circ Q$. That is precisely the deduction made immediately after defining the family. [Pasted text.txt, L134-L145]

1. What would fail with $\leq \rho$

If we instead defined

$$\mathcal{K}_{\leq} = \{Q : \deg Q_{ij} \leq D,\ |Q_{ij}(t)| \leq 1,\ \|Q(1) - Q(0)\|_\infty \leq \rho\}\,,$$

this would include every constant curve satisfying the coordinate bounds. For a constant curve,

$$\operatorname{osc}_{[0,1]} F(Q(t)) = 0.$$

Therefore we could not obtain the conclusion needed for the shell construction:

$$\boxed{\exists \eta > 0 \quad \operatorname{osc}_{[0,1]} F(Q(t)) \geq \eta \quad \text{for every } Q \in \mathcal{K}.}$$

The issue is **not that** $\mathcal{K}_{\leq}$ **fails to be compact**—it is compact. The issue is that the continuous oscillation functional has minimum zero on it.

2. Excluding constant curves separately would not suffice

Suppose we kept $\leq \rho$ but added "Q is nonconstant." Each individual curve would then have positive F-oscillation, by the preceding lemma, but those positive oscillations could approach zero.

For a concrete illustration, in $U_2(\mathbb{R})$ take

$$Q_\varepsilon(t) = \begin{pmatrix} 1 & \varepsilon t \\ 0 & 1 \end{pmatrix}, \qquad 0 < \varepsilon \leq \rho = \frac{1}{4}.$$

Every Q_ε is nonconstant and has endpoint displacement $\varepsilon \leq \rho$. In this group,

$$F\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} = \operatorname{dist}(x, \mathbb{Z})^2.$$

Hence

$$F(Q_\varepsilon(t)) = \varepsilon^2 t^2 \quad (0 \le t \le 1),$$

and

$$\operatorname{osc}_{[0,1]} F(Q_\varepsilon) = \varepsilon^2 \longrightarrow 0.$$

So **“nonconstant” gives pointwise positivity; the fixed displacement gives uniform positivity after compactness**. Removing the constant curves from $\mathcal{K}_\le$ loses closedness, since Q_ε converges to the constant identity curve.

3. Why equality fits the first-exit argument exactly

Later, we suppose that the polynomial lift escapes distance ρ from its starting point, and choose the **first** time T such that

$$\|Q(T) - Q(0)\|_\infty = \rho.$$

By the definition of first exit,

$$\|Q(t) - Q(0)\|_\infty \le \rho \qquad (0 \le t \le T).$$

Since $Q(0)$ has coordinates bounded by $1/2$, this bounds all coordinates on $[0, T]$ by

$$\frac{1}{2} + \rho = \frac{3}{4}.$$

Consequently, the rescaled curve

$$\widetilde{Q}(u) = Q(Tu), \qquad 0 \le u \le 1,$$

satisfies both the coordinate bounds and

$$\|\widetilde{Q}(1) - \widetilde{Q}(0)\|_\infty = \rho.$$

Thus it belongs to **exactly the normalized family** $\mathcal{K}$, and we can apply the uniform oscillation and sampling result. This is how the family is used in the proof. [Pasted text.txt, L245-L251]

Strictly speaking, equality is not the only possible choice: the condition

$$\|Q(1) - Q(0)\|_\infty \ge \rho$$

together with the same coordinate bounds would also give a compact family containing no constant curves. But equality is the most natural normalization because **a first-exit curve has endpoint displacement exactly** ρ.

In short: **the upper bound controls where the curve lies; the positive endpoint displacement prevents its motion from degenerating to zero. We need both, and the first-exit argument supplies them.**

---

147. User

2026-09-22 21:57:20 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-22T13:57:20.918000Z

In the paper that you prepared for me, it looks like you used/introduced a new Ramsey principle that implied the coloring result and the density result, instead of proving each of them separately. Can you extract from this paper an abstract principle that implies both coloring results and density results for ramsey theory on the integers?

148. Assistant

2026-09-22 21:57:49 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-22T13:57:49.821334Z

I'll isolate the abstract omitted-color Ramsey principle, independently of the polynomial machinery, and derive its coloring, density, and extremal consequences.

149. Assistant

2026-09-22 22:00:55 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-22T14:00:55.425914Z

The key observation is that every bad configuration meets every color class; deleting one class therefore yields both high-density avoidance sets and finite extremal bounds.

150. Assistant

2026-09-22 22:03:43 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-22T14:03:43.713628Z

Yes. The common principle is that the bad configurations admit colorings in which every bad configuration meets every color. Such a coloring simultaneously provides a binary obstruction, high-density sets avoiding all bad configurations, and the almost-lossless extremal comparison.

This is stronger than an ordinary coloring with no monochromatic bad configuration. Your manuscript makes precisely that distinction in Theorem 1.2: a progression violating the approximation bound must meet **every** color class, not merely two of them. [vdw_largeness_nilrecurrence.tex, L245-L267]

In hypergraph terminology, this is a **polychromatic coloring**: the integers are the vertices, the bad configurations are the edges, and every edge contains every color. This terminology is established; for example, it is the definition in the abstract of Bollobás–Pritchard–Rothvoß–Scott, *Cover-Decomposition and Polychromatic Numbers*. The abstract formulation below is an extraction of your argument, rather than a claim that this general coloring notion is new. arXiv

1. The basic abstract principle

Let $\mathscr{F}$ be **any collection of nonempty finite subsets of** $\mathbb{N}$. Think of its members as forbidden configurations. A set $A \subseteq \mathbb{N}$ is called $\mathscr{F}$-free if it contains no member of $\mathscr{F}$.

Proposition: omitted-color transfer

Suppose there is a coloring

$$c : \mathbb{N} \longrightarrow \{0, \dots, r-1\}, \qquad r \geq 2,$$

such that

$$\boxed{c(F) = \{0, \dots, r-1\} \qquad (F \in \mathscr{F}).} \tag{1}$$

For each color e, put

$$H_e = \{n \in \mathbb{N} : c(n) \neq e\}.$$

Then:

Coloring conclusion. There is a binary coloring of $\mathbb{N}$ with no monochromatic member of $\mathscr{F}$.

Finite-density conclusion. Every finite $E \subseteq \mathbb{N}$ has an $\mathscr{F}$-free subset E' satisfying

$$\boxed{|E'| \geq \left(1 - \frac{1}{r}\right)|E|.} \tag{2}$$

Upper-density conclusion. Every $E \subseteq \mathbb{N}$ has an $\mathscr{F}$-free subset E' satisfying

$$\boxed{\overline{d}(E') \geq \left(1 - \frac{1}{r}\right)\overline{d}(E).} \tag{3}$$

In particular, there is an $\mathscr{F}$-free set of upper density exactly δ for every

$$0 < \delta \leq 1 - \frac{1}{r}.$$

Proof

The essential observation is simply

$$\boxed{\text{Every } H_e \text{ is } \mathscr{F}\text{-free.}} \tag{4}$$

Indeed, a member of $\mathscr{F}$ contained in H_e would omit color e, contradicting (1).

For the coloring conclusion, partition $\mathbb{N}$ into

$$C_0 = \{n : c(n) = 0\}, \qquad H_0 = \{n : c(n) \neq 0\}.$$

The set H_0 is $\mathscr{F}$-free. So is C_0, because $C_0 \subseteq H_1$. These two sets give the required binary coloring.

For the finite-density conclusion, each integer belongs to exactly $r - 1$ of the sets H_e. Consequently,

$$\sum_{e=0}^{r-1} |E \cap H_e| = (r-1)|E|. \tag{5}$$

At least one intersection therefore has cardinality at least

$$\frac{r-1}{r}|E|.$$

Take that intersection as E'.

For the upper-density conclusion, choose $N_j \to \infty$ such that

$$\frac{|E \cap [1, N_j]|}{N_j} \longrightarrow \overline{d}(E).$$

Applying (5) to $E \cap [1, N_j]$, and passing to a subsequence on which the same color e is selected, gives

$$\overline{d}(E \cap H_e) \geq \frac{r-1}{r}\overline{d}(E).$$

Finally, take $E = \mathbb{N}$. If the resulting $\mathscr{F}$-free set H has upper density $\alpha \geq \delta$, enumerate it as

$$H = \{h_1 < h_2 < \cdots\}$$

and retain h_j when

$$\left\lfloor \frac{\delta j}{\alpha} \right\rfloor > \left\lfloor \frac{\delta (j-1)}{\alpha} \right\rfloor .$$

The resulting subset A satisfies

$$|A \cap [1,N]| = \left\lfloor \frac{\delta}{\alpha} |H \cap [1,N]| \right\rfloor ,$$

so $\overline{d}(A) = \delta$. It remains $\mathscr{F}$-free. □

This is the abstract version of the complement-of-one-color and thinning arguments in Section 5 of your manuscript. The additional formulation for an arbitrary ambient set E makes the same mechanism more reusable. [vdw_largeness_nilrecurrence.tex, L24-L72]

The single identity behind both density conclusions

Both conclusions come from

$$\boxed{\sum_{e=0}^{r-1} \mathbf{1}_{H_e}(n) = r-1 \qquad (n \in \mathbb{N}).} \tag{6}$$

There is **no assumption that the color classes have equal densities**, or even that their natural densities exist.

The same identity also proves a weighted version: for every finitely supported function $w : \mathbb{N} \to [0,\infty)$, some $\mathscr{F}$-free set H_e satisfies

$$\sum_{n \in H_e} w(n) \geq \left(1 - \frac{1}{r}\right) \sum_{n} w(n). \tag{7}$$

Thus the principle is not specifically tied to ordinary counting density.

2. A unified principle for restricted arithmetic progressions

We can now formulate exactly the abstraction relevant to your paper.

For each $k \geq 2$, let

$$B_k \subseteq \mathbb{N}$$

be a prescribed set of **good common differences**. The sets B_k may shrink with k; no monotonicity assumption is needed.

Assume the following property:

> **Polychromatic obstruction hypothesis.** For every $r \geq 2$, there are an integer K_r and one coloring
>
> $$c_r : \mathbb{N} \to \{0, \dots, r-1\}$$

> such that, for every $k \geq K_r$ and every $a, d \in \mathbb{N}$,
>
> $$\boxed{d \notin B_k \quad \Longrightarrow \quad \{c_r(a+jd) : 0 \leq j < k\} = \{0, \dots, r-1\}.} \tag{8}$$

The coloring is fixed simultaneously for all $k \geq K_r$.

Theorem: simultaneous vdW and Szemerédi transfer

Under hypothesis (8), the following hold.

Binary Ramsey conclusion. For every $k \geq K_2$,

$$\boxed{S \text{ is } (2,k)\text{-vdW-large} \quad \Longrightarrow \quad S \cap B_k \neq \varnothing.} \tag{9}$$

Fixed-density conclusion. Fix $0 < \delta < 1$, and choose $r \geq 2$ such that

$$1 - \frac{1}{r} \geq \delta.$$

There is one set $A_\delta \subseteq \mathbb{N}$, with

$$\overline{d}(A_\delta) = \delta,$$

such that

$$\boxed{\{a, a+d, \ldots, a+(k-1)d\} \subseteq A_\delta, \quad k \geq K_r \quad \Longrightarrow \quad d \in B_k.} \tag{10}$$

Consequently, for every $k \geq K_r$,

$$\boxed{S \text{ is } (\delta,k)\text{-Szemerédi-large} \quad \Longrightarrow \quad S \cap B_k \neq \varnothing.} \tag{11}$$

Proof

For (9), apply $(2,k)$-vdW-largeness to c_2. A monochromatic progression does not contain both colors, so its difference must belong to B_k.

For the density conclusion, apply the abstract proposition to the family

$$\mathscr{F}_r = \{\{a, a+d, \ldots, a+(k-1)d\} : k \geq K_r,\ d \notin B_k\}. \tag{12}$$

Hypothesis (8) says that every member of this family contains all r colors. The proposition supplies an $\mathscr{F}_r$-free set A_δ of upper density exactly δ, which is precisely (10).

Applying (δ,k)-Szemerédi-largeness to this particular A_δ proves (11). □

This is the common Ramsey principle you were noticing. Once (8) has been proved, neither the binary application nor the density application requires further information about the nature of the sets B_k.

In particular, for a *fixed* target set B, take $B_k = B$. We obtain

$$\boxed{\begin{aligned} &S \text{ is 2-vdW-large} && \Longrightarrow S \cap B \neq \varnothing, \\ &S \text{ is } \delta\text{-Szemerédi-large for some } \delta \in (0,1) && \Longrightarrow S \cap B \neq \varnothing. \end{aligned}} \tag{13}$$

The common hypothesis is a property of the **target** B, expressed through colorings of progressions whose differences lie outside it.

3. The extremal theorem is a third consequence of the same principle

Recall

$$r_k(\Omega) = \max\{|E| : E \subseteq \Omega \text{ contains no } k\text{-term AP}\},$$

and

$$r_k(\Omega; B_k) = \max\{|E| : E \subseteq \Omega \text{ contains no } k\text{-term AP with difference in } B_k\}.$$

Under (8), for every $r \geq 2$, every $k \geq K_r$, and every finite $\Omega \subseteq \mathbb{N}$,

$$\boxed{r_k(\Omega) \leq r_k(\Omega; B_k) \leq \frac{r}{r-1}\, r_k(\Omega).} \tag{14}$$

Indeed, let $E \subseteq \Omega$ contain no good k-term progression. The finite-density part of the proposition gives a subset

$$E' \subseteq E, \qquad |E'| \geq (1 - 1/r)|E|,$$

containing no **bad** k-term progression. Because $E' \subseteq E$, it also contains no good one. It therefore contains no k-term progression at all, giving

$$(1 - 1/r)|E| \leq r_k(\Omega).$$

The lower bound in (14) is immediate.

Choosing r so that

$$\frac{r}{r-1} \leq 1 + \varepsilon$$

gives the $1 + \varepsilon$ conclusion. This is exactly the deletion argument used in the manuscript's extremal section, now separated from generalized polynomials. [vdw_largeness_nilrecurrence.tex, L150-L167]

A useful way to summarize the proof is:

> **From a set containing no good configurations, delete a small part to eliminate the bad configurations too. What remains contains no configurations at all.**

The same argument works for any family of finite configurations partitioned into "good" and "bad," not only arithmetic progressions.

4. Precisely how your paper instantiates the principle

Fix special generalized polynomials

$$P_1, \ldots, P_m$$

of assigned weights d_i, and fix a target prefactor $c_* > 0$. Define

$$B_k(c_*) = \left\{ d \in \mathbb{N} : \|P_i(d)\|_{\mathbb{R}/\mathbb{Z}} \leq c_* k^{-d_i} \text{ for every } i \right\}. \tag{15}$$

Theorem 1.2 gives the polychromatic obstruction, initially with a prefactor depending on the palette size. Lemma 6.1 lets that prefactor be replaced by the prescribed c_*, while enlarging the threshold. Thus (8) holds for the **same target family** (15), with

$$K_r = K(r, m, D, c_*),$$

uniformly in the real coefficients. This is the precise role of the target-constant lemma; one should not silently ignore the original dependence of C on r. [vdw_largeness_nilrecurrence.tex, L124-L147]

The abstract principle then gives, without further polynomial calculations:

$$\begin{array}{c}(2,k)\text{-vdW-largeness}\\ \text{or}\\ (\delta,k)\text{-Szemerédi-largeness}\end{array} \quad\Longrightarrow\quad \exists d\in S:\|P_i(d)\|\le c_*k^{-d_i}.$$

It also gives

$$r_k(\Omega;B_k(c_*))\le(1+\varepsilon)r_k(\Omega)$$

for sufficiently large k.

For qualitative approximation, fix tolerances $\eta_i>0$ and choose k sufficiently large that

$$c_*k^{-d_i}<\eta_i.$$

This is the passage already made in the paper from its finite approximation statements to simultaneous approximation in any fixed neighborhood of zero. [vdw_largeness_nilrecurrence.tex, L75-L101]

So the logical architecture is

$$\boxed{\begin{array}{c}\text{elementary polynomial argument}\\ \Downarrow\\ \text{polychromatic obstruction for bad progressions}\\ \Downarrow\\ \text{binary recurrence, fixed-density recurrence, extremal transfer.}\end{array}}$$

5. There is also an abstract analytic mechanism behind the coloring

The combinatorial principle above explains the deductions. We can extract a second, more analytic principle explaining **why the shell construction produces the required colorings**.

Lemma: compact oscillation produces polychromatic samples

Let

$$\mathcal{C}\subseteq C([0,1],\mathbb{R})$$

be compact in the uniform norm, and suppose every $f\in\mathcal{C}$ is nonconstant.

For every $r\ge2$, there are $h>0$ and $T_0\ge1$ such that, for every $f\in\mathcal{C}$ and every real $T\ge T_0$, the integers

$$\left\lfloor\frac{f(j/T)}{h}\right\rfloor,\qquad 0\le j\le\lfloor T\rfloor,$$

represent every residue class modulo r.

Proof

The continuous functional

$$f\longmapsto\max f-\min f$$

is positive on the compact set $\mathcal{C}$. Hence

$$\eta:=\min_{f\in\mathcal{C}}\operatorname{osc}f>0.$$

Choose $h < \eta/(4r)$. The range of every f is an interval of length at least η, so it contains the midpoints of r consecutive complete bands

$$[jh, (j+1)h).$$

These bands have all r colors modulo r.

Compactness of $\mathcal{C}$ gives a common modulus of continuity. For sufficiently large T, the mesh $\{j/T\}$ samples within $h/4$ of each chosen midpoint value. Those samples remain inside their respective bands, proving the assertion. □

In your proof, this compact family is

$$\mathcal{C} = \{F \circ Q : Q \in \mathcal{K}\}.$$

The normalized displacement condition excludes constant curves, and the no-flat-shell lemma ensures that each $F \circ Q$ is nonconstant. These are exactly the compactness and sampling ingredients of the elementary source. [Pasted text.txt, L123-L159]

The remaining polynomial argument shows that a progression violating the desired approximation would generate such a normalized excursion, whose sampled shell colors agree with colors appearing in the progression. Hence that progression must contain every color. The manuscript's proof outline identifies these steps explicitly. [vdw_largeness_nilrecurrence.tex, L324-L350]

This separates two reusable ideas:

$$\boxed{\text{compact nonconstant excursions} \Longrightarrow \text{polychromatic obstructions} \Longrightarrow \text{coloring and density consequences.}}$$

The first implication needs a way of obtaining an excursion from a bad configuration. In your paper, the lifting and coefficient-extraction argument supplies that part. The abstract lemma alone does not supply it.

6. What this does not assert

It would be too strong to describe this as a general conversion of **any** coloring theorem into a density theorem.

For example, take

$$\mathscr{F} = \{\{n, n+1\} : n \geq 1\}.$$

Parity gives a binary coloring with both colors in every forbidden pair. But every $\mathscr{F}$-free set has upper density at most $1/2$, since it contains at most one element of each pair

$$\{1,2\}, \{3,4\}, \ldots.$$

A binary obstruction alone therefore does not provide forbidden-configuration-free sets of density close to one.

The crucial extra strength is

$$\boxed{\text{arbitrarily many colors, each required to occur in every sufficiently large bad configuration.}}$$

The threshold is allowed to increase with the number of colors.

Likewise, the principle does not prove that 2-vdW-largeness implies δ-Szemerédi-largeness. It says that **both hypotheses force intersection with targets having this stronger obstruction property**. That is a common consequence, not an equivalence between the two largeness notions.

What I would extract into the paper

I would isolate the proposition in Section 1 above as an **“omitted-color transfer lemma”**, and use it for both the density and extremal deductions. The central statement is just:

> If every forbidden configuration meets every class of an r-coloring, then deleting a suitable color leaves a forbidden-configuration-free subset retaining at least $1 - 1/r$ of any finite nonnegative weight, or at least that proportion of upper density.

That is the abstract principle doing the work. **The substantive contribution of your proof is establishing its polychromatic hypothesis for the bad generalized-polynomial progressions, with one coloring across all sufficiently large lengths and with coefficient-uniform shrinking targets.**

---

151. User

2026-09-22 22:10:39 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-22T14:10:39.157000Z

Selected text:

Selection 1

The basic abstract principle

My request:

is this principle stated explicitly in the paper that you referenced, or in some other paper?

---

152. Assistant

2026-09-22 22:11:20 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-22T14:11:20.114784Z

I’ll check whether the cited literature states this transfer principle explicitly, rather than merely defining polychromatic colorings.

---

153. Assistant

2026-09-22 22:18:18 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-22T14:18:18.794091Z

Yes—the finite combinatorial core is explicitly stated and proved in another paper, as Observation 4 of a paper by Dömötör Pálvölgyi. The full package I gave, including natural upper density on $\mathbb{N}$, is an elementary extension; I did not locate that entire package stated as one proposition.

The paper I referenced previously supplies the **polychromatic-coloring terminology and its interpretation as a partition into transversals**, rather than explicitly stating all the consequences I listed. Springer arXiv

1. An explicit reference for the finite principle

The reference is:

Dömötör Pálvölgyi, "Note on Polychromatic Coloring of Hereditary Hypergraph Families," ***Graphs and Combinatorics*** **40, article 131 (2024), Observation 4.**
DOI: **10.1007/s00373-024-02836-y**.
In the preprint **arXiv:2309.04603v1**, it is **Observation 4 on page 2**, with its proof immediately afterward. The published version retains the same numbering. Springer arXiv

Write $\alpha(\mathcal{H})$ for the largest size of a vertex set containing no entire hyperedge. His observation states

$$\frac{\alpha(\mathcal{H})}{|V(\mathcal{H})|} < \frac{r-1}{r} \quad \Longrightarrow \quad \mathcal{H} \text{ has no polychromatic } r\text{-coloring.}$$

Equivalently,

$$\boxed{\mathcal{H} \text{ admits a polychromatic } r\text{-coloring} \quad \Longrightarrow \quad \alpha(\mathcal{H}) \geq \left\lceil \frac{r-1}{r} |V(\mathcal{H})| \right\rceil .} \tag{1}$$

This is exactly our finite deletion principle. arXiv

His proof is also exactly the argument we used: a smallest color class C has size at most $\lfloor |V|/r \rfloor$; every edge meets C; therefore $V \setminus C$ is independent and has at least $\lceil (r-1)|V|/r \rceil$ vertices. **Observation 4 itself does not require a hereditary family or uniform hyperedge size.** Springer

To recover our assertion for an arbitrary finite set $E \subseteq \mathbb{N}$, apply (1) to

$$\mathcal{H}_E = \big(E, \{F \in \mathscr{F} : F \subseteq E\}\big).$$

The original coloring remains polychromatic on every edge of this finite hypergraph. Consequently, E has an $\mathscr{F}$-free subset of size at least $(1 - 1/r)|E|$.

Here we retain **only whole forbidden configurations contained in** E, not arbitrary intersections $F \cap E$.

2. What the originally cited paper explicitly says

In **Bollobás–Pritchard–Rothvoß–Scott, "Cover-Decomposition and Polychromatic Numbers," arXiv:1009.6144v3**, the relevant passage is:

Section 1.1, "Terminology and Notation," page 2.

They define a polychromatic r-coloring by requiring every hyperedge to contain every color. They immediately explain the equivalent formulation: the color classes partition the vertex set into sets each meeting every edge, called **vertex covers or transversals**. The same section specifies that the paper considers finite hypergraphs. arXiv

Thus that reference supports the terminology and the underlying transversal interpretation. **I did not find our complete coloring-and-density proposition explicitly stated there.** Pálvölgyi's Observation 4 is the more direct reference for the finite bound.

3. Which parts of our formulation go beyond the explicit observation?

The weighted and upper-density formulations are straightforward extensions of the same argument, but I did not find them explicitly stated in those papers.

For example, with

$$H_e = \{n : c(n) \neq e\},$$

every H_e is $\mathscr{F}$-free and

$$\sum_{e=0}^{r-1} \mathbf{1}_{H_e}(n) = r - 1.$$

For any finitely supported nonnegative weight w, this gives

$$\sum_{e=0}^{r-1} \sum_{n \in H_e} w(n) = (r-1) \sum_{n} w(n).$$

Therefore some H_e retains at least $1 - 1/r$ of the total weight.

For natural upper density, apply the finite counting identity along a sequence of cutoffs realizing $\overline{d}(E)$, and pass to a subsequence on which the same color is omitted. This produces

$$\overline{d}(E \cap H_e) \geq \left(1 - \frac{1}{r}\right) \overline{d}(E).$$

Thinning then gives an avoiding set of any smaller prescribed upper density.

These deductions explain why the finite observation suffices for the more expansive formulation in our discussion. **Failure to find that exact formulation in print is not evidence that these elementary extensions are new.**

4. How I would attribute it in your paper

I would introduce the lemma with wording such as:

```
We record an elementary consequence of polychromatic colorability.
Its finite unweighted form appears in
\cite[Observation~4]{Palvolgyi2024}.
We include the weighted and upper-density variants needed below.
```

Pálvölgyi provides a precise published reference, though I am not claiming it is the earliest occurrence of the observation.

I would not present the abstract deletion-and-averaging principle as a new Ramsey theorem. The place to look for novelty in your paper is the construction of polychromatic colorings for the bad generalized-polynomial progressions—with coefficient-uniform shrinking targets and one coloring across all sufficiently large lengths—not the elementary extraction of a large avoiding set once that coloring exists.

154. User

2026-09-22 22:29:48 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-22T14:29:48.841000Z

Great! Please do one more literature search to check if there are any other papers with similar statements that I can cite/reference before introducing the statement I need in my paper.

155. Assistant

2026-09-22 22:30:28 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-22T14:30:28.526955Z

I'll search the polychromatic-coloring and transversal literature for explicit deletion principles, including weighted variants, while distinguishing exact matches from broader related results.

156. Assistant

2026-09-22 22:33:37 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-22T14:33:37.973350Z

The strongest additional precedent is Cherkashin's 2018 paper, which explicitly uses the smallest-color-class deletion argument; older epsilon-net work gives related precedents.

157. Assistant

2026-09-22 22:41:02 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-22T14:41:02.547212Z

Yes. The most useful new find is an explicit coloring-to-density result on the integers, not just a finite-hypergraph observation. Together with the previously identified Pálvölgyi observation, it gives you a particularly good pair of references: one for the general finite deletion argument and one for its density interpretation on $\mathbb{Z}$. Springer arXiv

I also found older uses of the same argument in extremal graph theory and epsilon-net theory, and a paper specifically about polychromatic colorings of arithmetic progressions with restricted differences. Here are the precise references and the distinctions between them.

1. The closest new reference: *Polychromatic Colorings on the Integers*

Maria Axenovich, John Goldwasser, Bernard Lidický, Ryan R. Martin, David Offner, John Talbot, and Michael Young, "Polychromatic Colorings on the Integers," ***Integers*** **19 (2019), Paper A18.**
Preprint: **arXiv:1704.00042v2**. arXiv

Pinpoint reference

Lemma 3, stated on **page 2**, with its proof in **§3, pages 9–10** of that preprint. It states

$$\boxed{\alpha(S) \leq \frac{1}{p(S)}.}$$

Here $p(S)$ is the largest number of colors in a coloring of $\mathbb{Z}$ for which every translate of the finite set S contains every color. Their $\alpha(S)$ is the infimum of the densities of **blocking sets**—sets meeting every translate of S. It is **not** the independence number. arXiv

The proof replaces the coloring by a periodic one, selects a color class of density at most $1/p(S)$, and observes that this class is a blocking set. Its complement therefore contains no translate of S and has density at least

$$1 - \frac{1}{p(S)}.$$

This is the same deletion mechanism as ours, with an explicit density conclusion. arXiv

Scope: they treat translates of one fixed finite set. Their periodicity argument should not be imported into our setting of an arbitrary family of forbidden progressions with unbounded diameters.

This sharpens my previous answer: a density version is explicitly in the literature, although not in the full arbitrary-family formulation we discussed.

2. An older explicit extremal version: Alon–Krech–Szabó

Noga Alon, Anja Krech, and Tibor Szabó, "Turán's Theorem in the Hypercube," *SIAM Journal on Discrete Mathematics* 21(1) (2007), 66–72.
DOI: **10.1137/060649422**. SIAM

Pinpoint reference

In the **nine-page author manuscript**, see **§1, page 3, the paragraph immediately preceding Theorem 4**.

They define $pc(n, d)$ as the maximum number of colors in an edge-coloring of the hypercube Q_n such that every d-dimensional subcube contains every color. If $f(n, d)$ is the minimum number of edges whose deletion destroys every Q_d, they explicitly record

$$\boxed{f(n, d) \leq \frac{e(Q_n)}{pc(n, d)}.}$$

Equivalently,

$$\mathrm{ex}(Q_n, Q_d) \geq \left(1 - \frac{1}{pc(n, d)}\right) e(Q_n).$$

The same page records the analogous vertex-deletion inequality. Fu Berlin Page

This is exactly the small-color-class argument in a particular configuration hypergraph: delete one color, thereby destroy every forbidden subcube, and retain almost all edges.

Why it is useful: it is an older, explicit precedent for using polychromatic colorings to obtain large configuration-free sets. It does not state our arbitrary weighted or natural-upper-density version, but it clearly demonstrates that the underlying extremal mechanism is established.

3. Two explicit precedents in epsilon-net theory

These papers phrase the conclusion as finding a **small hitting set**, rather than its **large configuration-free complement**. Those are the two sides of the same argument.

Smorodinsky–Yuditsky

Shakhar Smorodinsky and Yelena Yuditsky, "Polychromatic Coloring for Half-Planes," *Journal of Combinatorial Theory, Series A* 119(1) (2012), 146–154.
DOI: **10.1016/j.jcta.2011.07.001**; preprint **arXiv:1006.3191v1**. SIAM

Pinpoint: §4, "Small Epsilon-Nets for Half-Planes," page 9 of the preprint, especially the paragraph immediately preceding **Theorem 4.1**.

They explicitly observe that every color class of the relevant polychromatic coloring is an epsilon-net. They then use pigeonhole to choose a color class of size at most n/r. Theorem 4.1 formulates a general conversion from linear bounds for polychromatic-coloring thresholds to $O(1/\epsilon)$-size epsilon-nets. arXiv

For our purpose, the preceding paragraph is the most direct citation:

$$\text{every color class hits every relevant edge} \quad \Longrightarrow \quad \text{one small color class hits them all.}$$

Taking its complement gives the large avoiding set.

Keszegh–Pálvölgyi

Balázs Keszegh and Dömötör Pálvölgyi, "An Abstract Approach to Polychromatic Coloring: Shallow Hitting Sets in ABA-Free Hypergraphs and Pseudohalfplanes," *Journal of Computational Geometry* 10(1) (2019), 1–26.
DOI: **10.20382/jocg.v10i1a1**; preprint **arXiv:1410.0258**. Journal of Computational Geometry

Pinpoint: Appendix B, "Small epsilon-nets for pseudohalfplanes," final paragraph, published page 26.

The authors explicitly say that their polychromatic coloring partitions the vertices into epsilon-nets, and that one of these nets is small by pigeonhole. The appendix also directs readers to Smorodinsky–Yuditsky for the corresponding arguments. Journal of Computational Geometry

Why it is useful: this is a very clean published passage for the interpretation "polychromatic color classes are disjoint hitting sets." It supports the conceptual explanation immediately before your lemma, even though Pálvölgyi's Observation 4 is a more direct citation for the precise independence-number inequality.

4. Related integer-Ramsey work: restricted arithmetic progressions

Balázs Bursics, Bence Csonka, and Luca Szepessy, "Hitting Sets and Colorings of Hypergraphs," *SIAM Journal on Discrete Mathematics* 40(3) (2026), 1123–1140.
DOI: **10.1137/23M1615796**; preprint **arXiv:2307.12154v3**. SIAM

Pinpoint references

In version 3, see **§1.3, "Arithmetic progressions," pages 6–8**, particularly **Definition 9 on page 7** and **Theorems 6 and 7 on page 8**. arXiv

They study polychromatic colorability for hypergraphs arising from arithmetic progressions with restricted common differences and restrictions on starting points. Their introduction explicitly connects this to **ladders**, meaning sets of allowed differences preserving van der Waerden's theorem. arXiv

This is relevant background for your paper's **polychromatic progression formulation**, but it is not the best reference for the deletion lemma itself. Also, their hyperedges are **intersections of progressions with finite vertex sets**, whereas our extremal formulation concerns complete progressions contained in the ambient

set. Their coloring results concern particular structured difference sets, not the generalized-polynomial shrinking targets in your theorem. arXiv

I would cite this paper in a short discussion of related polychromatic-Ramsey work, rather than attribute our abstract density deduction to it.

5. Which references should actually appear before your lemma?

For the lemma itself, I would use **Pálvölgyi's Observation 4 and Axenovich et al.'s Lemma 3**.

The former is the direct general finite statement:

$$\mathcal{H} \text{ admits a polychromatic } r\text{-coloring} \quad \Longrightarrow \quad \alpha(\mathcal{H}) \geq \left\lceil \frac{r-1}{r} |V(\mathcal{H})| \right\rceil .$$

The latter supplies the closely related density statement for translates on the integers. Springer arXiv

A suitable paragraph would be:

```
A polychromatic coloring partitions the vertex set into transversals;
the complement of each color class is therefore independent.
Deleting a smallest color class gives a large independent set.
The corresponding finite inequality is recorded in
\cite[Observation~4]{Palvolgyi2024}, while a density version for
translates of finite subsets of the integers appears in
\cite[Lemma~3]{AxenovichEtAl2019}.
We record the weighted and upper-density forms needed here,
together with their elementary proofs.
```

The two cited papers support the respective finite and integer-density claims in that paragraph. Springer arXiv

To give a little more historical context without lengthening the proof section, an optional preceding sentence is:

```
The use of polychromatic colorings to construct small hitting sets,
or equivalently large configuration-free sets, also appears in
extremal problems for the hypercube \cite{AlonKrechSzabo2007}
and in the construction of epsilon-nets
\cite[Section~4]{SmorodinskyYuditsky2012}.
```

The precise passages identified above substantiate those attributions. Fu Berlin Page arXiv

6. What I did—and did not—find stated explicitly

The search now gives explicit precedents for **both** components you care about: the finite $1 - 1/r$ retention bound and a density consequence on the integers. It would therefore be too cautious to describe all density interpretations as merely our own extrapolation. Springer arXiv

I did **not** locate one statement containing the entire package for an arbitrary family $\mathscr{F}$ of finite subsets of $\mathbb{N}$:

$$\text{polychromatic coloring} \Longrightarrow \begin{cases} \text{binary obstruction}, \\ \text{weighted finite retention}, \\ \text{relative natural-upper-density retention}, \\ \text{exact-density thinning}. \end{cases}$$

But that is a limitation of the search, not evidence that this elementary package is new.

For example, the weighted extension follows immediately by choosing a color C_e of smallest weight:

$$\sum_e w(E \cap C_e) = w(E) \quad \Longrightarrow \quad w(E \setminus C_e) \geq (1 - 1/r)w(E).$$

The upper-density version uses the same finite identity along a density-realizing subsequence. Those short deductions can simply be included after citing the established finite and translate-density precedents.

My recommendation is to present your abstract lemma as a useful formulation of a standard polychromatic-coloring argument—not as a new Ramsey principle. The theorem constructing the required colorings for bad generalized-polynomial progressions is the substantive result; the lemma explains how its coloring, density, and extremal consequences fit together.

I have collected the six references discussed above, including comments identifying the precise passages, in a bibliography file:

Download the BibTeX references

---

158. User

2026-09-24 12:46:29 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-24T04:46:29.762000Z

are there any other references that you think I should add to line 131 of the tex code of this draft?

> **Attachments recorded in source:**
>
> - Pasted text(20260924-044612).txt

---

159. Assistant

2026-09-24 12:46:54 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-24T04:46:54.653196Z

I'll check which references directly support that sentence about nilsystem obstructions, and distinguish them from broader background citations.

---

160. Assistant

2026-09-24 12:50:04 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-24T04:50:04.063605Z

Host–Kra–Maass's Proposition 7.9 is especially relevant: it gives explicit polynomial-phase sets failing largeness, while Huang–Shao–Ye supports the subsequent broader generalized-polynomial claim.

161. Assistant

2026-09-24 12:54:08 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-24T04:54:08.702039Z

Yes. I would add Host–Kra–Maass at that point, and give a more precise Huang–Shao–Ye citation in the following sentence. I also found two additional papers worth considering, although I would not put every reference into one long citation string.

I am referring to the sentence near line 131 beginning “Nilsystems are a common source of obstructions to largeness,” currently citing Furstenberg, the k-recurrence separation paper, and your van der Corput paper. [Pasted text.txt, L124-L132]

1. The most directly relevant addition: Host–Kra–Maass

Bernard Host, Bryna Kra, and Alejandro Maass, “Variations on topological recurrence,” *Monatshefte für Mathematik* 179 (2016), 57–89. This is already your bibliography entry `HKM`. Springer

There are **two particularly relevant passages**:

Corollary 5.10, pages 21–22 of arXiv:1408.2728. They construct hitting-time sets in minimal s-step nilsystems that are sets of $(s-1)$-topological recurrence but not s-topological recurrence. This directly supports your statement that nilsystems generate obstructions to largeness. arXiv

Proposition 7.9, page 30 of that preprint. This is even closer to the combinatorial formulation of your paper. For

$$R = \{n \in \mathbb{N} : \|n^\ell \alpha\| > \eta\}, \qquad m = \left\lceil 2^{\ell-1}\eta^{-1} \right\rceil,$$

they show that R is **not** m**-large**. Their proof explicitly constructs a finite coloring using a polynomial phase and then takes finite differences. arXiv

I would therefore add

```
\cite[Corollary~5.10 and Proposition~7.9]{HKM}
```

to the discussion. The corollary supplies the dynamical framework; the proposition supplies an explicit restricted-difference coloring obstruction.

2. The precise general reference: Huang–Shao–Ye, Theorem A(1)

For the **next sentence**, asserting that complements of nil-Bohr neighborhoods are not large, the most appropriate reference is:

Wen Huang, Song Shao, and Xiangdong Ye, “Nil Bohr-sets and almost automorphy of higher order,” Theorem A(1), §1.1.3, printed page 3 of arXiv:1407.1179.

They prove that every Nil_d-Bohr $_0$ set B contains

$$N^d(U) = \{n : U \cap T^{-n}U \cap \cdots \cap T^{-dn}U \neq \varnothing\}$$

for some minimal d-step nilsystem and nonempty open U. arXiv

Consequently, $\mathbb{N} \setminus B$ fails d-topological recurrence, hence fails largeness. This is the direct justification for that sentence—not merely general background about nilsystems.

Use

```
\cite[Theorem~A(1)]{HSY}
```

there. For their related **question**, the pinpoint is **Problem B-II(2), §1.1.2, printed page 2**; it is formulated as a fixed-order higher Bohr problem. That is more precise than attributing an unspecified question to the entire monograph. arXiv

3. An additional published paper: *Powers of sequences and recurrence*

Nikos Frantzikinakis, Emmanuel Lesigne, and Máté Wierdl, "Powers of sequences and recurrence," ***Proceedings of the London Mathematical Society*** **(3) 98 (2009), no. 2, 504–530.**
DOI: **10.1112/plms/pdn040**. London Mathematical Society (LMS)

The relevant passage is **Proposition 5.1 and the example immediately following it, §5.1, page 12 of arXiv:0711.3159v2**.

The proposition states that if R is a set of ℓ-recurrence and p is an integer polynomial with $p(0) = 0$ and $\deg p \leq \ell$, then

$$\inf_{r \in R} \|\alpha p(r)\| = 0 \qquad (\alpha \in \mathbb{R}).$$

Thus it extends the monomial approximation obstruction to general integer polynomials. The accompanying example treats $p(n) = n^2 + n$. arXiv

Qualification: the proposition is stated for measurable multiple recurrence, not directly for vdW-largeness. I would cite it as further background on **polynomial obstructions to multiple recurrence**, rather than use its bare statement as the justification for non-largeness.

This is a reasonable additional reference, but less essential to your exact sentence than HKM.

4. A more recent addition: Hernández on Weyl systems

Felipe Hernández, "Multiple Polynomial Recurrence in Weyl Systems," arXiv:2504.19899. The version I checked is **v2**, revised January 2026. It characterizes polynomial recurrence within Weyl systems and proves equivalence of the measurable and topological formulations in that class. arXiv

The most relevant pinpoints are **Remark 4.10 and Example 4.12, pages 27–28**.

Remark 4.10 constructs polynomial-phase sets

$$R = \{n : \|q(n)\alpha\| > 1/4\}$$

that distinguish recurrence in Weyl systems of different steps. Example 4.12 gives explicit polynomial-phase sets separating different polynomial recurrence schemes. arXiv

This is useful **broader context**, particularly if you want to indicate that nilpotent obstructions are still being developed beyond the classical monomial examples. It is not needed to justify the basic non-largeness assertion.

A suitable separate sentence would be:

```
For further polynomial recurrence obstructions in Weyl systems, see
\cite[Remark~4.10 and Example~4.12]{HernandezWeyl}.
```

5. I would also sharpen the references already present

Your existing Frantzikinakis–Lesigne–Wierdl citation can be made more informative:

```
\cite[Theorem~1.4 and Proposition~3.2]{KBut NotK+1Recurrence}
```

In the published paper, **Theorem 1.4 is on page 840**, and the paragraph following it explicitly notes failure of k-**topological** recurrence. **Proposition 3.2, pages 843–844, and Example 3.3, pages 844–845**, display the affine-torus mechanism producing the obstruction. Numdam

There is also a small attribution distinction concerning your van der Corput paper. In **arXiv:2303.11832**, Theorem 1.12 itself is a **positive recurrence theorem** strengthening the positive side of the FLW examples. The discussion immediately afterward, on printed page 7, explicitly describes the uniquely ergodic skew-shift obstruction. If this is your `LebesguevdCForZ` entry, the more precise citation for the sentence you are writing is

```
\cite[discussion following Theorem~1.12]{LebesguevdCForZ}
```

rather than citing Theorem 1.12 alone as an obstruction theorem. arXiv

Suggested replacement

I would replace the current citation-heavy sentence and the opening of the next sentence by the following:

```
Nilsystems provide classical obstructions to largeness. Examples
arising from affine transformations of tori appear in
\cite[p.~177]{FurstenbergBook} and
\cite[Theorem~1.4 and Proposition~3.2]{KBut NotK+1Recurrence};
see also \cite[discussion following Theorem~1.12]{LebesguevdCForZ}.
Host, Kra, and Maass give a general nilsystem construction in
\cite[Corollary~5.10]{HKM} and an explicit finite-color obstruction
for monomial phases in \cite[Proposition~7.9]{HKM}.

More generally, Huang, Shao, and Ye
\cite[Theorem~A(1)]{HSY} show that every finite-step nil-Bohr
neighborhood of zero contains a multiple return-time set for a
minimal nilsystem. Consequently, the complement of such a
neighborhood is not large.
```

The FLW paper, HKM, and HSY support these respective obstruction claims; the citation to your paper points to its discussion of the same examples. Numdam arXiv arXiv arXiv

My recommendation is to add HKM and the precise HSY theorem citation unconditionally. The 2009 FLW paper and Hernández's preprint are worthwhile additional references, but I would introduce them in a separate sentence about broader polynomial recurrence rather than make the original sentence carry every example.

I have put the **two genuinely additional bibliography entries**—`FLWPowers` and `HernandezWeyl`—in a file, with comments recording the pinpoints:

Download the additional BibTeX entries

162. User

2026-09-24 13:01:51 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-24T05:01:51.462000Z

were huang shao ye the first to put into print the fact that complements of nilbohr neighborhoods of 0 are not large?

---

163. Assistant

2026-09-24 13:02:15 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-24T05:02:15.393312Z

I'll trace the earlier references and distinguish the general nil-Bohr statement from the older polynomial-phase examples.

---

164. Assistant

2026-09-24 13:05:31 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-24T05:05:31.438879Z

The general theorem already appears in HSY's September 2011 preprint, and Host–Kra–Maass explicitly attribute it to them; I'm checking earlier sources for predecessors.

---

165. Assistant

2026-09-24 13:09:02 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-24T05:09:02.069184Z

For the statement in full generality, Huang–Shao–Ye are the earliest authors I could verify—and I found that their result was already publicly available in September 2011, not just in the 2014 preprint or the 2016 memoir. I would credit them for the general theorem, while avoiding an unqualified "first" unless you also obtain confirmation from someone familiar with its history. arXiv

There is particularly good evidence for this attribution: they state the relevant recurrence consequence explicitly in their 2011 paper, and Host–Kra–Maass subsequently attribute the underlying theorem to them.

1. The earliest explicit source I found

Wen Huang, Song Shao, and Xiangdong Ye,
***Nil Bohr $_0$-sets, Poincaré recurrence and generalized polynomials*,**
arXiv:1109.3636v1, submitted 16 September 2011. arXiv

There are two useful pinpoint references.

Theorem A, §1.3, printed page 3, states that if $B \subseteq \mathbb{Z}$ is a Nil_d-Bohr $_0$ set, then there are a minimal d-step nilsystem (X, T) and a nonempty open set $U \subseteq X$ such that

$$\boxed{\{n \in \mathbb{Z} : U \cap T^{-n}U \cap \cdots \cap T^{-dn}U \neq \varnothing\} \subseteq B.} \tag{1}$$

They present this as an affirmative solution to their Problem B-II(1), rather than as a previously known theorem. arXiv

Corollary D, printed page 4, explicitly records

$$\mathcal{F}_{\mathrm{Poi}_d} \subseteq \mathcal{F}_{\mathrm{Bir}_d} \subseteq \mathcal{F}^*_{d,0}.$$

The second inclusion says that **every set of d-topological recurrence meets every Nil_d-Bohr $_0$ set**. This is the recurrence formulation directly relevant to your question. arXiv

Thus they do not need to use the precise phrase “the complement is not large”: they explicitly state a stronger, order-sensitive result from which it follows.

2. Why their theorem gives exactly the finite-color obstruction

Here is the short deduction, to make clear that we are not conflating two different statements.

Let B be a Nil_d-Bohr $_0$ set, and choose $(X,T), U$ from (1). By minimality and compactness, finitely many translates cover X:

$$X = \bigcup_{i=1}^{r} T^{-t_i} U.$$

Fix $x \in X$, and color $n \in \mathbb{N}$ by an index i for which

$$T^n x \in T^{-t_i} U.$$

If

$$a, a+h, \dots, a+dh$$

is monochromatic with color i, then

$$T^{a+t_i} x,\ T^{a+t_i+h} x, \dots, T^{a+t_i+dh} x \in U.$$

Therefore

$$U \cap T^{-h} U \cap \cdots \cap T^{-dh} U \neq \varnothing,$$

and (1) forces $h \in B$.

Hence this finite coloring has **no monochromatic $(d+1)$-term progression whose difference belongs to $\mathbb{N} \setminus B$**. In particular,

$$\boxed{\mathbb{N} \setminus B \text{ is not large.}}$$

This is a direct consequence of their realization theorem; no additional nilmanifold result is needed for the deduction.

3. The later literature confirms the attribution

In **Host–Kra–Maass, *Variations on topological recurrence*, Theorem 5.2, printed page 19 of arXiv:1408.2728**, the authors reproduce (1) and explicitly attribute it to Huang–Shao–Ye, citing their Theorem A. Their Corollary 5.3 then derives the equivalence between pointwise recurrence for d-step nilsystems and the corresponding multiple-recurrence condition. arXiv

This is stronger evidence for the attribution than merely finding the result in HSY: **another paper explicitly credits them for it**.

The later HSY manuscript, *Nil Bohr-sets and almost automorphy of higher order*, was posted on 4 July 2014. Its arXiv record says that it combines the two earlier papers arXiv:1109.3636 and arXiv:1110.6599. In this combined version, the relevant result is **Theorem A(1), §1.1.3, printed page 3**. arXiv arXiv The combined work was published as *Memoirs of the American Mathematical Society* **241 (2016), no. 1143**. arXiv

So the important historical date for the publicly available result is **2011**.

4. What was known earlier?

Special cases certainly predate HSY. For example, Frantzikinakis–Lesigne–Wierdl's 2006 paper gives the polynomial-phase sets

$$S_k = \{n \in \mathbb{N} : \{n^k \alpha\} \in [1/4, 3/4]\}$$

and explicitly notes that they fail k-topological recurrence. See **Theorem 1.4 and the paragraph following it, printed pages 840–841**. The same discussion traces the earlier recurrence-separation example to **Furstenberg, pages 177–178**. Numdam

Those are important predecessors, but they concern particular polynomial-phase or affine-system obstructions, rather than the assertion for **every** finite-step nil-Bohr $_0$ neighborhood.

Host and Kra introduced the general nil-Bohr-set framework in their earlier paper *Nil-Bohr sets of integers*, posted in March 2009 and published in 2011. Its principal combinatorial results concern piecewise nil-Bohr structure and sums with gaps; I did not find the general complement-nonlargeness assertion stated there. arXiv arXiv

Accordingly, I would distinguish:

Earlier polynomial-phase examples:	Furstenberg and subsequent work;
General nil-Bohr framework:	Host–Kra;
General realization theorem giving nonlargeness of complements:	Huang–Shao–Ye, publicly available in 2011.

The sources above support that distinction; they do not establish an exhaustive priority claim ruling out every earlier implicit argument.

5. Recommended wording in your introduction

For the sentence in your draft saying that complements of nil-Bohr neighborhoods "were known to not be large," I would now give a direct attribution rather than leave the source implicit. [Pasted text.txt, L124-L132]

```
Huang, Shao, and Ye \cite[Theorem~A(1)]{HSY} proved that every
\(d\)-step nil-Bohr neighborhood of zero contains a set of
\(d\)-fold topological return times. Consequently, the complement
of every finite-step nil-Bohr neighborhood of zero is not large.
```

This accurately attributes the theorem and identifies nonlargeness as its consequence. The pinpoint refers to the combined memoir; for the original 2011 preprint, the corresponding references are **Theorem A, page 3, and Corollary D, page 4**. arXiv arXiv

My conclusion is that crediting the general result to Huang–Shao–Ye is well supported. Calling their September 2011 paper "the earliest explicit source we have located" is justified; claiming absolute historical priority would go beyond what this literature check establishes.

166. User

2026-09-24 13:26:58 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-24T05:26:58.006000Z

I guess they didn't published this 2011 arXiv article separately? They probably used it as the base of their later memoirs article?

---

167. Assistant

2026-09-24 13:27:31 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-24T05:27:31.304265Z

I'll check the arXiv comments and publication record to confirm how the 2011 preprint relates to the later memoir.

---

168. Assistant

2026-09-24 13:30:09 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-24T05:30:09.352744Z

Yes—your understanding is correct. The 2011 preprint was incorporated into the later memoir, together with a second 2011 paper. This is explicitly confirmed by the authors, rather than merely inferred from the overlap in content. arXiv

On the arXiv abstract page for **arXiv:1407.1179**, the **Comments** field says:

> "This paper consists of the following two papers arXiv:1109.3636 and arXiv:1110.6599."

It then states that the combined paper will be published in *Memoirs of the AMS*. arXiv

The two original papers were:

1. **"Nil Bohr$_0$-sets, Poincaré recurrence and generalized polynomials,"** arXiv:1109.3636, submitted **16 September 2011**—the preprint we were discussing. arXiv
2. **"Higher order almost automorphy, recurrence sets and the regionally proximal relation,"** arXiv:1110.6599, submitted **30 October 2011**. arXiv

The combined manuscript was posted on **4 July 2014**, and the published reference is:

Wen Huang, Song Shao, and Xiangdong Ye,
***Nil Bohr-sets and almost automorphy of higher order*,**
Memoirs of the American Mathematical Society 241 (2016), no. 1143. arXiv American Mathematical Society AMS Bookstore

I found **no separate journal publication of the September 2011 preprint under its original title**. Its arXiv record does not list one, while the combined manuscript explicitly identifies it as a constituent paper. Thus "combined into the later memoir" is more precise than saying it merely served as the memoir's basis. arXiv arXiv

For your bibliography, I would **cite the published memoir**, not add the 2011 preprint as a separate reference unless you specifically want to document when the result first became publicly available. The historical distinction is simply: **available as a preprint in 2011; published as part of the memoir in 2016**. arXiv arXiv American Mathematical Society

169. User

2026-09-25 16:14:03 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-25T08:14:03.249000Z

Can you help me correct the paragraph directly before section 2? I think that I am using some other property of ovelrine{d} in addition to subadditivity, but I am not quite sure how to describe it.

> **Attachments recorded in source:**
>
> - Pasted text(20260925-081341).txt

---

170. Assistant

2026-09-25 16:14:38 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-25T08:14:38.967571Z

I'll identify the precise density properties the argument uses, check them for the other densities mentioned, and revise the paragraph while preserving the intended generalizations.

---

171. Assistant

2026-09-25 16:18:06 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-25T08:18:06.902527Z

The missing ingredient is a partition-averaging property arising from finite additivity before taking limits, not merely subadditivity. It also holds for Richter's density.

---

172. Assistant

2026-09-25 16:19:31 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-25T08:19:31.415888Z

Yes. What you are using is a finite-partition averaging property, not merely finite subadditivity. For natural upper density, this can also be expressed as **superadditivity of the complementary lower density on disjoint sets**. The sentence claiming that finite subadditivity is the only property used should therefore be changed. [Pasted text.txt, L216-L219]

Your intended extensions to the other densities are still valid, including Richter's density, but they need this stronger justification.

1. The precise property being used

Let

$$\mathbb{N} = C_0 \sqcup \cdots \sqcup C_{r-1}$$

be the partition into color classes. The required conclusion is

$$\boxed{\max_{0 \le e < r} \overline{d}(\mathbb{N} \setminus C_e) \ge 1 - \frac{1}{r}.} \tag{1}$$

For natural upper density, this follows from **additivity of finite counting before taking the limsup**:

$$\sum_{e=0}^{r-1} \frac{|(\mathbb{N}\setminus C_e)\cap[1,N]|}{N} = r-1.$$

Consequently,

$$\sum_{e=0}^{r-1} \overline{d}(\mathbb{N}\setminus C_e) \geq \limsup_{N\to\infty} \sum_{e=0}^{r-1} \frac{|(\mathbb{N}\setminus C_e)\cap[1,N]|}{N} = r-1,$$

which proves (1).

An equivalent way to explain this calculation uses the ordinary lower density:

$$\underline{d}(A) = 1 - \overline{d}(\mathbb{N}\setminus A).$$

For a finite partition,

$$\sum_{e=0}^{r-1} \underline{d}(C_e) \leq \underline{d}(\mathbb{N}) = 1.$$

Thus some C_e has lower density at most $1/r$, and its complement has upper density at least $1-1/r$.

The important distinction is that subadditivity of $\overline{d}$ controls unions from above, whereas here we need a lower bound for the density of a union of all but one color.

2. Suggested replacement paragraph

I would state the exact property rather than introduce a general theory of upper and lower densities in the introduction:

```
Theorem~\ref{thm:2LargeIsNilrecurrent} follows directly from
Theorem~\ref{thm:omitted} by taking \(r=2\).
For Theorem~\ref{thm:deltaSzLargeIsNilrecurrent}, take
\(r=\lceil(1-\delta)^{-1}\rceil\), and write
\(C_0,\ldots,C_{r-1}\) for the color classes.
Since
\[
 \sum_{e=0}^{r-1}\overline d(\N\setminus C_e)
 \geq
 \limsup_{N\to\infty}
 \sum_{e=0}^{r-1}
 \frac{|(\N\setminus C_e)\cap[1,N]|}{N}
 =r-1,
\]
some \(\N\setminus C_e\) has upper density at least
\(1-1/r\geq\delta\). Every progression in this set omits color \(e\),
so the conclusion follows from Theorem~\ref{thm:omitted}.

More generally, the same deduction applies to any density
\(\mathfrak d:\mathcal P(\N)\to[0,1]\) satisfying the
finite-partition property
\[
 \N=\bigsqcup_{e=0}^{r-1}C_e
 \quad\Longrightarrow\quad
 \max_{0\leq e<r}\mathfrak d(\N\setminus C_e)\geq1-\frac1r.
\]
This property, rather than finite subadditivity alone, is what the
```

```
argument uses. It holds for upper densities along fixed additive
or multiplicative F{\o}lner sequences, additive or multiplicative
upper Banach density, and the density \(\dhash\) defined in
\cite[(1.3)]{RichterDensity}. Thus the corresponding versions of
Theorem~\ref{thm:deltaSzLargeIsNilrecurrent} hold when
\(\delta\)-Szemer\'edi-largeness is defined using
\(\mathfrak d(A)\geq\delta\).
```

This preserves your intended generalizations. It also uses your revised definition with **density at least** δ, so no thinning property needs to be assumed for the replacement density. [Pasted text.txt, L188-L195]

3. Why the other densities satisfy this property

For density along a fixed sequence of nonempty finite sets (F_N), the same proof works with

$$\frac{|A \cap F_N|}{|F_N|}$$

in place of $|A \cap [1,N]|/N$. **The Følner property is not needed for this particular argument.** It is simply a statement about normalized finite averages.

For additive or multiplicative upper Banach density, use the same argument along a fixed corresponding Følner sequence, together with the fact that upper Banach density dominates the resulting upper density.

A useful general sufficient condition is:

> $\mathfrak{d}$ dominates a finitely additive probability measure μ on $\mathbb{N}$.

Indeed, some color satisfies $\mu(C_e) \le 1/r$, and then

$$\mathfrak{d}(\mathbb{N} \setminus C_e) \ge \mu(\mathbb{N} \setminus C_e) = 1 - \mu(C_e) \ge 1 - \frac{1}{r}.$$

This explains why upper densities built from probability averages have the property.

Checking Richter's density

I checked the actual definition, rather than relying on its stated subadditivity. It is equation **(1.3)** of Richter's *Sums and products in sets of positive density*. The operation there is a supremum over subsequences of logarithmic averages, followed by an upper limit along increasingly divisible dilations. arXiv

Here is a direct verification for a finite partition. Choose a common subsequence (N_s) along which all the limits

$$L_e(m) := \lim_{s\to\infty} \mathbb{E}^{\log}_{n\in[N_s]} \mathbf{1}_{C_e}(mn)$$

exist, simultaneously for every $e < r$ and $m \in \mathbb{N}$. Such a subsequence exists by diagonal extraction. For every m,

$$\sum_{e=0}^{r-1} L_e(m) = 1.$$

Passing to a subsequence of factorial dilations, arrange that

$$L_e(j_t!) \longrightarrow p_e \qquad (e < r).$$

Then $\sum_e p_e = 1$, so some $p_e \le 1/r$.

For each fixed a, all sufficiently large $j_t!$ belong to $a\mathbb{N}$. Hence

$$\sup_{m \in a\mathbb{N}} (1 - L_e(m)) \ge 1 - p_e.$$

Using (1.3) gives

```
\[
\dhash(\mathbb N\setminus C_e)
\geq1-p_e\geq1-\frac1r.
\]
```

Thus **Richter's density does satisfy the required property**; the verification comes from its averaging definition, not from subadditivity alone. arXiv

4. Why simply adding "normalized and monotone" would not suffice

For clarity, even a normalized, monotone, finitely subadditive set function can fail (1). Define

$$\nu(A) = \begin{cases} 0, & A \text{ is finite}, \\ 1, & \mathbb{N} \setminus A \text{ is finite}, \\ \frac{1}{2}, & A \text{ and } \mathbb{N} \setminus A \text{ are infinite}. \end{cases}$$

This function is monotone, finitely subadditive, and satisfies $\nu(\mathbb{N}) = 1$. But for the partition into the three residue classes modulo 3, every complement has value $1/2$, rather than at least $2/3$.

So I would **not** repair the sentence merely by adding normalization or monotonicity. The finite-partition property in the replacement paragraph identifies exactly what your deduction needs.

---

173. User

2026-09-25 16:57:01 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-25T08:57:01.709000Z

Thanks! The results about modified szemeredi numbers, do they follow only from the polychromatic theorem, or also directly from the delta-szeremedi large implies nilrecurrence theorem?

---

174. Assistant

2026-09-25 16:57:35 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-25T08:57:35.913822Z

I'll distinguish the fixed-neighborhood extremal estimate from the shrinking-target version and identify exactly which one follows from the qualitative δ-Szemerédi theorem alone.

175. Assistant

2026-09-25 17:00:41 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-25T09:00:41.991166Z

For a fixed nil-Bohr neighborhood, the modified Szemerédi-number result follows from the qualitative theorem alone, by contraposition and translation averaging. The polychromatic theorem is needed in the current proof for the stronger **coefficient-uniform shrinking-target version**, although a quantitative density-test theorem would also suffice.

So there are two distinct conclusions to separate. Your draft already distinguishes the fixed-neighborhood asymptotic from the shrinking-target estimate, but currently derives both through the polychromatic theorem. [Pasted text.txt, L1233-L1248]

1. The fixed-neighborhood result follows directly from the qualitative theorem

Recall that $r_k(\Omega)$ is the maximum size of a subset of Ω containing no k-term arithmetic progression, whereas $r_k(\Omega;B)$ forbids only those k-term progressions whose common difference belongs to B.

Proposition

Assume the theorem

$$S \text{ is } \delta\text{-Szemerédi-large for some } \delta \in (0,1) \quad \Longrightarrow \quad S \text{ is nil-Bohr recurrent.}$$

Then, for every fixed finite-step nil-Bohr $_0$ set B and every $\varepsilon > 0$, there is an integer $K = K(B,\varepsilon)$ such that

$$\boxed{r_k(\Omega) \le r_k(\Omega;B) \le (1+\varepsilon) r_k(\Omega)} \tag{1}$$

for every $k \ge K$ and every finite $\Omega \subseteq \mathbb{N}$.

Here the qualitative input is exactly your theorem, with the current definition using upper density **at least** δ. [Pasted text.txt, L188-L200]

Proof

Set

$$\delta = \frac{1}{1+\varepsilon}, \qquad R = \mathbb{N} \setminus B.$$

Because R misses the nil-Bohr neighborhood B, it is not nil-Bohr recurrent. By the contrapositive of the qualitative theorem, R is not δ-Szemerédi-large.

Consequently, there are a set $H \subseteq \mathbb{N}$ and an integer $K \ge 2$ such that

$$\alpha := \overline{d}(H) \ge \delta$$

and H contains no K-term progression with difference in R. Equivalently,

$$\boxed{\text{every } K\text{-term progression in } H \text{ has its difference in } B.} \tag{2}$$

The same holds for every longer progression, by taking its first K terms.

Now fix $k \ge K$, a finite set Ω, and a subset $E \subseteq \Omega$ containing no k-term progression with difference in B.

For each $t \in \mathbb{N}$, consider

$$E_t = E \cap (H - t) = \{x \in E : x + t \in H\}.$$

Every E_t contains no k-term progression at all. Indeed, such a progression would translate into H, so (2) would force its difference into B. But it also lies in E, contradicting the choice of E. Therefore

$$|E_t| \le r_k(\Omega) \qquad (t \in \mathbb{N}). \tag{3}$$

Choose $N_j \to \infty$ with

$$\frac{|H \cap [1, N_j]|}{N_j} \longrightarrow \alpha.$$

For each fixed x, translating the averaging interval by x changes its count by a bounded amount, so

$$\frac{|H \cap [x+1, x+N_j]|}{N_j} \longrightarrow \alpha.$$

Since E is finite,

$$\begin{aligned} \frac{1}{N_j} \sum_{t=1}^{N_j} |E_t| &= \sum_{x \in E} \frac{|H \cap [x+1, x+N_j]|}{N_j} \\ &\longrightarrow \alpha |E|. \end{aligned}$$

Averaging (3) and passing to the limit gives

$$\delta |E| \le \alpha |E| \le r_k(\Omega).$$

Thus

$$|E| \le (1 + \varepsilon) r_k(\Omega).$$

Maximize over E. The lower bound in (1) is immediate. □

No coloring theorem is invoked in this deduction. The only inputs are the qualitative density-to-nilrecurrence theorem, translation invariance of the progression's difference, and finite averaging.

In particular, it proves

$$\boxed{\lim_{k \to \infty} \sup_{\substack{\Omega \subset \mathbb{N} \text{ finite} \\ |\Omega| > 0}} \frac{r_k(\Omega; B)}{r_k(\Omega)} = 1} \tag{4}$$

for every fixed nil-Bohr $_0$ set B.

The uniformity over all finite Ω, including all intervals $[N]$, costs nothing extra: H and K were chosen before Ω.

2. What does not follow from that qualitative deduction

The stronger result concerns

$$B_k(c) = \left\{ d \in \mathbb{N} : \|P_i(d)\| \le c k^{-d_i} \text{ for every } i \right\}, \tag{5}$$

and asserts

$$r_k(\Omega; B_k(c)) \le (1 + \varepsilon) r_k(\Omega)$$

for all sufficiently large k, with a threshold depending only on

$$\varepsilon, c, m, D,$$

not on the real coefficients. This is the quantitative comparison in the manuscript. [Pasted text.txt, L309-L327]

The qualitative argument does not give the necessary relationship between the shrinking neighborhood and the progression length.

To see the issue, fix an integer j and apply the proposition above to the fixed set $B_j(c)$. It gives

$$r_k(\Omega; B_j(c)) \le (1+\varepsilon) r_k(\Omega) \qquad \text{whenever } k \ge K(B_j(c), \varepsilon). \tag{6}$$

But the desired estimate requires ** $j = k$**. Nothing in the qualitative theorem guarantees

$$k \ge K(B_k(c), \varepsilon).$$

Nor does that theorem, as stated, supply uniformity over the coefficients defining $B_k(c)$.

Thus the qualitative deduction establishes the **fixed-target comparison**, but does not itself furnish the **same-length, coefficient-uniform shrinking-target comparison**.

3. Even for shrinking targets, the full polychromatic formulation is not essential

What the averaging proof really needs is a high-density test set.

Suppose, for a fixed $\delta \in (0,1)$, we know that there are k_0 and one set $H \subseteq \mathbb{N}$ with

$$\overline{d}(H) \ge \delta$$

such that

$$\boxed{\text{every } k\text{-term progression in } H,\ k \ge k_0, \text{ has its difference in } B_k(c).} \tag{7}$$

Then the same translation-averaging argument gives

$$r_k(\Omega; B_k(c)) \le \delta^{-1} r_k(\Omega) \qquad (k \ge k_0). \tag{8}$$

Taking $\delta = (1+\varepsilon)^{-1}$ proves the desired comparison.

Your quantitative density-test theorem gives (7), initially with the approximation constant $C(\delta, m, D)$. [Pasted text.txt, L1037-L1049] To prescribe an arbitrary constant $c > 0$, apply that theorem to

$$P_1, \dots, P_m,\ qP_1, \dots, qP_m$$

and use the same fixed-multiplier, no-wrap argument as in Lemma 6.1. That argument does not intrinsically require a coloring; it only requires simultaneous bounds for P_i and qP_i. [Pasted text.txt, L1182-L1198]

Consequently, the relevant hierarchy is

$$\boxed{\begin{aligned} &\text{qualitative density-to-nilrecurrence theorem} \Longrightarrow \text{fixed-neighborhood extremal comparison,} \\ &\qquad\qquad \text{quantitative density-test theorem} \Longrightarrow \text{uniform shrinking-target extremal comparison.} \end{aligned}}$$

The polychromatic theorem proves the quantitative density-test theorem in the current manuscript, and also gives a shorter finite deletion proof of the extremal inequality. But it is **not the only possible route**.

4. Implication for the presentation of your paper

The fixed-neighborhood $1+\varepsilon$ result can be advertised as a consequence of the qualitative δ-Szemerédi-largeness theorem. You do not need to retain the quantitative theorem merely to retain that application.

The additional quantitative contribution is specifically that the same comparison holds for neighborhoods shrinking at the stated rates, with a threshold uniform in the polynomial coefficients.

One related wording correction: the abstract currently says that the high-density test set avoids progressions whose differences lie **in** the nil-Bohr neighborhood. It should say **outside** the neighborhood. The test set H forces its long progressions to have differences **inside** B; intersecting its translates with a set that already avoids B-differences then eliminates all progressions. [Pasted text.txt, L91-L91]

---

176. User

2026-09-26 12:30:26 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-26T04:30:26.004000Z

please give the bibtex code for the papers that you suggested i cite regarding polychromatic ramsey theory in order to give credit for the idea that the polychromatic result implies the partition result and the density result.

---

177. Assistant

2026-09-26 12:31:09 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-26T04:31:09.546832Z

I'll verify the publication details and provide BibTeX entries, marking the relevant observation or lemma and distinguishing direct references from broader background.

---

178. Assistant

2026-09-26 12:34:18 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-26T04:34:18.886038Z

The clearest direct citations are Pálvölgyi's Observation 4 and Axenovich et al.'s Lemma 3; Bollobás et al. is useful for the partition-into-transversals terminology.

---

179. Assistant

2026-09-26 12:36:05 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-26T04:36:05.893516Z

For this attribution, I would use Pálvölgyi and Axenovich et al. as the two principal references. Bollobás–Pritchard–Rothvoß–Scott is a useful third reference for the terminology and the interpretation of polychromatic colorings as partitions into transversals. The relevant passages are **Observation 4**, **Lemma 3**, and **§1.1**, respectively. Springer arXiv arXiv

Download the BibTeX file, including these three and the four supplementary references from the earlier search

1. Pálvölgyi: the finite deletion principle

Observation 4 and its proof explicitly give the independent-set bound obtained by deleting a smallest color class of a polychromatic coloring. The published reference is *Graphs and Combinatorics* **40** (2024), Article 131. Springer

```
@article{Palvolgyi2024,
  author        = {P{\'a}lv{\"o}lgyi, D{\"o}m{\"o}t{\"o}r},
  title         = {Note on polychromatic coloring of hereditary
                   hypergraph families},
  journal       = {Graphs and Combinatorics},
  volume        = {40},
  number        = {6},
  year          = {2024},
  note          = {Article 131},
  doi           = {10.1007/s00373-024-02836-y},
  eprint        = {2309.04603},
  archivePrefix = {arXiv},
  primaryClass  = {math.CO}
}
```

Use:

```
\cite[Observation~4]{Palvolgyi2024}
```

2. Axenovich et al.: the density version on the integers

Lemma 3, proved in **Section 3**, relates polychromatic colorings to the density of blocking sets for translates of a fixed finite subset of $\mathbb{Z}$. Taking the complement of such a blocking set gives the corresponding large avoiding set. The published article is *Integers* **19** (2019), Paper A18; the DOI below is the one listed on the journal's volume page. arXiv Colgate Math

```
@article{AxenovichEtAl2019,
  author        = {Axenovich, Maria and Goldwasser, John
                   and Lidick{\'y}, Bernard and Martin, Ryan R.
                   and Offner, David and Talbot, John
                   and Young, Michael},
  title         = {Polychromatic colorings on the integers},
  journal       = {Integers},
  volume        = {19},
  year          = {2019},
  note          = {Paper No. A18},
  doi           = {10.5281/zenodo.10705168},
  eprint        = {1704.00042},
  archivePrefix = {arXiv},
  primaryClass  = {math.CO}
}
```

Use:

```
\cite[Lemma~3]{AxenovichEtAl2019}
```

3. Bollobás–Pritchard–Rothvoß–Scott: the transversal interpretation

Section 1.1, "Terminology and Notation," explicitly identifies a polychromatic coloring with a partition of the vertices into sets each meeting every hyperedge. This is the structural observation behind taking complements of color classes. The journal version appeared in *SIAM Journal on Discrete Mathematics* **27** (2013), 240–256. arXiv SIAM

```
@article{BollobasEtAl2013,
  author        = {Bollob{\'a}s, B{\'e}la and Pritchard, David
                   and Rothvo{\ss}, Thomas and Scott, Alex},
  title         = {Cover-decomposition and polychromatic numbers},
  journal       = {SIAM Journal on Discrete Mathematics},
  volume        = {27},
  number        = {1},
  pages         = {240--256},
  year          = {2013},
  doi           = {10.1137/110856332},
  eprint        = {1009.6144},
  archivePrefix = {arXiv},
  primaryClass  = {math.CO}
}
```

Use:

```
\cite[Section~1.1]{BollobasEtAl2013}
```

Suggested attribution in your paper

The following distinguishes the general terminology, the explicit finite bound, and the related density result, without suggesting that any one reference states your entire upper-density formulation verbatim. Those distinctions match the cited passages. arXiv Springer arXiv

```
The passage from a polychromatic obstruction to partition and density
obstructions uses a standard observation: every color class meets
every forbidden configuration, so its complement contains none.
See \cite[Section~1.1]{BollobasEtAl2013} for the interpretation of
polychromatic colorings as partitions into transversals.
Deleting a smallest color class gives the finite independent-set
bound recorded in \cite[Observation~4]{Palvolgyi2024}; a related
density statement for translates of a fixed finite subset of the
integers appears in \cite[Lemma~3]{AxenovichEtAl2019}.
We include the elementary upper-density argument needed here.
```

For a shorter introduction, the two principal references suffice:

```
The following deductions use the standard deletion-of-one-color
argument for polychromatic colorings; see
\cite[Observation~4]{Palvolgyi2024} for its finite form and
\cite[Lemma~3]{AxenovichEtAl2019} for a related density statement
on the integers.
```

180. User

2026-09-26 13:21:17 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-26T05:21:17.745000Z

to show that we really need this property of d bar and not just finite subadditivity, can you make a finitely subadditive counter example for me?

---

181. Assistant

2026-09-26 13:23:10 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-26T05:23:10.754133Z

Yes. There is a very simple counterexample that is not only finitely subadditive, but also normalized, monotone, translation-invariant, and zero on finite sets. Nevertheless, it fails the complement-of-one-color property used in your paragraph. [Pasted text.txt, L216-L219]

Define $\nu : \mathcal{P}(\mathbb{N}) \to [0,1]$ by

$$\boxed{\nu(A) = \begin{cases} 0, & A \text{ is finite}, \\ \frac{1}{2}, & A \text{ and } \mathbb{N} \setminus A \text{ are both infinite}, \\ 1, & \mathbb{N} \setminus A \text{ is finite}. \end{cases}}$$

1. Why ν is finitely subadditive

We verify

$$\nu(A \cup B) \le \nu(A) + \nu(B).$$

If one of the sets, say A, is finite, adding it does not change whether B is finite, infinite with infinite complement, or cofinite. Thus

$$\nu(A \cup B) = \nu(B) = \nu(A) + \nu(B).$$

If both A and B are infinite, then

$$\nu(A) + \nu(B) \ge \frac{1}{2} + \frac{1}{2} = 1 \ge \nu(A \cup B).$$

This proves subadditivity for two sets, and the finite-union version follows by induction.

Also, directly from the definition,

$$\nu(\varnothing) = 0, \qquad \nu(\mathbb{N}) = 1,$$

and ν is monotone and unchanged by finite modifications or translations. It even satisfies the complementary identity

$$\nu(A) + \nu(\mathbb{N} \setminus A) = 1.$$

2. Why the required partition property fails

Partition $\mathbb{N}$ into the three residue classes modulo 3:

$$\mathbb{N} = C_0 \sqcup C_1 \sqcup C_2, \qquad C_i = \{n : n \equiv i \pmod 3\}.$$

For each i, both C_i and its complement are infinite. Hence

$$\nu(\mathbb{N} \setminus C_i) = \frac{1}{2},$$

so

$$\boxed{\max_{0 \le i < 3} \nu(\mathbb{N} \setminus C_i) = \frac{1}{2} < \frac{2}{3} = 1 - \frac{1}{3}.}$$

Thus **finite subadditivity does not imply that the union of some $r-1$ classes of an r-partition has density at least $1 - 1/r$.**

For comparison, normalized subadditivity *does* give the desired assertion for $r = 2$, because

$$1 = \nu(\mathbb{N}) \le \nu(C_0) + \nu(C_1).$$

The failure first appears with three colors.

3. It even disproves the proposed extension of the recurrence theorem

This example does more than invalidate that particular deduction. It shows that the density-to-nilrecurrence theorem would actually be **false** for arbitrary normalized, finitely subadditive set functions.

Use your draft's definition of density largeness with the condition $\nu(A) \ge \delta$, rather than equality. [Pasted text.txt, L188-L200] Take

$$\delta = \frac{2}{3}, \qquad S = \{1\}.$$

Then

$$\nu(A) \ge \frac{2}{3} \quad \Longleftrightarrow \quad A \text{ is cofinite.}$$

Every cofinite set contains arbitrarily long intervals, which are arithmetic progressions with difference 1. Therefore S is $(2/3)$-Szemerédi-large **with respect to ν.**

But S is not even Bohr recurrent: it misses the Bohr neighborhood

$$B = \left\{ n \in \mathbb{N} : \left\| \frac{n}{2} \right\| < \frac{1}{4} \right\} = 2\mathbb{N}.$$

Consequently,

$$\boxed{\begin{array}{c} \nu \text{ is normalized, monotone, translation-invariant, and finitely subadditive,} \\ S \text{ is } (2/3)\text{-Szemerédi-large with respect to } \nu, \\ S \text{ is not Bohr recurrent.} \end{array}}$$

So the additional finite-partition property is genuinely doing work; finite subadditivity alone cannot replace it.

---

182. User

2026-09-26 13:24:28 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-26T05:24:28.736000Z

Please turn this counter example into a footnote with at most 2 sentences that I Can drop into the paper.

183. Assistant

2026-09-26 13:25:40 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-26T05:25:40.307813Z

```
\footnote{Define \(\nu(A)\) to be \(0\) if \(A\) is finite,
\(1\) if \(A\) is cofinite, and \(1/2\) otherwise.
Then \(\nu\) is normalized, monotone, translation-invariant, and
finitely subadditive, but for the partition
\(\N=C_0\sqcup C_1\sqcup C_2\) into residue classes modulo \(3\),
we have \(\max_{0\leq i<3}\nu(\N\setminus C_i)=1/2<2/3\).}
```

---

184. User

2026-09-26 14:23:10 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-26T06:23:10.242000Z

Selected text:

Selection 1

Then, for every fixed finite-step nil-Bohr $_0$ set B and every $\varepsilon > 0$, there is an integer $K = K(B, \varepsilon)$ such that

$$\boxed{r_k(\Omega) \leq r_k(\Omega; B) \leq (1+\varepsilon) r_k(\Omega)} \tag{1}$$

for every $k \geq K$ and every finite $\Omega \subseteq \mathbb{N}$.

My request:

can we also deduce an analogous result for van der Waerden type numbers and not just Szemeredi type numbers?

---

185. Assistant

2026-09-26 14:24:05 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-26T06:24:05.318682Z

I'll check the van der Waerden analogue, carefully distinguishing comparisons involving extra colors from any claimed multiplicative control of the interval length.

---

186. Assistant

2026-09-26 14:27:15 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-26T06:27:15.507946Z

A literal $1+\varepsilon$ interval-length comparison already fails for $B = 2\mathbb{N}$: the restricted threshold is $2W(r,k)-1$, reflecting a general divisibility obstruction.

187. Assistant

2026-09-26 14:28:41 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-26T06:28:41.564608Z

There is a van der Waerden analogue, but the literal $1+\varepsilon$ comparison for the usual van der Waerden numbers is false—even for the Bohr neighborhood $B = 2\mathbb{N}$. What your theorem gives is a comparison with **twice as many colors**:

$$\boxed{W(r,k) \le W(r,k;B) \le W(2r,k) \qquad (k \ge K_B).} \tag{1}$$

There is also a natural **colored extremal quantity** for which the $1+\varepsilon$ comparison does hold, directly from the inequality you selected. I will distinguish these two versions.

1. The usual van der Waerden numbers: a $1+\varepsilon$ ratio is impossible

Write

$$W(r,k) = \min\{N : \text{every } r\text{-coloring of } [N] \text{ has a monochromatic } k\text{-term AP}\},$$

and define the restricted version

$$W(r,k;B) = \min\left\{N : \begin{array}{l}\text{every } r\text{-coloring of } [N] \text{ has a monochromatic} \\ k\text{-term AP whose common difference belongs to } B\end{array}\right\}.$$

These restricted-difference thresholds are studied in the literature; for example, Wesley denotes them by $n(AP_B, k; r)$. Here I put the number of colors first, consistently with $W(r,k)$. arXiv

The obstruction to an almost-equality is elementary.

Proposition

For every $q \ge 2$, $r \ge 2$, and $k \ge 2$,

$$\boxed{W(r,k;q\mathbb{N}) = q\big(W(r,k)-1\big)+1.} \tag{2}$$

Proof

Put $W = W(r,k)$.

Upper bound. The interval

$$[\,q(W-1)+1\,]$$

contains the progression

$$1, 1+q, \dots, 1+(W-1)q.$$

Given any r-coloring, identify these W points with $[W]$. The definition of W gives a monochromatic k-term progression among them. Its common difference is divisible by q. Thus

$$W(r,k;q\mathbb{N}) \le q(W-1)+1.$$

Lower bound. Choose an r-coloring of $[W-1]$ containing no monochromatic k-term progression.

Partition $[q(W-1)]$ into its q residue classes modulo q, each containing $W-1$ points. Color each residue class using a copy of the chosen coloring, in its natural increasing order.

Any progression whose difference is divisible by q lies entirely in one residue class. After identifying that class with $[W-1]$, a monochromatic k-term progression would contradict our choice of coloring. Therefore

$$W(r,k;q\mathbb{N}) > q(W-1),$$

proving (2). □

Now $q\mathbb{N}$ is a Bohr neighborhood of zero:

$$q\mathbb{N} = \left\{ n \in \mathbb{N} : \left\| \frac{n}{q} \right\| < \frac{1}{2q} \right\}.$$

Nevertheless, (2) gives

$$\frac{W(r,k;q\mathbb{N})}{W(r,k)} = q - \frac{q-1}{W(r,k)} \longrightarrow q \qquad (k \to \infty). \tag{3}$$

In particular,

$$\boxed{\frac{W(r,k;2\mathbb{N})}{W(r,k)} \longrightarrow 2, \quad \text{not } 1.}$$

Thus the proposed statement

$$W(r,k;B) \le (1+\varepsilon)W(r,k)$$

for every sufficiently large k cannot hold for all fixed nil-Bohr neighborhoods B. The failure already occurs for a rational circle rotation.

2. What your qualitative two-color theorem does give

The valid comparison (1) follows from the **qualitative 2-vdW-largeness theorem alone**. It does not require the density theorem or the stronger polychromatic formulation.

Proposition

For every fixed finite-step nil-Bohr $_0$ set B, there is $K_B \ge 2$ such that

$$\boxed{W(r,k) \le W(r,k;B) \le W(2r,k)} \tag{4}$$

for every $r \ge 2$ and $k \ge K_B$.

The threshold K_B is independent of r.

Proof

Since $\mathbb{N} \setminus B$ misses a nil-Bohr neighborhood, your theorem implies that it is not 2-vdW-large. Hence there are a binary coloring

$$c_0 : \mathbb{N} \to \{0,1\}$$

and an integer K_B such that

$$\boxed{\text{every monochromatic } K_B\text{-term progression for } c_0 \text{ has difference in } B.} \tag{5}$$

This is exactly the obstruction used in the manuscript's nil-Bohr localization proof. [Pasted text.txt, L1154-L1165]

Fix $k \geq K_B$, and let

$$c : [W(2r,k)] \to \{1,\dots,r\}$$

be arbitrary. Refine it to the $2r$-coloring

$$n \longmapsto (c(n), c_0(n)).$$

There is a monochromatic k-term progression for this refined coloring. It is monochromatic for c, and its first K_B terms are monochromatic for c_0. By (5), its difference lies in B.

This proves the upper bound. The lower bound follows because requiring the difference to lie in B only makes finding a progression harder. □

For all $k \geq 2$, including the smaller lengths, the same proof gives

$$W(r,k;B) \leq W\big(2r, \max\{k, K_B\}\big). \tag{6}$$

The factor two is inside the color parameter. Equation (4) does not assert $W(r,k;B) \leq 2W(r,k)$.

The shrinking-target version also holds

For

$$B_k(c) = \{d : \|P_i(d)\| \leq ck^{-d_i} \text{ for every } i\},$$

the binary case of the quantitative theorem, together with the lemma prescribing the target constant, gives

$$\boxed{W(r,k) \leq W(r,k;B_k(c)) \leq W(2r,k)} \tag{7}$$

whenever

$$k \geq k_0(c,m,D).$$

Here the threshold is independent of the coefficients of the P_i, and also independent of the original number of colors r: the auxiliary obstruction uses only two colors. The required binary shrinking-target coloring follows from the manuscript's omitted-color theorem and target-constant lemma. [Pasted text.txt, L205-L214] [Pasted text.txt, L1175-L1198]

3. A genuine $1+\varepsilon$ analogue for colored extremal numbers

Your selected inequality concerns an **extremal cardinality**, rather than a least interval length. There is a natural coloring analogue at that same extremal level.

For a finite set $\Omega \subseteq \mathbb{N}$, define

$$v_{r,k}(\Omega;B) = \max\left\{ |E| : \begin{array}{l} E \subseteq \Omega \text{ admits an } r\text{-coloring with no} \\ \text{monochromatic } k\text{-term AP of difference in } B \end{array} \right\}, \tag{8}$$

and write

$$v_{r,k}(\Omega) = v_{r,k}(\Omega;\mathbb{N}).$$

I am introducing $v_{r,k}$ here as notation; it is not a notation already defined in your manuscript. When $r=1$, these are precisely your quantities $r_k(\Omega;B)$ and $r_k(\Omega)$.

Corollary

For every fixed finite-step nil-Bohr $_0$ set B and every $\varepsilon > 0$, there is $K = K(B, \varepsilon)$ such that

$$\boxed{v_{r,k}(\Omega) \le v_{r,k}(\Omega; B) \le (1+\varepsilon) v_{r,k}(\Omega)} \tag{9}$$

for every $k \ge K$, every $r \ge 1$, and every finite $\Omega \subseteq \mathbb{N}$.

The same threshold works for every number of colors.

Proof

Use the selected extremal theorem with the given B, ε. Let $E \subseteq \Omega$ admit an r-coloring avoiding monochromatic k-term progressions with difference in B, and write its color classes as

$$E = E_1 \sqcup \cdots \sqcup E_r.$$

Each E_i contains no k-term progression with difference in B. Therefore

$$|E_i| = r_k(E_i; B) \le (1+\varepsilon) r_k(E_i). \tag{10}$$

Choose a k-AP-free subset $F_i \subseteq E_i$ with

$$|F_i| = r_k(E_i) \ge \frac{|E_i|}{1+\varepsilon}.$$

Then

$$F = F_1 \sqcup \cdots \sqcup F_r$$

admits an r-coloring with no monochromatic k-term progression at all: give F_i color i. Moreover,

$$|F| = \sum_i |F_i| \ge \frac{1}{1+\varepsilon} \sum_i |E_i| = \frac{|E|}{1+\varepsilon}.$$

Hence

$$v_{r,k}(\Omega) \ge \frac{|E|}{1+\varepsilon}.$$

Maximizing over E gives the upper bound in (9); the lower bound is immediate. □

This deduction uses exactly the fact that your theorem is uniform over **every finite ambient set**: we apply it separately with ambient sets $E_1, \dots, E_r$. The fixed-neighborhood extremal statement appears at the end of the manuscript's application section. [Pasted text.txt, L1233-L1248]

Consequently, (9) also follows from the qualitative δ-Szemerédi-largeness-to-nilrecurrence theorem, through the translation-averaging argument we previously gave for the selected inequality. Its shrinking-target counterpart follows in precisely the same way from the shrinking-target extremal theorem.

4. Why (9) does not imply almost-equality of the usual van der Waerden numbers

The connection between the two quantities is

$$W(r, k; B) = \min\{N : v_{r,k}([N]; B) < N\}. \tag{11}$$

Thus a van der Waerden number detects the **first failure to color the entire interval**. An extremal comparison permits deletion of a small fraction of its points.

For instance, an estimate

$$v_{r,k}([N];B) \le (1+\varepsilon)v_{r,k}([N])$$

is compatible with

$$v_{r,k}([N];B) = N, \qquad v_{r,k}([N]) = N-1$$

whenever N is sufficiently large. It therefore need not preserve the threshold in (11).

This is why the $q\mathbb{N}$ example causes no contradiction: **large colorable subsets can have almost the same extremal size even though the interval thresholds differ by a factor tending to q.**

Conclusion

The distinctions are:

$$\boxed{\begin{aligned} &\text{Usual van der Waerden thresholds:} && W(r,k;B) \le W(2r,k) \quad \text{for large } k; \\ &\text{Colored extremal cardinalities:} && v_{r,k}(\Omega;B) \le (1+\varepsilon)v_{r,k}(\Omega) \quad \text{for large } k. \end{aligned}}$$

The second is a direct r-color generalization of your Szemerédi extremal comparison. **A $1+\varepsilon$ comparison of the usual van der Waerden numbers themselves is ruled out by the exact formula (2).**

---

188. User

2026-09-26 14:46:58 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-26T06:46:58.361000Z

Is there a term for 'Szemeredi numbers' in the literature, I.e., the N = N(delta,k) such that for any M >= N and any subset of [1,M] of density at least delta, there is a k term arithmetic progression?

---

189. Assistant

2026-09-26 14:47:34 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-26T06:47:34.670087Z

I'll check the terminology used in the literature for this threshold, especially how it is distinguished from the extremal function $r_k(N)$.

---

190. Assistant

2026-09-26 14:50:46 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-26T06:50:46.244703Z

"Szemerédi numbers" does occur, but ambiguously, sometimes meaning $r_k(N)$. I also found literature distinguishing the first valid size from the eventual threshold valid thereafter.

191. Assistant

2026-09-26 14:53:44 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-26T06:53:44.901156Z

Yes—"Szemerédi number" is explicitly used in the literature for this kind of density threshold. It is a legitimate term to use in your paper, but you should define it because the notation—and occasionally the quantity meant by the name—varies between authors. Indian Academy of Sciences ScienceDirect Cheriton School of Computer Science

Explicit references

Bidisha Roy and Subha Sarkar, "Regularity of certain diophantine equations," *Proceedings—Mathematical Sciences* 129 (2019), Article 19. In the passage immediately before **Lemma 2.1**, they denote the least integer in the finite Szemerédi theorem by $M(k,\delta)$ and explicitly call it a **"Szemerédi number."** The DOI is **10.1007/s12044-019-0463-2**. Indian Academy of Sciences Springer

Alexander Fish, "On product of difference sets for sets of positive density," arXiv:1702.02544. This is particularly useful for your precise convention. In the **proof of Lemma 2.1, printed page 3**, Fish introduces $N(\ell,\mu(A))$ with the property that **for every** $N \geq N(\ell,\mu(A))$, every subset of $[N]$ of density at least $\mu(A)$ contains an $(\ell+1)$-term progression. In **§4, "Further problems," printed page 5**, he calls this quantity **"Szemerédi's number."** Thus this reference uses exactly your "all sufficiently large intervals" formulation, apart from indexing the progression length by $\ell+1$. School of Mathematics and Statistics

A further published example is **Sadık Eyidoğan, Haydar Göral, and Mustafa Kutay Kutlu, "Arithmetic progressions in certain subsets of finite fields," *Finite Fields and Their Applications* 91 (2023), Article 102264**. They explicitly call the smallest threshold $N(\varepsilon,k)$ the **Szemerédi number**, denoted $S(\varepsilon,k)$. ScienceDirect ScienceDirect

The definition I would use in your paper

I would write **"the Szemerédi number $N_{\mathrm{Sz}}(k,\delta)$"** and define it by

$$N_{\mathrm{Sz}}(k,\delta)=\min\left\{N\in\mathbb{N}:\ \begin{array}{l}\text{for every integer } M\geq N \text{ and every } A\subseteq[M],\\ |A|\geq\delta M \implies A \text{ contains a nondegenerate } k\text{-term AP}\end{array}\right\}.$$

For example:

```
For \(k\geq2\) and \(0<\delta\leq1\), let
\(N_{\mathrm{Sz}}(k,\delta)\) be the least positive integer \(N\)
such that, for every \(M\geq N\), every subset of \([M]\) of
cardinality at least \(\delta M\) contains a nondegenerate
\(k\)-term arithmetic progression.
We call \(N_{\mathrm{Sz}}(k,\delta)\) the Szemer\'edi number.
```

For an integer satisfying this property without necessarily being the **least** one, I would use the descriptive phrase **"a valid Szemerédi threshold."**

Two distinctions worth preserving

First, this threshold is different from your manuscript's extremal function

$$r_k(M)=\max\{|A|:A\subseteq[M]\text{ contains no }k\text{-term AP}\}.$$

Your draft uses r_k in precisely this extremal-cardinality sense. [Pasted text.txt, L297-L303] The exact relationship is

$$\boxed{N_{\mathrm{Sz}}(k,\delta) = \min\{N : \ r_k(M) < \delta M \text{ for every } M \geq N\}.}$$

The strict inequality is appropriate because your density hypothesis is $|A| \geq \delta M$.

Second, **retaining "for every $M \geq N$" matters**. The property at one interval length need not persist at the next length. For example, at density $\delta = 0.7$, every qualifying subset of $[3]$ is all of $[3]$, so contains a three-term progression; but $\{1,2,4\} \subseteq [4]$ has density $3/4$ and contains none.

There is also genuine terminological variation: **Kevin O'Bryant, "Sets of Natural Numbers with Proscribed Subsets," §1, printed page 2**, uses "Szemerédi number" for the least N with $r_k(N) = n$, an inverse extremal quantity rather than your fixed-density threshold. Cheriton School of Computer Science

So "Szemerédi number" is supported by published usage; an explicit definition with your all-M convention is the right way to make it unambiguous.

---

192. User

2026-09-26 17:30:03 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-26T09:30:03.656000Z

Here is an updated draft of the paper. Can you give me a paragraph to add to the end of the introduction to say that the idea of deducing partition regular and density regular corollaries from polychromatic results was previously established in (some of) the 3 papers that you gave me earlier and that I already added to the bibtex. Give some precise page/theorem numbers for what you reference in those papers.

> **Attachments recorded in source:**
>
> - Pasted text(20260926-092902).txt

---

193. Assistant

2026-09-26 17:30:38 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-26T09:30:38.912254Z

I'll check the updated introduction and verify the citation pinpoints, then give you a paragraph to insert at the end.

---

194. Assistant

2026-09-26 17:33:49 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-26T09:33:49.250298Z

One citation detail needs correction: Axenovich et al.'s density result is Lemma 1 in the published *Integers* paper, although it is Lemma 3 on arXiv.

195. Assistant

2026-09-26 17:35:38 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-26T09:35:38.662061Z

I would place the following paragraph immediately before `\textbf{Organization:}`. It attributes the general mechanism without suggesting that the cited papers already contain your generalized-polynomial conclusions. Your introduction has already given the relevant complement-of-one-color argument. [Pasted text(20260926-092902).txt, L32-L44]

```
The passage from Theorem~\ref{thm:omitted} to its partition and
density consequences uses a standard principle from the theory of
polychromatic colorings. Such a coloring partitions the vertices
of a hypergraph into transversals, each meeting every edge; see
\cite[Section~1.1]{BollobasEtAl2013}. Consequently, merging the
colors into two nonempty groups gives a coloring with no
monochromatic edge, while the complement of each original color
class is independent. In the finite setting, deleting a smallest
color class retains at least a proportion \(1-1/r\) of the vertices;
this observation is recorded explicitly in
\cite[Observation~4, p.~2]{Palvolgyi2024}. A corresponding density
argument for translates of a fixed finite subset of \(\Z\) appears
in \cite[Lemma~1, p.~3, and its proof on p.~10]{AxenovichEtAl2019},
where polychromatic colorings yield blocking sets of small density.
Our upper-density deduction above applies the same
complement-of-one-color principle to the family of progressions
whose differences violate the prescribed approximation conditions.
```

Verified pinpoint references

Bollobás–Pritchard–Rothvoß–Scott, *Cover-Decomposition and Polychromatic Numbers*. Section **1.1, "Terminology and Notation,"** explicitly identifies polychromatic colorings with partitions into transversals. The passage is on **page 2 of arXiv:1009.6144v3**. This supports the interpretation used in the paragraph; the merging-of-colors conclusion follows immediately from it. arXiv

Pálvölgyi, *Note on Polychromatic Coloring of Hereditary Hypergraph Families*. Observation 4 and its proof, page 2 of the published article, explicitly give the independent-set bound by deleting a smallest color class. The numbering and page coincide with those in the preprint. D-NB

Axenovich et al., *Polychromatic Colorings on the Integers*. In the **published *Integers* version**, the relevant statement is **Lemma 1 on page 3**, and its complete proof is on **page 10 in Section 3**. They prove $\alpha(S) \leq 1/p(S)$, where $\alpha(S)$ is the infimum of blocking-set densities. The proof produces a periodic polychromatic coloring and selects a color class of small density; its complement contains no translate of S. Colgate Math

One correction to my earlier citation advice: I previously gave **Lemma 3**, which is the numbering in **arXiv:1704.00042v2**, not in the published paper. Since your bibliography cites the published article, **Lemma 1 is the correct reference to use**. arXiv Colgate Math

196. User

2026-09-26 18:16:30 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-26T10:16:30.494000Z

what is alpha(G) in graph theory?

197. Assistant

2026-09-26 18:17:58 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-26T10:17:58.307890Z

For a finite graph $G = (V, E)$, ** $\alpha(G)$ denotes its independence number**: the largest number of vertices one can select with no two selected vertices joined by an edge. Thus

$$\boxed{\alpha(G) = \max\{|A| : A \subseteq V,\ \{u, v\} \notin E \text{ for all distinct } u, v \in A\}.}$$

Such a set A is called an **independent set**.

For example, the path

$$1 - 2 - 3 - 4 - 5$$

has independence number 3: we can select $\{1, 3, 5\}$, but cannot select four vertices without including an adjacent pair.

The hypergraph version relevant to your paper

For a hypergraph $\mathcal{H} = (V, \mathcal{E})$, edges may contain more than two vertices. Here

$$\boxed{\alpha(\mathcal{H}) = \max\{|A| : A \subseteq V,\ e \nsubseteq A \text{ for every } e \in \mathcal{E}\}.}$$

In other words, an independent set contains **no entire hyperedge**. It may contain several vertices of an edge, provided it does not contain all of them.

This makes your extremal quantity $r_k(\Omega)$ an independence number. Define $\mathcal{H}_k(\Omega)$ to have vertex set Ω and edges consisting of all k-term arithmetic progressions contained in Ω. Then

$$\boxed{\alpha(\mathcal{H}_k(\Omega)) = r_k(\Omega),}$$

because both sides maximize the size of a progression-free subset of Ω. Your draft defines $r_k(\Omega)$ in precisely this way. [Pasted text.txt, L232-L239]

Why it appears in the polychromatic argument

If every hyperedge contains all r colors, deleting any one color class leaves an independent set. Deleting a *smallest* color class therefore gives

$$\boxed{\alpha(\mathcal{H}) \geq \left\lceil \frac{r-1}{r} |V| \right\rceil.}$$

That is exactly the independent-set bound we were discussing.

One notation distinction: **the $\alpha(S)$ in the Axenovich et al. blocking-density statement we discussed is a different quantity**, not this graph-theoretic independence number.

---

198. User

2026-09-26 18:22:01 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-26T10:22:01.214000Z

Please propose a better name for section 3.8.

> **Attachments recorded in source:**
>
> - Pasted text(20260926-102124).txt

---

199. Assistant

2026-09-26 18:23:23 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-26T10:23:23.711839Z

I would use **“Control of the weighted leading coefficients”**:

```
\subsection{Control of the weighted leading coefficients}
```

The current title, “The first-exit argument,” names only one part of the proof. The subsection first shows that the polynomial lift remains bounded, then extracts its leading coefficients and clears denominators to obtain

$$\delta_L \tau(p)\gamma = E, \qquad |E_{ij}| \ll k^{-w_{ij}}.$$

The proposed title describes the **conclusion of the whole subsection**, rather than just the argument used to establish boundedness. [Pasted text.txt, L717-L726] [Pasted text.txt, L769-L799]

It also leads naturally into the next subsection, **“Recovering the special generalized polynomials”**: first control the weighted leading coefficients, then interpret that control as the desired Diophantine approximation. [Pasted text.txt, L807-L826]

---

200. User

2026-09-26 20:04:24 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-26T12:04:24.907000Z

is there a natural extension of our theorem 1.9, which if true, would simultaneously solve the 2-large is large conjecture and the density variation that we ask?

---

201. Assistant

2026-09-26 20:05:00 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-26T12:05:00.874970Z

I’ll inspect Theorem 1.9’s quantifiers and seek a natural polychromatic extension strong enough to imply both conjectures without building either conclusion into the hypothesis.

---

202. Assistant

2026-09-26 20:08:13 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-26T12:08:13.448599Z

A natural candidate replaces nil-Bohr targets by difference sets supporting fixed-length progressions in positive-density sets; this would actually imply the stronger statement that every 2-vdW-large set is Szemerédi-large.

203. Assistant

2026-09-26 20:15:20 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-26T12:15:20.731895Z

Yes. A natural common extension is to replace the nil-Bohr target in the qualitative consequence of Theorem 1.9 by the set of differences supporting a prescribed-length progression in an arbitrary positive-density set.

This would solve both questions. **It would actually prove something stronger: every 2-vdW-large set is Szemerédi-large.** I would present it as a conjectural extension, not as an equivalent reformulation of the two questions.

Your Theorem 1.9 provides polychromatic colorings for progressions whose differences violate generalized-polynomial approximation conditions. The proposed extension keeps precisely that polychromatic conclusion but broadens the class of forbidden differences. [Pasted text.txt, L199-L204]

1. The proposed extension

For $A \subseteq \mathbb{N}$ and $\ell \geq 2$, define

$$R_\ell(A) = \{d \in \mathbb{N} : \exists b \in \mathbb{N},\ b, b+d, \dots, b+(\ell-1)d \in A\}.$$

Thus $R_\ell(A)$ is the set of common differences of ℓ-term progressions contained in A.

Conjectural extension: polychromatic transfer to positive-density sets

Let $A \subseteq \mathbb{N}$ satisfy $\overline{d}(A) > 0$, and let $\ell, r \geq 2$. There exist $K \geq 2$ and a coloring

$$c : \mathbb{N} \longrightarrow \{1, \dots, r\}$$

such that, for all $a, d \in \mathbb{N}$,

$$\boxed{\{c(a+jd) : 0 \leq j < K\} \neq \{1, \dots, r\} \quad \Longrightarrow \quad d \in R_\ell(A).} \tag{P}$$

In words:

> **A sufficiently long progression omitting a color has the same common difference as an ℓ-term progression somewhere in A.**

The two progressions need not have the same starting point, and the progression inside A need not be contained in the original progression. **Only their common difference must agree exactly.**

Equivalently,

$$d \notin R_\ell(A) \quad \Longrightarrow \quad \text{every } K\text{-term progression of difference } d \text{ contains all } r \text{ colors.}$$

The same coloring automatically works for every length $k \geq K$, by considering the first K terms. Here K is allowed to depend on A, ℓ, r; the conjecture does not ask for a quantitative bound depending only on $\overline{d}(A), \ell, r$.

2. Why this solves the 2-large-is-large conjecture

Assume (P), and let S be 2-vdW-large.

Fix **any** set $A \subseteq \mathbb{N}$ with $\overline{d}(A) > 0$ and any desired progression length $\ell \geq 2$. Apply (P) with $r = 2$, obtaining a binary coloring and a threshold K.

Since S is 2-vdW-large, there is a monochromatic K-term progression with common difference $d \in S$. A monochromatic progression omits one of the two colors, so (P) gives

$$d \in R_\ell(A).$$

Therefore A contains an ℓ-term progression with common difference in S.

Since A and ℓ were arbitrary,

$$\boxed{S \text{ is 2-vdW-large} \quad \Longrightarrow \quad S \text{ is Szemerédi-large.}} \tag{1}$$

This implies the Brown–Graham–Landman conjecture. Indeed, in any finite coloring

$$\mathbb{N} = C_1 \sqcup \cdots \sqcup C_q,$$

some C_i has positive upper density. Applying Szemerédi-largeness to that cell gives arbitrarily long monochromatic progressions with differences in S.

Thus

$$2\text{-vdW-large} \Longrightarrow \text{Szemerédi-large} \Longrightarrow \text{vdW-large}.$$

The first implication is stronger than the conjecture stated in your introduction. [Pasted text.txt, L102-L116]

3. Why it also solves the density-independence question

Assume (P), and suppose S is δ_0-Szemerédi-large for some fixed $0 < \delta_0 < 1$.

Again fix an arbitrary positive-upper-density set $A \subseteq \mathbb{N}$ and an arbitrary $\ell \geq 2$. Choose $r \geq 2$ such that

$$1 - \frac{1}{r} \geq \delta_0,$$

and apply (P) to A, ℓ, r.

Write $C_1, \ldots, C_r$ for the resulting color classes. The finite-palette argument already used in your paper gives an index e such that

$$\overline{d}(\mathbb{N} \setminus C_e) \geq 1 - \frac{1}{r} \geq \delta_0.$$

No equal-density or equidistribution assertion about the color classes is needed. [Pasted text.txt, L206-L218]

By δ_0-Szemerédi-largeness, $\mathbb{N} \setminus C_e$ contains a K-term progression with common difference $d \in S$. That progression omits color e, so (P) implies

$$d \in R_\ell(A).$$

Hence A contains an ℓ-term progression with common difference in S.

Since A and ℓ were arbitrary,

$$\boxed{S \text{ is } \delta_0\text{-Szemerédi-large for some } \delta_0 \in (0,1) \quad \Longrightarrow \quad S \text{ is Szemerédi-large.}} \tag{2}$$

This is exactly the positive answer to your density-independence question. [Pasted text(20260926-102124).txt, L24-L36]

4. An equivalent formulation that makes the connection to Theorem 1.9 especially clear

The proposed extension can also be stated entirely in terms of a set S of forbidden differences.

Polychromatic amplification of density obstructions

> **If $S \subseteq \mathbb{N}$ is not Szemerédi-large, then for every $r \geq 2$ there are $K \geq 2$ and an r-coloring of $\mathbb{N}$ such that every K-term progression with common difference in S contains all r colors.**

To see the equivalence with (P), suppose first that S is not Szemerédi-large. There are a positive-upper-density set A and an integer $\ell \geq 2$ such that A contains no ℓ-term progression with common difference in S. Thus

$$S \cap R_\ell(A) = \varnothing.$$

Apply (P).

Conversely, given A, ℓ, the set

$$S = \mathbb{N} \setminus R_\ell(A)$$

is not Szemerédi-large, with A itself witnessing this failure. Applying the amplification statement to this S gives (P).

Your current theorem establishes this conclusion when S has a nil-Bohr obstruction. More explicitly, suppose

$$\max_i \|P_i(d)\| \geq \eta \qquad (d \in S)$$

for some finite family and some $\eta > 0$. Given r, choose K sufficiently large that the estimates in Theorem 1.9 would force

$$\|P_i(d)\| < \eta$$

for every i. A K-term progression of difference $d \in S$ therefore cannot omit a color. [Pasted text.txt, L199-L204]

The proposed change is consequently:

$$\boxed{\begin{array}{c} S \text{ has a nil-Bohr obstruction} \\ \Downarrow \text{ replace the hypothesis by} \\ S \text{ has any positive-density progression obstruction.} \end{array}}$$

This is a **qualitative extension of the obstruction mechanism** in Theorem 1.9. It does not attempt to preserve its weight-dependent shrinking-target estimates for objects that have no assigned polynomial weight.

5. Why this is a genuine extension beyond the nilpotent case

There are two useful ways to see the relationship.

Nil-Bohr targets fit into the return-set framework

Huang–Shao–Ye's Theorem A(1) says that every s-step nil-Bohr $_0$ set B contains

$$N^s(U) = \{d : U \cap T^{-d}U \cap \cdots \cap T^{-sd}U \neq \varnothing\}$$

for a minimal nilsystem and a nonempty open set U. The same theorem is restated as Theorem 5.2 in Host–Kra–Maass. arXiv arXiv

Fix a point x, and set

$$A = \{n \in \mathbb{N} : T^n x \in U\}.$$

Minimality and compactness make A syndetic: finitely many inverse translates of U cover the space. In particular, $\overline{d}(A) > 0$. Also,

$$R_{s+1}(A) \subseteq N^s(U) \subseteq B.$$

Thus (P), applied to this A, would give the qualitative polychromatic obstruction for the complement of B.

General positive-density return sets need not contain a nil-Bohr neighborhood

Alweiss's Theorem 1.14 supplies a set S_0 that meets every finite-step nil-Bohr neighborhood but admits a finite coloring with no monochromatic three-term progression of difference in S_0. arXiv

Choose a positive-upper-density color class A. Then

$$R_3(A) \cap S_0 = \varnothing.$$

Consequently, $R_3(A)$ cannot contain a nil-Bohr neighborhood: otherwise S_0 would meet it.

So there are already concrete positive-density return targets to which your current nil-Bohr theorem cannot be applied simply by finding a smaller nil-Bohr neighborhood inside them. **The proposed extension would have to deal with genuinely new obstructions.**

6. The strength of the proposal—and a tempting weaker version that is insufficient

Under (P), all four properties would coincide:

$$\boxed{\begin{aligned} & S \text{ is 2-vdW-large} \\ \iff\ & S \text{ is vdW-large} \\ \iff\ & S \text{ is Szemerédi-large} \\ \iff\ & S \text{ is } \delta\text{-Szemerédi-large for some } \delta \in (0,1). \end{aligned}} \tag{3}$$

Thus this is an ambitious common conjecture. In particular, it bridges **partition largeness and density largeness**, not just the number of colors or the density parameter separately. The distinction between chromatic and density recurrence is explicitly discussed in Section 4 of Robertson's *Down the Large Rabbit Hole*; that discussion does not provide such a bridge. arXiv

A weaker proposal would be:

> If S is not vdW-large, then its sufficiently long progressions admit polychromatic colorings with arbitrarily many colors.

That would prove the Brown–Graham–Landman conjecture. The complement-of-one-color argument would also show

$$\delta\text{-Szemerédi-large for some } \delta \implies \text{vdW-large}.$$

But it would **not by itself** establish Szemerédi-largeness at every positive density: one would still need to pass from vdW-largeness to density largeness.

This is why I would use **positive-density progression obstructions**, rather than only finite-color obstructions, in the common extension.

Suggested formulation for the paper

I would pose it as a question, with its stronger consequence made explicit:

```
For \(A\subseteq\N\) and \(\ell\geq2\), write
\[
 R_\ell(A)
 =
 \{d\in\N:\text{there exists }b\in\N
   \text{ with }\{b,b+d,\ldots,b+(\ell-1)d\}\subseteq A\}.
\]

\begin{question}[Polychromatic transfer]\label{qu:polychromatic-transfer}
Let \(A\subseteq\N\) have positive upper density, and let
\(\ell,r\geq2\).
Do there exist \(K\geq2\) and a coloring
\(c:\N\to\{1,\ldots,r\}\) such that, for every \(a,d\in\N\),
\[
 \{c(a+jd):0\leq j<K\}\neq\{1,\ldots,r\}
 \quad\Longrightarrow\quad d\in R_\ell(A)?
\]
\end{question}

A positive answer would imply both the Brown--Graham--Landman
conjecture and Question~\ref{qu:density}.
Indeed, the case \(r=2\) would show that every \(2\)-vdW-large
set is Szemer\'edi-large, while taking \(r\) with
\(1-1/r\geq\delta\) and deleting a suitable color class would
show the same for every \(\delta\)-Szemer\'edi-large set.
```

This is the extension I would investigate first: it preserves the exact omitted-color structure of Theorem 1.9, has a concrete combinatorial target, and implies both desired conclusions by the same two deductions already used in your paper. Its truth, however, requires substantially more than the present nilpotent argument.

---

204. User

2026-09-28 13:04:12 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-28T05:04:12.200000Z

does our polychromatic result imply that every 2-syndetic-large set is also a set of nilbohr recurrence?

205. Assistant

2026-09-28 13:04:49 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-28T05:04:49.216283Z

I'll check the precise definition of 2-syndetic-largeness and whether the polychromatic obstruction produces the syndetic test set needed for nil-Bohr recurrence.

206. Assistant

2026-09-28 13:08:01 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-28T05:08:01.811173Z

Host–Kra–Maass's notion uses test sets meeting every pair of consecutive integers; high density alone does not ensure this, so the density deduction is insufficient by itself.

207. Assistant

2026-09-28 13:17:13 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-28T05:17:13.649033Z

The standard comparison does yield 3-syndetic-large $\Rightarrow$ 2-vdW-large, hence nil-Bohr recurrence. The 2-syndetic case needs additional control of consecutive omissions.

208. Assistant

2026-09-28 13:19:03 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-28T05:19:03.009646Z

I would not yet state this as a corollary. The existing argument proves nil-Bohr recurrence for 3-syndetic-large sets, but I have not established the implication for 2-syndetic-large sets. The obstacle is that deleting one color produces a set of high upper density, but does not necessarily produce a set with gaps at most two.

Here is the precise distinction, including the additional statement that would settle your question.

1. The definition and the direction of the known implications

Using Host–Kra–Maass's definitions, a set $A \subseteq \mathbb{N}$ is ** 2-syndetic** if

$$A \cap \{n, n+1\} \neq \varnothing \qquad (n \in \mathbb{N}).$$

A set $S \subseteq \mathbb{N}$ is ** 2-syndetic-large** if every such A contains arbitrarily long arithmetic progressions with common difference in S. See their Definitions 7.10–7.11. arXiv

Their Proposition 7.12 gives

$$\boxed{\text{3-syndetic-large} \implies \text{2-vdW-large} \implies \text{2-syndetic-large.}} \tag{1}$$

Thus 2-syndetic-largeness is the **weaker hypothesis** in this chain; we cannot apply our 2-vdW-largeness theorem to it directly. arXiv

2. Why deleting a color does not resolve the issue

For a prescribed nil-Bohr neighborhood B, our polychromatic theorem supplies, for each $r \geq 2$, a coloring c and a length K such that

$$d \notin B \quad \Longrightarrow \quad \text{every } K\text{-term progression of difference } d \text{ contains all } r \text{ colors.}$$

This is the fixed-target consequence of the approximation statement in the draft. [Pasted text.txt, L199-L204]

Consequently, every complement

$$A_e = \{n : c(n) \neq e\}$$

has the property

$$\text{every } K\text{-term progression in } A_e \text{ has difference in } B. \tag{2}$$

Moreover, some A_e has upper density at least $1 - 1/r$, which is the deduction already used in the introduction. [Pasted text.txt, L206-L218]

But to apply 2-syndetic-largeness, we would need

$$\boxed{A_e \cap \{n, n+1\} \neq \varnothing \quad \text{for every } n.} \tag{3}$$

Equivalently, the deleted color class $c^{-1}(e)$ must contain **no two consecutive integers**. The polychromatic conclusion does not assert that.

A concrete illustration, even for a Bohr target

Consider the periodic coloring

$$0, 0, 1, 1, 2, 2, \ 0, 0, 1, 1, 2, 2, \ldots$$

and take

$$B = 3\mathbb{N}.$$

Every six-term progression whose difference is not divisible by 3 contains all three colors. Indeed, modulo 6, its step either runs through all six residues or runs through the three residues of one parity; either way, all three colors occur.

Nevertheless, **none of the three color complements is 2-syndetic**, because each deleted color occurs in a pair of consecutive positions. Each complement even has natural density $2/3$, so existence and equality of the color-class densities do not solve the problem.

This is **not** a counterexample to your proposed recurrence implication. It shows only that the required gap bound does not follow from polychromaticity by taking a color complement.

3. The 3-syndetic conclusion does follow, with a short proof

The following deduction uses only the binary case of our theorem.

Proposition

Every 3-syndetic-large set is a set of nil-Bohr recurrence.

Proof

Fix a nil-Bohr neighborhood B. Take a binary coloring

$$c : \mathbb{N} \to \{0, 1\}$$

and $K \geq 2$ such that every monochromatic K-term progression has difference in B, as supplied by the manuscript's theorem. [Pasted text.txt, L199-L206]

Encode this coloring by the set

$$E = \{2n - 1 + c(n) : n \in \mathbb{N}\}.$$

There is exactly one point of E in each pair

$$\{2n - 1, 2n\}.$$

Successive selected points are therefore at distance at most 3, so E is 3-syndetic.

If S is 3-syndetic-large, choose a $2K$-term progression in E with common difference $d \in S$. Its even-indexed terms form

$$a, a + 2d, \ldots, a + 2(K - 1)d.$$

These terms have one fixed parity. Writing

$$a = 2b - 1 + e, \qquad e \in \{0, 1\},$$

their membership in E gives

$$c(b) = c(b + d) = \cdots = c(b + (K - 1)d) = e.$$

Hence $d \in B$. Since B was arbitrary, S is nil-Bohr recurrent. □

This is the encoding behind the relevant case of HKM's Proposition 7.12(ii). Notice exactly where the number 3 appears: selecting one point from each adjacent pair can produce consecutive selections $2n - 1$ and $2n + 2$, a gap of three. arXiv

4. The precise additional statement needed for 2-syndetic-largeness

Your proposed conclusion is equivalent to the following test-set statement:

> **For every finite-step nil-Bohr $_0$ set B, there exist a 2-syndetic set $A \subseteq \mathbb{N}$ and $K \geq 2$ such that every K-term arithmetic progression in A has its difference in B.**

$$\boxed{A \cap \{n, n + 1\} \neq \varnothing \ \ (\forall n), \qquad \{a, a + d, \ldots, a + (K - 1)d\} \subseteq A \Longrightarrow d \in B.} \tag{4}$$

The equivalence is elementary.

If (4) holds and S is 2-syndetic-large, apply its defining property to A. The resulting K-term progression has difference in $S \cap B$, proving nil-Bohr recurrence.

Conversely, suppose every 2-syndetic-large set is nil-Bohr recurrent. For a fixed B, its complement $\mathbb{N} \setminus B$ is not nil-Bohr recurrent, hence cannot be 2-syndetic-large. Negating that definition produces exactly the set A and length K in (4).

In the language of the polychromatic argument, we need a set $H = \mathbb{N} \setminus A$ which simultaneously:

$$\text{meets every sufficiently long progression with difference outside } B,$$
$$\text{contains no two consecutive integers.}$$

Our color classes provide the first property, but the present theorem does not provide the second.

So the conclusion for 2-syndetic-large sets remains unproved in this deduction—not disproved. The useful next strengthening is not another upper-density estimate: it is the construction of a progression-hitting set whose points are never consecutive, equivalently the 2-syndetic test set in (4).

---

209. User

2026-09-28 13:44:17 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-28T05:44:17.489000Z

Please try to prove that 2-syndetic large implies nil-bohr recurrence. Give it a serious try.

---

210. Assistant

2026-09-28 13:44:52 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-28T05:44:52.113485Z

I'll try to construct a 2-syndetic test set for each nil-Bohr neighborhood, checking whether the polychromatic polynomial argument can enforce the required gap bound.

---

211. Assistant

2026-09-28 13:51:03 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-28T05:51:03.681552Z

A promising workaround alternates two separated forbidden regions along a small root of the nilrotation, preventing consecutive exclusions. The key remaining check is preserving difference d, not merely a multiple.

---

212. Assistant

2026-09-28 14:00:15 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-28T06:00:15.682997Z

This now appears to prove the claim, after extending shell-and-rounding to triangular Mal'cev coordinates. Choosing the root after bounding auxiliary steps recovers exactly d via every $2L$-th point.

---

213. Assistant

2026-09-28 14:08:16 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-28T06:08:16.804766Z

Yes. I found a way to prove the full implication. The gap-of-three obstruction from the previous attempt can be overcome by repairing the binary encoding and making the underlying nilrotation sufficiently slow.

The repair introduces a small ambiguity: instead of obtaining an exactly monochromatic progression in the original coloring, we obtain a progression whose terms can each be moved forward by at most one integer to become monochromatic. A sufficiently deep root of the nilrotation makes these one-step changes small, and

the shell argument is robust enough to tolerate them.

This uses more than the stated polychromatic theorem as a black box. It uses the positive margins inside the shell bands that are already present in its proof. [Pasted text.txt, L597-L613]

1. The result

Recall that $A \subseteq \mathbb{N}$ is 2-syndetic when

$$A \cap \{n, n+1\} \neq \varnothing \qquad (n \geq 1).$$

A set $S \subseteq \mathbb{N}$ is 2-syndetic-large if every such A contains arbitrarily long arithmetic progressions with common difference in S. These are the conventions of Host–Kra–Maass, Definitions 7.10–7.11. arXiv

Theorem

Every 2-syndetic-large set is a set of nil-Bohr recurrence.

More precisely, the following test-set statement holds:

> For every finite-step nil-Bohr $_0$ set $B \subseteq \mathbb{N}$, there are a 2-syndetic set $A \subseteq \mathbb{N}$ and an integer $K \geq 2$ such that
>
> $$\boxed{\{a, a+d, \ldots, a+(K-1)d\} \subseteq A \quad \Longrightarrow \quad d \in B.} \tag{1}$$

Applying 2-syndetic-largeness to A immediately gives $S \cap B \neq \varnothing$.

I will first explain the new encoding, then state and prove the robust shell lemma needed to complete the construction.

2. Repairing the binary encoding

Let

$$c : \mathbb{N} \longrightarrow \{0, 1\}$$

be any binary coloring. Define

$$\boxed{A_c = \{2n-1 : c(n) = 0\} \ \cup \ \{2n : c(n) = 1 \text{ or } c(n+1) = 1\}.} \tag{2}$$

This set is **always 2-syndetic**.

Indeed, the pair $\{2n-1, 2n\}$ meets A_c: use the odd member when $c(n) = 0$, and the even member when $c(n) = 1$. Similarly, $\{2n, 2n+1\}$ meets A_c: use the odd member when $c(n+1) = 0$, and the even member when $c(n+1) = 1$.

Now suppose

$$a, a+d, \ldots, a+(2k-1)d \in A_c.$$

Consider the k even-indexed terms

$$a, a+2d, \ldots, a+2(k-1)d.$$

If $a = 2u - 1$ is odd, their membership in A_c gives

$$c(u + jd) = 0 \qquad (0 \leq j < k).$$

If $a = 2u$ is even, it gives

$$c(u + jd) = 1 \quad \text{or} \quad c(u + jd + 1) = 1 \qquad (0 \le j < k).$$

Thus in either case there are a color $e \in \{0, 1\}$ and choices $\epsilon_j \in \{0, 1\}$ such that

$$\boxed{c(u + jd + \epsilon_j) = e \qquad (0 \le j < k).} \tag{3}$$

The common difference after compression is still d. We have not replaced the allowed difference by $2d$, nor assumed any dilation closure of S.

The question is now whether our nilpotent coloring can tolerate the individual perturbations ϵ_j. The next lemma supplies exactly that.

3. A robust version of the shell argument

Let

$$Y = G/\Gamma$$

be a compact nilmanifold with G connected and simply connected, and fix a Riemannian metric d_Y.

Robust slow-motion lemma

There are a continuous function $F : Y \to \mathbb{R}$, numbers $h, \beta, C > 0$, and an integer $W \ge 1$, with the following property.

Put

$$\mathcal{I}_e = \bigcup_{\substack{j \in \mathbb{Z} \\ j \equiv e \pmod 2}} [jh, (j+1)h], \qquad e \in \{0, 1\},$$

and

$$\kappa(y) = \lfloor F(y)/h \rfloor \pmod 2.$$

Let

$$L_0 = \operatorname{lcm}(1, \dots, W).$$

Suppose $p(t) = g^t v$, with $g, v \in G$, and $k \ge 4W + 2$. If, for some $e \in \{0, 1\}$,

$$\boxed{\operatorname{dist}\big(F(p(n)\Gamma), \mathcal{I}_e\big) < \beta \qquad (0 \le n < k),} \tag{4}$$

then, for every positive integer L divisible by L_0, there is an integer b, $0 \le b < W$, such that

$$\boxed{d_Y\big(p(b + jL)\Gamma, p(b)\Gamma\big) \le C\,\frac{jL}{k}} \tag{5}$$

for all integers $j \ge 0$ with $jL \le k/2$.

All constants and the coloring depend only on the fixed nilmanifold and coordinate system. **They are independent of g, v, L, and k.**

Condition (4) is a thickened monochromaticity condition: the observations need only stay close to the scalar bands of one color.

I prove this lemma in Section 5 below. First, here is how it completes the recurrence argument.

4. Constructing the 2-syndetic test set

4.1. Realize the nil-Bohr target by multiple returns

Fix a finite-step nil-Bohr $_0$ set B. Huang–Shao–Ye's Theorem A(1) gives a minimal nilsystem (Z, T), a nonempty open set $U \subseteq Z$, and an integer $\ell \geq 1$ such that

$$\boxed{\{d \geq 1 : U \cap T^{-d}U \cap \cdots \cap T^{-\ell d}U \neq \varnothing\} \subseteq B.} \tag{6}$$

arXiv

Using the standard reduction recorded in their §3.1.2, realize this system as a compact invariant subset Z of an ambient nilmanifold

$$Y = G/\Gamma,$$

where G is connected and simply connected and T is the restriction of translation by some $\tau \in G$. We retain minimality on Z; **we do not assume minimality on all of** Y. arXiv

Write

$$\tau^t = \exp(t \log \tau).$$

There is a number $\lambda > 0$ such that

$$z \in Z, \qquad \max_{1 \leq j \leq \ell} d_Y(\tau^{jd} z, z) < \lambda \quad \Longrightarrow \quad d \in B. \tag{7}$$

To prove this, minimality gives a finite cover of Z by sets

$$T^{-s_1}U, \ldots, T^{-s_M}U.$$

Choose λ so that every λ-ball in Z is contained in one of these sets. If the points $z, T^d z, \ldots, T^{\ell d} z$ are all within λ of z, they lie in the same $T^{-s_i}U$. Applying T^{s_i} puts them all in U, and (6) gives $d \in B$.

We need a slight extension of (7), because roots of τ need not preserve Z. Set

$$Z_* = \bigcup_{0 \leq t \leq 1} \tau^t Z.$$

By uniform continuity of

$$(t, y) \longmapsto \tau^{-t} y \quad \text{on } [0,1] \times Y,$$

there is $\nu > 0$ such that

$$\boxed{y \in Z_*, \qquad \max_{1 \leq j \leq \ell} d_Y(\tau^{jd} y, y) < \nu \quad \Longrightarrow \quad d \in B.} \tag{8}$$

Indeed, write $y = \tau^t z$, with $z \in Z$ and $0 \leq t \leq 1$, and apply τ^{-t}. The transformations τ^{-t} and τ^{jd} commute, so (7) applies.

This handles the possible disconnectedness of the original nilsystem without requiring a root to preserve its minimal component.

4.2. Choose a sufficiently slow root

Apply the robust slow-motion lemma on Y, obtaining F, h, β, W, L_0, C.

Choose an integer L divisible by L_0 and sufficiently large that the translation

$$R(y) = \tau^{1/L} y$$

satisfies

$$\boxed{\sup_{y \in Y} |F(Ry) - F(y)| < \beta.} \tag{9}$$

This is possible because $\tau^{1/L} \to 1_G$, the action is continuous, and Y is compact.

Fix $x \in Z$, and define

$$c(n) = \kappa(R^n x).$$

Use this coloring in (2) to obtain the 2-syndetic set $A = A_c$.

Choose an integer k sufficiently large that

$$k \geq 4W + 2, \qquad \ell L \leq k/2, \qquad C\ell L/k < \nu. \tag{10}$$

We claim that A satisfies (1) with $K = 2k$.

4.3. A long progression in A forces its difference into B

Suppose

$$a, a + d, \ldots, a + (2k-1)d \in A.$$

By the encoding calculation, there are $u \geq 1$, a color e, and $\epsilon_j \in \{0, 1\}$ such that

$$\kappa(R^{u + jd + \epsilon_j} x) = e \qquad (0 \leq j < k).$$

Since the F-value of a point of color e belongs to $\mathcal{I}_e$, equation (9) implies

$$\operatorname{dist}\big(F(R^{u+jd}x), \mathcal{I}_e\big) < \beta \qquad (0 \leq j < k). \tag{11}$$

Choose a representative $\widetilde{x} \in G$ of x, and consider

$$p(t) = \tau^{(u+dt)/L} \widetilde{x}.$$

This has the form required by the robust lemma:

$$p(t) = \big(\tau^{d/L}\big)^t \big(\tau^{u/L}\widetilde{x}\big).$$

Equation (11) is its hypothesis (4).

Consequently, for some $0 \leq b < W$, the point

$$y = p(b)\Gamma$$

satisfies

$$\begin{aligned} d_Y(\tau^{jd} y, y) &= d_Y\big(p(b + jL)\Gamma, p(b)\Gamma\big) \\ &\leq Cj L/k \\ &< \nu \qquad (1 \leq j \leq \ell), \end{aligned} \tag{12}$$

where we used

$$\boxed{p(b+jL)\Gamma = \tau^{jd}p(b)\Gamma.} \tag{13}$$

This identity is crucial: **the index spacing L cancels the root $1/L$, leaving exactly the original difference d.**

Finally,

$$y = \tau^{(u+bd)/L}x \in Z_*.$$

Indeed, write $(u+bd)/L = m+t$, where $m \in \mathbb{Z}$ and $0 \le t < 1$; then

$$y = \tau^t(\tau^m x), \qquad \tau^m x \in Z.$$

Equations (8) and (12) now give $d \in B$.

This proves the test-set theorem, provided we establish the robust lemma.

5. Proof of the robust slow-motion lemma

The argument is an adaptation of the polynomial rounding and first-exit proof in your manuscript. The important changes are to use triangular Mal'cev coordinates and to retain a margin inside the shell bands. The corresponding ingredients in the draft are the polynomial-rounding lemma and the first-exit argument. [Pasted text.txt, L444-L494] [Pasted text.txt, L729-L767]

5.1. Triangular polynomial coordinates

Choose Mal'cev coordinates

$$\psi : G \longrightarrow \mathbb{R}^m$$

with $\psi(\Gamma) = \mathbb{Z}^m$. The multiplication law has the triangular form

$$\boxed{\psi_i(ab) = x_i + y_i + R_i(x_1,\dots,x_{i-1},y_1,\dots,y_{i-1}),} \tag{14}$$

where $x = \psi(a)$, $y = \psi(b)$, and the R_i are ordinary polynomials. Moreover, the coordinates of $g^t v$ are ordinary polynomials in t, with a degree bound depending only on the coordinate model, not on g, v. These facts follow from polynomial conversion between exponential and Mal'cev coordinates and the finite Baker–Campbell–Hausdorff formula; precise references are Green–Tao, Appendix A, Lemmas A.2–A.3, printed pages 526–528. Annals of Mathematics

To organize the degree bounds, choose positive integer weights $v_1,\dots,v_m$ recursively. Require each v_i to dominate the common degree bound for the coordinates of $g^t v$, and require every monomial in R_i to have weighted degree at most v_i, giving either occurrence of coordinate $j < i$ weight v_j.

This is possible because R_i uses only earlier coordinates.

Let $\mathcal{H}$ consist of curves $p : \mathbb{R} \to G$ satisfying

$$\deg \psi_i(p(t)) \le v_i.$$

It contains all curves $g^t v$, and (14) shows that it is closed under multiplication and affine substitutions. Put $D = \max_i v_i$.

These auxiliary weights need not be the intrinsic nilpotent weights: we only need a finite triangular degree bookkeeping scheme.

5.2. The shell function and robust sampling

Triangular reduction gives a canonical representative of each coset with coordinates in $[-1/2, 1/2)$. Define

$$F(g\Gamma) = \min_{\gamma \in \Gamma} \sum_{i=1}^{m} \psi_i(g\gamma)^2. \tag{15}$$

Exactly as in the matrix proof, the minimum is attained, F is continuous, and $0 \le F \le m/4$. On any compact set of representatives g, only finitely many γ can minimize: the coordinates of $g\gamma$ are bounded, hence so is $\gamma = g^{-1}(g\gamma)$, and Γ is discrete.

Also, if $Q(t)$ is a nonconstant curve with ordinary polynomial coordinates, then $F(Q(t)\Gamma)$ cannot be constant on an interval. For each of the finitely many possible minimizing γ,

$$\sum_i \psi_i(Q(t)\gamma)^2$$

is a nonconstant polynomial: multiplication by γ is invertible, and the leading nonzero even-degree coefficient of the sum of squares is positive. Finitely many nonconstant polynomials cannot take one prescribed value at every point of an interval.

Fix $\rho = 1/4$. Consider curves Q on $[0,1]$ with coordinate degrees at most D, coordinates bounded by 1, and

$$\|\psi(Q(1)) - \psi(Q(0))\|_\infty = \rho.$$

This is a compact family in coefficient space. The preceding nonconstancy therefore gives a uniform $\eta > 0$ such that

$$\operatorname{osc}_{[0,1]} F(Q(t)\Gamma) \ge \eta. \tag{16}$$

Set

$$h = \eta/20, \qquad \beta = h/8.$$

Define $\mathcal{I}_e, \kappa$ as in the lemma.

Uniform continuity gives $T_0 \ge 1$ such that every normalized curve, sampled at

$$Q(j/T), \qquad 0 \le j \le \lfloor T \rfloor, \quad T \ge T_0,$$

has an F-value within $h/8$ of a midpoint of a band of each parity.

In particular, for either prescribed e, some sampled value satisfies

$$\boxed{\operatorname{dist}\big(F(Q(j/T)\Gamma), \mathcal{I}_e\big) \ge 3h/8 > \beta.} \tag{17}$$

This is the robust form of the all-color sampling statement.

5.3. Polynomial rounding still works

Choose $M \ge D + 2$ sufficiently large that every degree- $\le D$ polynomial bounded by $1/2$ at

$$0, 1, \dots, M-1$$

moves by less than ρ on $[0, T_0]$. This is the interpolation estimate already proved in your paper.

Partition the canonical coordinate box into finitely many boxes of side length at most ξ, with

$$2^{D+1}\xi < 1.$$

Choose a finite van der Waerden number W for these box labels and progression length M.

Suppose p satisfies (4). Applying the auxiliary coloring to its first W samples gives b, q, with

$$0 \le b < W, \qquad 1 \le q \le W,$$

such that the canonical representatives of

$$p(b), p(b+q), \dots, p(b+(M-1)q)$$

lie in one small box.

We obtain curves $Q, C \in \mathcal{H}$ with

$$\boxed{p(b+qt)C(t) = Q(t), \qquad C(n) \in \Gamma \quad (n \in \mathbb{Z}),} \tag{18}$$

and with the first M values of Q canonical.

For completeness, here is why the rounding argument extends to (14). Write $A(t) = p(b+qt)$, and let $r_i(n)$ be the reduced coordinates on the initial samples. Construct the coordinates C_i, Q_i inductively. Once $C_1, \dots, C_{i-1}$ are constructed, put

$$f_i(t) = A_i(t) + R_i(A_{<i}(t), C_{<i}(t)).$$

Then $\deg f_i \le v_i$, and

$$r_i(n) - f_i(n) \in \mathbb{Z}.$$

Thus $\Delta^{v_i+1} r_i(n)$ is an integer. Since all $r_i(n)$ lie in an interval of length ξ, its absolute value is less than 1, so it vanishes. Interpolate r_i by a polynomial Q_i of degree at most v_i, and set

$$C_i = Q_i - f_i.$$

It is integer-valued at $0, \dots, v_i$, hence at all integers by Newton interpolation. The triangular multiplication formula then gives (18).

Put

$$J = \left\lfloor \frac{k-1-b}{q} \right\rfloor.$$

For every integer $0 \le n \le J$, equation (18) gives

$$Q(n)\Gamma = p(b+qn)\Gamma.$$

Consequently,

$$\operatorname{dist}\big(F(Q(n)\Gamma), \mathcal{I}_e\big) < \beta. \tag{19}$$

5.4. The first-exit contradiction tolerates the errors

I claim

$$\|\psi(Q(t)) - \psi(Q(0))\|_\infty < \rho \qquad (0 \le t \le J). \tag{20}$$

Otherwise, take the first time $T \le J$ when the displacement equals ρ. The initial interpolation estimate gives $T > T_0$.

Up to time T, all coordinates of Q are bounded by

$$1/2 + \rho = 3/4.$$

Therefore $u \mapsto Q(Tu)$ belongs to the compact normalized family.

The robust sampling statement (17), with mesh parameter T, produces an integer $n \le T$ satisfying

$$\operatorname{dist}\big(F(Q(n)\Gamma), \mathcal{I}_e\big) > \beta.$$

This contradicts (19). Hence (20) holds.

This is where robustness is used: the samples need not be exactly one color, because the opposite-color sample lies a definite distance inside its band.

5.5. Bounded polynomial lifts move slowly

Every coordinate of Q is now bounded by $3/4$ on $[0, J]$. Rescaled polynomial interpolation gives

$$\sup_{0 \le t \le J} \left| \frac{d}{dt} \psi_i(Q(t)) \right| \le \frac{C_D}{J}.$$

The map from the bounded coordinate box to Y is Lipschitz. Thus, for a constant C_0 depending only on the fixed model,

$$d_Y(Q(t)\Gamma, Q(0)\Gamma) \le C_0 t / J. \tag{21}$$

If L is divisible by L_0, then $q \mid L$. Hence (18), evaluated at the integer jL/q, yields

$$p(b + jL)\Gamma = Q(jL/q)\Gamma.$$

Moreover,

$$qJ \ge k - 1 - b - q \ge k - 1 - 2W \ge k/2.$$

Therefore $jL \le k/2$ guarantees $jL/q \le J$, and (21) gives

$$\begin{aligned} d_Y(p(b + jL)\Gamma, p(b)\Gamma) &\le C_0 \frac{jL}{qJ} \\ &\le 2C_0 \frac{jL}{k}. \end{aligned}$$

This proves the robust lemma with $C = 2C_0$. $\square$

6. What resolves the difficulty from the previous attempt

The earlier encoding selected one integer from each adjacent pair, which allowed gaps of three. The new set (2) fills exactly the troublesome gaps. The price is only the ambiguity

$$c(n) = 1 \quad \text{or} \quad c(n+1) = 1.$$

The root construction absorbs that ambiguity:

$$|F(R^{n+1}x) - F(R^n x)| < \beta.$$

The robust shell argument then converts approximate monochromaticity into slow motion at index spacing L, and

$$R^{Ld} = \tau^d$$

recovers the desired recurrence time.

The parameter order is

$$\boxed{\begin{gathered} B \longrightarrow (Z,T,U) \subseteq (Y,\tau) \longrightarrow \nu, \\ Y \longrightarrow F,\beta,W,L_0,C, \\ \longrightarrow L \in L_0\mathbb{N} \text{ sufficiently large} \longrightarrow k \text{ sufficiently large.} \end{gathered}}$$

There is no circular choice: the shell constants are fixed before choosing the root order.

Thus the desired implication is established by this argument. It strengthens the earlier 3-syndetic conclusion, and it does not require the general implication 2-syndetic-large $\Rightarrow$ 2-vdW-large. The latter is not being asserted.

The argument is qualitative: it does not yet give the coefficient-uniform shrinking-target theorem under 2-syndetic-largeness. It also uses the nilsystem realization theorem and basic Mal'cev-coordinate facts, so it is not simply a restatement of the purely special-generalized-polynomial proof.

Finally, the return times can be made arbitrarily large: apply the result to $B \cap q\mathbb{N}$, which is again a nil-Bohr $_0$ set by taking a product with a finite rotation, and then let q grow. Products of nilsystems likewise give simultaneous pointwise returns. The main conclusion is therefore

$$\boxed{\text{2-syndetic-large} \quad \Longrightarrow \quad \text{multiple pointwise recurrence for every finite-step nilsystem.}}$$

---

214. User

2026-09-28 16:10:10 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-28T08:10:10.637000Z

Please try to determine if either the alweiss example of a set of nilbohr recurrence but not double recurrence, or the modification of ben green of this example, is 2-syndetic large. Give it a serious try and work for at least 2 hours.

---

215. Assistant

2026-09-28 16:10:48 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-28T08:10:48.366140Z

I'll check both constructions as written, then look for a 2-syndetic avoiding set or an argument forcing recurrence in every such set.

---

216. Assistant

2026-09-28 16:14:01 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-28T08:14:01.561324Z

For Alweiss's original small-step set, the cosine obstruction appears robust under a repaired 2-syndetic encoding; Green's near-half-turn variant seems to require a different argument.

217. Assistant

2026-09-28 16:23:17 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-28T08:23:17.586994Z

The Alweiss case is now complete: its original set is not 2-syndetic-large. I'm still testing different obstructions for Green's variant.

---

218. Assistant

2026-09-28 16:50:31 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-28T08:50:31.526051Z

The Alweiss obstruction also admits a cleaner formulation: adding $n/2$ to the slowly varying cosine observable directly produces the required 2-syndetic witness.

---

219. Assistant

2026-09-28 17:53:27 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-28T09:53:27.397738Z

I can settle Alweiss's original example: it is not 2-syndetic-large, without modifying the example. I have not determined whether Green's half-turn modification is 2-syndetic-large.

The new ingredient is that the earlier cosine obstruction survives small perturbations of the individual terms of a progression. Combining this with the repaired binary encoding produces a set meeting **every pair of consecutive integers** but containing no sufficiently long progression with an allowed difference.

Download the complete standalone LaTeX proof
Read the compiled PDF — 6 pages

The note proves a general small-step version as well as the result for Alweiss's displayed parameters. Here is the proof for those displayed parameters, followed by the issue with Green's example.

1. Precise result for Alweiss's unchanged example

Let S_{A} be Alweiss's original set, with

$$a = \delta_1 = \frac{1}{100}, \qquad b = \delta_2 = 10^{-10}.$$

These are the parameter values explicitly given before Definition 2.1 of his paper. arXiv

Let (α_j) be its original rapidly decreasing frequency sequence, and put

$$\Lambda = \sum_{j \geq 1} |\alpha_j| < \infty.$$

The passage from the infinite-dimensional construction to this sequence and the integer set is in Section 2.3. arXiv

Theorem

Define

$$R = 1 + \lfloor 64\pi\Lambda \rfloor, \qquad K = 150R + 1.$$

There is an explicitly defined set $A \subseteq \mathbb{N}$ such that

$$\boxed{A \cap \{n, n+1\} \neq \varnothing \qquad (n \geq 1),} \tag{1}$$

but

$$\boxed{\{x, x+d, \ldots, x+(K-1)d\} \not\subseteq A \qquad (x \geq 1,\ d \in S_{\mathrm{A}}).} \tag{2}$$

Consequently, S_{A} is not 2-syndetic-large.

Here " 2-syndetic-large" has the Host–Kra–Maass meaning: every set satisfying (1) must contain arbitrarily long progressions with common difference in the prescribed set. arXiv

Neither the α_j's nor S_{A} are changed. The integer R is used only to construct the witnessing set A.

2. The explicit 2-syndetic set

For real t, define

$$H(t) = \sum_{j \geq 1} \bigl(\cos(2\pi\alpha_j t) - 1\bigr). \tag{3}$$

This converges absolutely, since

$$|\cos(2\pi\alpha_j t) - 1| \leq 2\pi|t|\,|\alpha_j|.$$

Moreover,

$$\boxed{|H(t+v) - H(t)| \leq 2\pi\Lambda|v| \qquad (t, v \in \mathbb{R}).} \tag{4}$$

Define a binary coloring by

$$c(n) = \lfloor 2H(n/R) \rfloor \pmod 2, \tag{5}$$

taking values in $\{0, 1\}$. Now put

$$\boxed{A = \{2n-1 : c(n) = 0\} \cup \{2n : c(n) = 1 \text{ or } c(n+1) = 1\}.} \tag{6}$$

The extra condition involving $c(n+1)$ is the repair that eliminates gaps of three.

Why A meets every consecutive pair

The pair $\{2n-1, 2n\}$ meets A: its odd member belongs when $c(n) = 0$, and its even member belongs when $c(n) = 1$.

The pair $\{2n, 2n+1\}$ also meets A: its odd member belongs when $c(n+1) = 0$, and its even member belongs when $c(n+1) = 1$.

Thus A is 2-syndetic for **every** binary coloring c.

What a progression in A tells us about c

Suppose

$$x, x+d, \ldots, x+150Rd \in A.$$

Select the 76 terms

$$x + 2Rtd, \qquad 0 \le t \le 75. \tag{7}$$

They have one fixed parity.

If $x = 2u - 1$, membership in A forces

$$c(u + Rtd) = 0 \qquad (0 \le t \le 75).$$

If $x = 2u$, membership forces

$$c(u + Rtd) = 1 \quad \text{or} \quad c(u + Rtd + 1) = 1.$$

Therefore, in either case, there are a color $e \in \{0, 1\}$ and choices

$$\epsilon_t \in \{0, 1\}$$

such that

$$\boxed{c(u + Rtd + \epsilon_t) = e \qquad (0 \le t \le 75).} \tag{8}$$

We do not necessarily obtain an exactly monochromatic progression for c. We obtain one with **independent one-step perturbations**. The rest of the proof shows that this is enough.

3. Allowed differences force a uniform real excursion

Fix $d \in S_{\mathrm{A}}$. Its defining conditions supply an index i and representatives σ_j of $d\alpha_j \bmod 1$ such that

$$\sigma_i = -a + \eta, \qquad |\eta| \le 2b,$$

and

$$|\sigma_j| \le 2b \quad (j \ne i), \qquad \sum_{j \ne i} |\sigma_j| \le a + 2b. \tag{9}$$

These are relaxed versions of Alweiss's strict inequalities; his additional lower bound on the total tail norm is not needed here. arXiv

In particular,

$$\sum_j |\sigma_j| \le 2a + 4b, \qquad \sum_{j \ne i} \sigma_j^2 \le 2b(a + 2b). \tag{10}$$

For an **arbitrary real starting parameter** v, write

$$y_t = H(v + td), \qquad 0 \le t \le 75.$$

We need two estimates:

$$\boxed{|y_{t+1} - y_t| < 0.126, \qquad \operatorname{osc}_{0 \le t \le 75} y_t > \frac{7}{10}.} \tag{11}$$

Both are uniform in v, d, and the distinguished index.

Small consecutive changes

By periodicity of cosine and (10),

$$\begin{aligned} |y_{t+1} - y_t| &\le 2\pi \sum_j |\sigma_j| \\ &\le 2\pi(2a + 4b) \\ &< 0.126. \end{aligned} \tag{12}$$

A large excursion, regardless of the initial phase

Consider

$$D_0 = y_0 - 2y_{25} + y_{50}, \qquad D_1 = y_{25} - 2y_{50} + y_{75}. \tag{13}$$

The reason for using 25 is

$$25a = \frac{1}{4}.$$

Thus the distinguished coordinate makes approximately a quarter-turn between these samples.

Put

$$\phi = 2\pi v \alpha_i.$$

Ignoring the tiny error η, the distinguished contributions at times $0, 25, 50, 75$ would be

$$\cos\phi - 1, \quad \sin\phi - 1, \quad -\cos\phi - 1, \quad -\sin\phi - 1.$$

Their two second differences are

$$-2\sin\phi, \qquad 2\cos\phi.$$

Consequently,

$$\boxed{\max\{2|\sin\phi|, 2|\cos\phi|\} \ge \sqrt{2}.} \tag{14}$$

This is why we use **two overlapping second differences**: either one alone can vanish at a special initial phase, but they cannot both be small.

Let us bound the errors.

At time $25j$, the distinguished cosine differs from its ideal value by at most

$$2\pi(25j)|\eta| \le 100\pi j b.$$

Therefore the error in either second difference is at most

$$800\pi b. \tag{15}$$

For a tail coordinate $j \neq i$, its second difference at spacing 25 has absolute value at most

$$|1 - e(25\sigma_j)|^2 \le 4\pi^2 25^2 \sigma_j^2, \qquad e(t) = e^{2\pi i t}.$$

Summing and using (10), the total tail contribution to either second difference is at most

$$8\pi^2 25^2 b(a + 2b). \tag{16}$$

Combining (14)–(16),

$$\begin{aligned}\max\{|D_0|,|D_1|\} &\ge \sqrt{2} - 800\pi b - 8\pi^2 25^2 b(a+2b)\\ &> \sqrt{2} - 10^{-6}\\ &> \frac{7}{5}.\end{aligned} \tag{17}$$

For any real sequence of oscillation L, every expression $y_s - 2y_t + y_u$ has absolute value at most $2L$. Thus (17) implies

$$\operatorname{osc}_{0\le t\le 75}\, y_t > \frac{7}{10},$$

which proves the second estimate in (11).

4. Why the obstruction tolerates small errors

The real-valued coloring

$$z \longmapsto \lfloor 2z \rfloor \pmod 2$$

uses alternating bands of width $1/2$.

Set

$$\beta = \frac{1}{32}.$$

For a fixed color e, thicken all its bands by β:

$$J_e = \bigcup_{m\in\mathbb{Z}} \left[m + \frac{e}{2} - \beta,\, m + \frac{e+1}{2} + \beta \right]. \tag{18}$$

The components of J_e have length

$$\frac{1}{2} + 2\beta = \frac{9}{16},$$

and the gaps between them have length

$$\frac{1}{2} - 2\beta = \frac{7}{16}.$$

Suppose that numbers z_t, all of color e, satisfy

$$|z_t - y_t| < \beta \qquad (0 \le t \le 75). \tag{19}$$

Then every y_t belongs to J_e.

But consecutive y_t's differ by less than $0.126 < 7/16$. They therefore cannot jump between different components of J_e. All of them must lie in one interval of length $9/16$.

This contradicts

$$\operatorname{osc} y_t > \frac{7}{10} > \frac{9}{16}.$$

We have proved the robust statement:

$$\boxed{\begin{array}{c} d \in S_{\mathrm{A}}, \quad v \in \mathbb{R} \\ \Longrightarrow \text{the 76 values } H(v+td) \text{ cannot all be moved by less than } 1/32 \\ \text{into bands of one color.}\end{array}} \tag{20}$$

This is stronger than the earlier exact-monochromaticity obstruction.

5. Complete the contradiction

Return to a hypothetical progression

$$x, x+d, \dots, x+150Rd \in A, \qquad d \in S_{\mathrm{A}}.$$

Equation (8), together with the definition of c, says that all the numbers

$$z_t = H\left(\frac{u+Rtd+\epsilon_t}{R}\right), \qquad 0 \le t \le 75,$$

have one common band color.

Define

$$y_t = H(u/R + td).$$

By (4) and the choice $R > 64\pi\Lambda$,

$$|z_t - y_t| \le \frac{2\pi\Lambda}{R} < \frac{1}{32}. \tag{21}$$

This contradicts (20), applied with $v = u/R$.

Therefore A has no K-term progression with difference in S_{A}. This completes the proof.

The crucial identity is

$$\boxed{\frac{u+Rtd+\epsilon_t}{R} = \frac{u}{R} + td + \frac{\epsilon_t}{R}.} \tag{22}$$

The selected index spacing R cancels the slowdown in the substantial part of the argument, while the perturbation remains small.

We have recovered the original difference d. At no point have we assumed that Rd or $2Rd$ belongs to S_{A}.

6. The argument works beyond the displayed numerical example

The accompanying note proves the same conclusion whenever

$$0 < a \le \frac{1}{100}, \qquad 0 < b \le a^3,$$

and the allowed differences satisfy (9).

Instead of 25, take

$$q = \left\lfloor \frac{1}{4a} + \frac{1}{2} \right\rfloor.$$

Then qa is close to $1/4$, so the same four-sample calculation works with times

$$0, q, 2q, 3q.$$

The extra quarter-turn error is controlled by $|qa - 1/4| \le a/2$. The resulting oscillation still exceeds $9/16$, uniformly over the stated parameter range.

The forbidden progression length becomes

$$\boxed{K = 6qR + 1.}$$

Thus the negative conclusion is not a coincidence peculiar to the exact rational value $a = 1/100$.

7. Why Green's modification is different

Green's Definition 2.1 uses a distinguished coordinate near $1/2$, with the other coordinates within

$$\delta = \frac{1}{200}$$

of zero and total tail norm less than 4:

$$\|d\alpha_i - 1/2\| \le \delta, \qquad \|d\alpha_j\| \le \delta \ (j \neq i), \qquad \sum_{j \neq i} \|d\alpha_j\| < 4.$$

His complex-valued second-difference argument gives a finite-color obstruction to three-term progressions. arXiv

There is a concrete reason why the real-cosine proof above cannot simply be reused.

In the ambient group

$$E = \ell^1(\mathbb{N}; \mathbb{T}),$$

the direction

$$s = (1/2, 0, 0, \ldots)$$

satisfies Green's coordinate conditions. Take

$$x = (1/4, 0, 0, \ldots).$$

Then

$$\cos\big(2\pi(1/4 + t/2)\big) = 0 \qquad (t \in \mathbb{Z}).$$

Consequently, for

$$H_E(z) = \sum_j (\cos(2\pi z_j) - 1),$$

we have

$$\boxed{H_E(x + ts) = -1 \qquad (t \in \mathbb{Z}).} \tag{23}$$

Thus the phase-uniform scalar oscillation statement fails **even when all the tail errors are zero**. This is not merely a loss in the numerical constants.

Replacing cosine by another continuous scalar statistic does not automatically fix it

There is a general half-turn obstruction. Let $F : E \to \mathbb{R}$ be continuous, and restrict it to the circle in the first coordinate. Define

$$g(t) = F(te_1) - F((t + 1/2)e_1).$$

Then

$$g(t+1/2) = -g(t).$$

By continuity, g vanishes somewhere. At such a point $x = te_1$,

$$F(x) = F(x+s),$$

and since $2s = 0$,

$$F(x+ns)$$

is constant for all integers n.

So **no continuous real-valued statistic can have positive oscillation on every half-turn orbit in this ambient model**.

This does not establish that Green's integer set is 2-syndetic-large. The ambient direction and starting point above are not asserted to occur exactly on its prescribed integer orbit. What it establishes is that Green's defining coordinate inequalities alone cannot supply the robust scalar oscillation lemma that settles Alweiss's original construction.

Nor does Green's many-color obstruction itself provide a 2-syndetic avoiding set: converting arbitrary finite-color obstructions to this bounded-gap form requires additional work.

8. Consequences and current status

Alweiss proves that his original set meets every finite-step nil-Bohr neighborhood. The construction above shows that this same set is not 2-syndetic-large. arXiv Hence we obtain the separation

$$\boxed{\text{nil-Bohr recurrence} \not\Longrightarrow \text{2-syndetic-largeness.}}$$

The result also strengthens the earlier negative conclusion about 2-vdW-largeness. Indeed, Host–Kra–Maass prove

$$\text{2-vdW-large} \Longrightarrow \text{2-syndetic-large.}$$

Here we have produced the stronger obstruction directly: an avoiding set with gaps at most two. arXiv

The outcome is therefore: Alweiss's original small-step example is not 2-syndetic-large; Green's half-turn modification remains undetermined in this investigation. The six-page note contains the complete proof of the first assertion, including the general parameter range, and explicitly separates that result from the limitation of the method for Green's construction.

---

220. User

2026-09-29 20:52:06 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-29T12:52:06.651000Z

can you extract from our paper a proof that 2-large sets are bohr recurrent? It should be simpler since I am only asking for bohr recurrent now and not nil bohr recurrent. I want the proof to use the shell colorings of our current proof, not the random colorings of the proof of alweiss.

221. Assistant

2026-09-29 20:52:38 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-29T12:52:38.476462Z

I'll specialize the shell argument to linear phases and check which lifting and denominator-clearing steps can be removed.

222. Assistant

2026-09-29 20:55:51 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-29T12:55:51.488570Z

The linear case simplifies substantially: pigeonhole replaces polynomial rounding, while an explicit quadratic shell calculation yields a fixed-coloring bound $\max_i \|d\alpha_i\| \ll_m k^{-1}$.

223. Assistant

2026-09-29 20:58:59 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-29T12:58:59.423746Z

Yes. The Bohr case admits a much shorter shell proof: pigeonhole replaces the auxiliary van der Waerden argument, and a quadratic calculation replaces the compactness argument. There are no noncommutative matrices, polynomial lattice corrections, or nested brackets to decode.

The shell function is exactly the one in your manuscript, specialized to

$$G = U_2(\mathbb{R})^m \cong \mathbb{R}^m, \qquad \Gamma = \mathbb{Z}^m : \qquad F(x) = \sum_{i=1}^{m} \|x_i\|_{\mathbb{R}/\mathbb{Z}}^2.$$

Thus the proof retains the manuscript's geometric idea: a monochromatic lifted progression cannot make a substantial excursion through alternating thin shells. [Pasted text.txt, L501-L524] [Pasted text.txt, L552-L586]

Download the standalone LaTeX proof
Read the two-page PDF

Here is the complete argument.

Theorem

Every 2-large set $S \subseteq \mathbb{N}$ is Bohr recurrent. Explicitly, for every

$$\alpha_1, \dots, \alpha_m \in \mathbb{R} \quad \text{and} \quad \varepsilon > 0,$$

there is $d \in S$ such that

$$\boxed{\max_{1 \le i \le m} \|d\alpha_i\|_{\mathbb{R}/\mathbb{Z}} < \varepsilon.}$$

In fact, the proof gives constants C_m, k_m, depending only on m, and one binary coloring for each $(\alpha_1, \dots, \alpha_m)$, such that every monochromatic k-term progression with $k \ge k_m$ has difference d satisfying

$$\max_i \|d\alpha_i\|_{\mathbb{R}/\mathbb{Z}} < \frac{C_m}{k}.$$

1. The shell function cannot stay almost constant on a substantial line segment

Write $|x|_\infty = \max_i |x_i|$, and define

$$F(x) = \sum_{i=1}^{m} \|x_i\|_{\mathbb{R}/\mathbb{Z}}^2 = \min_{z \in \mathbb{Z}^m} |x - z|_2^2.$$

This function is continuous and $\mathbb{Z}^m$-periodic. Moreover,

$$\boxed{|F(x) - F(y)| \le m|x - y|_\infty.} \tag{1}$$

Indeed, distance to $\mathbb{Z}$ is 1-Lipschitz and at most $1/2$, so its square is also 1-Lipschitz.

The degree-one version of your normalized-excursion lemma has a particularly simple proof.

Lemma

For every $x, v \in \mathbb{R}^m$ with $|v|_\infty = 1/4$,

$$\boxed{\operatorname{osc}_{0 \le t \le 1} F(x + tv) \ge \eta_m := \frac{1}{64(m+1)^2}.} \tag{2}$$

Proof

Each nonconstant coordinate $x_i + tv_i$ crosses at most one half-integer as t runs through $[0,1]$, since $|v_i| \le 1/4$. Therefore at most m breakpoints divide $[0,1]$ into at most $m+1$ intervals on which the nearest-integer vector is fixed.

Choose one such interval $[u, w]$ with

$$w - u \ge \frac{1}{m+1}.$$

On this interval, for some fixed $z \in \mathbb{Z}^m$,

$$f(t) := F(x + tv) = |x + tv - z|_2^2.$$

Ties at the endpoints do not affect the identity; a constant coordinate at a half-integer may use either nearest integer throughout.

Thus f is quadratic with leading coefficient

$$|v|_2^2 \ge |v|_\infty^2 = \frac{1}{16}.$$

The midpoint identity gives

$$f(u) - 2f\left(\frac{u+w}{2}\right) + f(w) = \frac{1}{2}|v|_2^2 (w-u)^2.$$

On the other hand, the left side is at most $2 \operatorname{osc}_{[u,w]} f$. Hence

$$\operatorname{osc}_{[0,1]} f \ge \frac{1}{4}|v|_2^2 (w-u)^2 \ge \frac{1}{64(m+1)^2}.$$

□

This replaces both “nonconstancy on polynomial curves” and compactness by an explicit quadratic calculation.

2. Choose one deterministic shell coloring

Fix the following parameters, all depending only on m:

$$h = \frac{\eta_m}{4}, \qquad M \in \mathbb{N} \text{ with } M > \frac{4m}{h}, \qquad Q = M^m, \qquad L = \operatorname{lcm}(1, \dots, Q). \tag{3}$$

For $\alpha = (\alpha_1, \dots, \alpha_m)$, color $n \in \mathbb{N}$ by

$$\boxed{c(n) = \left\lfloor \frac{F(n\alpha/L)}{h} \right\rfloor \pmod 2.} \tag{4}$$

Thus successive shells of squared distance have alternating colors.

The factor $1/L$ serves the same purpose as in the manuscript’s coloring $c(n) = \kappa(g(n/L))$: it will let us remove an auxiliary integer q and recover information about the original difference d. [Pasted text.txt, L863-L887]

Suppose now that

$$a, a + d, \dots, a + (k-1)d$$

is monochromatic, where

$$k \geq 2Q + 2.$$

We will prove

$$\boxed{\max_i \|d\alpha_i\|_{\mathbb{R}/\mathbb{Z}} < \frac{L}{2k}.} \tag{5}$$

3. Pigeonhole gives a slowly moving subprogression

Put

$$\theta = \frac{d\alpha}{L}.$$

Partition $[0,1)^m$ into $M^m = Q$ half-open boxes of side length $1/M$. Among

$$0, \theta, 2\theta, \dots, Q\theta \pmod{\mathbb{Z}^m},$$

two points lie in one box. Subtracting them gives

$$\boxed{1 \leq q \leq Q, \qquad z \in \mathbb{Z}^m, \qquad v = q\theta - z, \qquad |v|_\infty < \frac{1}{M}.} \tag{6}$$

This is the entire replacement for the auxiliary van der Waerden and polynomial-rounding steps.

Let

$$x = \frac{a\alpha}{L}, \qquad J = \left\lfloor \frac{k-1}{q} \right\rfloor.$$

By periodicity,

$$F(x+jv) = F\left(\frac{(a+jqd)\alpha}{L}\right) \qquad (0 \le j \le J). \tag{7}$$

These points correspond to a subprogression of the original monochromatic progression, so they have the same shell color.

Write

$$y_j = F(x+jv), \qquad b_j = \lfloor y_j/h \rfloor .$$

Equation (1) gives

$$|y_{j+1} - y_j| \le m|v|_\infty < \frac{m}{M} < \frac{h}{4}. \tag{8}$$

In particular,

$$|b_{j+1} - b_j| \le 1.$$

But monochromaticity says that all b_j have the same parity. Therefore

$$b_0 = b_1 = \cdots = b_J.$$

The samples do not merely lie in shells of the same color: they all lie in one and the same shell.

Indeed, changing to another shell of the same color would require skipping an intervening shell, which the small-step estimate forbids.

4. Shell confinement forces the step to be small on the scale $1/k$

Let b be the common value of the b_j. All integer samples satisfy

$$F(x+jv) \in [bh, (b+1)h).$$

For any real $t \in [0, J]$, choose $j = \lfloor t \rfloor$. Again by (1),

$$|F(x+tv) - F(x+jv)| \le m|v|_\infty |t-j| < \frac{h}{4}.$$

Consequently, the entire real segment satisfies

$$\boxed{\operatorname{osc}_{0\le t\le J} F(x+tv) \le \frac{3h}{2} < \eta_m.} \tag{9}$$

Suppose that

$$J|v|_\infty \ge \frac{1}{4}.$$

Then $v \neq 0$, and

$$T = \frac{1}{4|v|_\infty} \le J.$$

The rescaled segment

$$u \longmapsto x + uTv, \qquad 0 \le u \le 1,$$

has endpoint displacement

$$|Tv|_\infty = \frac{1}{4}.$$

By the lemma, its F-oscillation is at least η_m, contradicting (9).

Therefore

$$\boxed{|v|_\infty < \frac{1}{4J}.} \tag{10}$$

This also holds when $v = 0$.

Now use the fixed denominator L. Since $q \le Q$, we have $q \mid L$, and (6) gives

$$d\alpha = \frac{L}{q} z + \frac{L}{q} v, \qquad \frac{L}{q} z \in \mathbb{Z}^m.$$

Hence

$$\max_i \|d\alpha_i\|_{\mathbb{R}/\mathbb{Z}} \le \frac{L}{q} |v|_\infty < \frac{L}{4qJ}. \tag{11}$$

Finally,

$$qJ \ge k - 1 - q \ge k - 1 - Q \ge \frac{k}{2},$$

because $k \ge 2Q + 2$. Substituting into (11) proves (5):

$$\max_i \|d\alpha_i\|_{\mathbb{R}/\mathbb{Z}} < \frac{L}{2k}.$$

5. Apply 2-largeness

Given $\varepsilon > 0$, choose an integer

$$k \ge 2Q + 2, \qquad k > \frac{L}{2\varepsilon}.$$

Since S is 2-large, the binary coloring (4) has a monochromatic k-term progression with difference $d \in S$. Equation (5) yields

$$\max_i \|d\alpha_i\|_{\mathbb{R}/\mathbb{Z}} < \frac{L}{2k} < \varepsilon.$$

Thus S is Bohr recurrent. □

What survives from the general proof

The argument retains **the squared-distance shells, the obstruction to a normalized real excursion, and the fixed dilation that removes the auxiliary step**. These are the corresponding mechanisms in the manuscript's shell and first-exit argument. [Pasted text.txt, L729-L799]

Everything specific to higher degree disappears. In particular, **we never assume that** $qd \in S$. Pigeonhole controls

$$q\frac{d\alpha}{L} \pmod{\mathbb{Z}^m},$$

and multiplication by the integer L/q recovers $d\alpha$. That is why the final return time is the original allowed difference d, rather than an uncontrolled multiple of it.

The resulting proof uses only finite pigeonhole, elementary properties of distance to the integers, and a quadratic midpoint identity. It requires neither random colorings nor equidistribution, and the same coloring works for all sufficiently large progression lengths.

---

224. User

2026-09-30 12:22:27 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-30T04:22:27.193000Z

is there a particular reason the ell^2 norm is used instead of the ell^1 norm?

---

225. Assistant

2026-09-30 12:24:16 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-30T04:24:16.592384Z

Yes. The crucial reason is that ℓ^1 has flat shells: a nonconstant line segment can remain at exactly the same ℓ^1-distance from the lattice. That would break the oscillation argument. The squared ℓ^2-distance avoids this problem.

More precisely, the proof uses the **square** of the Euclidean distance:

$$F_2(x) = \min_{z \in \mathbb{Z}^m} |x - z|_2^2 = \sum_{i=1}^{m} \|x_i\|_{\mathbb{R}/\mathbb{Z}}^2.$$

Its counterpart in the full manuscript is the minimum of the sum of squares of the corrected matrix coordinates. [Pasted text.txt, L501-L507]

1. What goes wrong with ℓ^1

The natural replacement would be

$$F_1(x) = \min_{z \in \mathbb{Z}^m} |x - z|_1 = \sum_{i=1}^{m} \|x_i\|_{\mathbb{R}/\mathbb{Z}}.$$

Already in dimension two, consider

$$Q(t) = \left(\frac{1}{8} + \frac{t}{4},\ \frac{3}{8} - \frac{t}{4} \right), \qquad 0 \le t \le 1.$$

Both coordinates remain in $[1/8, 3/8]$, so their nearest integer is always zero. Consequently,

$$\boxed{F_1(Q(t)) = \left(\frac{1}{8} + \frac{t}{4} \right) + \left(\frac{3}{8} - \frac{t}{4} \right) = \frac{1}{2}}$$

throughout the interval.

Nevertheless,

$$|Q(1) - Q(0)|_\infty = \frac{1}{4} = \rho.$$

Thus this is **exactly the kind of normalized excursion used in our proof**, but its F_1-oscillation is zero. One coordinate increases while the other decreases, and their contributions cancel perfectly.

No matter how small we choose the shell width h, the coloring

$$x \longmapsto \lfloor F_1(x)/h \rfloor \pmod 2$$

is constant on this whole segment. The lift can therefore move distance ρ without crossing a single shell boundary. We lose the positive oscillation that drives the first-exit contradiction. The manuscript obtains that positive lower bound by applying its nonconstancy lemma to the compact normalized family. [Pasted text.txt, L552-L575]

2. Why squaring prevents this cancellation

For the same example,

$$\begin{aligned} F_2(Q(t)) &= \left(\frac{1}{8} + \frac{t}{4}\right)^2 + \left(\frac{3}{8} - \frac{t}{4}\right)^2 \\ &= \frac{1}{8} + \frac{1}{8}\left(t - \frac{1}{2}\right)^2. \end{aligned}$$

Its oscillation is $1/32$, rather than zero.

More generally, on an interval where the nearest-integer vector z is fixed,

$$F_2(x + tv) = |x + tv - z|_2^2,$$

and hence

$$\frac{d^2}{dt^2} F_2(x + tv) = 2|v|_2^2 > 0 \qquad (v \neq 0).$$

Every nonzero direction contributes positive curvature, regardless of whether its coordinates increase or decrease. This is the quadratic fact used in the simplified Bohr proof.

For the full generalized-polynomial proof, the corresponding fact is algebraic. If $p_1(t), \dots, p_\nu(t)$ are real polynomials, at least one nonconstant, let

$$b = \max_i \deg p_i \geq 1, \qquad a_i = [t^b] p_i(t).$$

Then

$$[t^{2b}] \sum_i p_i(t)^2 = \sum_i a_i^2 > 0.$$

So the sum of squares cannot be constant. Applying this to the coordinates of $Q(t)\gamma$, for the finitely many possible lattice minimizers γ, proves the manuscript's nonconstancy lemma. [Pasted text.txt, L527-L545]

By contrast, a sum of absolute values can be constant on an interval:

$$|t| + |1 - t| = 1 \qquad (0 \leq t \leq 1).$$

3. Is ℓ^2 uniquely necessary?

No—the essential property is nonconstancy along nonconstant polynomial curves, not specifically the exponent 2.

Using the Euclidean distance itself, $\sqrt{F_2}$, would preserve nonconstancy and allow the qualitative shell argument. Squaring is preferable because it produces ordinary polynomials on each lattice branch and, in the Bohr case, an immediate quadratic estimate.

Likewise, replacing the sum of squares by a sum of fixed positive even powers,

$$\sum_i p_i(t)^{2q}, \qquad q \geq 1,$$

would retain the positive-leading-coefficient argument. Squares are simply the lowest-degree and cleanest choice.

Although ℓ^1 and ℓ^2 are equivalent norms in finite dimensions, **norm equivalence does not preserve oscillation along curves**: comparable size functions can have very different level sets. That is why changing norms is harmless for many coordinate bounds in the paper, but not for the function defining the shells.

So the reason is structural rather than stylistic: **the ℓ^1 shells have flat directions that permit invisible motion; the squared ℓ^2 shells detect every nonconstant polynomial excursion.**

---

226. User

2026-09-30 14:51:20 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-30T06:51:20.044000Z

Please give me the tex code for the corrected and completed equation before section 3.10. I just want the concrete example worked out/written out in the 3 by 3 matrix case.

> **Attachments recorded in source:**
>
> - Pasted text(20260930-065049).txt

---

227. User

2026-09-30 14:51:31 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-30T06:51:31.602000Z

4 by 4 matrix case*

---

228. Assistant

2026-09-30 14:51:48 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-30T06:51:48.815566Z

Understood—I'll use the 4×4 case, checking the signs and entries against the example immediately before Section 3.10.

---

229. Assistant

2026-09-30 14:52:54 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-30T06:52:54.108663Z

The $(2,3)$-entry should be ** $-\beta n^v$**, and the innermost expression in both nested corrections should be ** γn^w**. Replace the unfinished example immediately before Section 3.10 with the following. [Pasted text(20260930-065049).txt, L54-L60]

```
For example, for
\[
 P(n)=\alpha n^u\nint{\beta n^v\nint{\gamma n^w}},
\]
the last-column calculation is
\begingroup
\setlength{\arraycolsep}{4pt}
\[
\begin{aligned}
&
\begin{pmatrix}
 1 & -\alpha n^u & 0 & 0\\
 0 & 1 & -\beta n^v & 0\\
 0 & 0 & 1 & \gamma n^w\\
 0 & 0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
 1 & 0 & 0 &
 -\nint{\alpha n^u\nint{\beta n^v\nint{\gamma n^w}}}\\
 0 & 1 & 0 &
 -\nint{\beta n^v\nint{\gamma n^w}}\\
 0 & 0 & 1 & -\nint{\gamma n^w}\\
 0 & 0 & 0 & 1
\end{pmatrix}
\\[6pt]
&\quad =
\begin{pmatrix}
 1 & -\alpha n^u & 0 &
 \alpha n^u\nint{\beta n^v\nint{\gamma n^w}}
 -\nint{\alpha n^u\nint{\beta n^v\nint{\gamma n^w}}}\\
 0 & 1 & -\beta n^v &
 \beta n^v\nint{\gamma n^w}
 -\nint{\beta n^v\nint{\gamma n^w}}\\
 0 & 0 & 1 &
 \gamma n^w-\nint{\gamma n^w}\\
 0 & 0 & 0 & 1
\end{pmatrix}.
\end{aligned}
\]
\endgroup
Thus the last column consists of the nearest-integer errors of the
successive nested expressions, and its top entry is
\(P(n)-\nint{P(n)}\), whose absolute value is
\(\tnorm{P(n)}\).
```

230. User

2026-09-30 14:54:56 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-30T06:54:56.672000Z

two things. Firstly, I should have clarified that the completed example has the right had side consisting of a matrix in the fundamental domain, i.e., all entries are between [-1/2,1/2). Secondly, don't write x-<x>. Maybe you should first propose for me a notation for x-<x>, then I can add it to the paper (if it is not already there), then you can incorporate it into the completed calcuation

231. Assistant

2026-09-30 14:55:30 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-30T06:55:30.898453Z

I'll first propose a signed centered-fractional-part notation and verify its endpoint convention; then we can use it to write the fully reduced fundamental-domain matrix.

---

232. Assistant

2026-09-30 14:56:45 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-30T06:56:45.243772Z

You are right: **my previous product reduced only the last column, whereas you want the canonical representative, with all six entries above the diagonal in $[-1/2, 1/2)$.** The integer correction matrix therefore needs additional nonzero entries in positions $(1,2)$, $(2,3)$, and $(1,3)$.

Suggested notation

I suggest

$$\boxed{\{x\}_{\mathrm{c}} := x - \langle x \rangle}$$

and call it the **centered fractional part** of x. The subscript distinguishes it from the ordinary fractional part $\{x\}$, which your introduction already uses. [Pasted text.txt, L123-L142]

It also distinguishes the **signed remainder** from the distance to the nearest integer:

$$\{x\}_{\mathrm{c}} \in [-1/2, 1/2), \qquad \bigl|\{x\}_{\mathrm{c}}\bigr| = \|x\|_{\mathbb{R}/\mathbb{Z}}.$$

For example,

$$\{3.7\}_{\mathrm{c}} = -0.3, \qquad \|3.7\|_{\mathbb{R}/\mathbb{Z}} = 0.3.$$

I do not see a separate notation for this signed remainder in the draft. You could add the following macro to the preamble:

```
\newcommand{\cpart}[1]{\left\{#1\right\}_{\mathrm c}}
```

Then replace the sentence introducing nearest-integer rounding by:

```
For \(x\in\R\), let \(\nint{x}:=\lfloor x+1/2\rfloor\) denote
its nearest integer, with ties resolved toward the larger integer,
and let
\[
 \cpart{x}:=x-\nint{x}\in[-1/2,1/2)
\]
denote its centered fractional part.
Thus \(|\cpart{x}|=\tnorm{x}\).
```

The displayed subtraction occurs only in the definition; the matrix calculation can then use `\cpart{...}` throughout.

One endpoint correction in the draft

Your current definition

$$\langle x\rangle = \lfloor x + 1/2\rfloor$$

already gives the interval you requested, $[-1/2, 1/2)$. However, the “Canonical representatives” subsection still says

$$I = (-1/2, 1/2].$$

Those conventions are inconsistent at half-integers. You should change that line to

```
Put \(I=[-1/2,1/2)\).
```

This keeps the canonical-reduction rule $\lambda_{ij} = -\langle f_{ij}\rangle$ consistent with the rounding convention in your introduction. [Pasted text.txt, L150-L151] [Pasted text.txt, L473-L484]

For the completed calculation, negative arguments should remain **inside** the centered-fractional-part notation—for example,

$$\{-\alpha n^u\}_{\mathrm{c}},$$

rather than $-\{\alpha n^u\}_{\mathrm{c}}$. These agree away from half-integers but not at them: both $\{1/2\}_{\mathrm{c}}$ and $\{-1/2\}_{\mathrm{c}}$ equal $-1/2$. Keeping the signs inside makes every entry land in the correct half-open fundamental interval without exceptions.

---

233. User

2026-09-30 15:02:13 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-30T07:02:13.135000Z

Great! Now give me the completed matrix calculation.

---

234. Assistant

2026-09-30 15:02:43 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-30T07:02:43.717650Z

I’ll include all six lattice corrections, keeping signs inside the rounding operators so the formula remains valid at half-integers.

---

235. Assistant

2026-09-30 15:03:48 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-30T07:03:48.523601Z

Using `\cpart{...}` for the centered fractional part, the following applies your canonical-reduction rule to **all six entries above the diagonal**, not just the last column. [Pasted text.txt, L479-L484]

```
For example, let
\[
```

```
 P(n)=\alpha n^u\nint{\beta n^v\nint{\gamma n^w}}.
\]
The full reduction to the fundamental domain is
\begingroup
\setlength{\arraycolsep}{3pt}
\renewcommand{\arraystretch}{1.25}
\[
\begin{aligned}
&
\begin{pmatrix}
 1 & -\alpha n^u & 0 & 0\\
 0 & 1 & -\beta n^v & 0\\
 0 & 0 & 1 & \gamma n^w\\
 0 & 0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
 1
 & -\nint{-\alpha n^u}
 & -\nint{\alpha n^u\nint{-\beta n^v}}
 & -\nint{\alpha n^u\nint{\beta n^v\nint{\gamma n^w}}}\\
 0
 & 1
 & -\nint{-\beta n^v}
 & -\nint{\beta n^v\nint{\gamma n^w}}\\
 0 & 0 & 1 & -\nint{\gamma n^w}\\
 0 & 0 & 0 & 1
\end{pmatrix}
\\[8pt]
&\quad =
\begin{pmatrix}
 1
 & \cpart{-\alpha n^u}
 & \cpart{\alpha n^u\nint{-\beta n^v}}
 & \cpart{\alpha n^u\nint{\beta n^v\nint{\gamma n^w}}}\\
 0
 & 1
 & \cpart{-\beta n^v}
 & \cpart{\beta n^v\nint{\gamma n^w}}\\
 0 & 0 & 1 & \cpart{\gamma n^w}\\
 0 & 0 & 0 & 1
\end{pmatrix}.
\end{aligned}
\]
\endgroup
The second matrix on the left belongs to \(\UT_4(\Z)\), and every
entry above the diagonal of the matrix on the right belongs to
\([-1/2,1/2)\). Thus the latter is the canonical representative
\(\rep(B(\alpha n^u,\beta n^v,\gamma n^w))\).
Its top-right entry is \(\cpart{P(n)}\), whose absolute value is
\(\tnorm{P(n)}\).
```

The negative argument in the $(1,3)$-entry is intentional: the $(2,3)$-entry of the integer correction is $-\langle -\beta n^v \rangle$, so multiplication produces the quantity

$$\alpha n^u \langle -\beta n^v \rangle$$

before centering. Keeping that sign inside the brackets makes the identity valid at half-integers as well.

236. User

2026-09-30 15:26:22 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-30T07:26:22.354000Z

on page 9, is it really necessary to specifiy that ||_infty refers to the off diagonal entries?

237. Assistant

2026-09-30 15:26:33 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-30T07:26:33.004518Z

No, I do **not** think it is necessary once you have already defined

```
\[
\coord{A-B}=\max_{i<j}|A_{ij}-B_{ij}|
\]
```

earlier. In the current draft that definition is explicit, and you even add that this is the coordinate distance rather than an invariant metric. [Pasted text.txt, L382-L389]

So when page 9 later says

> "Here `\(\coord{Q(t)}\)` means the maximum absolute off-diagonal entry,"

that is redundant. The notation has already been fixed. [Pasted text.txt, L597-L605]

I would simply delete that sentence and write:

```
Let $\Kpol$ be the family of polynomial matrix curves
$Q:[0,1]\to G$ whose coordinate degrees are at most $D$ and satisfy
\[
 \max_{0\leq t\leq1}\coord{Q(t)}\leq1,
 \qquad
 \coord{Q(1)-Q(0)}=\rho.
\]
For this auxiliary family we need only the common degree bound $D$.
```

There is also a conceptual reason the convention is natural: every matrix in G has diagonal entries equal to 1, while $Q(t)-Q(0)$ has diagonal entries equal to 0. The only coordinates that carry information are therefore the strictly upper-triangular ones. Re-explaining this each time makes the notation feel less settled than it really is.

I would keep the clarification only at the **first definition** of `\coord{·}`.

238. User

2026-09-30 15:27:54 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-30T07:27:54.055000Z

saying 'it is compact in coefficient space', is that just a roundabout way of saying it is compact with respect to the topology induced by ||_infty?

239. Assistant

2026-09-30 15:28:04 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-30T07:28:04.411704Z

Essentially yes, **once the degrees are uniformly bounded**.

What “compact in coefficient space” means is that we identify each polynomial matrix curve $Q \in \mathcal{K}$ with the finite vector of all coefficients of all its off-diagonal polynomial entries. Because every coordinate has degree at most D, this puts $\mathcal{K}$ inside some finite-dimensional space $\mathbb{R}^N$.

On this finite-dimensional space, the coefficient norm and the function norm

```
\[
\|Q\|_{\sup,\infty}
:=
\sup_{0\le t\le1}\coord{Q(t)}
\]
```

induce the same topology. So you could equally well say that $\mathcal{K}$ is compact for the uniform topology induced by

```
\[
Q\mapsto \sup_{t\in[0,1]}\coord{Q(t)}.
\]
```

The bounded-degree assumption is what makes this equivalence true. For arbitrary continuous functions or polynomials of unbounded degree, uniform boundedness would not imply compactness.

So I think a clearer sentence would be:

> Since the coordinate degrees are bounded by D, $\mathcal{K}$ lies in a finite-dimensional space of polynomial curves. The bound `\(\sup_{0\le t\le1}\coord{Q(t)}\le1\)` bounds all coefficients by interpolation, while the defining conditions are closed; hence $\mathcal{K}$ is compact.

That is more explicit than “compact in coefficient space,” and avoids making the reader wonder which topology is intended.

---

240. User

2026-09-30 15:32:18 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-30T07:32:18.969000Z

Is \osc_{[0,1]}F(Q(t)) := \sup_{x,y \in [0,1]}|F(Q(x))-F(Q(y))|. the correct definition of oscilation?

---

241. Assistant

2026-09-30 15:32:24 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-30T07:32:24.681039Z

Yes. That is the standard oscillation of the real-valued function $F \circ Q$ on $[0,1]$:

$$\boxed{\operatorname{osc}_{[0,1]} F(Q(t)) := \sup_{x,y\in[0,1]} |F(Q(x)) - F(Q(y))|\,.}$$

Since $F \circ Q$ is continuous on the compact interval $[0,1]$, you can equivalently write

$$\operatorname{osc}_{[0,1]} F(Q(t)) = \max_{t\in[0,1]} F(Q(t)) - \min_{t\in[0,1]} F(Q(t)).$$

I slightly prefer the second form in this paper, because later you explicitly talk about the range of $F \circ Q$ being an interval $[a, b]$, in which case the oscillation is simply $b - a$.

---

242. User

2026-09-30 15:36:34 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-30T07:36:34.232000Z

why is every member of mathcal{K} nonconstant?

---

243. Assistant

2026-09-30 15:36:39 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-30T07:36:39.296407Z

Because $\mathcal{K}$ was defined with the endpoint displacement condition

```
\[
\coord{Q(1)-Q(0)}=\rho,
\]
```

and $\rho = 1/4 > 0$.

So for every $Q \in \mathcal{K}$,

$$Q(1) \neq Q(0),$$

hence Q cannot be constant.

That is all that is being used there. In fact, this is exactly why the definition uses equality to a fixed positive escape radius rather than merely

```
\[
\coord{Q(1)-Q(0)}\le \rho.
\]
```

With only the latter condition, constant curves would belong to $\mathcal{K}$, and Lemma 3.5 would not give a positive uniform oscillation lower bound.

---

244. User

2026-09-30 16:09:48 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-30T08:09:48.617000Z

Please give me a revised section 3.1, a revised outline of the proof now that I have updated the paper.

> **Attachments recorded in source:**
>
> - Pasted text(20260930-080929).txt

245. Assistant

2026-09-30 16:12:03 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-30T08:12:03.226174Z

Replace Section 3.1 with the following. It connects the outline to the preceding Bohr-case argument and follows the updated proof through polynomial rounding, shell confinement, and last-column decoding. [Pasted text.txt, L560-L574] [Pasted text.txt, L832-L940]

```
\subsection{Outline of the argument}\label{ssec:Outline}

The proof extends the shell argument of the preceding section from
line segments on a torus to polynomial curves in products of
upper-unitriangular groups. For each special generalized polynomial
\(P_i\), we construct a band matrix \(g_{P_i}(t)\) whose first
superdiagonal contains its signed monomial factors. Let \(g(t)\)
be the tuple of these matrices, let \(G\) be the corresponding
product of matrix groups, and let \(\Gamma\) be its subgroup of
integer matrices. The nested nearest integers will be recovered
from the last columns after suitable integer corrections.

The primary coloring is defined using the continuous,
right-\(\Gamma\)-invariant function
\[
 F(A)=\min_{\gamma\in\Gamma}
      \sum_{i<j}(A\gamma)_{ij}^{\,2},
\]
where the sum includes all blocks. A nonconstant ordinary
polynomial matrix curve cannot remain in one level set of \(F\).
Indeed, on a compact parameter interval, \(F\) is the minimum of
finitely many nonconstant sums of squares of polynomials.
Compactness therefore gives a uniform positive oscillation for
polynomial curves of bounded degree making a fixed-size excursion
inside a bounded coordinate box. Coloring successive thin shells
cyclically with \(r\) colors produces a coloring \(\kappa\) under
which every sufficiently fine sample of such an excursion meets
all \(r\) colors. We then color the integers by
\(c(n)=\kappa(g(n/L))\), for a fixed integer \(L\) chosen below.

Suppose that a \(k\)-term progression
\(\{a,a+d,\ldots,a+(k-1)d\}\) omits a color, and put
\(p(t)=g((a+dt)/L)\).
An auxiliary finite coloring records the small box containing the
canonical representative of each \(p(j)\). Finite van der Waerden
provides a short subprogression, with bounded starting index \(b\)
and step \(q\), whose representatives lie in one box.
The polynomial rounding lemma then gives
\[
 p(b+qt)C(t)=Q(t),\qquad C(n)\in\Gamma\quad(n\in\Z),
\]
where \(C\) and \(Q\) have ordinary polynomial entries with the
required weighted degree bounds. Only the initial samples of
\(Q\) are known to be canonical, but all its relevant integer
samples omit the same primary color as the original progression.

A first-exit argument now keeps \(Q\) bounded on a real interval
\([0,J]\), where \(J\) is comparable to \(k/q\).
Otherwise, rescaling up to the first time \(Q\) moves the fixed
```

```
escape distance would produce a normalized excursion. The initial
interpolation estimate ensures that this exit occurs late enough
for the all-color sampling lemma to apply, contradicting the
omitted color. Polynomial interpolation on \([0,J]\) consequently
bounds the coefficient of degree \(w\) in each weight-\(w\)
coordinate by \(O(J^{-w})\).

The weighted coefficient-extraction map \(\tau\) removes the
starting point of the progression. The choice of \(L\) ensures
that applying the dilation \(\delta_{L/q}\) both removes the
auxiliary step and clears the factorial denominators in
\(\tau(C)\). We obtain
\[
 g(d)\gamma=E,\qquad \gamma\in\Gamma,\qquad
 |E_{ij}|\ll_{r,m,D}k^{-w_{ij}}.
\]
Reading each last column from bottom to top identifies the nested
nearest integers exactly. Its top entry is therefore
\(\cpart{P_i(d)}\), giving the required bound
\(\tnorm{P_i(d)}\ll_{r,m,D}k^{-d_i}\).
All parameters defining the shell coloring and \(L\) are chosen
independently of the real coefficients and the progression length.

Subsection~\ref{ssec:TheProof} assembles these steps to prove
Theorem~\ref{thm:omitted}.
Section~\ref{sec:Nilmanifolds} gives the nilsystem interpretation
and the answer to the question of Host, Kra, and Maass.
Section~\ref{sec:extremal} proves the extremal corollary, and
Section~\ref{sec:questions} discusses the density and
polychromatic extensions.
```

246. User

2026-09-30 16:24:27 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-30T08:24:27.943000Z

I generally like to have my tex code organized by having 1 complete english sentence as one line of tex code. Is it possible for you to update your internal settings to implement this style preference fro all future tex code that you give me in all of our chats?

247. Assistant

2026-09-30 16:25:04 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-30T08:25:04.966031Z

Yes—I've saved this preference for future TeX/LaTeX code across our chats: **one complete English sentence per source line**, without splitting a prose sentence across several lines or placing multiple sentences on the same line.

Paragraphs will remain separated by blank lines, while displayed equations, matrices, and mathematical environments will retain appropriate multiline formatting.

248. Assistant

2026-09-30 16:25:08 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-30T08:25:08.666286Z

I'll use one complete English prose sentence per TeX source line. I'm also checking whether that preference can be retained across future chats.

---

249. User

2026-09-30 17:12:36 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-30T09:12:36.146000Z

I generally like it. I am wondering if it worth somehow mentioning in this outline that we have to rescale G to make it a member of the compact family mathcal{K}. First give me a suggestion for how we could incorporate this idea into the outline, then give me your opinion on whether or not you think it is even a good idea to try and incorporate this into the outline.

---

250. Assistant

2026-09-30 17:13:54 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-30T09:13:54.111221Z

Suggested wording

I would replace the paragraph beginning "A first-exit argument now keeps Q bounded..." with the following. **The object being rescaled is the parameter of the polynomial lift Q, not the group G**: the proof uses $\widetilde{Q}(u) = Q(Tu)$, while G remains fixed. [Pasted text.txt, L866-L882]

```
A first-exit argument now keeps \(Q\) bounded on a real interval \([0,J]\), where \(J\) is comparable
to \(k/q\).
Suppose instead that \(Q\) first moves the fixed coordinate distance \(\rho\) from its starting point
at some time \(T\leq J\).
Rescaling the parameter gives the curve \(\widetilde Q(u)=Q(Tu)\), \(0\leq u\leq1\), which belongs to
the compact family \(\Kpol\) of normalized excursions: its coordinates remain in the prescribed
bounded box, its degree bounds are unchanged, and its endpoint displacement is exactly \(\rho\).
The initial interpolation estimate ensures that \(T\) is large enough for the all-color sampling lemma
to apply.
Since \(\widetilde Q(j/T)=Q(j)\), the resulting fine sample of the normalized curve consists of
integer samples of the original lift, and these must receive all \(r\) colors, contradicting the
omitted color.
Thus \(Q\) remains bounded, and polynomial interpolation on \([0,J]\) bounds the coefficient of degree
\(w\) in each weight-\(w\) coordinate by \(O(J^{-w})\).
```

Is it worth including?

Yes—I think making this normalization explicit improves the outline. The earlier version already mentioned "rescaling up to the first time," but did not explain precisely what was rescaled or how the normalized samples relate to the original ones. The two identities

$$\widetilde{Q}(u) = Q(Tu), \qquad \widetilde{Q}(j/T) = Q(j)$$

make that connection transparent.

This is an important conceptual step rather than just bookkeeping: **the compact family supplies a uniform sampling statement, but the particular lift Q does not initially belong to that family**. Stopping at the first escape time supplies the boundedness and prescribed endpoint displacement; reparametrizing then puts the curve on $[0,1]$, exactly as required in the definition of $\mathcal{K}$. [Pasted text.txt, L668-L690] [Pasted text.txt, L870-L876]

I would keep this explanation within the first-exit paragraph, rather than introduce rescaling as a separate stage of the outline. It is also worth calling it **“rescaling the parameter”** or **“reparametrizing the curve”**, because the proof later uses the genuinely different operation $\delta_{L/q}$ to change the matrix coordinates and clear denominators. [Pasted text.txt, L890-L907]

My preference is therefore to include the explicit reparametrization and the sample identity, while leaving the numerical bounds and the detailed verification of membership in $\mathcal{K}$ to the proof.

---

251. User

2026-09-30 17:20:05 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-30T09:20:05.414000Z

also, why did you call it wieghted coefficient extraction instead of just coefficient extraction?

---

252. Assistant

2026-09-30 17:20:55 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-30T09:20:55.166912Z

I used **“weighted” because the power of t being extracted depends on the weight assigned to the matrix entry**. It does **not** mean that we multiply the coefficients by weights.

The definition is

$$\tau(A)_{ij} = [t^{w_{ij}}]A_{ij}(t),$$

so each entry undergoes ordinary coefficient extraction, but potentially at a different degree. This is exactly how τ is defined in your draft. [Pasted text.txt, L458-L474]

For example, with adjacent weights $w_1 = 2$ and $w_2 = 1$, we have

$$w_{12} = 2, \qquad w_{23} = 1, \qquad w_{13} = 3,$$

and therefore

$$\tau(A)_{12} = [t^2]A_{12}(t), \qquad \tau(A)_{23} = [t]A_{23}(t), \qquad \tau(A)_{13} = [t^3]A_{13}(t).$$

Thus we are neither extracting one common coefficient from every entry nor necessarily taking each entry’s actual leading nonzero coefficient. For instance, if

$$A_{12}(t) = 5t + 1,$$

then $\tau(A)_{12} = 0$, because its assigned weight is 2, despite its actual leading coefficient being 5.

The assigned weights also explain why the extraction respects matrix multiplication: $w_{ij} = w_{ik} + w_{kj}$ makes the selected degrees match the cross terms in the product. That compatibility is proved in the draft’s algebra lemma. [Pasted text.txt, L497-L501]

Stylistically, though, “coefficient extraction” is perfectly adequate in the outline. “Weighted coefficient extraction” was a descriptive label for this particular entry-dependent choice, not a different kind of coefficient-extraction operation. I would favor explaining the choice rather than making the terminology heavier:

```
The coefficient-extraction map \(\tau\), which selects from each entry the coefficient of the power
prescribed by its weight, removes the starting point of the progression.
```

After that, simply calling τ **“the coefficient-extraction map”** is clear and sufficient.

---

253. User

2026-09-30 17:25:33 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-30T09:25:33.222000Z

how about leading coefficient extraction?

---

254. Assistant

2026-09-30 17:25:38 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-30T09:25:38.324775Z

Yes, **“leading coefficient extraction” is better than “weighted coefficient extraction”** for this paper.

The reason is that for every entry $A_{ij}(t)$, the map

$$\tau(A)_{ij} = [t^{w_{ij}}]A_{ij}(t)$$

extracts the coefficient of the **largest degree that is allowed** in that entry, since by definition

$$\deg A_{ij} \le w_{ij}.$$

So $[t^{w_{ij}}]A_{ij}(t)$ is precisely the leading coefficient whenever the entry actually has degree w_{ij}, and it is 0 if its degree is smaller. [Pasted text.txt, L466-L474]

That makes “leading coefficient extraction” very natural, especially because your subsection is already titled **“Control of the leading coefficients”**. [Pasted text.txt, L832-L840]

The only tiny caveat is that, literally speaking, if

$$A_{ij}(t) = 5t + 1$$

and $w_{ij} = 2$, then

$$[t^2]A_{ij}(t) = 0,$$

whereas the ordinary leading coefficient of A_{ij} is 5. So τ is not the usual leading-coefficient map applied independently to each polynomial entry; it extracts the coefficient at the **assigned top degree**.

Still, in context I think that distinction is already clear from the definition. I would use:

> “the leading-coefficient extraction map τ”

or even simply

> “the leading-coefficient map τ”

after it has been defined.

For the outline, I would write:

```
The leading-coefficient map \(\tau\) removes the starting point of the progression.
```

That is cleaner than either “weighted coefficient-extraction map” or the longer explanatory phrasing.

---

255. User

2026-09-30 17:31:59 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-30T09:31:59.500000Z

isn't it weird that J is comparable to k/q, so it is bounded, so T is bounded, so we are rescaling by a bounded quantity in order to enter a compact set, but the original Q does not come from a bounded set. Something in my understanding is not adding up.

---

256. Assistant

2026-09-30 17:32:07 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-30T09:32:07.267106Z

Yes — the point that fixes the apparent paradox is that ** J and T are not uniformly bounded**. In fact, q is uniformly bounded, while k is allowed to grow.

You have

$$J = \left\lfloor \frac{k-1-b}{q} \right\rfloor,$$

with

$$b, q \le W,$$

and W was fixed before k was supplied. [Pasted text.txt, L844-L859] Thus

$$J \asymp k/q \asymp k$$

up to constants depending only on W. So as $k \to \infty$, $J \to \infty$. There is no uniform upper bound on the first-exit time $T \le J$.

That said, there is a second subtlety here that is probably the real source of your discomfort.

The proof does **not** say that Q comes from some globally bounded family and that a bounded rescaling puts it into $\mathcal{K}$. Instead, it argues as follows.

Suppose Q escapes. Let T be the **first** time for which

```
\[
\coord{Q(T)-Q(0)}=\rho.
\]
```

Then, by definition of first exit,

```
\[
\coord{Q(t)-Q(0)}\le \rho
\qquad(0\le t\le T).
\]
```
Since $Q(0)$ is canonical,

```
\[
\coord{Q(0)}\le \frac12,
\]
```
and hence

```
\[
\coord{Q(t)}\le \frac12+\rho=\frac34
\qquad(0\le t\le T).
\]
```
This is exactly the spatial boundedness needed in the definition of $\mathcal{K}$. [Pasted text.txt, L866-L876]

Now define

$$\widetilde{Q}(u) = Q(Tu), \qquad 0 \le u \le 1.$$

Then three things hold:

$$\deg \widetilde{Q}_{ij} = \deg Q_{ij} \le D,$$

```
\[
\max_{0\le u\le1}\coord{\widetilde Q(u)}
=
\max_{0\le t\le T}\coord{Q(t)}
\le\frac34<1,
\]
```
and

```
\[
\coord{\widetilde Q(1)-\widetilde Q(0)}
=
\coord{Q(T)-Q(0)}
=\rho.
\]
```
Therefore

$$\boxed{\widetilde{Q} \in \mathcal{K}.}$$

The size of T is irrelevant to membership in $\mathcal{K}$. In fact, T may be arbitrarily large.

There is a superficially troubling feature: if

$$Q(t) = a_0 + a_1 t + \cdots + a_D t^D,$$

then

$$Q(Tu) = a_0 + (Ta_1)u + \cdots + (T^D a_D)u^D,$$

so a large T appears to blow up the coefficients.

But the **first-exit bound prevents those rescaled coefficients from actually becoming large**. The polynomial $Q(Tu)$ stays uniformly bounded by $3/4$ for every $u \in [0,1]$, and because its degree is at most D, interpolation then forces all of its coefficients to be uniformly bounded in terms of D. This is precisely why the family

```
\[
\{R:[0,1]\to G:\deg R_{ij}\le D,\ \sup_u\coord{R(u)}\le1,\
\coord{R(1)-R(0)}=\rho\}
\]
```

is compact. [Pasted text.txt, L668-L690]

So the logic is:

$$
\begin{array}{c}
Q \text{ may have arbitrary coefficients and live on a very long interval} \\
\Downarrow \\
\text{assume an escape and stop at the } \textit{first} \text{ escape time } T \\
\Downarrow \\
Q([0,T]) \text{ is automatically contained in a fixed bounded box} \\
\Downarrow \\
u \mapsto Q(Tu) \text{ has bounded degree, bounded range, and fixed displacement} \\
\Downarrow \\
\widetilde{Q} \in \mathcal{K}.
\end{array}
$$

The normalization is therefore not coming from the fact that T is bounded. Quite the opposite: ** T can be huge; first-exit normalization converts an arbitrarily long, potentially wild polynomial curve into a unit-time curve whose range is uniformly bounded.**

There is then a very nice role for the fact that T is large rather than bounded. You need

$$T > T_0$$

so that the sampling mesh

$$0, \frac{1}{T}, \frac{2}{T}, \dots, \frac{\lfloor T \rfloor}{T}$$

is sufficiently fine for the compact-family sampling lemma. Under the reparametrization,

$$\widetilde{Q}(j/T) = Q(j),$$

so those fine mesh points are exactly the integer samples of the original lift. [Pasted text.txt, L870-L881]

So I would actually sharpen what we wrote in the outline: **do not say or suggest that the rescaling factor is controlled.** What is controlled is the *range before first exit*. That is the key compactness mechanism.

---

257. User

2026-09-30 17:32:50 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-30T09:32:50.201000Z

but the minimum k does not depend on the coefficients of the P_i, which determine Q.

258. Assistant

2026-09-30 17:32:57 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-30T09:32:57.406919Z

Exactly — and that is the important point: **the argument was designed so that the coefficient dependence disappears before the compactness step.**

The polynomial lift Q certainly depends on the coefficients of the P_i. What is uniform is that, after the auxiliary van der Waerden step and polynomial rounding, its first M integer samples are canonical. Hence every coordinate of

$$Q(0), Q(1), \dots, Q(M-1)$$

lies in $[-1/2, 1/2)$, independently of the coefficients. [Pasted text.txt, L844-L856]

Because every coordinate polynomial has degree at most D, your interpolation lemma says that if $M = M(D, T_0, \rho)$ is chosen sufficiently large, then

```
\[
\coord{Q(t)-Q(0)}<\rho
\qquad(0\le t\le T_0),
\]
```

again **uniformly over all possible coefficients of** Q. [Pasted text.txt, L734-L748]

That is the crucial coefficient-uniform statement.

So if Q ever escapes the ρ-ball around $Q(0)$, its first escape time T automatically satisfies

$$T > T_0.$$

Now, up to that first escape time,

```
\[
\coord{Q(t)}
\le \coord{Q(0)}+\rho
\le \frac12+\frac14
=\frac34,
\]
```

so after reparametrizing

$$\widetilde{Q}(u) = Q(Tu), \qquad 0 \le u \le 1,$$

we have

```
\[
\max_{u\in[0,1]}\coord{\widetilde Q(u)}\le\frac34,
\qquad
\coord{\widetilde Q(1)-\widetilde Q(0)}=\rho,
\]
```

and the degree remains at most D. Hence

$$\widetilde{Q} \in \mathcal{K}.$$

Notice what happened: we never needed a coefficient bound for the original Q. The coefficient bound for the **rescaled** polynomial $\widetilde{Q}$ is a consequence of its bounded range and bounded degree. That is why $\mathcal{K}$ is a coefficient-independent compact family. [Pasted text.txt, L668-L690]

The logic is therefore

$$\text{arbitrary coefficients} \longrightarrow \text{canonical values at } M \text{ fixed integer points} \longrightarrow \text{uniform control on } [0, T_0] \longrightarrow T > T_0$$

if an escape occurs, and then

$$\text{first exit} \longrightarrow \text{bounded range before } T \longrightarrow \widetilde{Q} \in \mathcal{K}.$$

So T_0, M, W, L, and ultimately k_0 depend only on the degree/weight patterns and r, **not on the coefficients**.

There is one more useful way to see it. Suppose a coefficient of Q were astronomically large. Since $Q(0), \dots, Q(M-1)$ all lie in the fixed box and $\deg Q \le D$, those large monomial coefficients must undergo severe cancellation. Finite-dimensional interpolation says that such cancellation cannot cause arbitrary motion immediately after 0: the first M bounded samples already uniformly control the polynomial on the fixed interval $[0, T_0]$.

That is precisely why you chose M only after fixing D, T_0, ρ. [Pasted text.txt, L812-L830]

So your instinct is right that there would be a problem if the compactness argument relied directly on the raw coefficients of Q. It does not. **The coefficient-uniformity comes from bounded canonical samples plus bounded degree.**

---

259. User

2026-10-02 11:09:40 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-02T03:09:40.713000Z

I don't understand the argument about some line segment intersecting some cube in length Omega_d(L). Can you explain this to me in more detail?

> **Attachments recorded in source:**
>
> - Pasted text(20261002-030914).txt

---

260. Assistant

2026-10-02 11:10:01 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-02T03:10:01.577799Z

The geometric claim is simpler than the surrounding torus discussion makes it look: it is just an averaging argument over the L-cubes crossed by one straight segment in $\mathbb{R}^d$.

---

261. Assistant

2026-10-02 11:10:18 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-02T03:10:18.852280Z

Yes. The sentence in the draft is compressed, and I think it is obscuring a very elementary argument. The right way to think about it is in the **universal cover** $\mathbb{R}^d$, partitioned into translates of the fundamental cube, rather than as a statement about "cells of the torus." [Pasted text.txt, L334-L344]

Let

$$u = f(x), \qquad v = f(cy),$$

and, since $c \mid N$, put

$$J = \frac{N}{c}.$$

Then the points

$$f(x), f(x+cy), f(x+2cy), \dots, f(x+Ny)$$

lie on the straight segment

$$\Sigma = \{u + tv : 0 \le t \le J\} \subseteq \mathbb{R}^d.$$

Now partition $\mathbb{R}^d$ into the translates

$$C_z = Lz + [-L/2, L/2)^d, \qquad z \in \mathbb{Z}^d.$$

The claim is that if Σ is sufficiently long compared with L, then **one of these cubes contains a portion of Σ of length comparable to L.**

The cleanest way to see this is to measure length using ℓ^1, which you are already using nearby.

Let

$$M = \mathrm{length}_{\ell^1}(\Sigma) = J\|v\|_1.$$

As the segment moves through $\mathbb{R}^d$, it changes cubes only when it crosses one of the coordinate hyperplanes

$$x_i = L\left(m + \frac{1}{2}\right), \qquad m \in \mathbb{Z}.$$

For the i-th coordinate, the total displacement along the segment is

$$J|v_i|.$$

Therefore it can cross at most

$$\frac{J|v_i|}{L} + 1$$

such hyperplanes perpendicular to the i-th coordinate axis.

Summing over i, the total number of boundary crossings is at most

$$\sum_{i=1}^{d} \left(\frac{J|v_i|}{L} + 1 \right) = \frac{M}{L} + d.$$

Hence the segment intersects at most

$$\frac{M}{L} + d + 1$$

different L-cubes.

Now the intersections of Σ with those cubes partition the segment into subsegments, and their total ℓ^1-length is M. Therefore one cube C_z satisfies

$$\mathrm{length}_{\ell^1}(\Sigma \cap C_z) \ge \frac{M}{M/L + d + 1}.$$

If, say,

$$M \geq (d+1)L,$$

then

$$\frac{M}{M/L + d + 1} \geq \frac{M}{2M/L} = \frac{L}{2}.$$

So in fact you get the very concrete conclusion

$$\boxed{\text{length}_{\ell^1}(\Sigma \cap C_z) \geq \frac{L}{2}}$$

for some cube, provided the whole segment has ℓ^1-length at least $(d+1)L$.

That is your $\Omega_d(L)$ statement; with the ℓ^1 convention the constant can even be taken as $1/2$ once M is sufficiently large relative to dL.

Why is the whole segment sufficiently long?

Earlier you showed that some coordinate of $f(cy)$, say $v_i = cy'$, has a nontrivial residue modulo L. Writing

$$cy' = kL + \eta,$$

your divisibility $c \mid L$ and the assumption

$$\|y'\|_{\mathbb{R}/\mathbb{Z}} > \varepsilon$$

imply in particular

$$|\eta| > \varepsilon.$$

Thus certainly

$$\|v\|_1 \geq |v_i| \geq |\eta| > \varepsilon.$$

Consequently

$$M = J\|v\|_1 > \frac{N}{c}\,\varepsilon.$$

Since $c < (8dK)^d$ is bounded in terms of the previously chosen parameters, choosing

$$N \gg \frac{(8dK)^d\, dL}{\varepsilon}$$

ensures

$$M \geq (d+1)L.$$

So the logical chain is simply

$$\boxed{\begin{aligned}&\text{one coordinate of the step is nontrivial}\\ &\Longrightarrow \text{the lifted segment has length } \gg (N/c)\varepsilon\\ &\Longrightarrow \text{the segment is much longer than } L\\ &\Longrightarrow \text{it crosses only } O(M/L) \text{ cubes}\\ &\Longrightarrow \text{one cube contains } \Omega(L) \text{ length of it.}\end{aligned}}$$

There is one wording change I strongly recommend: **do not say “some cell of the torus.”** The torus $(\mathbb{R}/L\mathbb{Z})^d$ is represented by one fundamental cube. What you are doing here is lifting to $\mathbb{R}^d$ and looking at the translates of that fundamental cube.

I would replace the relevant part of the draft by something like:

```
Since \(c\mid N\), put \(J=N/c\) and consider the line segment
\[
\Sigma=\{f(x)+t f(cy):0\leq t\leq J\}\subseteq\R^d.
\]
Its \(\ell^1\)-length is \(M=J\|f(cy)\|_1\), and the preceding argument gives \(\|f(cy)\|_1>\eps\), so
\(M>(N/c)\eps\).
Partition \(\R^d\) into the translates \(Lz+[-L/2,L/2)^d\), \(z\in\Z^d\).
In the \(i\)-th coordinate, \(\Sigma\) crosses at most \(J|f(cy)_i|/L+1\) boundaries of these cubes.
Hence \(\Sigma\) meets at most \(M/L+d+1\) cubes.
The intersections of \(\Sigma\) with these cubes have total \(\ell^1\)-length \(M\), so one of them
has length at least
\[
\frac{M}{M/L+d+1}.
\]
Taking \(N\) large enough that \(M\ge(d+1)L\), this length is at least \(L/2\).
```

I think this is substantially clearer than the current $O(M/L)$/ $\Omega_d(L)$ paragraph.

---

262. User

2026-10-02 12:24:56 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-02T04:24:56.526000Z

is it clear that the thickness of each band is at least L/(4K)?

---

263. Assistant

2026-10-02 12:25:14 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-02T04:25:14.791264Z

Not in the form currently written. For arbitrary d, the claim that every shell has thickness at least $L/(4K)$ is false.

Your coloring is determined by

$$\left\lfloor \frac{K}{L^2}\sum_{i=1}^{d} x_i^2 \right\rfloor,$$

so the shell boundaries are spheres

$$\sum_{i=1}^{d} x_i^2 = \frac{jL^2}{K}.$$

Thus the j-th radial shell has Euclidean thickness

$$\Delta r_j = \frac{L}{\sqrt{K}}\left(\sqrt{j+1}-\sqrt{j}\right) = \frac{L}{\sqrt{K}\left(\sqrt{j+1}+\sqrt{j}\right)}.$$

Inside the fundamental cube

$$[-L/2, L/2)^d,$$

the radius is at most

$$\frac{\sqrt{d}\,L}{2}.$$

Hence, for every shell that meets the cube,

$$\sqrt{\frac{j+1}{K}} \lesssim \frac{\sqrt{d}}{2},$$

and consequently

$$\sqrt{j+1} + \sqrt{j} \le \sqrt{dK}$$

up to an immaterial endpoint adjustment. Therefore

$$\boxed{\Delta r_j \gtrsim \frac{L}{K\sqrt{d}}.}$$

More concretely, a safe bound is

$$\Delta r_j \ge \frac{L}{K(\sqrt{d} + K^{-1/2})},$$

and for $K \ge 1$,

$$\Delta r_j \ge \frac{L}{K(\sqrt{d}+1)}.$$

So the natural lower bound is

$$\boxed{\Omega_d(L/K),}$$

not $L/(4K)$ uniformly in d.

This matters because immediately afterward you use $L/(8K)$-density to guarantee sampling inside each band. If the shell thickness is really only on the order of

$$\frac{L}{K\sqrt{d}},$$

then $L/(8K)$ is too coarse once d is large.

The clean fix is to change the auxiliary box size earlier from

$$\frac{L}{8dK}$$

to something smaller so that the resulting orbit samples are, say,

$$\frac{L}{100K\sqrt{d}}$$

dense along the relevant line segment. Then this is comfortably below the minimum shell thickness.

Alternatively, because you are already using ℓ^1 estimates, you can sacrifice another factor of d and use a deliberately crude bound like

$$\text{shell thickness} \gg_d \frac{L}{K},$$

with all constants explicitly allowed to depend on d. That would make the exposition much cleaner.

So I would rewrite the sentence as something like:

> Every shell meeting the fundamental cube has Euclidean thickness at least $c_d L/K$ for some $c_d > 0$ depending only on d.

Then choose the earlier mesh/density parameter small enough compared with $c_d L/K$.

That is both correct and sufficient for the qualitative Bohr argument.

---

264. User

2026-10-07 19:39:40 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-07T11:39:40.165000Z

I am editing section 3.1. I want to add a sentence somewhere that explains why we need L. More concretely, why we need to clear the factorial denominators of the leading coefficients of the entries of C. Please explain it to me in chat, then give me some suggestions for what sentence(s) I could add to section 3.1.

> **Attachments recorded in source:**
>
> - Pasted text(20261007-113849).txt

---

265. Assistant

2026-10-07 19:40:10 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-07T11:40:10.532961Z

The reason for L is a little subtler than merely undoing the auxiliary step q.

The polynomial rounding lemma gives

$$p(b+qt)C(t) = Q(t), \qquad C(n) \in \Gamma \quad (n \in \mathbb{Z}).$$

The important point is that ** $C(n)$ is an integer matrix at integer arguments, but the coefficients of the polynomial entries of $C(t)$ need not be integers**. Your finite-difference argument gives only

$$w_{ij}!\,[t^{w_{ij}}]C_{ij}(t) \in \mathbb{Z}.$$

Equivalently,

$$\tau(C)_{ij} = \frac{z_{ij}}{w_{ij}!} \qquad (z_{ij} \in \mathbb{Z}).$$

So although $C(n) \in \Gamma$ for every integer n, in general

$$\tau(C) \notin \Gamma.$$

This is exactly the distinction your draft points out after the polynomial-rounding lemma. [Pasted text.txt, L589-L613]

The simplest scalar example is

$$C(t) = \binom{t}{2} = \frac{t^2 - t}{2}.$$

It is integer-valued on every $t \in \mathbb{Z}$, but

$$[t^2]C(t) = \frac{1}{2}.$$

Thus extracting a leading coefficient from an integer-valued polynomial can introduce a factorial denominator.

Why do we care? After applying τ, you obtain

$$(\delta_q \tau(p))\tau(C) = \tau(Q).$$

Ultimately you want an equation of the form

$$g(d)\gamma = E$$

with

$$\boxed{\gamma \in \Gamma,}$$

because the integer nature of γ is what lets the last-column argument identify the nearest-integer corrections exactly. If we merely had a rational matrix γ, the decoding step would fail: a small residual would no longer determine which **integer** was subtracted. [Pasted text.txt, L925-L942]

This is where L enters. You chose

$$L = D! \operatorname{lcm}(1, \dots, W).$$

The auxiliary step satisfies $q \le W$, so $q \mid L$, and in fact

$$\frac{L}{q}$$

is an integer multiple of $D!$. Since every $w_{ij} \le D$, it follows that

$$w_{ij}! \mid \frac{L}{q}.$$

Therefore applying the dilation $\delta_{L/q}$ gives

$$\gamma := \delta_{L/q}\tau(C),$$

whose entries are

$$\gamma_{ij} = \left(\frac{L}{q}\right)^{w_{ij}} \frac{z_{ij}}{w_{ij}!} \in \mathbb{Z}.$$

Hence

$$\gamma \in \Gamma.$$

At the same time,

$$\delta_{L/q}\delta_q \tau(p) = \delta_L \tau(p),$$

so the **same dilation does two jobs**:

$$\boxed{\text{remove the auxiliary step } q \quad + \quad \text{clear the factorial denominators of } \tau(C).}$$

Then, because you originally defined

$$p(t) = g((a + dt)/L),$$

the final factor L also restores the desired scale:

$$\delta_L \tau(p) = g(d).$$

So there is a nice three-stage story for L:

$$p(t) = g((a + dt)/L) \longrightarrow \delta_q \tau(p) \longrightarrow \delta_L \tau(p) = g(d),$$

while simultaneously

$$\tau(C) \in \text{rational matrices} \longrightarrow \delta_{L/q} \tau(C) \in \Gamma.$$

Your present outline already says that $\delta_{L/q}$ "removes the auxiliary step q and clears the factorial denominators," but it does not explain **why there are factorial denominators in the first place**. [Pasted text.txt, L470-L479] I think adding one sentence about integer-valued polynomials would make this much easier to understand.

Possible sentences for Section 3.1

- **My preferred version, two sentences:**

```
Although \(C(n)\in\Gamma\) for every integer \(n\), the polynomial entries of \(C\) need only be
integer-valued, so their leading coefficients may have factorial denominators.
We choose \(L\) sufficiently divisible so that the dilation \(\delta_{L/q}\) both removes the
auxiliary step \(q\) and clears these denominators, making \(\delta_{L/q}\tau(C)\) an element of \
(\Gamma\).
```

- **A slightly more concrete version:**

```
The fact that \(C(n)\in\Gamma\) for every \(n\in\Z\) does not imply that \(\tau(C)\in\Gamma\), since
an integer-valued polynomial of degree \(w\) may have leading coefficient in \(\frac{1}{w!}\Z\).
We therefore choose \(L\) so that \(L/q\) is divisible by every \(w!\) with \(w\leq D\), which makes \
(\delta_{L/q}\tau(C)\in\Gamma\) while simultaneously removing the auxiliary step \(q\).
```

- **If you want only one sentence:**

```
We choose \(L\) sufficiently divisible because \(C\) is only integer-valued, so the entries of \
(\tau(C)\) may have factorial denominators, and the dilation \(\delta_{L/q}\) must clear these
denominators as well as remove the auxiliary step \(q\).
```

I prefer the **second two-sentence version** for this paper. The phrase "C is only integer-valued" immediately tells the reader where the unexpected denominators come from, and the statement

$$\tau(C) \notin \Gamma \text{ in general}$$

is conceptually important enough that I would not hide it entirely.

266. User

2026-10-07 20:05:01 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-07T12:05:01.875000Z

do I still have a remark somewhere in the paper about the polynomial entries of C not having integer coefficients, or was that deleted?

267. Assistant

2026-10-07 20:05:10 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-07T12:05:10.025583Z

Yes — the point is still in the paper, but **not as a standalone `remark` environment**.

You currently have the sentence

> "The polynomial $\binom{t}{2} = (t^2 - t)/2$ shows why integer-valuedness does not mean that the monomial coefficients are integers."

immediately after the Newton-interpolation discussion. [Pasted text.txt, L589-L597]

You also state the more directly relevant consequence later:

> "The matrix $\tau(C)$ need not belong to Γ, even though $C(n)$ does for every integer n."

right before clearing the factorial denominators. [Pasted text.txt, L925-L940]

So the idea has **not** been deleted. In fact, I think these two appearances are enough; adding another formal remark would probably be redundant.

268. User

2026-10-07 20:14:55 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-07T12:14:55.954000Z

Please audit Section 3.1 for me.

> **Attachments recorded in source:**
>
> - Pasted text(20261007-121446).txt

269. Assistant

2026-10-07 20:15:06 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-07T12:15:06.554582Z

I'll audit only Section 3.1, focusing on substantive mathematical and grammatical errors rather than stylistic preferences.

270. Assistant

2026-10-07 20:15:39 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-07T12:15:39.058258Z

I've already found several substantive issues: the outline currently misdefines the shell coloring, misidentifies $\mathcal{K}$, and has a couple of variable/identity errors in the rounding paragraph.

---

271. Assistant

2026-10-07 20:16:34 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-07T12:16:34.458858Z

Section 3.1 is conceptually in good shape, but I found **seven genuine issues**, four of them mathematically substantive.

1. **Line 410: wrong notation L instead of L.**
   You currently write

$$P = L(\alpha_1 t^{w_1}, \dots, \alpha_s t^{w_s}),$$

but L is also your large divisible integer. This should be

$$P = \mathsf{L}(\alpha_1 t^{w_1}, \dots, \alpha_s t^{w_s}).$$

[Pasted text.txt, L408-L423]

2. **Lines 444–445: S should be F.**
   You write that a polynomial curve cannot remain in one level set of S, and that S is locally a minimum of sums of squares. The function just defined is F, so both occurrences should be F. [Pasted text.txt, L429-L445]
3. **Lines 447–449: $\mathcal{K}$ is not a compact subset of G, and the definition of κ is currently wrong.**
   The actual $\mathcal{K}$ is a compact **family of polynomial curves**

$$Q : [0,1] \to G,$$

not a subset of G. More importantly, you currently write

$$\kappa(A) = \left\lfloor \frac{F(A)}{h} \right\rfloor$$

with codomain $\{0, \dots, r-1\}$. You have omitted the reduction modulo r, and you say h should be "sufficiently large," whereas the proof takes h **small**, specifically $h = \eta/(20r)$. [Pasted text.txt, L447-L452]

I would replace that passage by:

```
Working with a suitable compact family \(\mathcal K\) of normalized polynomial curves gives a uniform
positive lower bound for the oscillation of \(F\) along polynomial curves of bounded degree making a
fixed-size excursion inside a bounded coordinate box.
We choose a sufficiently small \(h>0\) and define
\[
\kappa(A)=\left\lfloor\frac{F(A)}{h}\right\rfloor\pmod r.
\]
Coloring successive thin shells cyclically in this way ensures that every sufficiently fine sample of
a normalized excursion meets all \(r\) colors.
```

4. **Line 464 contains several mathematical errors and needs replacement.**
   You currently write

$$F(p(b+nt)) = F(Q(n)),$$

but the correct sampled argument is $p(b+qn)$, not $p(b+nt)$. Moreover, κ is a coloring of matrices, so $\kappa(F(Q(n)))$ is not well-typed: it should be $\kappa(Q(n))$. Most importantly, you need the omitted-color statement for **all** $0 \leq n \leq J$, not merely for the first M values; otherwise the later first-exit contradiction does not follow. [Pasted text.txt, L454-L470]

I would replace the sentence by:

```
Since \(C(n)\in\Gamma\) for every \(n\in\Z\) and \(\kappa\) is right-\(\Gamma\)-invariant, we have
\[
\kappa(Q(n))=\kappa(p(b+qn))
\]
for every integer \(n\).
Thus all the samples \(Q(n)\) with \(0\leq n\leq J\) omit the same color as the original progression.
```

This is the most important correction in the outline.

5. **Lines 467–468 should explicitly use the *first* escape time, and explain why the normalized curve belongs to $\mathcal{K}$.**
   Saying merely that Q "escapes ... at some time T" is insufficient: membership in $\mathcal{K}$ comes from taking the **first** such T, which guarantees that Q remains in the bounded coordinate box up to time T. [Pasted text.txt, L466-L471]

I suggest:

```
Suppose instead that \(T\leq J\) is the first time at which \(Q\) moves the fixed coordinate distance
\(\rho\) from \(Q(0)\).
Up to time \(T\), the first-exit property keeps \(Q\) in a fixed bounded coordinate box, so the
reparametrized curve \(\widetilde Q(u)=Q(Tu)\), \(0\leq u\leq1\), belongs to the compact family \
(\mathcal K\).
```

That makes the compactness mechanism much clearer.

6. **Line 463 uses $w_{ij} := \sum_{k=i}^{j-1} w_k$ as though there were one exponent list for the whole product.**
   There are m different blocks, each with its own exponent list. The later rigorous section handles this blockwise, but the outline formula is formally ambiguous. [Pasted text.txt, L454-L464]

I would avoid defining w_{ij} there and simply write:

```
where \(C\) and \(Q\) have ordinary polynomial entries with the prescribed weighted degree bounds,
namely the \((i,j)\)-entry in each block has degree at most the sum of the adjacent exponents from \
(i\) through \(j-1\).
```

7. **Lines 478–483 incorrectly suggest that $\{P_i(d)\}_c$ automatically appears as an entry of E.**
   This is not the same situation as line 427. At line 427 you are free to choose a special canonical lattice correction γ_t. Later, however,

$$E = g(d)\gamma_d$$

comes with the particular integer matrix

$$\gamma_d = \delta_{L/q}\tau(C).$$

You do **not** know a priori that its top-right entry is $\{P_i(d)\}_{\mathrm{c}}$. What proves that is the **last-column decoding lemma**, using the fact that the entries of E are sufficiently small. [Pasted text.txt, L473-L484]

This should instead say something like:

```
Applying the leading-coefficient map and then the dilation \(\delta_{L/q}\) gives
\[
g(d)\gamma_d=E,\qquad \gamma_d\in\Gamma,\qquad |E_{ij}|\ll_{r,m,D}k^{-w_{ij}}.
\]
For sufficiently large \(k\), the last-column entries of \(E\) have absolute value less than \(1/2\),
so the last-column decoding lemma identifies the upper-right entry of the block associated to \(P_i\)
with \(\cpart{P_i(d)}\).
Hence
\[
\tnorm{P_i(d)}\ll_{r,m,D}k^{-d_i}.
\]
```

That is both correct and better explains the final step.

There is also one **minor accuracy issue** at line 440: you say you minimize “the sum of squares of its entries,” but F sums only the **off-diagonal** entries. Since the diagonal entries are always 1, this would only differ by a constant, but I would still write “off-diagonal entries.” [Pasted text.txt, L429-L440]

After those corrections, I think Section 3.1 gives a mathematically faithful outline of the proof. The two changes I would regard as essential before moving on are **line 464** and the **last-column-decoding correction at lines 478–483**.

---

272. User

2026-10-07 20:42:42 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-07T12:42:42.385000Z

Selected text:

Selection 1

pact subset of G, and the definition of κ is currently wrong. The actual $\mathcal{K}$ is a compact family of polynomial curves

$$Q : [0,1] \to G,$$

My request:

i think this is trivial. I am obviously restricting to [0,1] to get the identification.

273. Assistant

2026-10-07 20:42:50 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-07T12:42:50.932021Z

Yes, I agree. That part of my audit was too pedantic.

In an outline, writing “working on a suitable compact set $\mathcal{K} \subseteq G$” when the intended object is the normalized family obtained by restricting the polynomial curves to $[0,1]$ is not a substantive issue if the identification is clear from the subsequent discussion. I would **not flag that as a genuine mistake**.

The real issue in that paragraph is only the definition of κ: you currently have

$$\kappa(A) = \left\lfloor \frac{F(A)}{h} \right\rfloor$$

with values in $\{0, \dots, r-1\}$, but it needs the reduction modulo r, and h should be chosen sufficiently **small**, not large.

So I would revise my audit item 3 to flag only those two points.

---

274. User

2026-10-07 20:55:11 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-07T12:55:11.886000Z

Selected text:

Selection 1

ine 463 uses $w_{ij} := \sum_{k=i}^{j-1} w_k$ as though there were one exponent list for the whole product. There are m different blocks, each with its own exponent list. The later rigorous section handles this blockwise, but the outline formula is formally ambiguous. Pasted text I would avoid defining w_{ij} there and simply write:

My request:

I agree that this is a problem, but I want a better fix than what you proposed.

---

275. Assistant

2026-10-07 20:55:33 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-07T12:55:33.335457Z

Yes. I think the clean fix is to **define the matrix-coordinate weights once near the beginning of the outline**, rather than trying to paraphrase them later.

Right after the sentence defining G and Γ, I would add:

```
For the block associated to \(P_i\), assign the matrix coordinate \((u,v)\), with \(u<v\), the weight
\[
w_{i;uv}:=w_{i,u}+w_{i,u+1}+\cdots+w_{i,v-1}.
\]
```

Of course, this requires writing

$$P_i(t) = \mathsf{L}(\alpha_{i,1}t^{w_{i,1}}, \dots, \alpha_{i,s_i}t^{w_{i,s_i}})$$

somewhere beforehand. I think that is worthwhile anyway, because Section 3.1 is already discussing all m expressions simultaneously.

Then your problematic sentence becomes very clean:

```
where \(C\) and \(Q\) have ordinary polynomial entries satisfying
\[
\deg C_{i;uv},\deg Q_{i;uv}\leq w_{i;uv}
\]
in the block corresponding to \(P_i\), and \(Q(n)\) is the canonical representative of \(p(b+qn)\) for
all \(0\leq n<M\).
```

And later you can write the error estimate as

$$|E_{i;uv}| \ll_{r,m,D} k^{-w_{i;uv}}.$$

I think this is much better than my previous suggestion because the coordinate weights are **structurally important throughout the argument**: they govern the polynomial degree bounds, the dilation δ_u, the map τ, the factorial denominators, and the final k^{-w} estimate. They deserve explicit notation in the outline.

If you want to avoid semicolons in subscripts, an equally good notation is

$$w_{uv}^{(i)} := w_{i,u} + \cdots + w_{i,v-1}.$$

I slightly prefer $w_{uv}^{(i)}$; it visually separates the **block index** i from the **matrix coordinates** u, v.

---

276. User

2026-10-07 20:59:29 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-07T12:59:29.741000Z

Please audit section 3.1 once again.

> **Attachments recorded in source:**
>
> - Pasted text(20261007-125912).txt

---

277. Assistant

2026-10-07 20:59:53 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-07T12:59:53.669000Z

I'm re-checking only Section 3.1 against the current proof, and I'll ignore the compact-family wording you already said was intentionally informal.

278. Assistant

2026-10-07 21:00:19 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-07T13:00:19.733475Z

The revised Section 3.1 is much better. I now see **five genuine issues** worth fixing; I would not flag anything else.

1. **Line 449: S should be F.** You write "cannot remain in one level set of S," but the function defined above is F. [Pasted text.txt, L434-L450]
2. **Line 454: h should be sufficiently small, not sufficiently large.** The proof later takes

$$h = \frac{\eta}{20r},$$

and the whole point is to make the shells thin enough that an excursion of F-oscillation at least η crosses many successive colors. A sufficiently large h could make the entire range lie in one shell. [Pasted text.txt, L452-L457] [Pasted text.txt, L737-L745]

So simply change "for a sufficiently large h" to "for a sufficiently small $h > 0$."

3. **Line 468 accidentally reintroduces the bad weight notation you just fixed.** You correctly defined

$$w_{i;uv} := w_{i,u} + \cdots + w_{i,v-1}$$

at lines 427–429, but line 468 then says

$$w_{i;uv} := \sum_{k=i}^{j-1} w_k,$$

which mixes the block index i with matrix-coordinate indices and contains an undefined j. [Pasted text.txt, L426-L432] [Pasted text.txt, L459-L469]

I would simply write:

```
where \(C\) and \(Q\) have ordinary polynomial entries satisfying \(\deg(C_{i;uv}),\deg(Q_{i;uv})\leq
w_{i;uv}\), and \(Q(n)\) is the canonical representative of \(p(b+qn)\) for all \(0\leq n<M\).
```

4. **Line 469 still states the omitted-color conclusion only for the first M samples, whereas the first-exit argument needs it through J.** You already have the stronger identity

$$\kappa(p(b+qn)) = \kappa(Q(n)) \qquad (n \in \mathbb{Z}).$$

Since $b + qn < k$ for every $0 \le n \le J$, the correct conclusion is that

$$\{\kappa(Q(n)) : 0 \le n \le J\}$$

omits the same color as the original progression. This is what makes the contradiction at line 475 work when T can be much larger than M. [Pasted text.txt, L459-L475]

I would replace line 469 by:

```
Since \(\kappa(p(b+qn))=\kappa(Q(n))\) for every \(n\in\Z\), all of the samples \(Q(n)\) with \(0\leq
n\leq J\) omit the same color as the original progression.
```

This is the most important remaining issue in the first-exit paragraph.

5. **Lines 485–488 still make the incorrect claim that $\{P_i(d)\}_c$ appears as an entry of E.** The particular matrix γ_d obtained from $\tau(C)$ is not known to be the canonical correction, so $E = g(d)\gamma_d$ is not known to be the canonical representative. What the proof actually does is use the small last-column entries of E to show, recursively, that γ_d must contain exactly the relevant nearest-integer corrections. Only then do you conclude

```
\[
\tnorm{P_i(d)}=|E_{\mathrm{top},i}|.
\]
```

[Pasted text.txt, L478-L489] The rigorous decoding lemma says exactly this. [Pasted text.txt, L976-L1004]

I would replace the final part by:

```
In particular, applying \(\delta_{L/q}\circ\tau\) to Equation~\eqref{eq:Outline} yields
\[
g(d)\gamma_d=E,\qquad \gamma_d\in\Gamma,\qquad |E_{i;uv}|\ll_{r,m,D}k^{-w_{i;uv}}.
\]
For sufficiently large \(k\), the last-column entries of \(E\) are small enough that the nearest-
integer corrections encoded by \(\gamma_d\) can be recovered recursively.
The upper-right entry in the block corresponding to \(P_i\) then gives
\[
\tnorm{P_i(d)}\ll_{r,m,D}k^{-d_i}.
\]
```

I prefer this wording because it actually tells the reader **what the final decoding step does**, rather than falsely making E sound canonical.

One very minor notation consistency point accompanies item 5: since you introduced $w_{i;uv}$ in the outline, the displayed estimate at line 485 should use $E_{i;uv}$ and $w_{i;uv}$, rather than reverting to E_{ij} and w_{ij}. [Pasted text.txt, L478-L488]

Other than these, I think Section 3.1 is now mathematically faithful to the detailed proof.

---

279. User

2026-10-07 21:02:18 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-07T13:02:18.842000Z

Selected text:

Selection 1

Line 469 still states the omitted-color conclusion only for the first M samples, whereas the first-exit argument needs it through J. You alr

My request:

J is mentioned later, so I can just go up to k instead, right?

280. Assistant

2026-10-07 21:02:49 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-07T13:02:49.250296Z

Not quite. You **cannot** say "for all $0 \le n < k$," because the corresponding original sample is

$$p(b + qn),$$

and for $q > 1$, $b + qn$ may already exceed $k - 1$.

What you can do, without introducing J yet, is write:

```
Since \(\kappa(p(b+qn))=\kappa(Q(n))\) for every \(n\in\Z\), every sample \(Q(n)\) for which \(0\leq
b+qn<k\) omits the same color as the original progression.
```

Then later, when you define

$$J = \left\lfloor \frac{k - 1 - b}{q} \right\rfloor,$$

this condition becomes exactly

$$0 \le n \le J.$$

I think that is actually cleaner than introducing J earlier.

---

281. User

2026-10-07 21:11:03 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-07T13:11:03.432000Z

Selected text:

Selection 1

Lines 485–488 still make the incorrect claim that $\{P_i(d)\}_c$ appears as an entry of E. The particular matrix γ_d obtained from $\tau(C)$ is not known to be the canonical correction, so $E = g(d)\gamma_d$ is not known to be the canonical representative. What the proof actually does is use the small last-column entries of E to show, recursively, that γ_d must contain exactly the relevant nearest-integer corrections. Only then do you conclude

My request:

but the example that we worked out seems to always have the generalized polynomial in the upper right corner, regardless of the value of k/the precision, no?

---

282. Assistant

2026-10-07 21:11:23 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-07T13:11:23.408744Z

Yes — but only because in the example we multiplied by a **very particular** integer matrix, namely the canonical correction matrix. That distinction is the whole issue.

For the depth- 3 example, you had

$$B(x_1,x_2,x_3)\begin{pmatrix}1 & * & * & -\langle x_1\langle x_2\langle x_3\rangle\rangle\rangle\\ 0 & 1 & * & -\langle x_2\langle x_3\rangle\rangle\\ 0 & 0 & 1 & -\langle x_3\rangle\\ 0 & 0 & 0 & 1\end{pmatrix},$$

and the upper-right entry of the product is indeed

$$\{x_1\langle x_2\langle x_3\rangle\rangle\}_{\mathrm{c}}.$$

That is an exact algebraic identity and has nothing to do with k or precision. [Pasted text.txt, L1007-L1055]

But it is **not true for an arbitrary** $\gamma\in\Gamma$.

The depth- 2 case makes this completely transparent. Put

$$B(x_1,x_2)=\begin{pmatrix}1 & -x_1 & 0\\ 0 & 1 & x_2\\ 0 & 0 & 1\end{pmatrix}$$

and take an arbitrary

$$\gamma=\begin{pmatrix}1 & a & b\\ 0 & 1 & c\\ 0 & 0 & 1\end{pmatrix}\in \mathrm{UT}_3(\mathbb{Z}).$$

Then

$$(B\gamma)_{13}=b-x_1c.$$

The generalized polynomial is

$$P=x_1\langle x_2\rangle.$$

There is no reason for

$$b-x_1c$$

to equal P modulo an integer unless

$$c=-\langle x_2\rangle.$$

For example, if $\gamma=I$, then

$$(B\gamma)_{13}=0$$

regardless of P.

What the canonical correction does is choose

$$c=-\langle x_2\rangle,$$

after which

$$(B\gamma)_{13}=b+x_1\langle x_2\rangle,$$

and then choosing

$$b=-\langle P\rangle$$

makes the upper-right entry exactly $\{P\}_{\mathrm{c}}$.

That is precisely what the last-column decoding lemma reconstructs. If

$$B(x_1, \dots, x_\ell)\gamma = E$$

for some arbitrary integer γ, but all the last-column errors are $< 1/2$, then the bottom equation forces

$$\gamma_{\ell,\ell+1} = -\langle x_\ell \rangle.$$

The next equation then forces the next nested correction, and so on upward, until the top-right entry is forced to be the centered fractional part of the generalized polynomial. [Pasted text.txt, L976-L1004]

So there are two statements:

$$\boxed{\text{For every } t, \text{ there exists a special canonical } \gamma_t \text{ for which the upper-right entry is } \{P(t)\}_{\mathrm{c}}.}$$

That is what your worked example shows.

But in the proof we first obtain

$$g(d)\gamma_d = E$$

for a **different, a priori arbitrary** $\gamma_d = \delta_{L/q}\tau(C) \in \Gamma$. Only after proving E is sufficiently small can we conclude that this γ_d must have the same relevant last-column entries as the canonical correction.

So I do think the sentence in the outline

> "Recalling that the values $\{P_i(d)\}_{\mathrm{c}}$ appear as entries of E..."

is logically premature. A very small change would preserve your intuition:

> "The smallness of E forces the relevant entries of γ_d to be the canonical nearest-integer corrections, so that the upper-right entry in each block is $\{P_i(d)\}_{\mathrm{c}}$."

That is exactly what is happening.

---

283. User

2026-10-07 21:12:37 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-07T13:12:37.116000Z

can you give me an example with a larger matrix that shows that we really need this smallness, and that the example that I have is just a lucky accident?

---

284. Assistant

2026-10-07 21:13:25 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-07T13:13:25.964530Z

Yes. A 5×5 example makes the role of the smallness hypothesis very transparent.

One correction first: your earlier 4×4 example is not really a "lucky accident." It works because the integer matrix multiplying B was **specifically constructed from the nested nearest-integer corrections**. What fails is the claim that an arbitrary $\gamma \in \Gamma$ will automatically put $\{P\}_{\mathrm{c}}$ in the upper-right corner.

Take

$$P = x_1\Big\langle x_2\big\langle x_3\langle x_4\rangle\big\rangle\Big\rangle$$

and

$$B = \begin{pmatrix} 1 & -x_1 & 0 & 0 & 0 \\ 0 & 1 & -x_2 & 0 & 0 \\ 0 & 0 & 1 & -x_3 & 0 \\ 0 & 0 & 0 & 1 & x_4 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

For an arbitrary $\gamma \in \mathrm{UT}_5(\mathbb{Z})$, write only its last column as

$$\gamma_{\bullet,5} = \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \\ 1 \end{pmatrix}, \qquad z_1, z_2, z_3, z_4 \in \mathbb{Z}.$$

Then the last column of

$$E = B\gamma$$

is

$$\begin{aligned} E_{45} &= x_4 + z_4, \\ E_{35} &= z_3 - x_3 z_4, \\ E_{25} &= z_2 - x_2 z_3, \\ E_{15} &= z_1 - x_1 z_2. \end{aligned}$$

Now choose

$$x_1 = 0.3, \qquad x_2 = 0.8, \qquad x_3 = 0.7, \qquad x_4 = 0.6.$$

Then

$$\langle x_4 \rangle = 1,$$

so

$$\langle x_3\langle x_4\rangle\rangle = \langle 0.7\rangle = 1,$$

and

$$\langle x_2\,\langle x_3\langle x_4\rangle\rangle\rangle = \langle 0.8\rangle = 1.$$

Therefore

$$P = 0.3, \qquad \{P\}_{\mathrm{c}} = 0.3.$$

Now take the completely innocent choice $\gamma = I$

Then

$$z_1 = z_2 = z_3 = z_4 = 0,$$

and hence

$$E = B.$$

Its last column is

$$\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0.6 \\ 1 \end{pmatrix}.$$

In particular,

$$\boxed{E_{15} = 0,}$$

whereas

$$\boxed{\{P\}_{\mathrm{c}} = 0.3.}$$

So the generalized polynomial does **not** automatically appear in the upper-right corner.

Notice something even stronger: the upper-right entry itself is already very small:

$$|E_{15}| = 0 < \frac{1}{2}.$$

Thus **smallness of the upper-right entry alone is not enough**. The problem is lower down the last column:

$$|E_{45}| = 0.6 > \frac{1}{2}.$$

What happens if all last-column entries are required to be small?

Suppose now

$$|E_{i5}| < \frac{1}{2}, \qquad i = 1, 2, 3, 4.$$

Start at the bottom. Since

$$E_{45} = 0.6 + z_4$$

and $z_4 \in \mathbb{Z}$, the condition

$$|0.6 + z_4| < \frac{1}{2}$$

forces

$$z_4 = -1 = -\langle 0.6 \rangle.$$

Then

$$E_{35} = z_3 - 0.7(-1) = z_3 + 0.7.$$

Smallness forces

$$z_3 = -1 = -\left\langle 0.7 \langle 0.6 \rangle \right\rangle.$$

Next,

$$E_{25} = z_2 - 0.8(-1) = z_2 + 0.8,$$

so smallness forces

$$z_2 = -1 = -\langle 0.8\,\langle 0.7\langle 0.6\rangle\rangle\rangle\,.$$

Finally,

$$E_{15} = z_1 - 0.3(-1) = z_1 + 0.3.$$

Smallness forces

$$z_1 = 0 = -\langle 0.3\rangle,$$

and consequently

$$\boxed{E_{15} = 0.3 = \{P\}_{\mathrm{c}}\,.}$$

So this example displays the exact logic of the decoding lemma:

$$\boxed{\text{small bottom entry} \Rightarrow z_4 \text{ canonical} \Rightarrow \text{small next entry} \Rightarrow z_3 \text{ canonical} \Rightarrow \cdots \Rightarrow E_{15} = \{P\}_{\mathrm{c}}\,.}$$

Without those smallness assumptions, the integers z_4, z_3, z_2, z_1 can be arbitrary, and there is no reason whatsoever for the upper-right entry to encode P.

This is why your worked canonical-reduction example can be slightly misleading pedagogically: there you **put in the correct z_i's by hand from the outset**. In the proof, $\gamma_d = \delta_{L/q}\tau(C)$ drops out of a completely different construction, so you must use the smallness of E to prove retrospectively that its last column contains exactly those canonical integers. [Pasted text.txt, L976-L1004]

---

285. User

2026-10-07 21:17:37 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-07T13:17:37.787000Z

Selected text:

Selection 1

Thus smallness of the upper-right entry alone is not enough. The problem is lower down the last column:

$$|E_{45}| = 0.6 > \frac{1}{2}.$$

My request:

exactly. Isn't this a bad example because the translated matrix does not land in the fundamental domain. I want an example where you translate the matrix into the fundamental domain, but you still don't get the centered fractional part in the upper right corner because we didn't assume the necessary smallness. Can you please construct that instead?

286. Assistant

2026-10-07 21:18:05 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-07T13:18:05.847802Z

You are right. **Such an example cannot exist.** My previous proposed example was answering the wrong question.

If

$$E = B(x_1, \dots, x_\ell)\gamma, \qquad \gamma \in \mathrm{UT}_{\ell+1}(\mathbb{Z}),$$

actually lies in the fundamental domain—that is, **every off-diagonal entry of E lies in**

$$[-1/2, 1/2),$$

then E is by definition the unique canonical representative of the coset $B\Gamma$. [Pasted text.txt, L567-L582] And for this band matrix, its upper-right entry is necessarily

$$\{\mathsf{L}(x_1, \dots, x_\ell)\}_{\mathrm{c}}\,.$$

Indeed, landing in the fundamental domain already gives exactly the "smallness" needed for the recursive decoding:

$$|E_{\ell,\ell+1}| < 1/2$$

forces

$$\gamma_{\ell,\ell+1} = -\langle x_\ell \rangle\,,$$

then

$$|E_{\ell-1,\ell+1}| < 1/2$$

forces the next nested nearest integer, and so on, until the top-right entry is

$$\{\mathsf{L}(x_1, \dots, x_\ell)\}_{\mathrm{c}}\,.$$

That is precisely the mechanism of your last-column decoding lemma. [Pasted text.txt, L976-L1004]

So there are really only two situations:

$$\boxed{B\gamma \text{ lies in the fundamental domain}} \quad \Longrightarrow \quad \boxed{\text{upper-right entry } = \{P\}_{\mathrm{c}}}$$

with no further hypothesis needed.

By contrast,

$$\boxed{\gamma \in \Gamma \text{ arbitrary}}$$

does **not** imply that the upper-right entry is $\{P\}_{\mathrm{c}}$; then you need the smallness of the relevant last-column entries to prove that γ has in fact chosen the canonical nearest-integer corrections.

And this makes me revise something important from my audit. In your actual proof, once you have

$$g(d)\gamma = E$$

and then choose k large enough that **all off-diagonal entries of E have absolute value** $< 1/4$, you actually know

$$E \in [-1/2, 1/2)^\nu,$$

so E is the canonical representative of $g(d)$. [Pasted text.txt, L1086-L1100]

Therefore, if you have already established in general that the canonical representative of $g_P(d)$ has $\{P(d)\}_c$ in the upper-right corner, then you could bypass the separate last-column decoding lemma entirely at this point.

The reason the current manuscript still uses that lemma is that it proves exactly the needed general statement directly from last-column smallness and does **not require** proving a separate general theorem about the full canonical representative. But conceptually:

> **" E is sufficiently small" means precisely that E has landed in the canonical fundamental domain.**

So your instinct about the 4×4 example was correct: it was not a lucky accident at all. The canonical reduction always puts the centered generalized polynomial in the upper-right corner. The only question in the proof is why the particular γ_d produced by coefficient extraction gives that canonical reduction; the answer is that the error matrix E is small enough to lie in the fundamental domain.

---

287. User

2026-10-07 21:20:28 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-07T13:20:28.690000Z

instead of |E_{i,ell+1}| < 1/2, we just need E to be in the fundamental domain, right? for example, some coordinates can be exactly -1/2, right?

---

288. Assistant

2026-10-07 21:20:40 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-07T13:20:40.541852Z

Yes. In fact, the current hypothesis

$$|E_{i,\ell+1}| < \frac{1}{2}$$

is slightly stronger than necessary.

With your convention

$$\langle x \rangle = \lfloor x + 1/2 \rfloor,$$

the canonical fundamental interval is

$$I = [-1/2, 1/2).$$

For every real x, there is a **unique** integer m such that

$$x + m \in I,$$

namely

$$m = -\langle x \rangle .$$

So an entry equal to $-1/2$ is perfectly allowable; $+1/2$ is the excluded endpoint. [Pasted text.txt, L567-L582]

Thus the decoding lemma can be strengthened to assume merely

$$E_{i,\ell+1} \in [-1/2, 1/2) \qquad (1 \le i \le \ell).$$

You do **not actually need all of** E to lie in the fundamental domain; only the relevant last-column entries need to lie in the canonical interval.

For example, at the bottom,

$$E_{\ell,\ell+1} = x_\ell + \gamma_{\ell,\ell+1} \in [-1/2, 1/2)$$

immediately implies, by uniqueness of the integer correction,

$$\gamma_{\ell,\ell+1} = -\langle x_\ell \rangle ,$$

even if

$$E_{\ell,\ell+1} = -\frac{1}{2}.$$

Then the same argument proceeds recursively upward. The current proof uses the stronger $< 1/2$ absolute-value condition. [Pasted text.txt, L976-L1004]

So I would revise the lemma to:

$$B(x_1, \dots, x_\ell)\gamma = E, \qquad E_{i,\ell+1} \in [-1/2, 1/2) \quad (1 \le i \le \ell).$$

And in the proof replace phrases like

> "Its absolute value is less than $1/2$, so..."

with

> "Since $E_{\ell,\ell+1} \in [-1/2, 1/2)$, uniqueness of the canonical integer correction gives..."

If the proof has already established that **all** off-diagonal entries of E lie in the fundamental domain, then of course this hypothesis follows automatically.

---

289. User

2026-10-07 21:28:29 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-07T13:28:29.581000Z

Please audit section 3.1 again.

> **Attachments recorded in source:**
>
> - Pasted text(20261007-132728).txt

---

290. Assistant

2026-10-07 21:28:59 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-07T13:28:59.045396Z

I'm checking the revised outline against the detailed proof as it now stands, and I'll only flag points that are still mathematically or grammatically substantive.

291. Assistant

2026-10-07 21:30:32 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-07T13:30:32.208089Z

The section is now much cleaner. I see **two genuine issues remaining**, plus one very minor accuracy point.

1. **Lines 426–429 use s_i and $w_{i,u}$ without ever defining them.** Earlier you only introduce a generic

$$P = \mathsf{L}(\alpha_1 t^{w_1}, \dots, \alpha_s t^{w_s}),$$

but then suddenly use s_i and $w_{i,u}$ for the individual P_i. [Pasted text.txt, L410-L429]

I would fix this right at line 410 by writing something like

$$P_i(t) = \mathsf{L}(\alpha_{i,1} t^{w_{i,1}}, \dots, \alpha_{i,s_i} t^{w_{i,s_i}})$$

before introducing the associated matrices. Then your definition

$$w_{i;uv} := w_{i,u} + \cdots + w_{i,v-1}$$

is completely well-defined.

2. **Line 488 is still slightly logically misleading.** After line 487 you have correctly explained that the last-column corrections are recovered recursively. But then

> "Recalling that the values $\{P_i(d)\}_{\mathrm{c}}$ appear as entries of E..."

sounds as though this follows merely from the earlier canonical-representative discussion. It does not: it follows from the recursive recovery you have just performed. [Pasted text.txt, L478-L490]

I would change only that sentence to something like:

```
The recursive recovery shows that the upper-right entry of the block corresponding to \(P_i\) is \
(\cpart{P_i(d)}\), and hence
\[
\tnorm{P_i(d)}\ll_{r,m,D}k^{-d_i}.
\]
```

That makes the logic exact.

There is also one **minor accuracy point** at line 445: you say you minimize "the sum of squares of its entries," whereas F sums the squares of the **off-diagonal entries**. [Pasted text.txt, L434-L445] Since the diagonal entries are all fixed at 1, this does not change the minimizer, so I would not regard it as a mathematical error, but "off-diagonal entries" would be more accurate.

Other than those, I do **not** see another genuine problem in Section 3.1. In particular, the J-range, the first-exit paragraph, the h-definition, the blockwise weights, and the role of L now all look consistent with the detailed proof.

292. User

2026-10-07 21:38:51 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-07T13:38:51.482000Z

Selected text:

Selection 1

sounds as though this follows merely from the earlier canonical-representative discussion. It does not: it follows from the recursive recovery you have just performed.

My request:

but I thought we just discussed this, the canonical representative has this property, so it does just follow from the fact that we have a canonical representative, right?

---

293. Assistant

2026-10-07 21:39:05 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-07T13:39:05.993066Z

Yes. You are right, and my last audit comment was mistaken.

Once you know that E is the canonical representative of $g(d)\Gamma$, the conclusion follows immediately from the property stated earlier in Section 3.1: the canonical representative of each block has $\{P_i(d)\}_{\mathrm{c}}$ in its upper-right corner. [Pasted text.txt, L426-L432]

From

$$g(d)\gamma_d = E, \qquad |E_{i;uv}| \ll_{r,m,D} k^{-w_{i;uv}},$$

for sufficiently large k, **all** off-diagonal entries of E lie in

$$(-1/2, 1/2) \subset [-1/2, 1/2).$$

Since $\gamma_d \in \Gamma$, E belongs to the coset $g(d)\Gamma$, and because all its off-diagonal coordinates lie in the fundamental domain, uniqueness of the canonical representative gives

$$E = \mathrm{rep}(g(d)).$$

[Pasted text.txt, L478-L489]

Therefore you do not need to describe the recursive decoding in the outline at all. I think the cleaner ending is:

```
In particular, applying \(\delta_{L/q}\circ\tau\) to Equation~\eqref{eq:Outline} yields
\[
g(d)\gamma_d=E,\qquad \gamma_d\in\Gamma,\qquad |E_{i;uv}|\ll_{r,m,D}k^{-w_{i;uv}}.
\]
For sufficiently large \(k\), every off-diagonal entry of \(E\) lies in \([-\frac12,\frac12)\), so \
(E\) is the canonical representative of \(g(d)\) in \(G/\Gamma\).
Consequently, the upper-right entry in the block corresponding to \(P_i\) is \(\cpart{P_i(d)}\), and
hence
```

```
\[
\tnorm{P_i(d)}\ll_{r,m,D}k^{-d_i}.
\]
```

I actually prefer this for Section 3.1. The later last-column decoding lemma is a more economical rigorous tool—it only needs the relevant last-column coordinates to be canonical—but at the **outline level**, saying that E is entirely inside the fundamental domain makes the geometry much easier to understand.

---

294. User

2026-10-07 21:41:00 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-07T13:41:00.322000Z

Please audit it once again.

> **Attachments recorded in source:**
>
> - Pasted text(20261007-134048).txt

---

295. Assistant

2026-10-07 21:41:39 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-07T13:41:39.561964Z

I'm checking only the current Section 3.1 and will avoid re-flagging points you've already decided are harmless.

---

296. Assistant

2026-10-07 21:42:17 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-07T13:42:17.209487Z

The current Section 3.1 is essentially sound. I find **one definite error and one minor mathematical inaccuracy**.

1. **Lines 422–423: the matrix uses s where it should use s_i.** You define

$$P_i = \mathsf{L}(\alpha_{i,1}t^{w_{i,1}}, \dots, \alpha_{i,s_i}t^{w_{i,s_i}}),$$

but the last two nonzero entries in the displayed matrix are written

$$-\alpha_{i,s-1}t^{w_{i,s-1}}, \qquad \alpha_{i,s}t^{w_{i,s}}.$$

They should be

$$-\alpha_{i,s_i-1}t^{w_{i,s_i-1}}, \qquad \alpha_{i,s_i}t^{w_{i,s_i}}.$$

[Pasted text.txt, L413-L426]

2. **Line 448: "sum of squares of its entries" should be "sum of squares of its off-diagonal entries."** The formula defining F omits the diagonal entries, whereas the prose says that for each block you minimize the sum of squares of all its entries. Since the diagonal entries are fixed at 1, this does not change the

minimizer, but it does change the value of the stated sum, so the “In other words” sentence is not literally correct. [Pasted text.txt, L437-L448]

I would simply change it to:

> “we multiply it by some element of $\mathrm{UT}_{s_i+1}(\mathbb{Z})$ to minimize the sum of squares of its **off-diagonal** entries…”

Everything else that we discussed previously now looks fixed. In particular, the blockwise weights are properly defined, the omitted-color conclusion is carried through J, h is chosen small, and the final statement that sufficiently small E is the canonical representative is correct. [Pasted text.txt, L429-L435] [Pasted text.txt, L455-L491]

I would consider Section 3.1 mathematically clean after those two edits.

---

297. User

2026-10-07 22:23:12 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-07T14:23:12.392000Z

please give me a bibtex entry for the recent open ai result that improves the bounds on szemeredi numbers.

---

298. Assistant

2026-10-07 22:23:58 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-07T14:23:58.943877Z

I’m rechecking only the current Section 3.1 against the detailed proof and the conventions we’ve now settled, without re-flagging the compact-family shorthand.

---

299. User

2026-10-07 22:24:50 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-07T14:24:50.701466Z

don't recheck the paper, just give me the bibtex code I asked for.

---

300. Assistant

2026-10-07 22:25:36 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-07T14:25:36.727688Z

Here is the official BibTeX entry from OpenAI’s release directory: GitHub

```
@misc{OAI:Quasipolynomial-Bounds-for-Arithmetic-Progressions-September-23-2026,
  author = {{OpenAI}},
  title = {{Quasipolynomial Bounds for Arithmetic Progressions}},
  howpublished = {OpenAI Math Release preprint
                  \href{https://github.com/openai/math/blob/main/preprints/Quasipolynomial-Bounds-for-
Arithmetic-Progressions-September-23-2026/paper.pdf}{OAI:Quasipolynomial-Bounds-for-Arithmetic-
Progressions-September-23-2026}},
  year = {2026}
}
```

This is the result giving

$$r_k(N) \le C_k N \exp\bigl(-c_k(\log N)^{\varepsilon_k}\bigr)$$

for every fixed $k \ge 3$.

---

301. User

2026-10-11 13:49:39 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-11T05:49:39.200000Z

what are the conditions on a nilmanifold that make it isomorphic to G/Gamma with G a subgroup of upper triangular matrices with 1s on the diagonal and real entries above the main diagonal, and Gamma = G intersect (integer matrices)?

---

302. Assistant

2026-10-11 13:49:53 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-11T05:49:53.562687Z

This is essentially a Mal'cev-theory question. The key distinction is between merely embedding the Lie group into a unipotent upper-triangular group and arranging that the lattice becomes exactly the intersection with the integer points.

---

303. Assistant

2026-10-11 13:50:01 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-11T05:50:01.364245Z

For the standard notion of a compact nilmanifold, essentially **no extra hypothesis is needed beyond the usual Mal'cev hypotheses**—but there is an important distinction between choosing a suitable matrix realization and using an arbitrary one.

Let

$$X = G_0/\Gamma_0,$$

where G_0 is a connected, simply connected nilpotent Lie group and $\Gamma_0 \le G_0$ is a discrete cocompact subgroup. Then Γ_0 determines a rational structure on the Lie algebra

$$\mathfrak{g}_{\mathbb{Q}} = \operatorname{span}_{\mathbb{Q}} \log \Gamma_0.$$

Equivalently, G_0 admits a lattice exactly when its Lie algebra has a basis with rational structure constants—Mal'cev's criterion. MathNet Annals of Mathematics

After choosing a **strong Mal'cev basis adapted to** Γ_0, one has coordinates in which

$$\Gamma_0 = \exp(\mathbb{Z}X_1)\cdots\exp(\mathbb{Z}X_m).$$

Such a basis always exists. MathNet One can then realize this rational nilpotent group faithfully as a unipotent upper-triangular matrix group

$$G \le UT_N(\mathbb{R}).$$

With a suitably chosen **integral** realization, the lattice can be arranged to be

$$\Gamma = G \cap UT_N(\mathbb{Z}),$$

and hence

$$X \cong G/\Gamma.$$

So, in the form relevant to your paper, the useful statement is roughly:

$$\boxed{\text{Every compact nilmanifold } G_0/\Gamma_0 \text{ with } G_0 \text{ connected and simply connected can be represented this way.}}$$

The lattice condition supplies exactly the rationality required for the integral upper-triangular model.

There is one caveat I would emphasize. If you take an **arbitrary** faithful embedding

$$G_0 \hookrightarrow UT_N(\mathbb{R})$$

that happens to send Γ_0 into $UT_N(\mathbb{Z})$, you need not have equality

$$\Gamma_0 = G_0 \cap UT_N(\mathbb{Z}).$$

For example, $2\mathbb{Z} \subset \mathbb{R}$ can be represented inside

$$UT_2(\mathbb{R}) = \left\{ \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} : x \in \mathbb{R} \right\}$$

as the even integral matrices; then the intersection with $UT_2(\mathbb{Z})$ is larger. You have to choose the representation **adapted to the given lattice**.

So if what you want for the exposition is a concise sentence, I would say:

> Every compact nilmanifold may, after choosing a suitable lattice-adapted Mal'cev realization, be written as G/Γ, where G is a closed subgroup of some $UT_N(\mathbb{R})$ and $\Gamma = G \cap UT_N(\mathbb{Z})$.

I would add "after choosing a suitable lattice-adapted realization," because without that qualifier the exact equality with the integer matrices is stronger than merely saying that nilpotent Lie groups are linearizable.

---

304. User

2026-10-11 14:21:35 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-11T06:21:35.090000Z

Please audit the updated paper for mathematical and grammatical mistakes.

> **Attachments recorded in source:**
>
> - Pasted text(20261011-062126).txt

---

305. Assistant

2026-10-11 14:22:13 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-11T06:22:13.753806Z

I'll audit the updated arguments and grammar, focusing on unresolved or newly introduced issues, with line references and concrete proposed fixes.

306. Assistant

2026-10-11 14:25:26 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-11T06:25:26.995865Z

The main rounding and first-exit argument appears consistent. Two issues stand out: the constant-improvement threshold mishandles the ceiling in q, and the final question has mismatched parameters.

---

307. Assistant

2026-10-11 14:34:42 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-11T06:34:42.087005Z

The revised endpoint convention is also consistent. One additional gap appears in Section 2: the argument linking path length to the number of shell crossings needs justification.

---

308. Assistant

2026-10-11 14:37:58 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-11T06:37:58.982964Z

I did not find a flaw in the main polynomial-rounding, shell-coloring, and coefficient-extraction argument in Section 3. Section 3.1 now describes that argument consistently, including the final use of the canonical representative. The remaining issues are mostly in the introductory discussion, the quantitative Bohr comparison, the constant-improvement lemma's cutoff, and the final question.

Below, line numbers refer to the **newly attached TeX file**. I distinguish errors from assertions that need additional justification.

Mathematical corrections

1. Line 1217: the stated cutoff does not match the preceding $1/4$ estimate

You choose

$$q = \left\lceil \frac{C_0}{c} \right\rceil$$

and then establish the no-wrap estimate under

$$k \geq 4qC_0.$$

But the concluding expression uses $4C_0^2/c$, which need not be at least $4qC_0$. The ceiling cannot simply be dropped. Also, the parameter list $k_1(r, m, d, c)$ should use D, not d. [Pasted text.txt, L1203-L1217]

The simplest replacement, preserving your existing $1/4$ argument, is:

```
Both \(q\) and the coloring were fixed independently of \(k\), so it suffices to take
\[
k_1(r,m,D,c)
=
\max\left\{
k_0(r,2m,D),
```

```
\left\lceil 4C(r,2m,D)\left\lceil\frac{C(r,2m,D)}{c}\right\rceil\right\rceil
\right\}.
\]
```

This is a mismatch in the written justification, not a counterexample to the lemma or even to your original cutoff. Your original expression can be retained with a slightly different argument: when $c < C_0$,

$$q < \frac{2C_0}{c},$$

so $k \geq 4C_0^2/c$ gives

$$|qu_i| < \frac{1}{2},$$

which is already sufficient for no wrap. When $c \geq C_0$, the original estimate already implies the desired bound. Nevertheless, the displayed replacement above is the more direct repair.

2. Lines 1253–1256: the final question uses K where the prescribed length ℓ should appear

The question introduces $\ell, r \geq 2$, but then never uses ℓ: both the tested progression and the progression sought inside A have length $K+1$. [Pasted text.txt, L1253-L1256]

This is not the transfer principle discussed earlier. The intended formulation allows a sufficiently long progression omitting a color to force an ** ℓ-term** progression in A with the same difference. In particular, the subsequent explanation using the case $r = 2$ does not directly establish arbitrary prescribed lengths from the question as currently written. [Pasted text.txt, L1259-L1262]

I would replace the question by:

```
\begin{question}\label{qu:polychromatic-transfer}
Let \(A\subseteq\N\) have positive upper density, and let \(\ell,r\geq2\).
Do there exist \(K\geq2\) and a coloring \(c:\N\to\{1,\ldots,r\}\) such that, for every \(a,d\in\N\),
if \(\{a,a+d,\ldots,a+(K-1)d\}\) omits a color, then
\[
A\cap(A-d)\cap\cdots\cap(A-(\ell-1)d)\neq\emptyset?
\]
\end{question}
```

Here K is the length of the progression used as a test, and ℓ is the independently prescribed length of the progression to be found in A.

3. Lines 363–409: the Bohr proof needs the geometric justification connecting a long path to shell crossings

The conclusion is correct, but the current proof jumps from “this path has length $\gg 1$” to “this path must completely cross at least $\Omega_d(K)$ shells.” A long path on a torus can wind repeatedly, so its length alone does not explain why the squared-distance function has a substantial oscillation. The fact that this is the projection of a **straight line segment** is essential. [Pasted text.txt, L359-L364] [Pasted text.txt, L409-L410]

Here is a complete justification that also supports your quantitative estimate.

Put

$$H(z) = \sum_{i=1}^{d} \|z_i\|_{\mathbb{R}/\mathbb{Z}}^2$$

and consider the lifted segment

$$\Sigma = \{f(x) + tv : 0 \le t \le J\} \subseteq \mathbb{R}^d.$$

Let $R = J\|v\|_1$ be its ℓ^1-length. The preceding divisibility argument gives

$$R > \frac{k\varepsilon}{L}.$$

If $R \ge d+1$, the segment crosses at most $R+d$ coordinate boundaries of the unit-cube tiling, so it is divided into at most $R+d+1$ pieces. Some piece therefore has ℓ^1-length at least

$$\frac{R}{R+d+1} \ge \frac{1}{2}.$$

Its Euclidean length is at least $1/(2\sqrt{d})$. After translating its cube to the fundamental cube, H on that segment is an ordinary squared Euclidean norm.

For a segment with endpoints u, w,

$$|u|_2^2 - 2\left|\frac{u+w}{2}\right|_2^2 + |w|_2^2 = \frac{1}{2}|u-w|_2^2.$$

Consequently, its H-oscillation is at least

$$\frac{1}{4}|u-w|_2^2 \ge \frac{1}{16d}.$$

This is the missing reason that sufficiently many shells are crossed.

The sampling assertion can then be justified **without any claim about dimension-independent Euclidean shell thickness**. The function H is 1-Lipschitz for the torus ℓ^1 metric:

$$|H(z) - H(z')| \le \sum_i \|z_i - z_i'\|_{\mathbb{R}/\mathbb{Z}}.$$

Thus your d/M-dense samples come within d/M in H-value of each shell midpoint. Taking

$$\frac{d}{M} < \frac{1}{2K}$$

ensures that a sample near a midpoint has that shell's color.

The existing choice $M = K^2$ is therefore adequate once K is sufficiently large compared with d. No further change to the basic coloring is needed.

4. Lines 412–422: distinguish the bound supported by your proof from the unprovided calculation for the random coloring

Your own estimate

$$k \le \frac{\exp((Cd^2)^d)}{\varepsilon}$$

can be justified using the preceding calculation. For example, for $0 < \varepsilon < 1/2$, take

$$K = 64d, \qquad M = K^2, \qquad L = (M^d)!, \qquad k = L\left\lceil \frac{d+1}{\varepsilon} \right\rceil.$$

These choices ensure the required path length and sampling precision, and

$$\log L \le M^d \log(M^d) \le (Cd^2)^d$$

after enlarging the absolute constant C. Thus your first displayed bound is supportable; I would add the parameter choices rather than leave it as "a careful analysis." [Pasted text.txt, L412-L417]

For the comparison

$$k \le \frac{\exp((Cd\log(2d))^d)}{\varepsilon},$$

I have not verified the asserted dependence on d. The cited version of Alweiss's proof leaves dimension-dependent constants implicit in its orbit-counting and sampling arguments; it does not state this numerical estimate. arXiv

This is not a claim that your comparison is false. It means that lines 418–422 need either a short derivation tracking those constants or a precise reference to one. [Pasted text.txt, L418-L422]

5. Line 133: the coloring argument is missing the hypothesis $S \cap B = \varnothing$

The preceding definition introduces an arbitrary $S \subseteq \mathbb{N}$. The next paragraph says that the torus coloring has no same-colored pair with difference in S, but this requires S to avoid the particular Bohr neighborhood being used. Without that hypothesis, the assertion is false—for example, for $S = \mathbb{N}$. [Pasted text.txt, L128-L134]

Start that paragraph with:

```
Suppose that \(S\cap B=\emptyset\), where \(B\) is the Bohr neighborhood in Definition~\ref{def:Bohr}.
```

The rest of the coloring argument then has its necessary hypothesis.

6. Lines 168–170 and 210: the change from "weight" to "degree" is incomplete

You now define the **degree** of a special generalized polynomial to be the sum of its displayed exponents, but the examples immediately afterward and the main theorem still call this quantity its **weight**. [Pasted text.txt, L162-L170] [Pasted text.txt, L208-L212]

This is more than a stylistic preference: the theorem currently uses terminology that the revised introduction no longer defines.

A clean convention would be:

> Special generalized polynomials have the assigned **degree** defined in the introduction; matrix coordinates have **weights**.

Under that convention, change "weight" to "degree" in the examples and "of weights $1 \le d_i \le D$" to "of degrees $1 \le d_i \le D$" in Theorem 1.9. In the constant-improvement proof, change "same assigned weight" to "same assigned degree." [Pasted text.txt, L1203-L1205]

There is no need to change the matrix-coordinate terminology in Section 3.

7. Line 1178: an arbitrary unipotent matrix subgroup need not give a compact quotient by its integer matrices

The sentence about obtaining nilmanifolds from groups of upper-unitriangular matrices needs a lattice qualification. Merely taking a subgroup $G \leq \mathrm{UT}_N(\mathbb{R})$ and its integer matrices does not ensure cocompactness. [Pasted text.txt, L1177-L1179]

For example,

$$G = \left\{ \begin{pmatrix} 1 & t & \sqrt{2}\,t \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} : t \in \mathbb{R} \right\}$$

is a closed one-dimensional subgroup, but

$$G \cap \mathrm{UT}_3(\mathbb{Z}) = \{I\}.$$

Its quotient by that intersection is therefore noncompact.

The simplest correction, sufficient for the examples used in your paper, is:

```
Standard examples are the compact quotients \(\UT_n(\R)/\UT_n(\Z)\) and their finite products.
```

To retain the more general subgroup formulation, explicitly require that $G \cap \mathrm{UT}_N(\mathbb{Z})$ be cocompact in G.

Statements and references needing clarification

8. Footnote at line 173: HSY identify the generated filter, not literally your family of exact basic sets

Your definition requires

$$S = \{n : \|P_i(n)\| < \varepsilon \text{ for every } i\}.$$

The footnote says this definition “agrees” with HSY’s definition. [Pasted text.txt, L172-L174]

HSY define nil-Bohr $_0$ sets by **containing** appropriate return-time sets, and their Theorems B and 4.2.11 identify the corresponding upward-closed families. Thus what you have defined is a **base for the same filter**. This is entirely sufficient for your recurrence statements, but the footnote should express that distinction. arXiv

A suitable replacement is:

```
\footnote{By Theorems B and 4.2.11 of \cite{HSY}, these sets form a base for the nil-Bohr$_0$ filter,
so meeting every set in this definition is equivalent to nil-Bohr recurrence in the terminology of
\cite{HSY}.}
```

Your half-integer convention does not need to be changed.

9. Line 192: the assertion that Alweiss's original example is not 2-large needs a proof or a reference

This may be one of the additional results established in your separate work, but it is not proved in the current manuscript, and it does not follow from the main theorem: the example is nil-Bohr recurrent. [Pasted text.txt, L188-L192]

Moreover, the cited version of *New Obstacles to Multiple Recurrence* explicitly asks whether its constructed sets are 2-large in **Question 3.7**. arXiv

Therefore, cite the separate proof that settles this for the original construction, include that proof, or remove the assertion. **I am not identifying this as a false statement; I am identifying an unsupported mathematical assertion in this draft.**

10. Line 1191: HKM asked for the existence of some finite r, not for the exact value

You say that Question 7.8 asks whether one can determine the "exact $r \geq 2$" for which the return set fails to be r-large. [Pasted text.txt, L1190-L1192]

Their actual question asks whether **there exists some** $r \geq 2$ for which it fails. Your answer $r = 2$ is indeed sharp, but the attribution should distinguish their existence question from your sharp answer. arXiv

Replace the sentence by:

```
In Question 7.8 they ask whether \(N(x,U)\) fails to be \(r\)-vdW-large for some finite \(r\geq2\).
```

11. Line 294: zero entropy is not, by itself, a justification of the random-coloring extension

The sentence says the random coloring extends "as … nil-manifolds have zero entropy as tori do." [Pasted text.txt, L293-L294]

There are two distinct issues. Entropy belongs to a **dynamical system**, so the relevant objects are nilrotations and torus rotations, not the underlying manifolds. More importantly, the proposed probabilistic extension needs suitable **uniform counting and approximation estimates** for the relevant family of orbits; the phrase "zero entropy" does not supply those estimates.

Alweiss describes his torus argument as, "in some sense," a combinatorial interpretation of zero entropy, rather than deriving the needed bounds solely from the zero-entropy assertion. arXiv

Since this alternative proof is not used, my recommendation is simply to retain the first sentence contrasting shell and random colorings and delete the extra justification. That avoids introducing an unproved methodological claim into the exposition.

Grammatical and typographical corrections

These are genuine wording or typing problems, rather than suggestions about prose style.

Line(s)	Correction
93	`Szemer\edi should be`Szemer'edi`. Also, “the optimal bound for k-term Szemerédi” is incomplete: write “the extremal number for k-term arithmetic progressions,” or explicitly refer to the extremal form of Szemerédi’s theorem. [Pasted text.txt, L93-L93]
129	“there exists $\alpha_1, \dots, \alpha_k, \varepsilon$” → “there exist $\alpha_1, \dots, \alpha_k, \varepsilon$.” [Pasted text.txt, L129-L129]
137–138	“Grytzcuk” → **“Grytczuk”**; “Mass” → **“Maass.”** [Pasted text.txt, L137-L138]
144–145	“Th proof” → “The proof”; “take the preimage the two coloring” → “pull back the two-coloring.” [Pasted text.txt, L144-L145]
153	“The next natural class … are those” has a singular–plural mismatch. Write “The next natural class of non-large sets consists of those that are not sets of nil-Bohr recurrence.” [Pasted text.txt, L153-L153]
173	“a $\varepsilon > 0$” → “an $\varepsilon > 0$.” [Pasted text.txt, L173-L173]
177	“In language of ergodic theory” → “In the language of ergodic theory”; “nimanifolds” → “nilmanifolds”; “finite dimensional tori” → “finite-dimensional tori.” [Pasted text.txt, L177-L177]
189	“whether the nil-Bohr recurrence implies largeness” → “whether nil-Bohr recurrence implies largeness.” [Pasted text.txt, L189-L189]
409	“compeltely” → “completely.” In the later clause, the displayed set is singular, so change “$\{\dots\}$ contain points” to “$\{\dots\}$ contains points.” [Pasted text.txt, L409-L409]
776–780	Remove the full stop at the end of the displayed definition of oscillation: the sentence continues with “is continuous on this compact family.” As written it reads “The function … . is continuous.” [Pasted text.txt, L776-L780]
805	“samples every color..” → “samples every color.” [Pasted text.txt, L798-L805]
985–989	Insert **“the samples”** after “says that,” before the displayed identity, so that “receive all r colors” has an explicit subject. [Pasted text.txt, L985-L989]
1177	“a quotient space consisting of a … group G and a … subgroup Γ” → “the quotient of a finite-dimensional nilpotent Lie group G by a discrete cocompact subgroup Γ.” [Pasted text.txt, L1177-L1177]
1250, 1260	“solve 2-large is large conjecture” → “solve the 2-large-is-large conjecture”; “vdW-Large” → “vdW-large.” [Pasted text.txt, L1250-L1250] [Pasted text.txt, L1260-L1260]
1269	The editorial instructions beginning “Put in a lot more. Substantial AI statement.” are **uncommented and will appear in the paper**. Remove those directives and finalize the acknowledgements paragraph. [Pasted text.txt, L1264-L1269]

Overall assessment

The central theorem’s proof appears mathematically sound in this audit. In particular, I found no new problem with polynomial rounding, uniform oscillation, the first-exit normalization, clearing the factorial denominators, or the half-open-interval decoding.

The highest-priority edits are the **progression lengths in the final question**, the **cutoff justification in the constant-improvement lemma**, and the **missing geometric calculation in the Bohr proof**. The claimed sharper bound for the random coloring remains the quantitative assertion I have not verified; the available cited proof does not make its dimension dependence explicit.

---

309. User

2026-10-11 15:04:00 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-11T07:04:00.648000Z

Selected text:

Selection 1

“a quotient space consisting of a … group G and a … subgroup Γ” → “the quotient of a finite-dimensional nilpotent Lie group G by a discrete cocompact subgroup Γ.

My request:

give me a complete replacement sentence please.

---

310. Assistant

2026-10-11 15:04:08 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-11T07:04:08.421792Z

<WritingBlock id="58321" variant="standard">A *nilmanifold* is a quotient space $X = G/\Gamma$, where G is a finite-dimensional nilpotent Lie group and Γ is a discrete cocompact subgroup of G.</WritingBlock>

---

311. User

2026-10-11 15:06:13 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-11T07:06:13.477000Z

Please do one more full audit of the paper.

> **Attachments recorded in source:**
>
> - Pasted text(20261011-070608).txt

---

312. Assistant

2026-10-11 15:06:46 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-11T07:06:46.112688Z

I’ll audit the revised draft end to end, checking proofs and citations while focusing on unresolved issues and any mistakes introduced by the latest edits.

313. Assistant

2026-10-11 15:09:58 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-11T07:09:58.140307Z

The main matrix proof still appears consistent; the remaining substantive issues are the Bohr shell-crossing justification, constant-improvement cutoff, and progression lengths in the final transfer question.

314. Assistant

2026-10-11 15:19:09 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-11T07:19:09.862973Z

The endpoint convention and coefficient-uniform first-exit argument also check out. I'm now checking citation numbering, compiled output, and the new AI statement.

315. Assistant

2026-10-11 15:29:01 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-11T07:29:01.584505Z

I did not find a flaw in the main proof in Section 3. The polynomial-rounding argument, the uniform sampling lemma, the first-exit argument, the denominator clearing, and the final identification of the canonical representative remain consistent. In particular, I would not make further mathematical changes to Section 3.1 on the basis of this audit. [Pasted text.txt, L987-L1036] [Pasted text.txt, L1160-L1186]

There are still several unresolved issues elsewhere. The line numbers below refer to the **newly attached file,** `Pasted text(20261011-070608).txt`.

Mathematical corrections and missing justifications

1. Lines 1267–1276: the final question still does not use ℓ

The question introduces $\ell, r \geq 2$, but the conclusion asks for a progression of length $K+1$ inside A, rather than the independently prescribed length ℓ. Thus ℓ is unused, and the subsequent deduction concerning arbitrarily long progressions is not justified by the stated formulation. [Pasted text.txt, L1267-L1276]

The intended question is:

```
\begin{question}\label{qu:polychromatic-transfer}
Let \(A\subseteq\N\) have positive upper density, and let \(\ell,r\geq2\).
Do there exist an integer \(K\geq2\) and a coloring \(c:\N\to\{1,\ldots,r\}\) such that, for every \
(a,d\in\N\), if \(\{a,a+d,\ldots,a+(K-1)d\}\) omits a color, then
\[
A\cap(A-d)\cap\cdots\cap(A-(\ell-1)d)\neq\emptyset?
\]
\end{question}
```

Here K is the test-progression length, while ℓ is the progression length sought inside A. With this replacement, the deductions immediately afterward work.

2. Line 1231: the cutoff still does not follow from the preceding $1/4$ calculation

The proof establishes its no-wrap estimate under

$$k \geq 4qC_0, \qquad q = \left\lceil \frac{C_0}{c} \right\rceil .$$

It then concludes with the potentially smaller threshold $4C_0^2/c$. That expression does not automatically satisfy the hypothesis of the preceding calculation, because of the ceiling in q. The parameter list also still has lowercase d, where it should have D. [Pasted text.txt, L1217-L1231]

The shortest repair uses the quantities already defined:

```
Both \(q\) and the coloring were fixed independently of \(k\), so it suffices to take
\[
k_1(r,m,D,c)=\max\{k',\lceil4qC_0\rceil\}.
\]
```

To be precise, I am not saying that your original cutoff is false. When $c < C_0$, one has $q < 2C_0/c$, so the original cutoff ensures

$$|qu_i| \leq \frac{qC_0}{k} < \frac{1}{2},$$

which is sufficient for no wrap. When $c \geq C_0$, the original estimate already gives the desired conclusion. Thus you could retain the original cutoff by changing the justification, but the replacement above preserves your existing argument without adding cases.

3. Lines 370–420: the Bohr proof still needs the geometric step explaining the shell crossings

The transition from the long torus path to "completely cross at least $\Omega_d(K)$ of the shells" remains unsupported in the written proof. The claim is repairable; the relevant fact is that the path comes from a **straight segment in the universal cover**, not merely that its length is large. [Pasted text.txt, L370-L375] [Pasted text.txt, L420-L421]

Here is the calculation that should be inserted before the shell-crossing conclusion.

Write

$$H(z) = \sum_{i=1}^{d} \|z_i\|_{\mathbb{R}/\mathbb{Z}}^2, \qquad \Sigma = \{f(x) + tv : 0 \leq t \leq J\},$$

and let

$$R = J\|v\|_1$$

be the ℓ^1-length of the lifted segment. The preceding divisibility calculation gives $R > k\varepsilon/L$, so choose k large enough that $R \geq d + 1$.

The coordinate boundaries of the unit-cube tiling divide Σ into at most

$$R + d + 1$$

pieces. Indeed, the i-th coordinate crosses at most $J|v_i| + 1$ boundaries. One piece therefore has ℓ^1-length at least

$$\frac{R}{R+d+1} \geq \frac{1}{2},$$

and hence Euclidean length at least $1/(2\sqrt{d})$.

On the closure of that piece, after translating its cube to the fundamental cube, H is an ordinary squared Euclidean norm. If u, w are the endpoints, then

$$H(u) - 2H\left(\frac{u+w}{2}\right) + H(w) = \frac{1}{2}|u-w|_2^2.$$

It follows that

$$\boxed{\operatorname{osc}_\Sigma H \geq \frac{1}{16d}.}$$

This supplies the required shell crossings.

For the sampling step, use

$$|H(z) - H(z')| \leq \sum_{i=1}^{d} \|z_i - z_i'\|_{\mathbb{R}/\mathbb{Z}}.$$

Consequently, a sample within d/M of a point whose H-value is a shell midpoint remains in the same shell whenever

$$\frac{d}{M} < \frac{1}{2K}.$$

For example, ** $K = 64d$ and $M = K^2$ suffice**: the oscillation is at least $4/K$, and the sampling error is comfortably below half the shell width.

This is a missing justification, not a failure of the shell-coloring method.

4. Lines 424–434: the two quantitative bounds have different verification status

Your shell bound

$$k \leq \frac{\exp((Cd^2)^d)}{\varepsilon}$$

is supported by the preceding calculation. For $0 < \varepsilon < 1/2$, take

$$K = 64d, \qquad M = K^2, \qquad L = (M^d)!, \qquad k = L\left\lceil \frac{d+1}{\varepsilon} \right\rceil.$$

These choices give the necessary divisibility, path length, and sampling precision. Moreover,

$$\log L \leq M^d \log(M^d) \leq (Cd^2)^d$$

after enlarging the absolute constant. I would record these choices, rather than leave the first bound entirely to “a careful analysis.” [Pasted text.txt, L424-L429]

I still have not verified the sharper bound claimed for the random coloring

$$k \leq \frac{\exp((Cd\log(2d))^d)}{\varepsilon}.$$

The cited proof leaves its relevant orbit-counting and sampling estimates with implicit dimension- and radius-dependent constants; it does not state this formula. arXiv

Thus lines 430–434 need a derivation, perhaps in the supplementary material, or a reference to a derivation. This is an **unverified assertion**, not a bound I have disproved. [Pasted text.txt, L430-L434]

5. Line 133: the Bohr-coloring argument still needs $S \cap B = \varnothing$

The definition immediately beforehand introduces an arbitrary S. The following paragraph then claims that the torus coloring forbids same-colored pairs with differences in S, without explicitly assuming that S avoids the Bohr neighborhood being used. [Pasted text.txt, L128-L134]

Add this sentence at the beginning of the paragraph:

```
Suppose that \(S\cap B=\emptyset\), where \(B\) is the Bohr neighborhood in Definition~\ref{def:Bohr}.
```

The coloring argument is then correctly grounded.

6. Lines 168–170, 210, and 1219: “degree” and “weight” remain inconsistent for special generalized polynomials

You now define the sum of the displayed exponents as the **degree** of P, but the next examples still call it its **weight**. The main theorem likewise says “of weights,” and the constant-improvement proof says “same assigned weight.” [Pasted text.txt, L168-L170] [Pasted text.txt, L208-L212] [Pasted text.txt, L1217-L1219]

The clean repair is to use **degree for the special generalized polynomials** and retain **weight for matrix coordinates**. In particular:

```
For example, \(\alpha n\) has degree \(1\), while \(\beta n^2\) and \(\gamma n\nint{\theta n}\) have
degree \(2\).
```

Change “of weights $1 \le d_i \le D$” to “of degrees $1 \le d_i \le D$,” and change “same assigned weight” to “same assigned degree” in the constant-improvement proof.

There is no reason to change the matrix-coordinate terminology in Section 3.

7. Line 1192: the matrix-subgroup example still needs a cocompactness qualification

The revised definition of a nilmanifold is correct. The next sentence remains too broad: an arbitrary real upper-unitriangular subgroup need not have a cocompact subgroup consisting of its integer matrices. [Pasted text.txt, L1191-L1193]

For example,

$$G = \{I + tE_{12} + \sqrt{2}\,tE_{13} : t \in \mathbb{R}\} \le \mathrm{UT}_3(\mathbb{R})$$

is a closed subgroup isomorphic to $\mathbb{R}$, but its only integer matrix is I. The resulting quotient is noncompact.

A complete replacement that covers the examples actually used in your proof is:

```
Standard examples of nilmanifolds are the quotients \(\UT_n(\R)/\UT_n(\Z)\) and their finite products.
```

Alternatively, retain the subgroup formulation but explicitly assume that $G \cap \mathrm{UT}_n(\mathbb{Z})$ is cocompact in G.

Citation and attribution corrections

8. Footnote at line 173: say “a base for the same filter,” not that the definitions literally agree

Your definition specifies exact sets of the form

$$\{n : \|P_i(n)\| < \varepsilon \text{ for every } i\}.$$

The footnote still says that this definition agrees with HSY’s definition. [Pasted text.txt, L172-L174]

HSY use an upward-closed family of sets containing appropriate return neighborhoods. The relevant results identify the **generated filters**, which is sufficient for your recurrence statements but is not literal equality of the two definitions. arXiv

Replace the footnote by:

```
\footnote{By Theorems B and 4.2.11 of \cite{HSY}, these sets form a base for the nil-Bohr$_0$ filter,
so meeting every set in this definition is equivalent to nil-Bohr recurrence in the terminology of
\cite{HSY}.}
```

This preserves your definition and fixes only the comparison.

9. Line 192: cite the separate proof that Alweiss’s original example is not 2-large

The assertion remains in the introduction without a proof or a reference to your separate argument. [Pasted text.txt, L188-L192]

The cited version of *New Obstacles to Multiple Recurrence* asks about the 2-largeness of its constructions in **Question 3.7**, rather than proving the negative assertion made here. arXiv

I am not disputing the separate result you developed. The manuscript needs to **cite that result**, include its proof, or explicitly point to supplementary material containing it. Any parameter restrictions in that result should also be reflected in the sentence.

10. Line 1205: HKM asked whether some finite r exists, not for the exact value

The sentence still says that Question 7.8 asks for the “exact $r \geq 2$.” [Pasted text.txt, L1204-L1206]

Their question asks whether there exists **some** $r \geq 2$ for which the return set fails to be r-large. Your answer $r = 2$ is sharp, but the attribution should distinguish their question from your stronger conclusion. arXiv

Use:

```
In Question 7.8 they ask whether \(N(x,U)\) fails to be \(r\)-vdW-large for some finite \(r\geq2\).
```

11. Line 305: zero entropy is not the stated justification for the random-coloring extension

The sentence still treats the extension as following from the fact that “nil-manifolds have zero entropy as tori do.” [Pasted text.txt, L304-L305]

Entropy belongs to the **dynamical system**, rather than the underlying manifold. More importantly, the probabilistic argument needs quantitative control of the relevant orbit families; zero entropy alone does not establish those estimates. Alweiss describes the connection with zero entropy as an interpretation of his combinatorial argument, not as a substitute for it. arXiv

Since the new AI statement announces supplementary material containing the alternative proof, you could replace the explanation by a direct reference to that material:

```
An alternative proof extending the random-coloring method of \cite{AlweissBohr} to nilrotations is
described in the supplementary material announced above.
```

That sentence should be retained only once the supplementary proof has been completed and made available. Otherwise, keep the initial contrast between shell and random colorings and omit the unproved explanation.

Grammatical and editorial corrections

The new AI statement introduces a few additional corrections. There are also a small number of older wording issues still present.

Location	Correction
Line 93	"the optimal bound for k-term Szemerédi" is incomplete. Specify the extremal quantity, rather than treating "Szemerédi" as its name. A replacement is given below. [Pasted text.txt, L93-L93]
Line 145	"the two coloring" → **"the two-coloring."** [Pasted text.txt, L145-L145]
Lines 294–295	Both occurrences of **"Mal'cev basis"** should be **"Mal'cev bases"** in these sentences. At line 294, **"Thoerem"** should be **"Theorem."** [Pasted text.txt, L294-L295]
Line 420	The second occurrence of the displayed set is followed by "contain points." Since the subject is a set, change this to **"contains points."** [Pasted text.txt, L420-L420]
Line 1234	Write **"k-AP-free"** in both occurrences. [Pasted text.txt, L1234-L1234]
Line 1264	"solve 2-large is large conjecture" → **"solve the 2-large-is-large conjecture."** [Pasted text.txt, L1264-L1264]

For the abstract's sentence at line 93, a mathematically precise replacement is:

```
We also show that, for every fixed nil-Bohr neighborhood \(B\) of \(0\) and every \(\eps>0\),
restricting the common difference to \(B\) increases the extremal number for \(k\)-term arithmetic
progressions by at most a factor of \(1+\eps\), uniformly in the size of the ambient interval, once \
(k\) is sufficiently large in terms of \(B\) and \(\eps\).
```

Line 298 also needs the repository name checked. For submission to the Chinese preprint repository, the name is **ChinaXiv**, not `ChinarXiv`. **ChinaRxiv** is a different site providing English access to papers from ChinaXiv and other sources. These names are not interchangeable. [Pasted text.txt, L298-L299] english.las.cas.cn ChinaRxiv

I have treated the account of your private work and its chronology in the AI statement as the authors' account; I have not independently verified that history.

Overall assessment

The principal theorem's proof still passes this audit. The most important remaining mathematical edits are the prescribed length ℓ in the final question, the cutoff justification in the constant-improvement lemma, and the geometric justification in the Bohr proof.

The sharper numerical bound attributed to the random-coloring argument remains unverified here. The other outstanding citation issues concern how results are stated or attributed, rather than a failure of the central shell-coloring construction.

---

316. User

2026-10-12 09:46:46 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-12T01:46:46.956000Z

Please do one more full audit of the paper.

> **Attachments recorded in source:**
>
> - Pasted text(20261012-014643).txt

---

317. Assistant

2026-10-12 09:47:21 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-12T01:47:21.434098Z

I'll audit the latest draft end to end, checking the revised proofs, quantitative claims, citations, and cross-references, and report only remaining or newly introduced issues.

---

318. Assistant

2026-10-12 09:50:30 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-12T01:50:30.556413Z

The cutoff and degree/weight issues are now fixed. The final question can also remain as written, but its claimed consequence needs a residue-class thinning argument.

---

319. Assistant

2026-10-12 09:59:40 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-12T01:59:40.139419Z

I found no new flaw in the main proof, and the sharper random-coloring bound is now justifiable. The remaining proof-writing issue is the Bohr path-to-shell-crossing step.

320. Assistant

2026-10-12 10:14:59 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-12T02:14:59.292982Z

I again found no flaw in the main proof in Section 3. In particular, the polynomial-rounding lemma, uniform sampling, first-exit normalization, denominator clearing, and final use of the canonical representative remain consistent. The revised cutoff in the constant-improvement lemma is also correct. [Pasted text.txt, L673-L723] [Pasted text.txt, L988-L1037] [Pasted text.txt, L1218-L1232]

This audit also resolves two points that I previously left problematic: **your final question can remain in its current form**, and **the sharper random-coloring bound can be justified**. Both would benefit from an additional explanation, given below.

All line numbers refer to `Pasted text(20261012-014643).txt`, not the earlier draft.

1. Lines 1268–1277: the final question is sufficient as written, but its consequences need one additional argument

You have removed the independently prescribed length ℓ. Your question now asks whether, for each positive-upper-density set A and each $r \geq 2$, there are K and a coloring such that an omitted-color progression of length $K+1$ forces a progression of that same length in A. [Pasted text.txt, L1268-L1277]

My earlier insistence that the question must contain an independent ℓ was too strong. The current formulation does imply the consequences you claim. The missing observation is that **we can force the resulting K to be arbitrarily large by first restricting A to one residue class modulo a sufficiently large integer**.

Here is the argument.

Fix a desired progression length $\ell \geq 2$ and a number of colors r. Let $W(r,\ell)$ be a finite van der Waerden number, and choose

$$q > W(r,\ell).$$

Some residue-class intersection

$$A' = A \cap (b + q\mathbb{Z})$$

has positive upper density. Apply your question to A', obtaining K and c.

Suppose that $K+1 \leq \ell$. By the definition of $W(r,\ell)$, the coloring c has a monochromatic ℓ-term progression inside $[1, W(r,\ell)]$, with some difference

$$0 < d < q.$$

Its first $K+1$ terms omit a color, so your proposed property produces a $(K+1)$-term progression in A' with difference d. But any two terms of A' are congruent modulo q, forcing $q \mid d$, a contradiction. Therefore

$$K+1 > \ell.$$

Now apply the largeness assumption on S: for $r=2$, 2-vdW-largeness supplies the required monochromatic progression; for the density version, choose r with $1 - 1/r \geq \delta$ and use the complement of a suitable color class. The transferred progression in $A' \subseteq A$ has more than ℓ terms.

Thus **keep the question**, but insert the following paragraph before the explanation of its consequences:

```
To obtain progressions of any prescribed length \(\ell\geq2\), choose an integer \(q>W(r,\ell)\),
where \(W(r,\ell)\) is a van der Waerden number.
Some residue-class intersection \(A'=A\cap(b+q\Z)\) has positive upper density.
Apply Question~\ref{qu:polychromatic-transfer} to \(A'\), obtaining \(K\) and a coloring \(c\).
If \(K+1\leq\ell\), a monochromatic \(\ell\)-term progression in \([1,W(r,\ell)]\) has a common
difference \(d\) satisfying \(0<d<q\), and its first \(K+1\) terms force a progression with difference
\(d\) in \(A'\).
This contradicts the fact that every difference between two elements of \(A'\) is divisible by \(q\).
Hence \(K+1>\ell\), which allows the proposed transfer principle to produce arbitrarily long
progressions.
```

This is an additional justification, **not a correction to the question itself**.

2. Lines 376–423: the Bohr proof still omits the geometric justification for crossing many shells

The written proof still passes directly from a long torus path to the assertion that it completely crosses $\Omega_d(K)$ shells. The straightness of its lift is what makes this true, and that connection should be stated. [Pasted text.txt, L372-L377] [Pasted text.txt, L422-L423]

The following calculation fills the gap without changing the construction.

Put

$$H(z)=\sum_{i=1}^{d}\|z_i\|_{\mathbb{R}/\mathbb{Z}}^2,\qquad \Sigma=\{f(x)+tv:0\le t\le J\},\qquad R=J\|v\|_1.$$

Your preceding divisibility argument gives $R>k\varepsilon/L$. Choose k so that $R\ge d+1$.

The unit-cube boundaries divide Σ into at most $R+d+1$ pieces: the i-th coordinate crosses at most $J|v_i|+1$ boundaries. Some piece therefore has ℓ^1-length at least

$$\frac{R}{R+d+1}\ge\frac12,$$

and hence Euclidean length at least $1/(2\sqrt{d})$.

On the closure of that piece, H agrees with squared Euclidean distance from the center of its fundamental cube. If u,w are its endpoints, then

$$H(u)-2H\left(\frac{u+w}{2}\right)+H(w)=\frac12|u-w|_2^2.$$

Consequently,

$$\operatorname{osc}_\Sigma H\ge\frac14|u-w|_2^2\ge\frac{1}{16d}.$$

That is the needed positive oscillation. Since the shells have width $1/K$ in ** H-value**, taking $K\ge64d$ supplies consecutive full shells of both colors.

For sampling, use the elementary inequality

$$|H(z)-H(z')|\le\sum_{i=1}^{d}\|z_i-z_i'\|_{\mathbb{R}/\mathbb{Z}}.$$

Your d/M-dense samples therefore come within d/M in H-value of each shell midpoint. The condition

$$\frac{d}{M} < \frac{1}{2K}$$

ensures that such a sample remains in the midpoint's shell.

This is preferable to reasoning about Euclidean shell thickness, and it confirms that the existing choice $M = K^2$ works once K is sufficiently large in terms of d.

For the first quantitative estimate at lines 426–431, concrete sufficient choices are

$$K = 64d, \qquad M = K^2, \qquad L = (M^d)!, \qquad k = L\left\lceil \frac{d+1}{\varepsilon} \right\rceil$$

for $0 < \varepsilon < 1/2$. They give

$$k \le \frac{\exp((Cd^2)^d)}{\varepsilon}$$

after enlarging the absolute constant C. Thus that estimate is correct; the manuscript needs the geometric calculation and parameter tracking, not a different coloring. [Pasted text.txt, L426-L431]

3. Lines 432–436: I can now justify the sharper random-coloring bound

The claimed estimate

$$k \le \frac{\exp((Cd\log(2d))^d)}{\varepsilon}$$

can also be supported. I therefore withdraw my previous classification of it as unverified. The cited paper supplies the random-grid method, although it does not explicitly state this dimension-dependent formula. [Pasted text.txt, L432-L436] arXiv

Here is an independent constant-tracking argument within that method.

Let $m_0 \ge 8$ be an integer, and independently color the unit cubes of

$$(\mathbb{R}/m_0\mathbb{Z})^d$$

red or blue. Consider all m_0-term progressions

$$x + jv \pmod{m_0\mathbb{Z}^d}, \qquad 0 \le j < m_0, \qquad \frac{1}{4} \le |v|_\infty \le \frac{1}{2}.$$

Each such progression visits at least $m_0/8$ distinct cubes. Indeed, one coordinate moves monotonically by at least $(m_0 - 1)/4$, with consecutive changes at most $1/2$ and total displacement less than one period.

There are at most

$$(Cm_0^4)^d$$

possible sequences of visited cubes. To see this, in each coordinate the sequence of floors is determined by the lines

$$x_i + jv_i = a, \qquad a \in \mathbb{Z}, \quad 0 \le j < m_0,$$

in the two-dimensional (x_i, v_i)-parameter rectangle. There are $O(m_0^2)$ relevant lines and therefore $O(m_0^4)$ cells, edges, and vertices in their arrangement. Including lower-dimensional faces handles samples on cube boundaries.

A union bound now gives a suitable coloring whenever

$$(Cm_0^4)^d\, 2^{1-m_0/8} < 1.$$

Thus

$$m_0 = \lceil C_0 d \log(2d) \rceil$$

works for a sufficiently large absolute C_0.

To transfer this coloring to the integers, set

$$R = (4m_0)^d, \qquad L = R!,$$

and color n by the cube containing

$$\frac{m_0 n \alpha}{L} \pmod{m_0 \mathbb{Z}^d}.$$

For a difference y with $\max_i \|y\alpha_i\| > \varepsilon$, pigeonhole in the grid of side length $1/4$ gives $1 \le q \le R$ and a representative

$$v = q\frac{m_0 y \alpha}{L} - m_0 z, \qquad |v|_\infty < \frac{1}{4}.$$

Because $q \mid L$, the bad-difference condition implies

$$|v|_\infty > \frac{m_0 q \varepsilon}{L}.$$

Choose

$$s = \left\lceil \frac{1}{4|v|_\infty} \right\rceil.$$

Then

$$\frac{1}{4} \le |sv|_\infty < \frac{1}{2}, \qquad qs(m_0 - 1) < \frac{L}{2\varepsilon}.$$

Hence a sufficiently long original progression contains an m_0-term subprogression with index spacing qs that receives both colors.

Finally,

$$\log L \le R \log R \le (Cd \log(2d))^d,$$

which yields the claimed estimate.

The bound can stay. I recommend adding this calculation, or a reference to a supplementary version of it, rather than leaving the explicit dependence on d entirely to “a careful analysis.”

4. Line 133: the Bohr-coloring argument still lacks its hypothesis on S

The paragraph claims that the finite torus coloring forbids same-colored pairs whose difference is in S, but this requires S to avoid the Bohr neighborhood being used. The immediately preceding definition does not impose that condition on S. [Pasted text.txt, L128-L134]

Insert:

```
Suppose that \(S\cap B=\emptyset\), where \(B\) is the Bohr neighborhood in Definition~\ref{def:Bohr}.
```

Without such a hypothesis, the assertion is false, for example when $S = \mathbb{N}$.

5. Line 137: there is an earlier written source for the Bohr-recurrence question

The sentence says that you and Grytczuk "first raised" the question in the cited joint paper. [Pasted text.txt, L137-L138]

However, **your 2016 master's thesis already states exactly this as Conjecture 8.13, on printed page 39**. Moreover, the joint paper explicitly points to Conjecture 8.13 of that thesis in its discussion of finite unions of lonely sets. Computer Science at UMD Colgate Mathematics

I cannot determine the history of private discussions from those sources, but the written history should acknowledge the thesis rather than suggesting that the question first appeared in the joint paper.

With a thesis entry under a key such as `FarhangiThesis`, a replacement is:

```
The question of whether every \(2\)-large set is a set of Bohr recurrence appears as Conjecture~8.13
in the second author's master's thesis \cite{FarhangiThesis} and was subsequently discussed by the
second author and Grytczuk \cite{FarhangiGrytczuk}.
```

Alternatively, simply removing "first" avoids the problematic priority claim without adding a reference.

6. Line 192: the negative 2-largeness assertion still needs a reference to its separate proof

The manuscript says that Alweiss's original example is not 2-large, but the cited paper itself asks about the 2-largeness of its constructions in Question 3.7. [Pasted text.txt, L188-L192] arXiv

This is not a claim that your assertion is false. It is a citation gap: the current paper does not contain the separate argument discussed in this chat.

An explicit reference to the supplementary proof would suffice. That reference should identify the parameter choices covered by the argument, rather than silently asserting the result for every variant of the construction.

7. Line 1193: clarify whether you mean the full unitriangular group

The nilmanifold definition immediately beforehand is now correct. The next sentence still says that examples are obtained from "a finite-dimensional group of upper triangular matrices" and its integer matrices. [Pasted text.txt, L1192-L1196]

For the **full** group $\mathrm{UT}_n(\mathbb{R})$, this is correct. For an arbitrary real subgroup, its integer matrices need not be cocompact. For example,

$$G = \{I + tE_{12} + \sqrt{2}\,tE_{13} : t \in \mathbb{R}\}$$

has only the identity as an integer matrix, so the corresponding quotient is noncompact.

The shortest unambiguous replacement is:

```
Standard examples of nilmanifolds are the quotients \(\UT_n(\R)/\UT_n(\Z)\) and their finite products.
```

Alternatively, retain the subgroup formulation but expressly require that the integer matrices form a lattice.

8. Lines 1215 and 1238–1239: a minor mismatch in the range of c

The constant-improvement lemma is stated only for

$$0 < c < C(r, m, D),$$

whereas the corollary permits any $c > 0$ and invokes the lemma for that prescribed value. [Pasted text.txt, L1214-L1215] [Pasted text.txt, L1237-L1239]

This is an easy formal adjustment, **not a substantive gap**. Your proof of the lemma already works for every $c > 0$: the choice $q = \lceil C_0/c \rceil$ is always a positive integer and satisfies $C_0/q \le c$.

Therefore change only

```
any prescribed \(c\in(0,C)\)
```

to

```
any prescribed \(c>0\)
```

The proof and corrected cutoff can remain unchanged.

Grammatical and typesetting corrections

Only a few of the previously identified wording issues remain.

Location	Correction
Line 93	"the optimal bound for k-term Szemerédi" does not identify the quantity being compared. Specify the **extremal number**; otherwise a reader might understand a threshold $N(\delta, k)$, for which your displayed multiplicative comparison is not the stated result. [Pasted text.txt, L93-L93]
Line 145	"the two coloring" → **"the two-coloring."** [Pasted text.txt, L145-L145]
Line 300	The preprint repository is spelled **ChinaXiv**, not `ChinarXiv`. [Pasted text.txt, L300-L301] National Science Library
Line 1235	Use **"k-AP-free"** in both occurrences. [Pasted text.txt, L1235-L1235]

For the abstract, a complete replacement is:

```
We also show that, for every fixed nil-Bohr neighborhood \(B\) of \(0\) and every \(\eps>0\),
restricting the common difference to \(B\) increases the extremal number for \(k\)-term arithmetic
progressions by at most a factor of \(1+\eps\), uniformly in the size of the ambient interval, once \
(k\) is sufficiently large in terms of \(B\) and \(\eps\).
```

The local TeX compilation also produced one small overfull line—about **9.8 pt**—in the supplementary-materials paragraph at lines 300–301, caused by displaying the full URL as ordinary hyperlink text. Using `\url{...}` or a shorter descriptive hyperlink would address it. The mathematical displays, including the canonical-reduction example, did not produce comparable overflow.

Points that are now resolved

I would **not** repeat the earlier objections concerning the cutoff, degree/weight terminology, HSY footnote, HKM question, or canonical representative. The relevant changes now say what they need to say: “degree (or weight)” explicitly makes the terminology synonymous; the HSY footnote compares recurrence notions rather than exact families of sets; and the HKM attribution asks for the existence of an r, not its exact value. [Pasted text.txt, L168-L174] [Pasted text.txt, L1205-L1207]

I also checked the three pinpoint polychromatic references: **Section 1.1** of Bollobás–Pritchard–Rothvoß–Scott, **Observation 4** of Pálvölgyi, and **Lemma 1** of the published Axenovich et al. paper support the uses made of them. The last is numbered differently in the arXiv version, but your citation to Lemma 1 matches the published version. arXiv D-NB Colgate Mathematics

Overall: the principal proof remains sound in this audit. The main additions still needed are the short geometric justification in the Bohr proof and the residue-class argument explaining the final question’s consequences. The remaining corrections are local hypotheses, references, and wording. The supplied TeX compiled successfully with internal cross-references resolved; the current `.bib` database was not included in this upload, so the final formatted bibliography was not part of the compilation check.

---

321. User

2026-10-12 10:46:27 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-12T02:46:27.766000Z

Selected text:

Selection 1

. Lines 1268–1277: the final question is sufficient as written, but its consequences need one additional argument You have removed the independently prescribed length ℓ. Your question now asks whether, for each positive-upper-density set A and each $r \geq 2$, there are K and a coloring such that an omitted-color progression of length $K+1$ forces a progression of that same length in A.

My request:

can i instead just change the question to make the coloring depend on K, which will be chosen alongside A and r?

322. Assistant

2026-10-12 10:47:22 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-12T02:47:22.387951Z

I'll check the quantifiers carefully: allowing the coloring to depend on K differs from prescribing K while requiring both progression lengths to equal K.

---

323. Assistant

2026-10-12 10:50:29 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-12T02:50:29.913837Z

The coloring can depend on K, but requiring the same-length conclusion for every prescribed K would make the question false. Your current wording already allows K and the coloring to be chosen together; what would change is making K an input rather than an output. [Pasted text.txt, L1268-L1272]

Why we cannot prescribe an arbitrary K

Fix $r, K \geq 2$, let $W = W(r, K+1)$ be a van der Waerden number, and choose an integer $q > W$. Take

$$A = q\mathbb{N},$$

which has positive density $1/q$.

Every r-coloring of $\mathbb{N}$ has a monochromatic $(K+1)$-term progression inside $[1, W]$. Its common difference satisfies

$$0 < d < q.$$

That progression omits a color, but

$$A \cap (A - d) = \varnothing,$$

because $q \nmid d$. In particular,

$$A \cap (A - d) \cap \cdots \cap (A - Kd) = \varnothing.$$

Thus **no coloring works for this prescribed triple** (A, r, K), regardless of its permitted dependence on the parameters.

The simplest revision that keeps only one length parameter

You can instead ask for **arbitrarily large K, allowing a different coloring for each such K**. This preserves your same-length formulation and makes the consequences immediate, without putting the residue-class argument into the paper:

```
\begin{question}\label{qu:polychromatic-transfer}
Let \(A\subseteq\N\) have positive upper density, and let \(r\geq2\) be an integer.
Do there exist arbitrarily large integers \(K\geq2\) for which there is a coloring \(c:\N\to\
{1,\ldots,r\}\) such that, for every \(a,d\in\N\), if \(\{a,a+d,\ldots,a+Kd\}\) omits a color, then
\[
A\cap(A-d)\cap\cdots\cap(A-Kd)\neq\emptyset?
\]
\end{question}
```

There is no requirement that the coloring be independent of K. You could make that explicit with the additional sentence:

```
The coloring may depend on \(A\), \(r\), and \(K\).
```

This formulation gives both intended deductions directly. For a 2-vdW-large set S, apply the question with $r = 2$: for each available K, its coloring has a monochromatic $(K+1)$-term progression with difference $d \in S$, and hence A has a progression of that length with difference in S. Since the available K's are arbitrarily large, this proves Szemerédi-largeness.

For a δ-Szemerédi-large set S, choose r with $1 - 1/r \geq \delta$. For each available K, the complement of some color class has upper density at least δ, so it contains a $(K+1)$-term progression with difference $d \in S$. This progression omits a color and therefore transfers to A.

I recommend this revision. It retains one length parameter, explicitly permits the dependence you want, and avoids the obstruction to prescribing a small K. In fact, the residue-class argument from my previous answer shows that, when asserted for every positive-density A, your current question already implies this arbitrarily-large- K version. So this revision makes the useful consequence explicit rather than introducing a genuinely stronger universal principle.

---

324. User

2026-10-12 10:50:39 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-12T02:50:39.090000Z

Selected text:

Selection 1

the optimal bound for k-term Szemerédi" does not identify the quantity being compared. Specify the extremal number; otherwise a reader might understand a threshold $N(\delta, k)$, for which your displayed multiplicative comparison is not the stated result.

My request:

but doesn't the analogous result for N(delta,k) follow as a corollary? That was my intent.

---

325. Assistant

2026-10-12 10:52:33 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-12T02:52:33.842439Z

There **is** an analogous corollary for the Szemerédi threshold $N(\delta, k)$, but it is weaker than a multiplicative $1 + \varepsilon$ comparison. In fact, the latter is false, even when the allowed differences form an ordinary Bohr neighborhood of 0.

1. What your corollary implies

Let $N(\delta, k)$ be the smallest N such that every subset of $[N]$ with at least δN elements contains a k-term arithmetic progression. Define $N_B(\delta, k)$ analogously, requiring the common difference to belong to B.

Your extremal result says that, for every fixed nil-Bohr neighborhood B and every $\varepsilon > 0$, provided k is sufficiently large,

$$r_k(N) \le r_k(N;B) \le (1+\varepsilon) r_k(N)$$

for every N.

Consequently,

$$\boxed{N(\delta,k) \le N_B(\delta,k) \le N\left(\frac{\delta}{1+\varepsilon},k\right).}$$

Indeed, at $N = N(\delta/(1+\varepsilon),k)$, we have

$$r_k(N) < \frac{\delta}{1+\varepsilon} N,$$

so

$$r_k(N;B) \le (1+\varepsilon) r_k(N) < \delta N.$$

Thus your theorem does yield a natural comparison of Szemerédi thresholds.

2. Why a multiplicative $1+\varepsilon$ comparison for $N(\delta,k)$ is false

Consider the Bohr neighborhood

$$B = q\mathbb{N}$$

for a fixed integer $q \ge 2$.

Write $N_0 = N(\delta,k)$. By minimality, there exists a set

$$A \subseteq [1, N_0 - 1]$$

of density at least δ containing no k-term arithmetic progression.

Construct

$$A' = \{q(n-1) + j : n \in A,\ 1 \le j \le q\}.$$

Then $A' \subseteq [1, q(N_0-1)]$ has density at least δ. Moreover, it contains no k-term progression with common difference divisible by q, since such a progression would project to a k-term progression in A.

Therefore

$$N_B(\delta,k) \ge q(N_0-1) + 1.$$

Conversely, every subset of $[1, qN_0]$ of density at least δ has density at least δ in one residue class modulo q. That residue class contains a k-term progression whose difference is divisible by q. Hence

$$N_B(\delta,k) \le qN_0.$$

Combining the two inequalities gives

$$\boxed{q(N(\delta,k)-1) + 1 \le N_B(\delta,k) \le qN(\delta,k).}$$

Since $N(\delta,k) \ge k$, it follows that

$$\boxed{\lim_{k\to\infty}\frac{N_B(\delta,k)}{N(\delta,k)}=q.}$$

In particular, for $B=2\mathbb{N}$, the ratio approaches **2, not 1**. For general q, it can approach any prescribed positive integer.

3. What I recommend for the abstract

Your result is therefore stronger in the **extremal-number formulation** than in the threshold formulation. I would explicitly state the extremal result, but you can also mention its threshold consequence.

For example:

```
We also show that, for every fixed nil-Bohr neighborhood \(B\) of \(0\) and every \(\eps>0\),
restricting the common difference to \(B\) increases the extremal number for \(k\)-term arithmetic
progressions by at most a factor of \(1+\eps\), provided \(k\) is sufficiently large.
As a consequence, the corresponding Szemer\'edi thresholds satisfy \(N_B(\delta,k)\le
N(\delta/(1+\eps),k)\).
```

So your intuition about a threshold corollary was correct. The distinction is that the $1+\varepsilon$ factor applies to the *extremal density or cardinality*, not to the *length of the ambient interval* needed to force a progression.

---

326. User

2026-10-12 11:05:42 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-12T03:05:42.791000Z

Selected text:

Selection 1

The local TeX compilation also produced one small overfull line—about 9.8 pt—in the supplementary-materials paragraph at lines 300–301, caused by displaying the full URL as ordinary hyperlink text. Using \url{...} or a shorter descriptive hyperlink would address it. The mathematical displays, including the canonical-reduction example, did not produce comparable overflow.

My request:

give me the replacement tex code using url{} please

---

327. Assistant

2026-10-12 11:05:55 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-12T03:05:55.340841Z

```
After this article is submitted to arXiv and ChinarXiv, the website \url{https://sohailfarhangi.com/2-
large-sets-are-sets-of-nil-bohr-recurrence/} will be updated with supplementary materials for this
article.
```

Your existing `\usepackage{hyperref}` already provides `\url`, so no additional package is needed.

328. User

2026-10-12 11:11:41 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-12T03:11:41.181000Z

Selected text:

Selection 1

The bound can stay. I recommend adding this calculation, or a reference to a supplementary version of it, rather than leaving the explicit dependence on d entirely to “a careful analysis.”

My request:

We are choosing to ignore this.

---

329. Assistant

2026-10-12 11:12:00 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-12T03:12:00.706061Z

Understood. In future audits, I will not flag the absence of a detailed derivation or supplementary reference for the sharper random-coloring bound. I will revisit it only if I find a substantive mathematical error.

---

330. User

2026-10-12 11:19:05 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-12T03:19:05.950000Z

Selected text:

Selection 1

The sentence says that you and Grytczuk “first raised” the question in the cited joint paper. Pasted text However, your 2016 master’s thesis already states exactly this as Conjecture 8.13, on printed page 39. Moreover, the joint paper explicitly points to Conjecture 8.13 of that thesis in its discussion of finite unions of lonely sets. Computer Science at UMD+1

My request:

out of respect for Grytczuk, we are choosing to cite our joint paper rather than my master's thesis.

---

331. Assistant

2026-10-12 11:19:20 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-12T03:19:20.494505Z

Understood. I’ll respect your decision to cite the joint paper with Grytczuk rather than your master’s thesis. I will not raise this citation choice again in future manuscript audits.

332. User

2026-10-12 11:20:10 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-12T03:20:10.525000Z

Selected text:

Selection 1

The manuscript says that Alweiss's original example is not 2-large, but the cited paper itself asks about the 2-largeness of its constructions in Question 3.7.

My request:

this is a valid point. Ryan said that he will update his preprint after it is accepted for publication, so we will ignore this point for now and fix it later.

---

333. Assistant

2026-10-12 11:20:22 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-12T03:20:22.581372Z

Understood. I'll exclude this citation issue from future manuscript audits for now. We'll revisit it after Ryan updates his preprint following its acceptance for publication.

---

334. User

2026-10-12 11:48:50 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-12T03:48:50.497000Z

Please do one more full audit of the paper. I am still working on point 3 of your last audit, so you don't have to repeat that point.

> **Attachments recorded in source:**
>
> - Pasted text(20261012-034823).txt

---

335. Assistant

2026-10-12 11:49:24 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-12T03:49:24.436421Z

I'll audit the updated arguments, notation, and prose, leaving aside point 3 and the citation issues we agreed to defer.

336. Assistant

2026-10-12 11:52:35 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-12T03:52:35.325805Z

The revised cutoff and "arbitrarily large K" formulation now work as intended. I'm checking the new degree footnote and tracing the remaining parameter dependencies through the proofs.

---

337. Assistant

2026-10-12 12:01:53 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-12T04:01:53.508798Z

I found no new gap in the main proof in Section 3 or in the extremal corollary. The updated cutoff, the extension to every $c > 0$, and the arbitrarily-large- K formulation of the final question now work as intended. [Pasted text.txt, L1215-L1233] [Pasted text.txt, L1269-L1278]

The remaining changes I recommend are below. Line numbers refer to `Pasted text(20261012-034823).txt`.

1. Line 169: revise the new footnote about degree

The new footnote says:

> "the notion of degree is well defined for special generalized polynomials, but not arbitrary generalized polynomials."

That is **too categorical**. Huang–Shao–Ye explicitly define the classes GP_d of generalized polynomials of degree at most d in **Definition 4.1.1**. Thus there is an established degree filtration for general generalized polynomials; it is not correct to suggest that no degree notion is available. [Pasted text.txt, L163-L171] arXiv

The distinction relevant to your proof is between **the degree assigned to a displayed expression** and a notion intrinsic to the function after allowing arbitrary rewritings and cancellations. Your definition by the sum of the exponents is entirely suitable for the proof.

I would simply remove the warning and write:

```
We define the degree of the expression \(P\) in \eqref{eq:sgp-definition} by \
(\deg(P)=w_1+\cdots+w_s\), and its \emph{depth} by \(d(P)=s\).
```

This also accommodates zero coefficients without ambiguity: a zero expression may still carry its assigned exponent pattern, exactly as your proof permits.

To distinguish this from a more serious objection: **for a nonzero special generalized polynomial of your prescribed form, the exponent sum does agree with its polynomial growth exponent**. Indeed, writing $d = w_1 + \cdots + w_s$, successive rounding errors give

$$P(n) = \left(\prod_{j=1}^{s} \alpha_j \right) n^d + O(n^{d-1}).$$

So the degree convention itself is natural; the issue is only the footnote's blanket claim about arbitrary generalized polynomials.

2. Line 112: qualify the polynomial-image example

The introduction currently lists “the image of a polynomial P with integer coefficients and $P(0)=0$” as an example of a large set. This needs the usual positivity and nondegeneracy qualifications, because your large sets consist of **positive integers**. For example, $P(n)=-n^2$ satisfies the stated conditions but has no positive values. [Pasted text.txt, L108-L112]

The relevant statement in Brown–Graham–Landman explicitly assumes a **positive leading coefficient** and intersects the image with the positive integers; see their **Theorem 3.1, page 32**, with $a=0$. Cambridge University Press

A replacement for the middle sentence of that paragraph is:

```
Brown, Graham, and Landman give examples of large sets in their paper, including \(P(\N)\cap\N\) for
every nonconstant polynomial \(P\in\Z[t]\) with positive leading coefficient and \(P(0)=0\), as well
as several examples of sets that are not large.
```

This is a small missing qualification in the introduction, not an issue with your own theorem.

3. Line 1272: “omits a colors” should be “omits a color”

The revised quantifiers in the final question are correct, but the edit introduced this grammatical error. [Pasted text.txt, L1269-L1272]

The complete question can read:

```
\begin{question}\label{qu:polychromatic-transfer}
Let \(A\subseteq\N\) have positive upper density, and let \(r\geq2\) be an integer.
Do there exist arbitrarily large integers \(K\geq2\) for which there is a coloring \(c:\N\to\
{1,\ldots,r\}\) such that, for every \(a,d\in\N\), if \(\{a,a+d,\ldots,a+Kd\}\) omits a color, then
\[
A\cap(A-d)\cap\cdots\cap(A-Kd)\neq\emptyset?
\]
\end{question}
```

No additional residue-class argument is needed for the consequences of this revised formulation. Since K can be arbitrarily large, the argument already written immediately afterward produces arbitrarily long progressions. [Pasted text.txt, L1275-L1278]

4. Line 301: the repository name is still misspelled

The name remains `ChinarXiv`; it should be **ChinaXiv**. My earlier replacement using `\url` unfortunately preserved that spelling error. [Pasted text.txt, L301-L302] ChinaXiv

Here is the corrected complete sentence:

```
After this article is submitted to arXiv and ChinaXiv, the website \url{https://sohailfarhangi.com/2-
large-sets-are-sets-of-nil-bohr-recurrence/} will be updated with supplementary materials for this
article.
```

The `\url` change **does fix the previous overfull line** in my local compilation.

Minor TeX-reference issue

At **line 380**, you have

```
\begin{figure}[H]\label{fig:ShellColoring}
```

but the caption occurs later, followed by the correctly positioned label `fig:shell-progression`. [Pasted text.txt, L380-L385] [Pasted text.txt, L419-L421]

The first label is set before the figure counter is advanced by the caption. In the compiled auxiliary file, `fig:ShellColoring` **therefore refers to Section 2, not Figure 1**. It is not currently used in the body, so this does not create a visible incorrect reference, but the simplest repair is to delete that first label:

```
\begin{figure}[H]
```

Keep the existing label after the caption:

```
\caption{The coloring of $[-1/2,1/2]^2$ with $K=48$, alternately colored shells, and marked points
equally spaced along a segment crossing the square.}
\label{fig:shell-progression}
```

The compilation also reports a duplicate PDF destination named `figure.1`. Moving `float` before `hyperref` removed that warning in my test. Replace the present three-package sequence at lines 15–17 with:

```
\usepackage{tikz}
\usepackage{float}
\usepackage{hyperref}
```

This is a hyperlink/typesetting issue, not a mathematical one. [Pasted text.txt, L15-L17]

Overall assessment

The main argument still appears sound: the rounded lift has the required weighted degree bounds and integer-valued correction; the omitted-color hypothesis controls all relevant integer samples; the real first-exit normalization gives the uniform bound; and the extracted, dilated error matrix lies in the canonical fundamental domain for sufficiently large k. [Pasted text.txt, L674-L723] [Pasted text.txt, L984-L1038] [Pasted text.txt, L1166-L1188]

The source compiled successfully, with internal cross-references resolved and **no overfull boxes** in the source-only test. I did not build the final bibliography from an updated `.bib` file, so this was not a check of the formatted reference list. Beyond the points above, I found no further mathematical or grammatical correction to recommend in this pass.

338. User

2026-10-12 12:05:30 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-12T04:05:30.589000Z

Selected text:

Selection 1

The name remains ChinarXiv; it should be ChinaXiv. My earlier replacement using \url unfortunately preserved that spelling error

My request:

are you sure? I went to https://chinarxiv.org/ and it says ChinaRxiv.

---

339. User

2026-10-12 12:05:57 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-12T04:05:57.712000Z

are you sure? I went to https://chinarxiv.org/ and it says ChinaRxiv.

---

340. Assistant

2026-10-12 12:06:20 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-12T04:06:20.515376Z

You're right that **https://chinarxiv.org/** identifies itself as **ChinaRxiv**. I checked the website directly. My earlier correction was too categorical. <Cite ref="turn574544view0"/>

There are actually two different websites:

- **<Link url="https://chinarxiv.org/" title="ChinaRxiv"/>** — an English-language platform providing access to research papers from the Chinese preprint repository, including translated versions.
- **<Link url="https://chinaxiv.org/" title="ChinaXiv"/>** — the Chinese Academy of Sciences' preprint repository, which accepts original preprint submissions. <Cite refs={["turn574544view1","turn574544search3"]}/>

For your manuscript, if you're referring to the site you visited, **ChinaRxiv** is the correct capitalization, not ChinaXiv.

However, since your sentence concerns *submitting* the paper, **ChinaXiv** appears to be the appropriate repository. ChinaRxiv primarily republishes and translates papers from ChinaXiv.

So my earlier assertion that `chinarxiv.org` was misspelled was incorrect. The distinction is between the two services, not a typo in the domain.

341. User

2026-10-12 12:12:05 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-12T04:12:05.138000Z

Selected text:

Selection 1

4. Line 301: the repository name is still misspelled

My request:

so can I submit the english paper to ChinaXiv?

---

342. Assistant

2026-10-12 12:12:21 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-12T04:12:21.099483Z

Yes. ChinaXiv explicitly accepts research papers written in English, so you can submit your English-language mathematics preprint without translating it into Chinese.

I verified this from ChinaXiv's own submission instructions: authors can select **Chinese, English, or Russian** as the manuscript language. The platform also accepts PDF submissions. <Cite refs={["turn430860search0","turn430860search2"]}/>

There is, however, one practical issue concerning your account registration.

ChinaXiv's registration instructions prioritize Chinese institutional email addresses ending in `.ac.cn` or `.edu.cn`. If you use a BIMSA address ending in `.bimsa.cn` or a personal Gmail address, you may need administrator verification before submitting. The contact address is **eprint@mail.las.ac.cn**. <Cite ref="turn415577view1"/>

There is also a particularly convenient option for your planned submission to both repositories: **ChinaXiv supports one-click submission of an existing arXiv preprint**. After logging in, go to *My Center → My Submissions → arXiv Submit* and enter the arXiv identifier. <Cite ref="turn415577view1"/>

I would therefore submit your English paper to arXiv first, then use ChinaXiv's arXiv import feature. Acceptance on ChinaXiv is still subject to its preliminary screening.

Official website: <Link url="https://chinaxiv.org/" title="ChinaXiv"/>.

343. User

2026-10-12 12:21:29 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-12T04:21:29.070000Z

Selected text:

Selection 1

2. Lines 376–423: the Bohr proof still omits the geometric justification for crossing many shells

My request:

earlier when I said I am still addressing point 3, I meant to say that I am addressing point 2. I have not yet addressed it, so please do one more full audit of the paper, and repeat the contents of point 2 so that I don't have to scroll up everytime I want to see it.

> **Attachments recorded in source:**
>
> - Pasted text(20261012-042125).txt

---

344. Assistant

2026-10-12 12:22:04 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-12T04:22:04.310736Z

I'll audit the updated draft and repeat the full shell-crossing justification with current line numbers, while leaving aside the issues you've chosen to defer.

---

345. Assistant

2026-10-12 12:25:13 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-12T04:25:13.463522Z

I have not found a new gap in the central argument. The remaining Bohr step can use squared-distance oscillation directly, avoiding any shell-thickness estimate.

---

346. Assistant

2026-10-12 12:34:13 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-12T04:34:13.179790Z

I found no new mathematical or grammatical mistake to add in this version. The main argument in Section 3 remains consistent, as do the constant-improvement lemma, the extremal corollary, and the consequences of the revised arbitrarily-large- K question. The outstanding issue is the geometric justification in the Bohr proof, which I reproduce below. [Pasted text.txt, L963-L1035] [Pasted text.txt, L1212-L1252] [Pasted text.txt, L1266-L1275]

Line numbers here refer to `Pasted text(20261012-042125).txt`.

The outstanding Bohr step — your point 2

At **lines 370–375**, you obtain a long linear path on the torus and show that its integer samples are d/M-dense. At **line 420**, you then assert that the path completely crosses $\Omega_d(K)$ shells. The missing justification is the passage from **the length of the lifted straight segment** to **a definite oscillation of the squared-distance function defining the shells**. [Pasted text.txt, L370-L375] [Pasted text.txt, L420-L421]

A. A sufficiently long lifted segment spends definite length in one cube

Define the periodic function

$$H(z) = \sum_{i=1}^{d} \|z_i\|_{\mathbb{R}/\mathbb{Z}}^2, \qquad z \in \mathbb{R}^d,$$

and consider the lifted segment

$$\Sigma = \{f(x) + tv : 0 \le t \le J\} \subseteq \mathbb{R}^d.$$

Its ℓ^1-length is

$$R = J\|v\|_1.$$

Your preceding calculation gives $|v_i| > c\varepsilon/L$ for at least one coordinate, and $J = k/c$. Consequently,

$$R > \frac{k\varepsilon}{L}.$$

Choose k large enough that $R \ge d + 1$. [Pasted text.txt, L362-L374]

Partition $\mathbb{R}^d$ into the cubes

$$z + [-1/2, 1/2)^d, \qquad z \in \mathbb{Z}^d.$$

The i-th coordinate of Σ crosses at most $J|v_i| + 1$ coordinate boundaries. Thus these boundaries divide Σ into at most

$$R + d + 1$$

pieces.

Their total ℓ^1-length is R, so one piece has ℓ^1-length at least

$$\frac{R}{R + d + 1} \ge \frac{1}{2}.$$

Its Euclidean length is therefore at least

$$\frac{1}{2\sqrt{d}}.$$

Taking the closure of this piece handles endpoints on cube boundaries without changing its length.

B. On that piece, the squared-distance function has definite oscillation

Let u, w be the endpoints of the piece's closure. On its closed cube, H agrees with squared Euclidean distance from the cube's center. Therefore,

$$H(u) - 2H\left(\frac{u+w}{2}\right) + H(w) = \frac{1}{2}\|u - w\|_2^2.$$

If Δ denotes the oscillation of H along that piece, the left-hand side is at most 2Δ. Hence

$$\Delta \geq \frac{1}{4}\|u-w\|_2^2 \geq \frac{1}{16d}.$$

In particular,

$$\boxed{\operatorname{osc}_\Sigma H \geq \frac{1}{16d}.}$$

The range of H along Σ is an interval. Since your shells correspond to successive intervals of width $1/K$ in ** H-value**, this range contains at least

$$\frac{K}{16d}-2$$

complete consecutive shell intervals. For $K \geq 64d$, there are at least two, and therefore shells of both colors. This also gives the stated $\Omega_d(K)$ conclusion.

C. The integer samples actually enter those shells

The useful estimate is

$$\boxed{|H(z)-H(z')| \leq \sum_{i=1}^{d} \|z_i - z_i'\|_{\mathbb{R}/\mathbb{Z}}.}$$

Indeed, the distance-to-the-nearest-integer function is 1-Lipschitz on the circle and takes values in $[0,1/2]$, so

$$\begin{aligned}\left|\|z_i\|_{\mathbb{R}/\mathbb{Z}}^2 - \|z_i'\|_{\mathbb{R}/\mathbb{Z}}^2\right| &= \left|\|z_i\|_{\mathbb{R}/\mathbb{Z}} - \|z_i'\|_{\mathbb{R}/\mathbb{Z}}\right| \left(\|z_i\|_{\mathbb{R}/\mathbb{Z}} + \|z_i'\|_{\mathbb{R}/\mathbb{Z}}\right) \\ &\leq \|z_i - z_i'\|_{\mathbb{R}/\mathbb{Z}}.\end{aligned}$$

Now choose a point on the path whose H-value is the midpoint of a completely crossed shell interval. Your d/M-density estimate supplies an integer sample whose H-value differs from that midpoint by at most d/M.

Thus the sample remains in the same shell whenever

$$\frac{d}{M} < \frac{1}{2K}.$$

Your choice $M = K^2$ satisfies this once $K > 2d$, so it certainly works with $K \geq 64d$.

This proves the sampling claim without needing an estimate for the Euclidean thickness of a shell. The relevant width is $1/K$ in H-value, and the Lipschitz estimate converts your sampling precision into control of that value.

Complete replacement for lines 420–421

The following incorporates the whole argument into the proof. It uses your existing notation and macros.

```
We take \(K\geq64d\), retain \(M=K^2\), and choose \(k\) to be a multiple of \(L\) such that \
(k\eps/L\geq d+1\).
Put
\[
H(z)=\sum_{i=1}^d\tnorm{z_i}^2,
\qquad
\Sigma=\{f(x)+tv:0\leq t\leq J\},
```

```
\qquad
R=J\|v\|_1.
\]
The preceding estimate gives \(R>k\eps/L\geq d+1\).
Partition \(\R^d\) into the cubes \(z+[-1/2,1/2)^d\), \(z\in\Z^d\).
The \(i\)-th coordinate of \(\Sigma\) crosses at most \(J|v_i|+1\) boundaries, so these boundaries
divide \(\Sigma\) into at most \(R+d+1\) pieces.
One piece therefore has \(\ell^1\)-length at least
\[
\frac{R}{R+d+1}\geq\frac12,
\]
and hence Euclidean length at least \(1/(2\sqrt d)\).
Let \(u,w\) be the endpoints of the closure of this piece.
On the corresponding closed cube, \(H\) is squared Euclidean distance from the cube's center, so
\[
H(u)-2H\!\left(\frac{u+w}{2}\right)+H(w)
=\frac12\|u-w\|_2^2.
\]
Consequently,
\[
\osc_{\Sigma}H\geq\frac14\|u-w\|_2^2\geq\frac1{16d}.
\]
Since the range of \(H\) along \(\Sigma\) is an interval, it contains at least \(K/(16d)-2\geq2\)
consecutive complete shell intervals of width \(1/K\), including shells of both colors.
Moreover, for \(z,z'\in\R^d\),
\[
|H(z)-H(z')|
\leq\sum_{i=1}^d\tnorm{z_i-z_i'}.
\]
Thus the \(d/M\)-density of the sampled orbit ensures that, for every midpoint of one of these shell
intervals, some sampled \(H\)-value is within \(d/M\) of it.
Since \(d/M=d/K^2<1/(2K)\), this sampled value lies in the same shell.
Therefore \(\{h(x),h(x+cy),\ldots,h(x+ky)\}\) receives both colors, and so does the original
progression \(\{x,x+y,\ldots,x+ky\}\).
```

These conditions are compatible with all the earlier choices. For example, when $0<\varepsilon<1/2$, one may take

$$K=64d,\qquad M=K^2,\qquad L=(M^d)!,\qquad k=L\left\lceil\frac{d+1}{\varepsilon}\right\rceil.$$

Thus the missing justification does not require changing the construction or the stated shell-coloring bound.

Compilation check

The updated source compiled successfully. The source-only run had **no overfull boxes, no unresolved internal cross-references, and no duplicate figure-destination warning**. I did not build the formatted bibliography from an updated `.bib` file, so this is not a verification of the final reference list.

Apart from the shell-crossing justification above, I have no further substantive correction to recommend in this pass.

347. User

2026-10-12 13:36:32 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-12T05:36:32.928000Z

Selected text:

Selection 1

The outstanding Bohr step — your point 2

My request:

we are choosing not to show this extra detail. Can you just confirm that the current choice of parameters is correct?

---

348. Assistant

2026-10-12 13:37:31 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-12T05:37:31.198680Z

Yes. I confirm that the current parameter choices are mathematically sufficient. In particular, the following explicit choices work for $0 < \varepsilon \leq 1/2$:

$$\boxed{\begin{aligned} K &= 64d, \\ M &= K^2, \\ L &= (M^d)!, \\ k &= L\left\lceil \frac{d+1}{\varepsilon} \right\rceil. \end{aligned}}$$

They satisfy all the requirements:

- **Divisibility:** Every $1 \leq c \leq M^d$ divides both L and k.
- **Sufficient path length:** $k\varepsilon/L \geq d+1$, which guarantees the required oscillation.
- **Enough shell crossings:** The oscillation is at least $1/(16d) = 4/K$, ensuring shells of both colors.
- **Fine enough sampling:** $d/M = d/K^2 < 1/(2K)$, so the integer samples detect both colors.

These choices also yield your stated quantitative bound

$$\boxed{k \leq \frac{\exp((Cd^2)^d)}{\varepsilon}}$$

for a sufficiently large absolute constant C. For $\varepsilon > 1/2$, the assertion is vacuous.

You can retain the current brief geometric argument without changing the parameters. I will also refrain from repeatedly flagging the omitted geometric details in future audits, unless I discover an actual mathematical error.

349. User

2026-10-12 13:41:14 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-12T05:41:14.732000Z

Please do one more full audit of the paper. If there are no more recommended changes, then give me a pdf file with an itemized list of every change I made to this paper from the first version that I uploaded to you this morning. This report is intended for Ryan so that he knows what I changed.

> **Attachments recorded in source:**
>
> - Pasted text(20261012-054035).txt

---

350. Assistant

2026-10-12 13:41:39 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-12T05:41:39.897534Z

I'll audit the latest draft, then compare it with this morning's first upload and prepare Ryan's change report if no further corrections are needed.

---

351. Assistant

2026-10-12 13:44:51 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-12T05:44:51.040805Z

The first upload today is `20261012-014643`; the latest is `20261012-054035`. Their core Section 3 proof is unchanged.

---

352. Assistant

2026-10-12 13:53:59 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-12T05:53:59.742883Z

I found no further corrections; the latest source compiles without overfull boxes. I'm finalizing Ryan's report, distinguishing net edits from temporary changes later removed.

---

353. Assistant

2026-10-12 13:57:57 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-12T05:57:57.216619Z

I found no further mathematical or grammatical changes to recommend. The main proof, constant-improvement lemma, extremal corollary, and deductions from the final question remain consistent in this audit. [Pasted text.txt, L952-L1035] [Pasted text.txt, L1212-L1275]

The latest source compiled successfully, with internal cross-references resolved, **no overfull boxes**, and no duplicate figure-destination warning. The updated `.bib` database was not included, so the final formatted bibliography was not part of this check.

Report for Ryan

Download the five-page revision report

The report compares `Pasted text(20261012-014643).txt`, your first upload this morning, with `Pasted text(20261012-054035).txt`, the latest upload. It contains **24 itemized net changes**, with before-and-after source line numbers, including mathematical wording, terminology, email address, source formatting, and figure adjustments. It also separately records the two temporary edits introduced and subsequently removed or corrected in the intermediate drafts.

Importantly, the comparison confirms that **the entire Section 3—including the outline and proofs—is unchanged between those two versions**. The report also distinguishes changes already present in this morning's first upload from changes made afterward, so Ryan will not be given an inflated account of today's revisions.

---

354. User

2026-10-12 19:19:39 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-12T11:19:39.463000Z

give me the msc classes for this paper for arxiv

---

355. Assistant

2026-10-12 19:20:22 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-10-12T11:20:22.483911Z

For your paper *2-large sets are sets of nil-Bohr recurrence*, I recommend the **MSC2020 classifications already listed in your manuscript**:

MSC code	Classification
05D10 (Primary)	Ramsey theory
37B20 (Secondary)	Recurrence in topological dynamical systems
11J54 (Secondary)	Small fractional parts of polynomials and generalizations
11B25 (Secondary)	Arithmetic progressions

These are all valid MSC2020 codes and appropriately reflect the paper's results. <Cite refs={["turn978825search24","turn822454search0","turn822454search1"]}/>

Copy and paste into arXiv's MSC classification field:

<box background="surface-secondary" radius="lg" padding={3} gap={2}> <text preserveWhitespace size="sm">05D10 (Primary), 37B20, 11J54, 11B25 (Secondary)</text> <row justify="end"> <button color="secondary" variant="outline" size="sm" onClick={()=>GenUI.copy("05D10 (Primary), 37B20, 11J54, 11B25 (Secondary)")}> <icon name="copy" size="xs"/> Copy </button> </row> </box>
I would retain **05D10 as primary**, since the main theorem concerns van der Waerden largeness and restricted arithmetic progressions.

Saved alternate messages

356. User

2026-09-16 19:04:28 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-16T11:04:28.210000Z

The applications, such as Theorem 2.2 of the pdf, seem nice. Can you do a literature search and make sure that Theorem 2.2 is not a trivial corollary of some other existing result in the literature, and that is 'genuinely needs' the work of my and Ryan to be obtained?

---

357. Assistant

2026-09-16 19:04:56 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-16T11:04:56.131397Z

I'll compare Theorem 2.2 with the closest coloring, recurrence, and approximation results, focusing on its factor-two transfer, unchanged progression length, and coefficient-uniform shrinking targets.

---

358. Assistant

2026-09-16 19:07:59 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-16T11:07:59.779953Z

I'm also testing a potentially simpler route: averaging translates of a large coordinate region, then asking whether classical quantitative equidistribution already yields the nested generalized-polynomial bounds.

---

359. Assistant

2026-09-16 19:17:03 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-16T11:17:03.413063Z

There is substantial overlap: classical quantitative Weyl theory already gives the same-length k^{-s} transfer for one monomial. The potentially novel content is simultaneous nested generalized-polynomial control.

---

360. Assistant

2026-09-16 19:19:29 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-16T11:19:29.136867Z

The literature check changes my assessment: I would not claim that Theorem 2.2 genuinely requires your new two-color theorem. I found an alternative derivation of the same coefficient-uniform extremal statement from **classical quantitative Weyl theory, averaging, and elementary triangular-matrix algebra**. In fact, the factor 2 can be replaced by any prescribed factor greater than 1, after adjusting the constants.

There are two different issues here. **I did not locate the exact extremal inequality explicitly stated in a published paper.** But that does not establish the stronger claim you asked about—that obtaining it genuinely needs your and Ryan's new result. The alternative proof below is a reason not to make that claim.

I should therefore revise my earlier presentation: **Theorem 2.2 is a useful consequence of your theorem, but it is not a convincing application demonstrating the necessity of the new two-color machinery.** The full nested-expression deduction is not literally a one-line corollary of an older theorem; nevertheless, it can be obtained without proving your binary-coloring result.

1. The precise statement being checked

Theorem 2.2 of the applications note concerns

$$B_k = \left\{ d \geq 1 : \|P_i(d)\| \leq Ck^{-d_i} \quad (1 \leq i \leq m) \right\},$$

where the P_i are special generalized polynomials of assigned weights $d_i \leq D$. It states

$$\boxed{r_k(\Omega; B_k) \leq 2r_k(\Omega)} \tag{1}$$

for every finite $\Omega \subseteq \mathbb{N}$ and every sufficiently large k, with C and the threshold depending only on m, D. The proof in the note partitions an extremal set according to your binary coloring.
[quantitative_nilrecurrence_applications.tex, L150-L203]

The alternative argument gives the following.

Proposition: an independent extremal bound

For every $m, D \geq 1$ and every $0 < \rho < 1$, there are constants

$$C = C(m, D, \rho), \qquad k_0 = k_0(m, D, \rho)$$

such that, for every family of special generalized polynomials $P_1, \ldots, P_m$ of weights $d_i \leq D$, the sets

$$\widetilde{B}_k = \left\{ d \geq 1 : \|P_i(d)\| \leq Ck^{-d_i} \quad \text{for every } i \right\}$$

satisfy

$$\boxed{r_k(\Omega; \widetilde{B}_k) \leq \rho^{-1} r_k(\Omega)} \tag{2}$$

for all $k \geq k_0$ and all finite $\Omega \subseteq \mathbb{N}$.

Taking $\rho = 1/2$ gives the form of Theorem 2.2. Taking $\rho = (1+\eta)^{-1}$ gives the factor $1 + \eta$.

The constants may be larger than those produced by your proof. This is an alternative derivation at the same level of parameter dependence, not a claim about identical numerical constants. Since neither theorem specifies numerical constants or claims optimality, that distinction does not provide a clear novelty advantage for the application.

The important older input is **Proposition 4.3 of Green–Tao, "The quantitative behaviour of polynomial orbits on nilmanifolds," Annals of Mathematics 175 (2012), 465–540**. Despite the paper's title, the proposition we need is an ordinary one-variable polynomial version of quantitative Weyl theory—not the full nilmanifold factorization theorem. Annals of Mathematics

2. What the literature already supplies

Here are the closest results I checked and how they bear on the question.

The scalar linear case is already an immediate consequence of older work. The half-circle coloring in Host–Kra–Maass, Lemma 7.4, and its effective formulation in your paper with Grytczuk give the $O(k^{-1})$ obstruction for αd. Splitting an arbitrary finite set according to that coloring gives the corresponding factor-two extremal inequality immediately. Your 2021 paper explicitly identifies its Theorem 2 as an effective version of the earlier lemma. Colgate Math

The familiar many-color localization results do not, by themselves, give the desired full statement. Host–Kra–Maass's Proposition 7.7 has a color cost depending on the dimension, while Proposition 7.9 gives a monomial obstruction whose number of colors depends on the tolerance. Applying those results mechanically to shrinking tolerances would lose the constant factor 2. arXiv

Coefficient-uniform recurrence for bracket polynomials is also older. Tointon's Theorem 4.10 proves recurrence estimates uniform over realizations of fixed bracket forms. His Section 6 discusses the upper-triangular matrix representation, including the exact 3×3 matrix representing $\alpha n[\beta n]$. These statements do not directly state your extremal inequality, but they show that coefficient-uniformity and this matrix encoding are not, in themselves, new features of the application. arXiv

Ordinary simultaneous Diophantine approximation results are not direct substitutes. For example, Maynard's theorem produces an approximating integer below a prescribed bound. It does not locate an arithmetic progression inside an arbitrary finite set, so it does not immediately yield (1). arXiv

The decisive issue is therefore not a direct quotation of one of these recurrence theorems. It is an **averaging argument that bypasses the need for a two-coloring**. I will give the derivation, including the nested-expression step, so the assessment does not rest on an unverified assertion that the result "should follow from standard methods."

3. The classical scalar ingredient

Write $\{x\} \in [0,1)$ for the ordinary fractional part. This notation is used only in the auxiliary argument; the brackets defining your special generalized polynomials will still mean nearest-integer rounding.

Fix D and $0 < \sigma < 1/2$, and put

$$I = [\sigma, 1-\sigma].$$

We need the following consequence of quantitative Weyl theory.

Lemma: polynomial lifting away from the boundary

There are integers Q, N_0, depending only on D, σ, such that the following holds.

Suppose $f \in \mathbb{R}[t]$ has degree at most D, $N \geq N_0$, and

$$\{f(n)\} \in I \qquad (0 \leq n < N).$$

Then there are an integer $1 \leq q \leq Q$ and an integer-valued polynomial $c(t)$ of degree at most D such that

$$\boxed{f(qt) + c(t) = \{f(qt)\} \in I} \tag{3}$$

for every integer

$$0 \leq t \leq \left\lfloor \frac{N-1}{q} \right\rfloor.$$

Here "integer-valued" means $c(\mathbb{Z}) \subseteq \mathbb{Z}$; its coefficients need not be integers.

Proof

The sequence $f(n) \bmod 1$ misses the fixed open arc around zero. A Lipschitz bump supported in that arc shows that this sequence fails a fixed level of equidistribution depending only on σ.

Write

$$f(t) = \sum_{j=0}^{D} \alpha_j \binom{t}{j}.$$

Green–Tao's Proposition 4.3 then supplies a bounded integer q_0 such that

$$\alpha_j = \frac{b_j}{q_0} + e_j, \qquad b_j \in \mathbb{Z}, \qquad |e_j| \ll_{D,\sigma} N^{-j} \quad (1 \le j \le D). \tag{4}$$

This is precisely the coefficient control encoded by their $C^\infty[N]$-norm. The harmless change from indices $1, \dots, N$ to $0, \dots, N-1$ only changes constants. arXiv

Set

$$q = q_0 D!,$$

and decompose

$$f(qt) = R(t) + E(t),$$

where

$$R(t) = \sum_{j=1}^{D} \frac{b_j}{q_0} \binom{qt}{j}, \qquad E(t) = \alpha_0 + \sum_{j=1}^{D} e_j \binom{qt}{j}.$$

The polynomial R is integer-valued. Indeed,

$$\frac{1}{q_0}\binom{qt}{j} = \frac{D!}{j!}\, t\,(qt-1)\cdots(qt-j+1) \in \mathbb{Z} \qquad (t \in \mathbb{Z}). \tag{5}$$

On the real interval $0 \le t \le (N-1)/q$, equation (4) gives

$$|E'(t)| \ll_{D,\sigma} N^{-1}.$$

For sufficiently large N,

$$|E(t+1) - E(t)| < 2\sigma. \tag{6}$$

At every relevant integer t, the value $E(t)$ belongs to

$$\bigcup_{z \in \mathbb{Z}} [z + \sigma, z + 1 - \sigma],$$

because $R(t)$ is integral and $\{f(qt)\} \in I$. Distinct components of this union are separated by gaps of length 2σ. By (6), consecutive samples cannot move between components.

Thus there is one integer z such that

$$E(t) - z \in I$$

at all the relevant samples. Taking

$$c(t) = -R(t) - z$$

proves (3). □

The point is that avoiding a fixed boundary strip lets us replace the lattice corrections by one ordinary integer-valued polynomial after passing to a bounded-step subprogression.

4. The same argument works entry by entry for the matrices encoding SGPs

This is the part needed to check the *full* statement, rather than only ordinary monomials.

Fix the weight patterns of $P_1, \dots, P_m$, and use the product G of upper-unitriangular matrix groups from your draft, with integer lattice Γ. If the adjacent weights in a block are $w_1, \dots, w_\ell$, set

$$w_{ij} = w_i + \cdots + w_{j-1}.$$

Let $\mathcal{H}$ consist of polynomial matrices whose (i,j)-entry has degree at most w_{ij}. These are the same elementary weighted matrices used in the shorter proof. [Pasted text.txt, L77-L107]

Every right coset modulo Γ has a unique representative whose off-diagonal entries lie in $[0,1)$. Let r be the total number of these entries, and let

$$K_\sigma \subseteq G/\Gamma$$

be the set whose canonical representatives have **every** off-diagonal entry in $I = [\sigma, 1-\sigma]$.

Matrix lifting claim

There are Q_*, N_*, depending only on the weights and σ, such that if

$$p(n)\Gamma \in K_\sigma \qquad (0 \le n < N), \quad p \in \mathcal{H},$$

and $N \ge N_*$, then some $1 \le q \le Q_*$ admits polynomial matrices $C, R \in \mathcal{H}$ satisfying

$$\boxed{p(qt)C(t) = R(t), \qquad C(\mathbb{Z}) \subseteq \Gamma,} \tag{7}$$

where $R(t)$ is the canonical representative at all integer samples

$$0 \le t < J, \qquad J = \left\lfloor \frac{N-1}{q} \right\rfloor + 1.$$

In particular,

$$|R_{ij}(t)| \le 1$$

at these samples.

Proof of the claim

Process entries in increasing matrix gap $j - i$. Suppose we have reached a subsequence $p(st)$, and the corrections $C_{kj}(t)$, $i < k < j$, have already been constructed as integer-valued polynomials agreeing with the canonical lattice corrections on this subsequence.

The next unreduced coordinate is the **ordinary polynomial**

$$f_{ij}(t) = p_{ij}(st) + \sum_{i<k<j} p_{ik}(st) C_{kj}(t). \tag{8}$$

Its degree is at most w_{ij}, because the weights add. Its fractional part is the canonical (i,j)-coordinate, so it belongs to I at every current sample.

Apply the scalar lemma to f_{ij}. Pass to the resulting bounded-step subprogression, compose all previously constructed corrections with that dilation, and use the lemma's integer-valued correction for C_{ij}.

The earlier coordinates remain correct after restriction. Repeating this for all r coordinates proves (7). There are only r applications of the scalar lemma, so the accumulated step is bounded, for example by Q^r. Taking N_* sufficiently large ensures that every intermediate sequence remains long enough.

No two-coloring, shell function, or nilmanifold factorization theorem has been used.

Extracting the weighted leading coefficients

Define

$$\tau(A)_{ij} = [t^{w_{ij}}]A_{ij}(t),$$

with diagonal entries equal to one, and let δ_u multiply entry (i,j) by $u^{w_{ij}}$. The additive weights give

$$\tau(AB) = \tau(A)\tau(B), \qquad \tau(A(qt)) = \delta_q\tau(A), \qquad \tau(v) = I \tag{9}$$

for every constant matrix v.

Since C_{ij} is integer-valued,

$$w_{ij}!\,\tau(C)_{ij} \in \mathbb{Z}. \tag{10}$$

This follows by taking its w_{ij}-th finite difference.

Fix, **before the coefficients or N are specified**,

$$L = D!\,\mathrm{lcm}(1,\dots,Q_*). \tag{11}$$

Applying τ to (7), and then dilating by L/q, gives

$$\boxed{\delta_L\tau(p)\,\gamma = E, \qquad \gamma \in \Gamma,} \tag{12}$$

where

$$\gamma = \delta_{L/q}\tau(C), \qquad E = \delta_{L/q}\tau(R).$$

Integrality follows from (10), since L/q is a multiple of $D!$.

The coordinates of R are bounded at J consecutive integers. Ordinary polynomial interpolation therefore gives

$$|\tau(R)_{ij}| \ll_D J^{-w_{ij}}.$$

Consequently,

$$|E_{ij}| \ll_D \left(\frac{L}{qJ}\right)^{w_{ij}} \ll_{D,L} N^{-w_{ij}}, \tag{13}$$

using $qJ \geq N-1$.

This is the required weight-dependent estimate, obtained from the older scalar Weyl input.

5. Averaging replaces the binary coloring

Here is why the preceding confinement statement is sufficient for the extremal application.

For

$$P(t) = a_1 t^{w_1}\left[a_2 t^{w_2}[\cdots[a_\ell t^{w_\ell}]\cdots]\right],$$

use the band matrix $g(t)$ whose first superdiagonal is

$$-a_1 t^{w_1}, \ldots, -a_{\ell-1} t^{w_{\ell-1}}, +a_\ell t^{w_\ell},$$

with all higher entries zero. For the family, take the product of these matrices.

The 3×3 instance of this encoding already appears explicitly in Tointon's treatment of the Bergelson–Leibman representation. For the special nested expressions here, the general encoding is checked directly by the same last-column recursion. arXiv

Choose σ sufficiently small that

$$\mu(K_\sigma) = (1-2\sigma)^r \geq \rho, \tag{14}$$

where μ is Haar probability on G/Γ.

The measure formula is elementary in these matrix groups: off-diagonal matrix coordinates give Lebesgue Haar measure, and their unit cube is a fundamental domain for the integer lattice.

For any finite $A \subseteq \mathbb{N}$ and $x \in G/\Gamma$, define

$$A_x = \{n \in A : g(n/L)x \in K_\sigma\}. \tag{15}$$

Every left translation preserves Haar measure, so

$$\boxed{\int_{G/\Gamma} |A_x|\, d\mu(x) = \mu(K_\sigma)|A| \geq \rho|A|.} \tag{16}$$

We now show that every sufficiently long progression in **any** A_x has the desired difference.

Suppose

$$a, a+d, \ldots, a+(k-1)d \in A_x,$$

and write $x = v\Gamma$. Apply (12)–(13) to

$$p(t) = g\big((a+dt)/L\big)v.$$

The constant right factor disappears under τ, and weighted homogeneity gives

$$\delta_L \tau(p) = g(d).$$

Therefore

$$g(d)\gamma = E, \qquad |E_{ij}| \ll k^{-w_{ij}}. \tag{17}$$

For sufficiently large k, every off-diagonal error is less than $1/4$. The last-column equations then identify the nearest-integer roundings uniquely.

For example, for $P(d) = \alpha d[\beta d]$,

$$g(d) = \begin{pmatrix} 1 & -\alpha d & 0 \\ 0 & 1 & \beta d \\ 0 & 0 & 1 \end{pmatrix}.$$

The last-column equations are

$$\beta d + \gamma_{23} = E_{23}, \qquad \gamma_{13} - \alpha d\,\gamma_{23} = E_{13}.$$

The first gives $\gamma_{23} = -[\beta d]$, so the second becomes

$$\alpha d[\beta d] + \gamma_{13} = E_{13}.$$

Hence

$$\|\alpha d[\beta d]\| \le |E_{13}| \ll k^{-2}.$$

For a longer nested expression, the same bottom-to-top recursion gives

$$P_i(d) + z_i = E_{\mathrm{top},i}, \qquad z_i \in \mathbb{Z},$$

and thus

$$\boxed{\|P_i(d)\| \ll k^{-d_i}.} \tag{18}$$

This agrees with the arithmetic decoding in your draft, but the **confinement input has been obtained from scalar Weyl theory rather than the two-color theorem**. [Pasted text.txt, L325-L349]

Finally, suppose $A \subseteq \Omega$ contains no k-term progression with difference in $\widetilde{B}_k$. Equation (18) implies that every A_x is entirely k-AP-free. Therefore

$$|A_x| \le r_k(\Omega) \qquad \text{for every } x.$$

Integrating and using (16),

$$\rho|A| \le \int |A_x|\, d\mu(x) \le r_k(\Omega).$$

Taking the maximum over A proves

$$r_k(\Omega; \widetilde{B}_k) \le \rho^{-1} r_k(\Omega).$$

There are only finitely many weight patterns for m SGPs of weights at most D. Maximizing the constants and thresholds over those patterns gives exactly the asserted dependence on m, D, ρ. □

6. Why this does not recover your main theorem

The distinction is substantial.

Your theorem constructs **one binary coloring** of all the integers such that *both* color classes exclude all bad progressions, simultaneously over the relevant lengths. That is the input required to apply an arbitrary $(2, k)$-largeness hypothesis. [quantitative_nilrecurrence_applications.tex, L134-L148]

The alternative proof constructs a **family of overlapping good subsets** A_x. Averaging shows that one can retain a large proportion of any given finite set, but it does not partition the integers into two good classes.

In particular, it gives no reason why the complement of a chosen good class should also be good.

That is exactly why the alternative argument suffices for an **extremal density bound**, yet does not prove

$$(2,k)\text{-large} \quad \Longrightarrow \quad \text{the prescribed approximation.}$$

Another way to phrase the distinction is:

> **Averaging can replace a coloring in the extremal application, but not in the restricted two-color Ramsey theorem.**

Thus this finding does **not** show that your principal theorem is an old result in disguise. It shows that Theorem 2.2 discards enough of the principal theorem's content that an older, weaker mechanism becomes sufficient.

7. What I would change in the applications note and the paper's positioning

I would make three revisions.

First, remove the suggestion that Theorem 2.2 demonstrates a genuinely new consequence unavailable from the classical toolkit. Its proof from your theorem is exceptionally short and pleasant, but that is a statement about the efficiency of *your proof of the corollary*, not the necessity of your main theorem for obtaining it.

Second, retain it only with an appropriate description. For example:

> "The quantitative coloring theorem gives an immediate same-length extremal comparison for generalized-polynomial constraints. Related comparisons can also be obtained by quantitative equidistribution and averaging; the coloring formulation additionally applies to arbitrary finitely large sets of allowed differences."

The first sentence can be sharpened or the second expanded after writing out the alternative argument in a permanent note. I would not attribute the precise extremal inequality to Green–Tao or Tointon: I have not found it explicitly stated there.

Third, distinguish the applications according to their hypotheses. The finite Szemerédi and sparse-ambient consequences downstream of Theorem 2.2 inherit the same alternative route. By contrast, statements that actually start with

$$S \text{ is } (2,k)\text{-large}$$

or

$$S \text{ is } (2r,k)\text{-large}$$

are not supplied by this averaging argument. The finite Ramsey localization and polynomial-length return results therefore need a separate novelty audit; this investigation does not establish that they are old or new.

Bottom line

I cannot support the claim that Theorem 2.2 "genuinely needs" your and Ryan's new theorem. The decisive reason is not merely that closely related literature exists: the argument above recovers its full SGP extremal conclusion from an older quantitative Weyl theorem and elementary algebra, with the same type of coefficient-uniform constants and no increase in progression length.

I would still call it a useful corollary. I would **not** use it as a principal argument for expanding the paper around the quantitative generalization.

The result whose distinctive content survives this check is the **binary-coloring statement itself—and applications that truly need its control of arbitrary sets of allowed differences**, rather than only a density estimate in an ambient finite set.

---

361. User

2026-09-16 19:22:56 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-16T11:22:56.443000Z

if we use the fact that delta large for some delta in (0,1) implies nilbohr recurrence, then we can also improve the constant from 2 to 1+epsilon, right?

---

362. Assistant

2026-09-16 19:23:34 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-16T11:23:34.146782Z

Yes—the factor $1+\varepsilon$ should follow, but one must distinguish the omitted-color quantitative argument from qualitative density-to-nilrecurrence, which only handles fixed nil-Bohr targets.

---

363. Assistant

2026-09-16 19:26:31 (Asia/Shanghai, UTC+08:00)
Source timestamp: 2026-09-16T11:26:31.501781Z

Yes. Your fixed-density strengthening gives exactly the $1+\varepsilon$ improvement, and I should have pointed out this much shorter route.

There is one distinction to retain:

The qualitative implication gives the improvement for each fixed nil-Bohr neighborhood, for all sufficiently large progression lengths. The quantitative omitted-color theorem gives the improvement for the shrinking targets of Theorem 2.2, at the same progression length k.

Here are both deductions.

1. The quantitative improvement is an immediate counting argument

Your fixed-density note contains the following strengthening: for any prescribed $r \geq 2$, there is an r-coloring such that **every sufficiently long progression omitting even one color** has difference d satisfying

$$\|P_i(d)\| \leq C(r,m,D)k^{-d_i}.$$

The constants are independent of the coefficients, and one coloring works for all sufficiently large k. [fixed_density_nilrecurrence.tex, L75-L99]

That immediately improves the extremal constant from 2 to

$$\frac{r}{r-1}.$$

Proposition

For every $\varepsilon > 0$ and $m, D \geq 1$, there are constants

$$C = C(\varepsilon, m, D), \qquad k_0 = k_0(\varepsilon, m, D)$$

such that, for every family of special generalized polynomials $P_1, \ldots, P_m$ of assigned weights $d_i \leq D$, the sets

$$B_{k,\varepsilon} = \left\{ d \in \mathbb{N} : \|P_i(d)\| \leq C k^{-d_i} \quad (1 \leq i \leq m) \right\}$$

satisfy

$$\boxed{r_k(\Omega) \leq r_k(\Omega; B_{k,\varepsilon}) \leq (1+\varepsilon) r_k(\Omega)} \tag{1}$$

for every finite $\Omega \subseteq \mathbb{N}$ and every $k \geq k_0$.

Proof

Choose

$$r = \left\lceil \frac{1}{\varepsilon} \right\rceil + 1,$$

so that

$$\frac{r}{r-1} \leq 1 + \varepsilon.$$

Let

$$c : \mathbb{N} \longrightarrow \{1, \ldots, r\}$$

be the omitted-color coloring for the prescribed family.

Suppose $A \subseteq \Omega$ contains no k-term progression with difference in $B_{k,\varepsilon}$. For each j, put

$$A_j = \{ n \in A : c(n) \neq j \}.$$

Each A_j is k-AP-free. Indeed, a k-term progression in A_j omits color j, so the omitted-color theorem puts its difference in $B_{k,\varepsilon}$. Since it also lies in A, this contradicts our assumption.

Consequently,

$$|A_j| \leq r_k(\Omega) \qquad (1 \leq j \leq r).$$

Every element of A belongs to exactly $r-1$ of these sets, so

$$(r-1)|A| = \sum_{j=1}^{r} |A_j| \leq r \, r_k(\Omega).$$

Thus

$$|A| \leq \frac{r}{r-1} r_k(\Omega) \leq (1+\varepsilon) r_k(\Omega).$$

Taking the maximum over A proves the upper bound. The lower bound is immediate. □

So **the omitted-color theorem gives the desired improvement without any new analytic or geometric argument**.

The qualification about constants matters: C and k_0 may now depend on ε. We are not claiming the improvement for every ε while leaving the original target constant $C(m, D)$ unchanged.

2. Even the qualitative fixed-density implication gives a $1+\varepsilon$ bound for fixed targets

Your proposed reasoning also works directly from the qualitative statement, through a short translation-averaging lemma.

Translation-averaging lemma

Suppose $E \subseteq \mathbb{N}$ has upper density

$$\overline{d}(E) = \rho > 0$$

and every k-term progression in E has difference in B. Then

$$\boxed{r_k(\Omega; B) \le \rho^{-1} r_k(\Omega)} \tag{2}$$

for every finite $\Omega \subseteq \mathbb{N}$.

Proof

Let $A \subseteq \Omega$ contain no k-term progression with difference in B. For each $t \ge 0$, define

$$A_t = A \cap (E - t) = \{a \in A : a + t \in E\}.$$

Every A_t is k-AP-free: translating a progression in A_t by t puts it in E, without changing its difference. That difference would therefore belong to B, contradicting the choice of A. Hence

$$|A_t| \le r_k(\Omega) \qquad (t \ge 0). \tag{3}$$

On the other hand,

$$\begin{aligned} \frac{1}{N}\sum_{t=0}^{N-1} |A_t| &= \sum_{a \in A} \frac{|E \cap [a, a+N-1]|}{N} \\ &= |A| \frac{|E \cap [1, N]|}{N} + o(1), \end{aligned} \tag{4}$$

because A is fixed and finite.

Taking $N \to \infty$ along a subsequence realizing $\overline{d}(E) = \rho$, equations (3)–(4) give

$$\rho |A| \le r_k(\Omega).$$

This proves (2). □

Notice that **natural upper density is enough**. We do not need ordinary density, uniform density, or equidistribution.

Apply the qualitative theorem

Fix a finite-step nil-Bohr $_0$ set B and set

$$\rho = \frac{1}{1+\varepsilon}.$$

Since $\mathbb{N} \setminus B$ is not nil-Bohr recurrent, the implication

$$\rho\text{-density large} \Longrightarrow \text{nil-Bohr recurrent}$$

shows that $\mathbb{N} \setminus B$ is not ρ-density large.

Unpacking that negation, there are $E \subseteq \mathbb{N}$ and K such that

$$\overline{d}(E) = \rho$$

and every K-term progression in E has difference in B. This is also precisely the high-density nil-Bohr obstruction recorded in the fixed-density note. [fixed_density_nilrecurrence.tex, L120-L131]

Every longer progression in E has the same property, by considering its first K terms. Applying the averaging lemma at each $k \geq K$, we obtain

$$\boxed{r_k(\Omega; B) \leq (1+\varepsilon) r_k(\Omega) \qquad (k \geq K(B, \varepsilon)).} \tag{5}$$

This is the direct formal consequence of the qualitative theorem that you suggested.

3. What distinguishes the qualitative and quantitative versions

The only missing information in (5) is control of the threshold $K(B, \varepsilon)$.

For a **fixed** nil-Bohr neighborhood B, equation (5) is exactly what we want. But for

$$B_k = \{d : \|P_i(d)\| \leq C k^{-d_i}\},$$

the target itself changes with k. Applying (5) separately to B_k gives a threshold $K(B_k, \varepsilon)$, with no assurance that

$$k \geq K(B_k, \varepsilon).$$

That is why the qualitative implication alone does not automatically reproduce the same-length shrinking-target statement of Theorem 2.2.

The quantitative fixed-density note supplies the missing uniformity: it constructs one set E of any prescribed upper density $\rho < 1$, on which every sufficiently long progression satisfies the bounds Ck^{-d_i}, with coefficient-independent constants. Applying the translation-averaging lemma to that E gives (1) as well. [fixed_density_nilrecurrence.tex, L44-L58] The omitted-color counting proof above is simply an even shorter way to obtain the same conclusion.

Thus you are right: once we use your fixed-density strengthening, the factor 2 should not be presented as the best extremal conclusion. The natural formulation is $1+\varepsilon$, with ε-dependent constants and threshold. This gives a much simpler deduction within your framework than the independent argument in my previous answer; it does not, by itself, settle the separate question of the corollary's novelty.