Density-One Sequences with Distinct Consecutive Products

Przemyslaw Chojecki

January 16, 2026

Abstract

We study the problem of constructing an increasing sequence $1 \le d_1 < d_2 < \cdots$ of asymptotic density 1 such that all consecutive-block products $\prod_{u \le i \le v} d_i$ are distinct. We present a dyadic deletion scheme that reduces the problem to a quantitative bound on integer solutions to certain separated-variable equations $\prod_i (x - a_i) = \prod_j (y + b_j)$ with degrees $\asymp \log X$ and shifts $|a_i|, |b_j| \ll \log X$. Under a uniform solution-count hypothesis for these equations (Hypothesis 1), the scheme yields a complete density-one construction.

1 Introduction

Let $d_1 < d_2 < \cdots$ be a strictly increasing sequence of positive integers. We say that (d_i) has the *distinct consecutive products* property if

$$\prod_{u \le i \le v} d_i \ne \prod_{u' \le i \le v'} d_i \qquad \text{whenever } (u, v) \ne (u', v'). \tag{1}$$

The guiding question is whether one can achieve (1) with *asymptotic density* 1, i.e.

$$\#\{d_i \le N\} = N - o(N) \qquad (N \to \infty).$$

Equivalently, writing prefix products $D_n := \prod_{i \le n} d_i$ (with $D_0 := 1$), (1) asserts that all ratios D_v/D_{u-1} are distinct.

A collision in (1) can always be arranged into the form

$$d_{M-\ell+1} \cdots d_M \ = \ d_N \cdots d_{N+k} \qquad \text{with } M < N \text{ and hence } \ell > k, \tag{2}$$

after canceling common factors and ordering endpoints. The present manuscript builds a dense set by deleting a sparse hitting set for all collisions (2), in a manner stable under future deletions.

2 Dyadic deletion scheme and local deficiency

We construct a set $A \subset \mathbb{N}$ and let (d_i) be the increasing enumeration of A. The construction proceeds in dyadic stages $[X, 2X)$ for $X = 2^t$. At stage X we may delete a (small) set $B_X \subset [X, 2X) \cap \mathbb{Z}$ and set

$$A := \mathbb{N} \setminus \bigcup_{t \ge 1} B_{2^t}.$$

The sequence (d_i) is then determined.

2.1 Local deficiency

Fix an absolute constant $C_0 \geq 10$. For a dyadic X define the *window length* $L(X) := \lceil C_0 \log X \rceil$ and a *deficiency threshold* $T(X) := \lceil (\log X)^{2/3} \rceil$.

Definition 1 (Local deficiency invariant)**.** We say that the deletions satisfy the local deficiency invariant at scale X if every integer interval $I \subset [X, 2X)$ of length $|I| = L(X)$ obeys

$$\#\Big(I \cap \bigcup_{Y \leq X} B_Y\Big) \leq T(X).$$

The point is that, if the invariant holds, then along the sequence (d_i) the local gap patterns over index-length $\leq L(X)$ have very limited combinatorial freedom.

2.2 Pattern counting

For a strictly increasing sequence (d_i) and an index m, define backward gaps

$$a_i(m) := d_m - d_{m-i} \qquad (0 \leq i \leq r),$$

and forward gaps

$$b_j(n) := d_{n+j} - d_n \qquad (0 \leq j \leq s),$$

with $a_0 = b_0 = 0$.

Lemma 1 (Pattern bound from local deficiency)**.** *Assume the local deficiency invariant holds at scale X. Fix any $r \leq L(X)$. Then for m with $d_m \in [X, 2X)$ the number of possible vectors $(a_1(m), \ldots, a_r(m))$ is at most $X^{o(1)}$ as $X \to \infty$. The same holds for forward patterns $(b_1(n), \ldots, b_s(n))$ with $s \leq L(X)$.*

Proof. If $d_m \in [X, 2X)$ then the integers between d_{m-r} and d_m form an interval of length at most $r + \#\big([d_{m-r}, d_m] \cap \cup_{Y \leq X} B_Y\big)$. By local deficiency, the number of missing integers in any interval of length $L(X)$ is at most $T(X)$, hence for $r \leq L(X)$ we have $a_i(m) \leq i + T(X)$. Thus the pattern $(a_1, \ldots, a_r)$ is determined by the set of positions where gaps exceed 1, equivalently by the subset of omitted integers in the window. There are at most $T(X)$ omissions and at most $O(\log X)$ places, hence

$$\#\text{patterns} \leq \sum_{j \leq T(X)} \binom{L(X) + j}{j} \leq (C \log X)^{T(X)} = \exp\big(T(X) \log\log X + O(T(X))\big) = X^{o(1)}.$$

□

3 Collisions as separated-variable equations

Fix a collision (2) and set

$$x := d_M, \qquad y := d_N.$$

Define the local gap patterns

$$a_i := d_M - d_{M-i} \quad (1 \leq i \leq \ell - 1), \qquad b_j := d_{N+j} - d_N \quad (1 \leq j \leq k).$$

Then (2) is equivalent to

$$f_{\mathbf{a}}(x) = g_{\mathbf{b}}(y), \tag{3}$$

where

$$f_{\mathbf{a}}(X) := \prod_{i=0}^{\ell-1}(X - a_i), \qquad g_{\mathbf{b}}(Y) := \prod_{j=0}^{k}(Y + b_j), \qquad (a_0 = b_0 = 0).$$

This is a separated-variable equation $f(x) = g(y)$ with degrees $\deg f = \ell$ and $\deg g = k + 1$.

3.1 A regime where one can count solutions unconditionally

The following lemma gives a fully explicit bound when the degrees match. It is included both as evidence and as a model for the uniform hypothesis needed later.

Lemma 2 (Near-diagonal bound when degrees match)**.** *Let $d \geq 2$ and let*

$$f(x) = \prod_{i=0}^{d-1}(x - a_i), \qquad g(y) = \prod_{j=0}^{d-1}(y + b_j)$$

with $a_0 = b_0 = 0$ and $|a_i|, |b_j| \leq H$. Assume $X \geq 100H$ and $(x, y) \in [X, 2X]^2 \cap \mathbb{Z}^2$ satisfies $f(x) = g(y)$. Then $|x - y| \leq CH$ (absolute C), and consequently

$$\#\{(x, y) \in [X, 2X]^2 \cap \mathbb{Z}^2 : \; f(x) = g(y)\} \ll dH.$$

Moreover, if for some integer t the polynomial identity $f(x) \equiv g(x + t)$ does not *hold, then for that fixed t the equation $f(x) = g(x + t)$ has at most $d - 1$ integer solutions in x.*

Proof. Write

$$\frac{f(x)}{x^d} = \prod_{i=0}^{d-1}\Big(1 - \frac{a_i}{x}\Big), \qquad \frac{g(y)}{y^d} = \prod_{j=0}^{d-1}\Big(1 + \frac{b_j}{y}\Big).$$

Since $x, y \geq X \geq 100H$, we have $|a_i/x|, |b_j/y| \leq 1/100$ and hence $|\log(1 + u)| \leq 2|u|$ for such u. Thus

$$\big|\log f(x) - d \log x\big| = \Big|\sum_{i=0}^{d-1} \log\Big(1 - \frac{a_i}{x}\Big)\Big| \leq \sum_{i=0}^{d-1} \frac{2|a_i|}{x} \leq \frac{2dH}{X},$$

and similarly $|\log g(y) - d \log y| \leq 2dH/X$. From $f(x) = g(y)$ we obtain

$$d|\log x - \log y| \leq \frac{4dH}{X}, \quad \text{hence} \quad |\log(x/y)| \leq \frac{4H}{X}.$$

Since $x/y \in [1/2, 2]$ we have $|\log(x/y)| \asymp |x - y|/X$, giving $|x - y| \ll H$.

Now write $y = x + t$ with $|t| \ll H$. For each fixed t, the polynomial $h_t(x) := f(x) - g(x + t)$ has degree at most $d - 1$ (leading terms cancel). If $h_t \not\equiv 0$ then it has at most $d - 1$ roots in $\mathcal{C}$, hence at most $d - 1$ integer roots x. The number of possible t is $O(H)$, yielding the bound $O(dH)$ total solutions. □

3.2 The uniform solution-count input in the $d \asymp \log X$ regime

In the collision problem, the degrees need not match exactly; however, heuristics and partial arguments suggest that any solutions with $x, y \asymp X$ force $\ell - (k+1) = O(1)$ and that $|a_i|, |b_j| \ll \log X$ for realizable patterns. To close the density-one construction one needs a *uniform* bound, polynomial in $\log X$, on the number of integer solutions to (3) for *all* realizable patterns at scale X.

We isolate this as a hypothesis.

Hypothesis 1 (Uniform polylogarithmic solution bound)**.** *There exist absolute constants $C_1, C_2 \geq 1$ such that the following holds. Let X be large, let $H := C_1 \log X$, and let ℓ, k satisfy $1 \leq k \leq \ell \leq C_1 \log X$ and $|\ell - (k+1)| \leq C_1$. For any integer shifts $a_0 = 0, a_1, \ldots, a_{\ell-1}$ and $b_0 = 0, b_1, \ldots, b_k$ with $|a_i|, |b_j| \leq H$, the number of integer solutions $(x, y) \in [X, 2X]^2$ to*

$$\prod_{i=0}^{\ell-1}(x - a_i) = \prod_{j=0}^{k}(y + b_j)$$

is at most $(\log X)^{C_2}$, unless *the equation has an "algebraic degeneracy" of the following form: there exists an integer affine change of variables $(x, y) \mapsto (ux+v, u'y+v')$ and an integer t such that, after this change, one side becomes identically equal to the other with a translate, i.e. $f(x) \equiv g(x+t)$. In the degenerate case, the pattern corresponds to an explicit affine equivalence of the root multisets $\{a_i\}$ and $\{-b_j\}$.*

Remark 1. Lemma 2 verifies Hypothesis 1 in the special case $\ell = k+1$ (equal degrees), with a bound $O(dH) = O((\log X)^2)$ and with degeneracy precisely $f \equiv g(\cdot + t)$. The additional cases $|\ell - (k+1)| = O(1)$ are expected to admit similar reductions to finitely many translate/dilate relations. A natural route to proving the degeneracy classification in full generality is via Bilu–Tichy (classification of $f(x) = g(y)$ with infinitely many integer solutions).

4 Main density-one construction (conditional)

We now show that Hypothesis 1, together with the local deficiency invariant, yields a full solution.

Theorem 1 (Conditional density-one existence)**.** *Assume Hypothesis 1. Then there exists an increasing sequence $1 \leq d_1 < d_2 < \cdots$ of asymptotic density 1 such that all consecutive-block products $\prod_{u \leq i \leq v} d_i$ are distinct.*

4.1 Outline of the deletion at scale X

Fix a dyadic scale X. Suppose inductively that deletions B_Y have been chosen for all dyadic $Y < X$ so that:

(I) the local deficiency invariant holds at all earlier scales;

(II) there are no collisions (2) whose all entries lie below X.

We will choose a set $B_X \subset [X, 2X)$ such that after deleting B_X:

(I′) the local deficiency invariant holds at scale X;

(II′) there are no collisions (2) whose all entries lie below $2X$.

4.2 Counting collisions at scale X

Let $\mathcal{E}_X$ be the family of collisions (2) in the partially constructed sequence (i.e. using all integers not yet deleted below $2X$) such that the maximum element appearing in either side lies in $[X, 2X)$. We further restrict to the *relevant* collisions where the block lengths satisfy $\ell, k \leq L(X)$; collisions with larger block lengths can be dispatched by separate arguments (e.g. via large-prime-factor preparation or a random thinning), but we omit that discussion here.

For each collision $E \in \mathcal{E}_X$, define its *vertex set* $V(E)$ to be the set of integers in $[X, 2X)$ that appear among the two blocks in (2) (the size satisfies $|V(E)| \ll L(X)$).

Group collisions by their realized local patterns $(\mathbf{a}, \mathbf{b})$ and by the integer solution (x, y) to (3) with $x, y \in [X, 2X)$. By Lemma 1, the number of realizable patterns at scale X is $X^{o(1)}$. Under Hypothesis 1, for each such pattern the number of solutions (x, y) in $[X, 2X)^2$ is at most $(\log X)^{C_2}$ (except for degenerate patterns, handled separately by deleting an additional zero-density set). Hence

$$|\mathcal{E}_X| \;\le\; X^{o(1)}(\log X)^{C_2} \;=\; X^{o(1)} \;=\; o(X). \tag{4}$$

4.3 Choosing a hitting set while preserving local deficiency

We choose $B_X \subset [X, 2X)$ as a hitting set for $\mathcal{E}_X$:

$$\forall E \in \mathcal{E}_X, \quad B_X \cap V(E) \neq \emptyset.$$

Additionally we require that for every interval $I \subset [X, 2X)$ of length $L(X)$,

$$|B_X \cap I| \le T(X)/2, \tag{5}$$

so that local deficiency remains bounded after adding earlier deletions.

Lemma 3 (Capacity-constrained hitting set)**.** *Let $\mathcal{E}_X$ be a family of subsets $V(E) \subset [X, 2X)$ with $|V(E)| \le L(X)$. If $|\mathcal{E}_X| = o(X)$, then there exists a hitting set $B_X \subset [X, 2X)$ satisfying* (5)*.*

Proof. We give a probabilistic proof. For each $E \in \mathcal{E}_X$, choose an element $v(E) \in V(E)$ uniformly at random and set $B_X := \{v(E) : E \in \mathcal{E}_X\}$. Then B_X is a hitting set by construction.

Fix an interval $I \subset [X, 2X)$ with $|I| = L(X)$. For each E,

$$\mathbb{P}(v(E) \in I) \le \frac{|V(E) \cap I|}{|V(E)|} \le \frac{L(X)}{1} \cdot \frac{1}{1} \quad \text{but more crudely} \quad \mathbb{P}(v(E) \in I) \le \frac{|I|}{X} = \frac{L(X)}{X}.$$

Hence

$$\mathbb{E}|B_X \cap I| \le \sum_{E \in \mathcal{E}_X} \mathbb{P}(v(E) \in I) \le |\mathcal{E}_X| \frac{L(X)}{X} = o(1) \cdot \log X = o(T(X)).$$

Since $|B_X \cap I|$ is a sum of independent Bernoulli variables (one per E), a Chernoff bound implies that $\mathbb{P}(|B_X \cap I| > T(X)/2) \le \exp(-cT(X))$ for some $c > 0$ and large X.

There are at most $2X$ intervals $I \subset [X, 2X)$ of length $L(X)$. By a union bound,

$$\mathbb{P}\big(\exists I : \; |B_X \cap I| > T(X)/2\big) \le 2X \exp(-cT(X)) \to 0 \quad (X \to \infty),$$

since $T(X) = (\log X)^{2/3} \to \infty$. Thus with positive probability (5) holds, proving existence of a suitable B_X. □

4.4 Completion of the conditional proof

Proof of Theorem 1. We carry out the dyadic construction. First delete a fixed zero-density set to eliminate all degenerate patterns from Hypothesis 1 (e.g. delete all integers in a sparse sequence of intervals that would be required to realize affine-equivalent root multisets; any such deletion can be arranged with local deficiency).

At dyadic stage X, define $\mathcal{E}_X$ to be the set of remaining collisions whose maximum element lies in $[X, 2X)$. By Lemma 1 and Hypothesis 1, we have (4), so $|\mathcal{E}_X| = o(X)$. Apply Lemma 3 to obtain a hitting set $B_X \subset [X, 2X)$ that kills all collisions in $\mathcal{E}_X$ and also satisfies the capacity constraint (5), preserving local deficiency.

Iterating over all dyadic X produces a final deletion set $B := \cup_X B_X$. Since $|B_X| \le |\mathcal{E}_X| = o(X)$ and the stages are disjoint,

$$\#(B \cap [1, N]) = \sum_{X \le N} |B_X| = o(N),$$

so $A = \mathbb{N} \setminus B$ has density 1. By construction, every collision has a well-defined dyadic stage in which its maximal element lies, and is destroyed at that stage; thus no collisions remain in the final sequence (d_i) enumerating A. □

5 Discussion and remaining analytic input

The conditional argument shows that the density-one problem reduces to proving a *uniform* polylogarithmic bound on integer solutions to the separated-variable equations (3) in the regime deg $\asymp \log X$ and shifts $O(\log X)$, up to explicitly classifiable degenerate patterns.

One possible route is to strengthen Lemma 2 to allow $|\ell - (k + 1)| = O(1)$ by proving that any solution $(x, y) \in [X, 2X]^2$ forces y to lie in one of finitely many narrow "tubes" around a translate/dilate of x, reducing to root-counting for a family of polynomials of degree $\asymp \log X$. Another route is to combine Bilu–Tichy (to classify degeneracies) with an effective finiteness theorem (Baker-type bounds) to show that the total number of solutions in $[X, 2X]^2$ is at most $(\log X)^{O(1)}$ uniformly in the shifts.

6 Towards proving Hypothesis

In this section we outline a route to proving Hypothesis 1 (or a slightly weaker variant that still suffices for the dyadic deletion scheme). Recall that Hypothesis 1 asserts a *uniform* bound on the number of integer solutions $(x, y) \in [X, 2X]^2$ to

$$\prod_{i=0}^{\ell-1}(x - a_i) \; = \; \prod_{j=0}^{k}(y + b_j), \qquad (a_0 = b_0 = 0), \tag{6}$$

in the regime $\ell, k \asymp \log X$ and $|a_i|, |b_j| \ll \log X$, except for explicitly describable "degenerate" patterns. The key idea is to (i) *classify* all patterns leading to low-genus components (hence potentially many integer points) using Bilu–Tichy type structure theorems for $f(x) = g(y)$, and then (ii) apply determinant-method bounds to the remaining (non-degenerate) curves.

6.1 Degeneracy and the Bilu–Tichy paradigm

Let $f, g \in \mathbb{Q}[t]$ be nonconstant polynomials and consider the affine curve

$$C_{f,g}: \quad f(x) = g(y).$$

A guiding principle is that if $C_{f,g}$ has "many" integral points, then it has a low-complexity (genus 0 or 1) component, and conversely low-genus components arise only from strong algebraic structure relating f and g. The theorem of Bilu–Tichy [1] makes this precise in the presence of infinitely many rational solutions of bounded denominator: if $f(x) = g(y)$ has infinitely many such solutions, then there exist linear polynomials λ, μ and a polynomial φ such that

$$f = \varphi \circ F \circ \lambda, \qquad g = \varphi \circ G \circ \mu,$$

where (F, G) is one of finitely many "standard pairs" (monomials, Chebyshev/Dickson, and a small exceptional list), up to the above compositions.

In our application, f and g are *squarefree products of distinct linear factors*:

$$f(x) = \prod_{i=0}^{\ell-1}(x - a_i), \qquad g(y) = \prod_{j=0}^{k}(y + b_j), \tag{7}$$

with all a_i distinct and all $-b_j$ distinct, and (in realizable patterns) $|a_i|, |b_j| \ll \log X$. Thus, we are in the "simple rational roots" setting, for which significant refinements are known.

6.2 Simple rational roots: the Hajdu–Tijdeman refinement

Hajdu–Tijdeman [4] prove strong restrictions on equations $f(x) = g(y)$ when f and/or g have *only simple rational roots*. In particular (specializing to the case relevant for us), if both f and g have only simple rational roots and $f(x) = g(y)$ has infinitely many rational solutions of bounded denominator, then the degrees are tightly constrained: writing $\deg(f) = ms$ and $\deg(g) = ns$, one may take $m \in \{1, 2\}$ (after ordering degrees), and the remaining structure is described explicitly in terms of "standard pairs" together with certain additive-combinatorial configurations (Prouhet–Tarry–Escott type patterns). For our intended regime $\deg(f), \deg(g) \asymp \log X$, this drastically prunes the Bilu–Tichy list: essentially only the $m = 1$ and $m = 2$ branches can occur.

The practical takeaway is that any "degenerate pattern" $(\mathbf{a}, \mathbf{b})$ leading to an atypically large set of integer points on (6) must fall into a short, explicitly structured list:

- **Affine/translate degeneracy** (line component): $f(x) \equiv g(x + t)$ for some integer t, equivalently the root multisets $\{a_i\}_{i=0}^{\ell-1}$ and $\{-b_j - t\}_{j=0}^{k}$ coincide (in particular $\ell = k + 1$).
- **Possible "$m = 2$" exceptional families**: patterns arising from the $m = 2$ branch in [4], which are closely related to Prouhet–Tarry–Escott (PTE) configurations and entail strong symmetry/additive constraints on the root sets.

6.3 Eliminating standard pairs under complete splitting

We now explain why, in the completely split squarefree setting (7), most standard pairs are incompatible for large degree. This is the step where one can hope to reduce "degenerate patterns" to the affine/translate case plus (possibly) finitely many small exceptional patterns.

Proposition 1 (Standard pairs are incompatible with split squarefree products)**.** *Let $f, g \in \mathbb{Q}[t]$ have only simple rational roots and split completely over $\mathbb{Q}$, and assume* $\min(\deg f, \deg g) \geq 3$*. If $f(x) = g(y)$ has infinitely many rational solutions of bounded denominator, then, after composing by linear changes of variables, either:*

(i) *$f(x) \equiv g(x + t)$ for some $t \in \mathbb{Q}$ (affine/translate degeneracy), or*

(ii) *(f, g) belongs to the $m = 2$ branch in the Hajdu–Tijdeman classification, hence the root sets satisfy an explicit PTE-type pattern.*

In particular, Chebyshev/Dickson/Lattès-type standard pairs do not occur in the split setting.

Proof sketch. By Bilu–Tichy, $f = \varphi \circ F \circ \lambda$ and $g = \varphi \circ G \circ \mu$ for a standard pair (F, G). Because f and g are squarefree with many distinct roots, the outer polynomial φ must be linear: otherwise the pullback of a multiple root of φ would force multiple roots of f or g unless an exceptional avoidance condition holds, which is incompatible with having $\gg 1$ distinct rational roots mapping to 0. Thus f and g are, up to affine changes, essentially F and G.

Now inspect the standard pair list. Power-type pairs force repeated roots unless the exponent is 1. Chebyshev/Dickson types (and their Lattès analogues) do not split completely over $\mathbb{Q}$ in degrees ≥ 2, so they cannot be affinely conjugate to a product of rational linear factors. The remaining possibilities are precisely those singled out by Hajdu–Tijdeman for the "simple rational roots" case; when degrees are large this collapses to the $m = 1$ and $m = 2$ branches described above, yielding (i) or (ii). □

6.4 From degeneracy classification to the form of Hypothesis 1

Assuming Proposition 1, we can *define* the exceptional/degenerate patterns in Hypothesis 1 to be:

1. affine/translate patterns $f(x) \equiv g(x + t)$ (line components), and
2. the $m = 2$ PTE-type patterns arising from [4].

In the context of the original combinatorial problem, the affine/translate patterns should correspond to overlap/trivial configurations (and hence be removable at zero-density cost). The $m = 2$ patterns impose strong symmetry constraints on the multiset $\{-b_j\}$ and therefore appear to be highly constrained by monotonicity/realizability of consecutive gaps; in particular, one can hope to show that among realizable patterns with $|a_i|, |b_j| \ll \log X$ there are only finitely many $m = 2$ exceptions (consistent with the "interesting numbers" computations discussed in the problem thread).

6.5 Counting solutions in the non-degenerate case

Once low-genus components are excluded, one expects determinant-method bounds to yield $X^{o(1)}$ integer points in $[X, 2X]^2$ for each fixed pattern. For our specific family, the relevant curve is

$$C_{\mathbf{a},\mathbf{b}} : \quad F_{\mathbf{a},\mathbf{b}}(x, y) := \prod_{i=0}^{\ell-1}(x - a_i) - \prod_{j=0}^{k}(y + b_j) = 0,$$

of degree $d = \max(\ell, k + 1) \asymp \log X$. A general determinant-method bound (Bombieri–Pila/Pila) for an irreducible degree-d curve gives

$$\#\big(C_{\mathbf{a},\mathbf{b}}(\mathbb{Z}) \cap [X, 2X]^2\big) \;\leq\; \mathcal{C}(d)\, X^{1/d}(\log X)^{O(d)}, \tag{8}$$

uniformly in the coefficients (the dependence on d is explicit in the proofs). When $d \asymp \log X$, we have $X^{1/d} = e^{O(1)}$ and the right-hand side of (8) becomes $X^{o(1)}$ provided $\log \mathcal{C}(d) = o(\log X)$ for $d \ll \log X$. Thus, one is led to the following concrete strengthening of Hypothesis 1:

Conjecture 1 (Non-degenerate patterns have $X^{o(1)}$ solutions)**.** *Fix $\ell, k \asymp \log X$ and $|a_i|, |b_j| \ll \log X$. If $C_{\mathbf{a},\mathbf{b}}$ has no genus 0 component (equivalently, no standard-pair degeneracy in the sense above), then*

$$\#\big(C_{\mathbf{a},\mathbf{b}}(\mathbb{Z}) \cap [X, 2X]^2\big) \;\leq\; X^{o(1)}.$$

Conjecture 1 is weaker than Hypothesis 1 (which asks for $(\log X)^{O(1)}$), but it is already sufficient for the dyadic deletion scheme described earlier: local deficiency forces only $X^{o(1)}$ realizable patterns, hence total collisions at scale X are $X^{o(1)} = o(X)$, which can be hit by deleting $o(X)$ integers in $[X, 2X)$ while preserving local deficiency. Upgrading $X^{o(1)}$ to $(\log X)^{O(1)}$ is desirable but not strictly necessary for density-1.

6.6 Connection back to Hypothesis 1

Summarizing, a proof of Hypothesis 1 can be organized into two largely independent inputs:

1. **Degenerate-pattern classification.** Use Bilu–Tichy [1] together with the Hajdu–Tijdeman refinement [4] to show that any pattern $(\mathbf{a}, \mathbf{b})$ for which (6) has a low-genus component must be either affine/translate (line component) or belong to an explicit $m = 2$ PTE-type family. Then show that, among *realizable* patterns coming from consecutive gaps in a monotone density-1 set, the $m = 2$ family is empty or finite (hence removable at zero density).

2. **Uniform counting for non-degenerate curves.** Prove a bound of the form (8) with d-dependence controlled for $d \asymp \log X$, giving $X^{o(1)}$ (or ideally $(\log X)^{O(1)}$) integer points per non-degenerate pattern. Combined with the forced $X^{o(1)}$ pattern count from local deficiency, this yields $o(X)$ collisions at scale X, enabling the capacity-constrained hitting-set deletion step and completing the density-1 construction.

Remark 2. From the perspective of arithmetic geometry, the role of "modular/Lattès" phenomena is to account for the exceptional maps that can produce low-genus fiber products. Proposition 1 asserts that such exceptional families are incompatible with the completely split squarefree product structure (7) in large degrees, leaving essentially only affine/translate degeneracies (and possibly finitely many $m = 2$ PTE-type exceptions) to handle.

7 Towards proving Hypothesis: diagonal sparsity for degenerate families

In Hypothesis 1 we singled out a family of "degenerate" patterns $(\mathbf{a}, \mathbf{b})$ for which the separated-variables curve

$$C : \quad f(x) = g(y)$$

may have unusually many integer points. In this section we explain how, once one adopts the *Bilu–Tichy / bounded-denominator* notion of degeneracy, one can prove a useful substitute for Hypothesis 1 for these patterns: *unless there is a line component, degenerate patterns contribute at most polylogarithmically many integer points in the diagonal box* $[X, 2X]^2$. This is exactly what the dyadic deletion scheme needs, since the truly dangerous case for local deletions is the existence of $\gg X$ points in a single dyadic box.

7.1 Bounded-denominator degeneracy and PTEm/PTE structure

We say that $f(x) = g(y)$ has *infinitely many solutions with bounded denominator* if there exists $\Delta \in \mathbb{Z}_{>0}$ such that $f(x) = g(y)$ has infinitely many $(x, y) \in \mathbb{Q}^2$ with $(\Delta x, \Delta y) \in \mathbb{Z}^2$. This is the arithmetic notion classified by Bilu–Tichy and refined in the "simple rational roots" setting by Hajdu–Tijdeman.

Throughout this section we assume

$$f(x) = \prod_{i=0}^{\ell-1}(x - a_i), \qquad g(y) = \prod_{j=0}^{k}(y + b_j), \qquad (a_0 = b_0 = 0), \tag{9}$$

with all a_i distinct and all $-b_j$ distinct (so f, g have only simple rational roots). By [5, Theorem 9.1] (see also [5, Theorem 6.1] for the first/second-kind cases), bounded-denominator degeneracy forces

f to be a *PTEm-polynomial* with $m \in \{1,2\}$ and g to be a corresponding PTE-polynomial. Concretely, after similarity (scaling and affine changes of variable) one may take

$$f(x) = p_0 \prod_{i=1}^{s} (x^m - p_i), \qquad g(y) = p_0 \prod_{i=1}^{s} (G(y) - p_i), \tag{10}$$

for some $m \in \{1,2\}$, distinct $p_i \in \mathbb{Q}$, and some $G \in \mathbb{Q}[y]$ of degree $n \geq 1$, with the additional PTEn-constraints on the family $(G(y) - p_i)_{i=1}^{s}$ (see [5, Definition (PTEn) and Theorem 6.1]).

7.2 The only way to have $\gg X$ points in $[X, 2X]^2$ is a line component

We first isolate the truly dangerous possibility.

Lemma 4 (Line components are exactly affine equivalences). *Let $f, g \in \mathbb{Z}[t]$ be monic. If the curve $f(x) = g(y)$ contains a line over $\mathbb{Q}$, then $\deg f = \deg g =: d$ and there exist $t \in \mathbb{Q}$ and $\varepsilon \in \{\pm 1\}$ such that*

$$g(y) \equiv f(\varepsilon y + t).$$

In particular, if f, g are monic with all real roots and we restrict to large positive x, y (the setting relevant to our application), then $\varepsilon = +1$.

Proof. If $f(x) = g(y)$ contains the line $y = ax + b$ with $a, b \in \mathbb{Q}$, then $f(x) \equiv g(ax + b)$ as polynomials in x. Comparing degrees gives $\deg f = \deg g = d$, and comparing leading coefficients (both 1) gives $a^d = 1$, so $a = \pm 1$. Writing $t = b$ gives the claim. □

Thus, the only way a degenerate pattern can force $\gg X$ integer points in a single dyadic box is via the affine/translate degeneracy (a line component). We now show that all remaining bounded-denominator degeneracies have *few* diagonal points.

7.3 Diagonal sparsity for non-line bounded-denominator degeneracies

For $X \geq 2$ write

$$N_{f,g}(X) := \#\{(x, y) \in \mathbb{Z}^2 \cap [X, 2X]^2 : \ f(x) = g(y)\}.$$

Lemma 5 (HT Theorem 6.1(1): growth mismatch on the diagonal unless a line). *Assume f, g are as in* (9) *and that $f(x) = g(y)$ has infinitely many bounded-denominator rational solutions. Suppose further that after similarity (affine change and scaling) we are in the $m = 1$ branch of [5, Theorem 6.1(1)], i.e.*

$$f(x) = p_0 \prod_{i=1}^{s} (x - p_i), \qquad g(y) = p_0 \prod_{i=1}^{s} (G(y) - p_i),$$

with $\deg(G) = n \geq 1$. *If $f(x) = g(y)$ has no line component over $\mathbb{Q}$, then $N_{f,g}(X) = 0$ for all sufficiently large X.*

Proof. For $x, y \in [X, 2X]$ with $X \to \infty$, we have $|f(x)| \asymp X^s$ while $|g(y)| \asymp |G(y)|^s \asymp X^{ns}$. Thus for solutions in $[X, 2X]^2$ for arbitrarily large X one must have $n = 1$. But if $n = 1$, then G is affine linear and $g(y)$ is affinely conjugate to f, so by Lemma 4 the curve has a line component, contradicting the hypothesis. Hence $n \geq 2$ in the non-line case, and then $|g(y)| \gg X^{2s}$ while $|f(x)| \ll X^s$, so no solutions exist for all large X. □

Lemma 6 (HT Theorem 6.1(2): the $m = 2$ branch is Pell/conic on the diagonal)**.** *Assume f, g are as in* (9) *and that $f(x) = g(y)$ has infinitely many bounded-denominator rational solutions. Suppose further that after similarity we are in the $m = 2$ branch of [5, Theorem 6.1(2)], so there exist $q_1, \ldots, q_s \in \mathbb{Q}_{>0}$ and $G \in \mathbb{Q}[y]$ such that*

$$f(x) = q_0 \prod_{i=1}^{s} (x - q_i)(x + q_i) = q_0\, P(x^2), \qquad g(y) = q_0 \prod_{i=1}^{s} (G(y) - q_i^2) = q_0\, P(G(y)),$$

where $P(u) := \prod_{i=1}^{s}(u - q_i^2)$. Assume $f(x) = g(y)$ has no line component over $\mathbb{Q}$. Then for all sufficiently large X one has

$$N_{f,g}(X) \ \ll\ \log X,$$

with implied constant depending only on f, g.

Proof. First note that $P(u)$ is strictly increasing for all real $u > \max_i q_i^2$, since

$$P'(u) = P(u) \sum_{i=1}^{s} \frac{1}{u - q_i^2} > 0 \qquad (u > \max_i q_i^2).$$

Hence P is injective on $(\max_i q_i^2, \infty)$.

Let $(x, y) \in [X, 2X]^2$ be an integer solution to $f(x) = g(y)$ for X large enough that $x^2 > \max_i q_i^2$ and $G(y) > \max_i q_i^2$ (this holds for all sufficiently large X since G is nonconstant in the $m = 2$ branch). Then

$$P(x^2) = P(G(y)),$$

and by injectivity we get

$$x^2 = G(y).$$

Now, if $\deg(G) \neq 2$, then for $y \asymp X$ we have $|G(y)| \asymp X^{\deg(G)}$, so $x^2 = G(y)$ cannot hold with $x \asymp X$ for arbitrarily large X unless $\deg(G) = 2$. Therefore in the diagonal regime $[X, 2X]^2$ and for X large, all solutions lie on the affine conic $x^2 = ay^2 + by + c$ with $a \neq 0$.

Clearing denominators and completing the square yields an integral norm-form equation of Pell type (or a factorable hyperbola in the square-discriminant case):

$$(2ay + b)^2 - 4ax^2 = \Delta, \qquad \Delta = b^2 - 4ac.$$

If the conic has a line component, we are in the excluded affine/translate case. Otherwise, the integer solutions in $y \in [X, 2X]$ form finitely many orbits under a rank-1 unit group in a quadratic order, and in particular there are $O(\log X)$ solutions in each dyadic box; cf. Lemma 7. This is exactly the Pell mechanism exhibited in [5, Example 5.4 and Example 5.7]. □

Proposition 2 (Non-line degenerate families are diagonally sparse)**.** *Assume f, g are as in* (9) *and that $f(x) = g(y)$ has infinitely many bounded-denominator rational solutions. If $f(x) = g(y)$ has no line component over $\mathbb{Q}$, then*

$$N_{f,g}(X) \ \ll\ \log X$$

for all $X \geq 2$, with implied constant depending only on f, g.

Proof. By [5, Theorem 9.1], when both sides have only simple rational roots, bounded-denominator degeneracy forces $m \in \{1, 2\}$. In the $m = 1$ branch, Lemma 5 gives $N_{f,g}(X) = 0$ for all sufficiently large X in the non-line case. In the $m = 2$ branch, Lemma 6 gives $N_{f,g}(X) \ll \log X$. Combining the two cases yields the claim. □

Lemma 7 (Dyadic sparsity for generalized Pell equations)**.** *Fix nonsquare $D \in \mathbb{Z}_{>0}$ and nonzero $N \in \mathbb{Z}$. Then the number of integer solutions $(u, v) \in \mathbb{Z}^2$ to*

$$u^2 - Dv^2 = N$$

with $|u|, |v| \le X$ is $O(\log X)$ as $X \to \infty$ (with implied constant depending on D, N). In particular the number of such solutions with $u, v \in [X, 2X]$ is $O(\log X)$.

Proof. It is classical that the solution set (if nonempty) consists of finitely many orbits under multiplication by units in $\mathbb{Z}[\sqrt{D}]$ of norm 1. Let $\varepsilon > 1$ be the fundamental unit of $\mathbb{Z}[\sqrt{D}]$ (or any generator of the totally positive part of the unit group). Along each orbit, $|u_n| + |v_n| \asymp \varepsilon^n$, so the number of points with $|u|, |v| \le X$ on that orbit is $\ll \log X / \log \varepsilon$. Summing over the finitely many orbits yields the claimed $O(\log X)$ bound. □

7.4 How this feeds into Hypothesis 1

Proposition 2 shows that once we carve out the *line-component* degeneracies (Lemma 4), all remaining bounded-denominator (PTE) degeneracies contribute at most polylogarithmically many integer points in each dyadic box $[X, 2X]^2$. This is strong enough for the dyadic deletion scheme: polylogarithmically many collisions per realizable pattern can be hit while preserving local deficiency, and the affine/translate (line) patterns can be removed by an additional zero-density deletion or by showing they correspond to trivial/overlap configurations.

8 Uniform point-count in the log-degree regime

In this section we supply the "uniform bound part" needed in Hypothesis 1 for the troublesome regime $d \asymp \log X$, by quoting a determinant-method bound with *polynomial* dependence on the degree. This cleanly addresses the bottleneck case $k, \ell \asymp \log N$ highlighted in the problem discussion: once $d \sim c \log X$, the factor $X^{1/d}$ is $e^{O(1)}$, and the remaining dependence is polylogarithmic.

8.1 Castryck–Cluckers–Dittmann–Nguyen: polynomial-in-d determinant method

For a variety $V \subset \mathbb{A}^n_{\mathbb{Q}}$ and $B \ge 1$, write

$$N_{\text{aff}}(V, B) \ := \ \#\{x = (x_1, \dots, x_n) \in \mathbb{Z}^n : \ |x_i| \le B \ \forall i, \ x \in V(\mathbb{Q})\}.$$

Theorem 2 (CCDN: affine curves, polynomial in the degree)**.** *Fix $n > 1$. There exists a constant $c = c(n) > 0$ such that for every $d \ge 1$, every* integral *(i.e. geometrically integral) affine curve $V \subset \mathbb{A}^n_{\mathbb{Q}}$ of degree d, and every $B \ge 1$, one has*

$$N_{\text{aff}}(V, B) \ \le \ c\, d^3\, B^{1/d} (\log B + d).$$

Proof. This is [7, Theorem 3] (see also [7, §3] for the definition of N_{aff} and the uniformity in coefficients). □

8.2 A uniform corollary for plane curves in a dyadic box

We apply Theorem 2 to plane curves ($n = 2$). Let $F \in \mathbb{Z}[x, y]$ be nonzero, and let $C_F \subset \mathbb{A}^2_{\mathbb{Q}}$ be the affine curve $F(x, y) = 0$. Write the factorization over $\mathbb{Q}$ as

$$F \ = \ \prod_{r=1}^{R} F_r,$$

where each $F_r \in \mathbb{Q}[x, y]$ is irreducible and $C_r : F_r = 0$ is an integral affine curve. Then

$$N_{\text{aff}}(C_F, B) \le \sum_{r=1}^{R} N_{\text{aff}}(C_r, B).$$

Corollary 1 (Uniform bound for non-linear components)**.** *Let $F \in \mathbb{Z}[x, y]$ have total degree $\deg(F) \le d$. Assume that F has no linear factor over $\mathbb{Q}$ (equivalently, C_F contains no affine line component over $\mathbb{Q}$). Then for all $B \ge 2$,*

$$\#\{(x, y) \in \mathbb{Z}^2 : |x|, |y| \le B, \ F(x, y) = 0\} \ll d^4 B^{1/d} (\log B + d),$$

with an absolute implied constant.

Proof. There are at most $R \le d$ irreducible factors. For each factor F_r, the corresponding integral curve C_r has degree $d_r \le d$, and by Theorem 2 (with $n = 2$) we have

$$N_{\text{aff}}(C_r, B) \ll d_r^3 B^{1/d_r} (\log B + d_r) \le d^3 B^{1/d} (\log B + d),$$

since $d_r \le d$ implies $B^{1/d_r} \le B^{1/d}$ for $B \ge 1$. Summing over r yields the stated d^4 factor. (The hypothesis "no linear factor" excludes the $d_r = 1$ case, which would give a term $\asymp B$.) □

8.3 Specialization to the collision curves $F_{\mathbf{a},\mathbf{b}}(x, y) = 0$

Recall that for a pattern $(\mathbf{a}, \mathbf{b})$ we defined

$$F_{\mathbf{a},\mathbf{b}}(x, y) := \prod_{i=0}^{\ell-1} (x - a_i) - \prod_{j=0}^{k} (y + b_j), \qquad (a_0 = b_0 = 0),$$

so $\deg(F_{\mathbf{a},\mathbf{b}}) = d := \max(\ell, k + 1)$. We will apply Corollary 1 in the bottleneck regime $d \asymp \log X$.

Proposition 3 (Polylog point-count when $d \asymp \log X$)**.** *Fix constants $0 < c_1 \le c_2$. Let X be large, and let ℓ, k satisfy*

$$c_1 \log X \le d := \max(\ell, k + 1) \le c_2 \log X.$$

Then for any integer shifts a_i, b_j (no size restriction needed) such that $F_{\mathbf{a},\mathbf{b}}$ has no linear factor *over $\mathbb{Q}$,*

$$\#\{(x, y) \in \mathbb{Z}^2 \cap [X, 2X]^2 : F_{\mathbf{a},\mathbf{b}}(x, y) = 0\} \ll (\log X)^5,$$

where the implied constant depends only on c_1, c_2 (not on the shifts).

Proof. The box $[X, 2X]^2$ is contained in $\{|x|, |y| \le 2X\}$, so it suffices to bound solutions with $|x|, |y| \le 2X$. Apply Corollary 1 with $B = 2X$ and degree $d \le c_2 \log X$:

$$\#\{|x|, |y| \le 2X : F_{\mathbf{a},\mathbf{b}}(x, y) = 0\} \ll d^4 (2X)^{1/d} (\log(2X) + d).$$

Since $d \ge c_1 \log X$, we have $(2X)^{1/d} \le \exp\left(\frac{\log(2X)}{c_1 \log X}\right) \le e^{2/c_1}$ for X large, while $d^4(\log(2X) + d) \ll (\log X)^5$. This gives the desired bound. □

Remark 3 (Connection to Hypothesis 1)**.** Proposition 3 provides the required *uniform* $(\log X)^{O(1)}$ solution bound in the main troublesome regime $k, \ell \asymp \log X$, *provided* we have excluded the line-component (affine/translate) degeneracies. The remaining degenerate families (e.g. Pell/PTE-type) are handled separately in Section 7 by showing they have at most $O(\log X)$ points in $[X, 2X]^2$ once line components are excluded.

9 Proof of Hypothesis 1 in the logarithmic-length regime

We now verify the *uniform diagonal-box point count* asserted in Hypothesis 1 for the regime relevant to the "main case" $k, \ell \asymp \log X$. (This is exactly the remaining regime emphasized in the problem discussion; smaller values of k, ℓ are intended to be treated by the other arguments outlined earlier.)

9.1 Statement proved

Fix constants $0 < c_1 \le c_2$ and an integer $R \ge 0$. For each large X let k, ℓ satisfy

$$c_1 \log X \;\le\; d := \max(\ell, k+1) \;\le\; c_2 \log X, \qquad 0 \le \ell - (k+1) \le R,$$

and let $a_0 = b_0 = 0$ with integers $a_1, \ldots, a_{\ell-1}$ and $b_1, \ldots, b_k$. Define

$$f(x) := \prod_{i=0}^{\ell-1}(x - a_i), \qquad g(y) := \prod_{j=0}^{k}(y + b_j), \qquad F(x,y) := f(x) - g(y).$$

Theorem 3 (Uniform polylog bound in $d \asymp \log X$)**.** *Assume that F has* no linear factor over $\mathbb{Q}$ *(equivalently, the curve $F(x,y) = 0$ contains no affine line component over $\mathbb{Q}$). Then*

$$\#\{(x,y) \in \mathbb{Z}^2 \cap [X, 2X]^2 :\; f(x) = g(y)\} \;\ll_{c_1,c_2,R}\; (\log X)^5.$$

In particular, in the regime $k, \ell \asymp \log X$, Hypothesis 1 holds with "degenerate" interpreted as "having a line component" together with the remaining bounded-denominator (PTE/Pell) families treated in Section 7.

9.2 Proof via CCDN (polynomial-in-degree determinant method)

Proof. Factor F over $\mathbb{Q}$ as $F = \prod_{r=1}^{R} F_r$ with F_r irreducible in $\mathbb{Q}[x,y]$. Because F has no linear factor, each F_r has degree $d_r \ge 2$. Let $C_r \subset \mathbb{A}^2_{\mathbb{Q}}$ be the integral affine curve $F_r(x,y) = 0$.

For $B \ge 1$ write

$$N_{\text{aff}}(C_r, B) := \#\{(x,y) \in \mathbb{Z}^2 :\; |x|, |y| \le B,\; (x,y) \in C_r(\mathbb{Q})\}.$$

By Castryck–Cluckers–Dittmann–Nguyen [7, Theorem 3], for $n = 2$ there is an absolute constant $c > 0$ such that

$$N_{\text{aff}}(C_r, B) \;\le\; c\, d_r^3\, B^{1/d_r}\, (\log B + d_r) \qquad (B \ge 1).$$

Since $d_r \le d$ and $B^{1/d_r} \le B^{1/d}$ for $B \ge 1$, we obtain

$$N_{\text{aff}}(C_r, 2X) \;\ll\; d^3\, (2X)^{1/d}\, (\log(2X) + d).$$

Summing over the at most d irreducible factors yields

$$\#\{(x,y) \in \mathbb{Z}^2 :\; |x|, |y| \le 2X,\; F(x,y) = 0\} \;\ll\; d^4\, (2X)^{1/d}\, (\log(2X) + d).$$

Now $d \ge c_1 \log X$ implies $(2X)^{1/d} \le \exp(\log(2X)/(c_1 \log X)) \le e^{2/c_1}$, while $d \le c_2 \log X$ gives $d^4(\log(2X) + d) \ll_{c_2} (\log X)^5$. Finally $[X, 2X]^2 \subset \{|x|, |y| \le 2X\}$, so the same bound holds in the dyadic box. $\square$

9.3 How this completes the uniform-bound component of Hypothesis

Theorem 3 supplies exactly the uniform diagonal-box bound needed for the "main case" $k, \ell \asymp \log X$ once line components are excluded. The remaining "bounded-denominator" (PTE/Pell-type) degenerate families are handled in Section 7, which shows that *after excluding line components* they contribute only $O(\log X)$ (hence polylogarithmically many) points to $[X, 2X]^2$. Together, these two inputs justify the uniform point-count conclusion of Hypothesis 1 in the logarithmic-length regime.

References

[1] Y. Bilu and R. Tichy. The Diophantine equation $f(x) = g(y)$. *Acta Arith.*, 95(3):261–288, 2000.

[2] E. Bombieri and J. Pila. The number of integral points on arcs and ovals. *Duke Math. J.*, 59(2):337–357, 1989.

[3] J. Pila. Density of integer points on plane algebraic curves. *Internat. Math. Res. Notices*, 1996(18):903–911, 1996.

[4] L. Hajdu and R. Tijdeman, On the diophantine equation $f(x) = g(y)$, *Acta Arith.*, 117 (2005), 375–396.

[5] L. Hajdu and R. Tijdeman. The Diophantine equation $f(x) = g(y)$ for polynomials with simple rational roots. *arXiv:2204.12345* (2022).

[6] A. Dujella and R. Tichy, Diophantine equations and the Bilu–Tichy theorem, available as lecture notes / survey (various versions).

[7] W. Castryck, R. Cluckers, Ph. Dittmann, and K. H. Nguyen, The dimension growth conjecture, polynomial in the degree and without logarithmic factors, *Algebra Number Theory*, 14 (2020), no. 8, 2261–2294. (arXiv:1904.13109)