DAWN PUTNAM SESSION

Putnam 2015 — Papers A and B

Putnam 120 Archive

Contents

Putnam 2015 — Paper A **5**
2015 A1 5
2015 A2 7
2015 A3 10
2015 A4 13
2015 A5 16
2015 A6 18

Putnam 2015 — Paper B **23**
2015 B1 23
2015 B2 25
2015 B3 27
2015 B4 30
2015 B5 34
2015 B6 38

Putnam 2015 — Paper A

2015 A1

2015 A1. Let A and B be points on the same branch of the hyperbola $xy = 1$. Suppose that P is a point lying between A and B on this hyperbola, such that the area of the triangle APB is as large as possible. Show that the region bounded by the hyperbola and the chord AP has the same area as the region bounded by the hyperbola and the chord PB.

Fields: Euclidean geometry, single-variable calculus, symmetry arguments **Position:** A1 — the opener **Answer:** Both regions equal $\frac{t_2 - t_1}{2\sqrt{t_1 t_2}} - \frac{1}{2}\ln\frac{t_2}{t_1}$ at the maximiser $P = (\sqrt{t_1 t_2}, 1/\sqrt{t_1 t_2})$; the equality follows because $xy = 1$ carries an area-preserving symmetry that swaps the two regions.

Layer 1 THE WHAT — surface analysis

The natural attempt. Put $A = (t_1, 1/t_1)$, $B = (t_2, 1/t_2)$ with $0 < t_1 < t_2$, and let $P = (t, 1/t)$ with $t_1 \le t \le t_2$. The area of triangle APB is a determinant,

$$\frac{1}{2}\begin{vmatrix} 1 & 1 & 1 \\ t_1 & t & t_2 \\ 1/t_1 & 1/t & 1/t_2 \end{vmatrix} = \frac{t_2 - t_1}{2t_1 t_2}\left(t_1 + t_2 - t - \frac{t_1 t_2}{t}\right).$$

With t_1, t_2 fixed, this is maximised when $t + t_1 t_2/t$ is minimised, and by AM-GM that happens exactly at $t = \sqrt{t_1 t_2}$, with equality forced by the equal-terms condition of AM-GM. So $P = (\sqrt{t_1 t_2}, 1/\sqrt{t_1 t_2})$.

Now find the two regions. The line AP has equation $y = \dfrac{t_1 + t - x}{t t_1}$, so the area between it and the hyperbola is

$$\int_{t_1}^{t}\left(\frac{t_1 + t - x}{t t_1} - \frac{1}{x}\right)dx = \frac{t}{2t_1} - \frac{t_1}{2t} - \ln\frac{t}{t_1},$$

and by the same computation with the roles of t_1, t replaced by t, t_2, the area between PB and the hyperbola is $\dfrac{t_2}{2t} - \dfrac{t}{2t_2} - \ln\dfrac{t_2}{t}$. Substituting $t = \sqrt{t_1 t_2}$ into both turns each into

$$\frac{t_2 - t_1}{2\sqrt{t_1 t_2}} - \ln\sqrt{\frac{t_2}{t_1}},$$

and the two expressions coincide term for term. This works. It also requires trusting two separate integrations, a square-root substitution, and a log-of-a-ratio manipulation to land on the same three-term expression twice, with no warning in advance that they would match.

The model chain. For any $\lambda > 0$, the map $T_\lambda : (x, y) \mapsto (\lambda x, y/\lambda)$ sends $xy = 1$ to itself, since $(\lambda x)(y/\lambda) = xy$. Its Jacobian is $\begin{pmatrix} \lambda & 0 \\ 0 & 1/\lambda \end{pmatrix}$, determinant 1, so T_λ preserves area exactly: for every region R, $\text{area}(T_\lambda R) = \text{area}(R)$.

Choose $\lambda = 1/\sqrt{t_1 t_2}$. Then $T_\lambda(A) = (t_1/\sqrt{t_1 t_2}, \sqrt{t_1 t_2}/t_1) = (\sqrt{t_1/t_2}, \sqrt{t_2/t_1})$ and $T_\lambda(B) = (\sqrt{t_2/t_1}, \sqrt{t_1/t_2})$: these two points are reflections of one another in the line $y = x$. Relabel so A now sits above $y = x$ and B below it.

As P moves along the branch from A to B, the area of APB increases and then decreases (this is exactly the AM-GM computation above, now transported by T_λ), so the maximiser is unique. Reflection in $y = x$, i.e. $(x, y) \mapsto (y, x)$, also fixes $xy = 1$ and also preserves area, and it swaps A and B. If P^* is the area-maximiser, its reflection is also an area-maximiser for the reflected (i.e. swapped) pair B, A – but maximising the area of APB and of BPA is the same problem, so by uniqueness P^* is its own reflection: $P^* = (1, 1)$.

Now reflection in $y = x$ fixes $P^* = (1, 1)$ and swaps $A \leftrightarrow B$, hence swaps the chord AP^* with the chord P^*B, and swaps the region cut off by the first with the region cut off by the second. Since reflection preserves area, the two regions have equal area. Undoing T_λ (which also preserves area) carries this conclusion back to the original A, B, P.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree up to the moment A, B are named with coordinates. The natural attempt then fixes $A = (t_1, 1/t_1)$, $B = (t_2, 1/t_2)$ and starts optimising and integrating immediately. The model chain instead asks, before writing a single integral, whether the configuration can be moved – without changing any area – into a more symmetric one.

> *This map $(x, y) \mapsto (\lambda x, y/\lambda)$ fixes the curve for every λ, and its Jacobian determinant is 1 – so I can use it to slide A and B into any position along the branch that this one-parameter family reaches, and every area in the picture stays exactly what it was.*

> *The problem asks me to prove two areas, defined by an asymmetric construction (a left piece and a right piece), are equal. An asymmetric construction producing a symmetric conclusion is exactly the signature of a hidden symmetry – so before computing either area, I should look for the map that swaps the two pieces.*

Layer 3 THE WHY

Why does the rescaling $(x, y) \mapsto (\lambda x, y/\lambda)$ occur to the model solver and not the direct one? Because the defining equation $xy = 1$ is a statement about a *product* being fixed, and the group of maps preserving a product xy is exactly the group $\{(x, y) \mapsto (\lambda x, y/\lambda) : \lambda > 0\}$ together with the swap $(x, y) \mapsto (y, x)$ – this is not a trick imported from outside the problem, it is read off directly from the shape of the defining equation. A curve given by a sum $x + y = 1$ has a different symmetry group (translations along the line, plus the same swap); a curve given by $x^2 + y^2 = 1$ has rotations. The algebraic form of the constraint *is* the symmetry group, and finding it is a matter of asking "what operations leave this equation unchanged?" rather than "what operations have I seen used on hyperbolas before?"

The second fact driving the model chain is that determinant-1 maps preserve area – this is not a property special to this problem but the definition of the change-of-variables Jacobian: $dA' = |\det J|\, dA$. Once $|\det J| = 1$ is checked (a one-line computation), every area statement about the original configuration transfers for free to the rescaled one, and back. So the two facts in play are (1) read the symmetry group off the defining equation, and (2) know which maps in that group also preserve area, and the combination replaces two definite integrals and a square root with a one-line reflection argument.

The remaining piece – that the maximiser P^* must itself be the fixed point of the swap – follows from uniqueness of the maximum. If the maximising P for A, B were not fixed by the reflection that exchanges A, B, its reflection would be a *different* point also maximising the same area (since reflection preserves area and swaps the problem $APB \leftrightarrow BPA$, which are the same

problem), contradicting uniqueness. Uniqueness of the extremiser is therefore not incidental bookkeeping; it is the second load-bearing fact, alongside the symmetry group itself.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the phrase "region bounded by the hyperbola and the chord" appearing twice, for an asymmetric split of the same configuration, with the claim that the two are equal. On seeing it, write down: "this equality has no reason to hold unless something swaps the two halves – find that something before computing."

2. That cue suggests identifying the automorphism group of the curve $xy = 1$ directly from its equation, rather than reaching for calculus first.

3. Hesitation sets in right after the AM-GM step locates P^*: the temptation is to keep the momentum of the calculus and push straight into the two integrals, because that route is guaranteed to work and the symmetry route is not yet guaranteed to be easier. What reinforces the switch away from momentum is noticing that the two integrals, if computed, are going to have to match anyway – so the work of confirming that match is being done twice, once by algebra and once (implicitly) by the symmetry that makes it inevitable.

4. The trap is treating "I found a method that works" (direct coordinates plus AM-GM plus integration) as license to stop looking for a better one. It works, but it spends two integrations and a square-root substitution to reconstruct a conclusion that a one-line reflection argument gives for free – and it is far more exposed to a sign error in the log-of-a-ratio step.

5. Recognise it faster by treating "prove two asymmetrically-defined quantities on a symmetric-looking curve are equal" as a standing instruction to find the curve's symmetry group first, specifically any subgroup of that group which also preserves the quantity being compared (area, length, whichever it is).

6. Reusable rule: read the symmetry group of a curve directly off the algebraic form of its defining equation (a product fixed $\Rightarrow$ reciprocal scalings; a sum fixed $\Rightarrow$ translations along the line; a sum of squares fixed $\Rightarrow$ rotations), then check which elements of that group also preserve the quantity in the problem (area, via determinant 1). Use that subgroup to move the configuration into the position where the claim is immediate, and only then, if at all, compute.

▷ **Cue:** a curve's defining equation is a product held constant, e.g. $xy = k$ $\longrightarrow$ **Move:** look for the two-parameter symmetry $(x, y) \mapsto (\lambda x, y/\lambda)$ and the swap $(x, y) \mapsto (y, x)$, and check whether either preserves the quantity being compared

▷ **Cue:** asked to prove an equality between two asymmetrically-constructed quantities on a symmetric figure $\longrightarrow$ **Move:** search for a symmetry of the figure that exchanges the two constructions, before attempting direct computation

▷ **Cue:** an extremiser is unique and the configuration has a symmetry exchanging two labelled objects (here A, B) $\longrightarrow$ **Move:** conclude the extremiser is a fixed point of that symmetry

2015 A2

2015 A2. Let $a_0 = 1$, $a_1 = 2$, and $a_n = 4a_{n-1} - a_{n-2}$ for $n \geq 2$. Find an odd prime factor of a_{2015}.

Fields: recurrences, algebraic number theory (Lucas-sequence divisibility) **Position:** A2 **Answer:** 181 is an odd prime factor of a_{2015}, because $a_5 = 362 = 2 \cdot 181$ and $a_5 \mid a_{2015}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. Solve the recurrence. The characteristic equation of $a_n = 4a_{n-1} - a_{n-2}$ is $x^2 - 4x + 1 = 0$, with roots $\alpha = 2 + \sqrt{3}$, $\beta = 2 - \sqrt{3}$; note $\alpha\beta = 1$ and $\alpha + \beta = 4$. Matching the

initial conditions $a_0 = 1 = (\alpha^0 + \beta^0)/2$ and $a_1 = 2 = (\alpha + \beta)/2$ gives the closed form

$$a_n = \frac{\alpha^n + \beta^n}{2}.$$

This is correct and not hard to reach. The trouble starts immediately after: a_{2015} is $(\alpha^{2015} + \beta^{2015})/2$, an integer with well over a thousand digits (the official solutions record that even after removing several known factors, an 824-digit cofactor of unknown factorisation remains). There is no prospect of computing a_{2015} by hand, listing its digits, and trial-dividing; nor is there any evident pattern in a_n mod p for small primes p that presents itself without already knowing what to look for. A student who has correctly found the closed form can still be stuck here, because the closed form answers "what is a_n", not "what divides a_{2015}" – and those are different questions.

The model chain. Do not try to know a_{2015}; try to know a *divisor* of it. Ask instead: for which pairs m, n does a_m divide a_n? Since $\alpha\beta = 1$, for any odd k,

$$\frac{a_{kn}}{a_n} = \frac{\alpha^{kn} + \beta^{kn}}{\alpha^n + \beta^n} = \alpha^{(k-1)n} - \alpha^{(k-2)n}\beta^n + \cdots - \alpha^n\beta^{(k-2)n} + \beta^{(k-1)n},$$

using the factorisation $u^k + v^k = (u+v)(u^{k-1} - u^{k-2}v + \cdots + v^{k-1})$ valid for odd k, with $u = \alpha^n, v = \beta^n$. The right-hand side is a $\mathbb{Z}$-linear combination of powers of α, β, hence has the form $A + B\sqrt{3}$ for integers A, B; but it is also equal to a_{kn}/a_n, a ratio of two known integers, hence rational. A number of the form $A + B\sqrt{3}$ that is rational must have $B = 0$, so $a_{kn}/a_n = A \in \mathbb{Z}$: that is, $a_n \mid a_{kn}$ whenever k is odd.

Now factor $2015 = 5 \cdot 13 \cdot 31$, all odd, so in particular $2015/5 = 403$ is odd. Taking $n = 5$, $k = 403$ gives $a_5 \mid a_{2015}$. Compute a_5 directly from the recurrence: $a_2 = 4(2) - 1 = 7$, $a_3 = 4(7) - 2 = 26$, $a_4 = 4(26) - 7 = 97$, $a_5 = 4(97) - 26 = 362 = 2 \times 181$. Since 181 is prime and divides a_5, and $a_5 \mid a_{2015}$, $181 \mid a_{2015}$: an odd prime factor, found without ever writing down a_{2015} itself.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains reach the same closed form $a_n = (\alpha^n + \beta^n)/2$ without difficulty. They split at the very next question asked. The natural attempt asks "what is a_{2015}, and what divides it?" and tries to answer both at once by brute force. The model chain asks only the second half, and asks it as a question about the structure of the *index* $n \mapsto a_n$, not about the value a_{2015}.

> *I don't need to know a_{2015} – I only need one prime factor of it, so if some earlier, computable term a_n can be shown to divide a_{2015}, the problem is solved by computing a_n instead, and a_n can be made as small as 2015's factorisation allows.*

> *The roots satisfy $\alpha\beta = 1$: that is exactly the signature of a Lucas-type sequence, and Lucas sequences are famous for one thing above all – index divisibility, $a_n \mid a_{kn}$. I have seen this shape before (it's the Fibonacci divisibility fact, $F_m \mid F_{mn}$); the same odd-power factorisation trick that proves it there should prove it here.*

Layer 3 THE WHY

Why does $a_n \mid a_{kn}$ for odd k deserve to be true, rather than being a coincidence discovered by luck? It comes from two independent facts about this specific closed form, and neither is an accident of this problem.

First, $\alpha\beta = 1$ is forced by the recurrence's own coefficients: for $a_n = 4a_{n-1} - a_{n-2}$, the product of the characteristic roots is the constant term of $x^2 - 4x + 1$, namely 1. Whenever a

linear recurrence has this shape – second order, integer coefficients, constant term 1 (equivalently ± 1) – its two roots are reciprocals, and $\alpha^n \beta^n = 1$ for every n. This is precisely what makes α^n, β^n behave like a single unit and its inverse, the same structural fact underlying Fibonacci and Lucas numbers ($x^2 - x - 1$, constant term -1) and Chebyshev-type recurrences generally. Recognising "constant term ± 1" as a standing invitation to look for index divisibility is the transferable content here, not a fact special to 4 and $\sqrt{3}$.

Second, the algebraic identity $u^k + v^k = (u+v)(u^{k-1} - u^{k-2}v + \cdots + v^{k-1})$ for odd k is the ordinary sum-of-odd-powers factorisation from school algebra ($u^3 + v^3 = (u+v)(u^2 - uv + v^2)$, generalised). Its role here is to turn "a_n divides a_{kn}" from a numerical claim into an algebraic one: the quotient is manifestly an algebraic integer in $\mathbb{Z}[\sqrt{3}]$ (a sum of powers of α, β with integer coefficients), and separately, manifestly rational (a quotient of two rational integers a_{kn}, a_n). An element of $\mathbb{Z}[\sqrt{3}]$ that is rational must be a rational integer, because $\{1, \sqrt{3}\}$ is a $\mathbb{Q}$-basis and rationality forces the $\sqrt{3}$-coefficient to vanish. This "integral and rational $\Rightarrow$ integer" argument is the standard mechanism by which algebraic number theory proves integrality results that look inaccessible from the recurrence alone, and it is the one fact in this solution that goes beyond A-level: it is the statement that $\mathbb{Z}[\sqrt{3}] \cap \mathbb{Q} = \mathbb{Z}$, a special case of $\mathbb{Z}$ being integrally closed.

> **Caveat.** The general fact "$\mathbb{Z}[\sqrt{d}]$ is integrally closed, so its rational elements are ordinary integers" is used here in the mild form needed (a linear combination of powers of α, β with integer coefficients that happens to be rational is an integer); this special case is elementary and directly verifiable by writing $A + B\sqrt{3} \in \mathbb{Q} \Rightarrow B = 0$, and does not require the general theory.

Once both facts are in hand, choosing $n = 5$ is simply reading off the smallest divisor of $2015 = 5 \cdot 13 \cdot 31$ for which the cofactor $2015/n$ is odd – every cofactor here is automatically odd, since 2015 itself is odd, so the smallest available $n > 1$ (namely 5) is the cheapest term to compute by hand.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the recurrence's constant term: $a_n = 4a_{n-1} - a_{n-2}$ has characteristic polynomial $x^2 - 4x + 1$, constant term 1. On seeing a second-order linear recurrence with integer coefficients and constant term ± 1, write down: "the roots are reciprocals; this sequence may have index-divisibility properties like Fibonacci."

2. That cue suggests hunting for a divisibility law $a_n \mid a_{kn}$ (for the appropriate parity of k) via the algebraic identity $u^k + v^k = (u+v)(\cdots)$ or $u^k - v^k = (u-v)(\cdots)$, rather than attempting to compute or characterise a_{2015} itself.

3. Hesitation occurs right after the closed form is found: it is tempting to treat "I have a formula for a_n" as "I am nearly done", and to start probing $a_{2015} \bmod$ small primes directly. What reinforces the correct turn is asking explicitly, before computing anything further, "what is the smallest possible object I could compute that would still answer the question?" – the question only demands one prime factor, not the value, and that reframing is what redirects effort toward the index-divisibility route.

4. The trap is the sheer size of a_{2015}: it invites a student to believe that finding a factor requires understanding the whole number, when in fact 2015's own factorisation ($5 \times 13 \times 31$) hands over a small, computable divisor of a_{2015} for free, via a structural property of the sequence rather than any arithmetic on the huge number itself.

5. Recognise it faster by treating "constant term ± 1 in a linear recurrence's characteristic polynomial" as a standing trigger for index-divisibility, exactly as "Fibonacci-like" should be. The moment a problem asks about a specific, enormous index of such a sequence and only wants a *factor* (not the value), factor the index instead of the term.

6. Reusable rule: for a linear recurrence with characteristic roots α, β satisfying $\alpha\beta = \pm 1$, and

n a divisor of N with N/n of the right parity, $a_n \mid a_N$ via the sum/difference-of-powers factorisation combined with the rationality-forces-integrality argument in $\mathbb{Z}[\sqrt{d}]$. When asked for a factor (not the value) of a distant term, factor the index and compute the smallest resulting term instead of the term itself.

- ▷ **Cue:** second-order linear recurrence, integer coefficients, characteristic polynomial with constant term ± 1 $\longrightarrow$ **Move:** recognise the roots are reciprocals ($\alpha\beta = \pm 1$) and look for index-divisibility, $a_n \mid a_{kn}$ for appropriate k
- ▷ **Cue:** asked for a factor, not the value, of a huge term a_N of such a sequence $\longrightarrow$ **Move:** factor N and search for a small divisor $n \mid N$ with the right parity of N/n; compute a_n instead of a_N
- ▷ **Cue:** a quantity is simultaneously visibly an algebraic integer in $\mathbb{Z}[\sqrt{d}]$ and visibly rational (a ratio of two known integers) $\longrightarrow$ **Move:** conclude it is an ordinary integer, since $\mathbb{Z}[\sqrt{d}] \cap \mathbb{Q} = \mathbb{Z}$

2015 A3

2015 A3. Compute

$$\log_2 \left(\prod_{a=1}^{2015} \prod_{b=1}^{2015} (1 + e^{2\pi i ab/2015}) \right)$$

Here i is the imaginary unit (that is, $i^2 = -1$).

Fields: roots of unity, cyclotomic polynomials, number theory (gcd sums, Dirichlet convolution) **Position:** A3 **Answer:** The answer is 13725, obtained as $\sum_{a=1}^{2015} \gcd(a, 2015) = (2 \cdot 5 - 1)(2 \cdot 13 - 1)(2 \cdot 31 - 1) = 9 \cdot 25 \cdot 61$.

Layer 1 THE WHAT — surface analysis

The natural attempt. Attack the factor $1 + e^{i\theta}$ directly. Since $1 + e^{i\theta} = 2\cos(\theta/2)\, e^{i\theta/2}$, each term in the double product contributes a magnitude $|2\cos(\pi ab/2015)|$ and a phase $\pi ab/2015 \pmod{2\pi}$ (up to a sign from cos being negative). To get $\log_2$ of the whole product, first argue the total product is a positive real number – not obvious termwise, since individual cosines can be negative – then take $\log_2$ of the product of the $2|\cos(\pi ab/2015)|$ factors, summed over all 2015^2 pairs (a, b).

This is technically sound in principle but it immediately produces a double sum of $\log_2 |2\cos(\pi ab/2015)|$ over more than four million pairs, with no evident closed form for $\sum_b \log_2 |2\cos(\pi ab/2015)|$ that isn't itself exactly as hard as the original product. Worse, establishing that the product is real and positive at all – so that $\log_2$ of it is even well-defined as a real number – requires tracking 4,060,225 phases $\theta_{ab} = \pi ab/2015$ and showing their total, after the sign corrections from negative cosines, lands on a multiple of 2π. This is the wall: the trigonometric decomposition is faithful to the algebra but gives no purchase on the sum.

The model chain. Fix a and evaluate the inner product $\prod_{b=1}^{2015}(1 + \omega^{ab})$ where $\omega = e^{2\pi i/2015}$, by relating it to a cyclotomic factorisation rather than to trigonometry. Write $d = \gcd(a, 2015)$, $a = da_1$, $2015 = dn_1$ with $\gcd(a_1, n_1) = 1$. Since $\gcd(a_1, n_1) = 1$, multiplication by a_1 permutes the residues mod n_1: as b runs over $1, \ldots, n_1$, $a_1 b \bmod n_1$ runs over $1, \ldots, n_1$ in some order. So with $\zeta = e^{2\pi i/n_1}$ (an n_1-th root of unity),

$$\prod_{b=1}^{n_1}(1 + e^{2\pi i ab/2015}) = \prod_{b=1}^{n_1}(1 + e^{2\pi i a_1 b/n_1}) = \prod_{b=1}^{n_1}(1 + \zeta^b).$$

The numbers $\zeta, \zeta^2, \ldots, \zeta^{n_1}$ are precisely the n_1 roots of $z^{n_1} - 1$, so $z^{n_1} - 1 = \prod_{b=1}^{n_1}(z - \zeta^b)$; setting

$z = -1$ and using that n_1 is odd (so $(-1)^{n_1} = -1$) gives

$$\prod_{b=1}^{n_1}(1+\zeta^b) = (-1)^{n_1+1}\big((-1)^{n_1}-1\big) = 2.$$

Since the full range $b = 1, \ldots, 2015$ consists of d repetitions of the pattern over a period-n_1 residue system (as ab mod 2015 only depends on b mod n_1), $\prod_{b=1}^{2015}(1+e^{2\pi iab/2015}) = 2^d = 2^{\gcd(a,2015)}$.

Therefore $\log_2 \prod_{b=1}^{2015}(1+e^{2\pi iab/2015}) = \gcd(a,2015)$ exactly, no approximation, and the whole sum collapses to

$$\log_2\left(\prod_{a=1}^{2015}\prod_{b=1}^{2015}(1+e^{2\pi iab/2015})\right) = \sum_{a=1}^{2015}\gcd(a,2015).$$

This last sum is standard: group by $d \mid 2015$, since exactly $\phi(2015/d)$ values of $a \in \{1, \ldots, 2015\}$ have $\gcd(a,2015) = d$, giving $\sum_{d|2015} d\,\phi(2015/d)$. With $2015 = pqr = 5 \cdot 13 \cdot 31$ squarefree, this convolution sum is multiplicative and evaluates on a prime p to $1 \cdot \phi(p) + p \cdot \phi(1) = (p-1) + p = 2p - 1$, so on the full product it is $(2p-1)(2q-1)(2r-1) = 9 \cdot 25 \cdot 61 = 13725$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains begin by isolating a single factor $1 + e^{i\theta}$ and both correctly rewrite it – the natural attempt as $2\cos(\theta/2)e^{i\theta/2}$, the model chain (implicitly) as a root of the polynomial $z^{n_1} - 1$ evaluated at $z = -1$. The split is which algebraic identity you reach for to describe that factor: a trigonometric half-angle identity, aimed at a magnitude and a phase, versus a polynomial-root identity, aimed at a product over a whole cyclotomic factor at once.

> *I am being asked for* $\log_2$ *of a product; that phrasing is telling me the product is going to collapse to a power of* 2*, so I should look for a mechanism that produces* $2^{something}$ *termwise, not a mechanism that produces a magnitude and a phase for each factor separately.*

> *The exponents* ab*, as* b *ranges over a full period mod* $n_1 = 2015/\gcd(a,2015)$*, and* $\gcd(a_1, n_1) = 1$*: multiplication by a unit permutes residues. So the set* $\{1+\omega^{ab} : b = 1, \ldots, n_1\}$ *is, as a* multiset*, exactly* $\{1+\zeta^b : b = 1, \ldots, n_1\}$ *– I can replace "the specific exponents that show up" with "all the* n_1*-th roots of unity", and a product over* all *roots of unity of a fixed shifted polynomial is something I already know how to evaluate.*

Layer 3 THE WHY

Why is the cyclotomic route the one that is natural here, rather than a stroke of luck? Because the object being multiplied, $1 + z$ evaluated at $z =$ a root of unity, is a linear polynomial in z, and a product of $(z -$ root$)$ over *all* roots of a polynomial is exactly what the polynomial itself factors into. The identity $z^n - 1 = \prod_{b=1}^{n}(z - \zeta^b)$ is not a special fact about this problem; it is the definition of $z^n - 1$ splitting over $\mathbb{C}$, restated. The moment you notice that $b \mapsto ab$ mod n_1 is a permutation of $\{1, \ldots, n_1\}$ (because $\gcd(a, n_1) = 1$ after factoring out $d = \gcd(a,2015)$), the product $\prod_b(1+\omega^{ab})$ *is* $\prod_b(1+\zeta^b)$, which is a product over literally every root of $z^{n_1} - 1$ – and that means it is $\pm(z^{n_1} - 1)$ evaluated at $z = -1$, up to a sign fixed by the degree. There is nothing to search for beyond recognising "a product over all the roots of a polynomial" as the phrase that triggers "evaluate the polynomial itself".

The trigonometric route, by contrast, treats each factor as an isolated complex number to be described (magnitude, phase) rather than as a root feeding into a polynomial it belongs to. That framing is not wrong, but it discards the one piece of structure – that the full set of exponents forms a complete, permuted residue system – which is precisely what makes the whole set of

factors summable in closed form. The instruction "the answer is a power of 2, find $\log_2$ of it" is itself a clue pointing at the polynomial route: $z^n - 1$ at $z = -1$ is $(-1)^n - 1$, which for odd n is -2, a power of 2 up to sign – exactly the kind of clean output the cyclotomic identity produces and the trigonometric decomposition does not.

The final gcd-sum, $\sum_{a=1}^{2015} \gcd(a, 2015)$, is where Kedlaya's solution is at its most terse: it simply states the identity $\sum_{d|n} d\phi(n/d)$ and evaluates it. Here is where that comes from: grouping the integers $1, \ldots, 2015$ by the value of $\gcd(a, 2015) = d$ is grouping by divisor, and for fixed $d \mid 2015$, the values $a = d, 2d, \ldots$ with $\gcd(a, 2015) = d$ correspond exactly to $a/d \in \{1, \ldots, 2015/d\}$ coprime to $2015/d$ – of which there are $\phi(2015/d)$ by the definition of Euler's totient. That $f(n) = \sum_{d|n} d\phi(n/d)$ is multiplicative follows because it is the Dirichlet convolution of the two multiplicative functions $\mathrm{id}(n) = n$ and $\phi(n)$, and a Dirichlet convolution of multiplicative functions is multiplicative – this is why it suffices to evaluate f at each prime power and multiply, rather than expanding the eight-term sum for pqr by hand (as the official solution does, for concreteness, but does not need to).

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the request for $\log_2$ of a product, not the product itself. On seeing it, write down: "the answer is meant to be an exact power of 2 – find the mechanism that produces 2^{integer} from each piece, rather than a general complex number."

2. That cue suggests treating each factor $1 + \omega^{ab}$ as a root feeding a polynomial (specifically, looking for the full set of roots of some $z^{n_1} - 1$ among the exponents present), rather than decomposing it trigonometrically into magnitude and phase.

3. Hesitation occurs at the very first step, choosing how to rewrite $1 + e^{i\theta}$: the half-angle form $2\cos(\theta/2)e^{i\theta/2}$ is the more familiar identity and the one most readily recalled under time pressure, so it is the default reach. What reinforces the switch to the cyclotomic view is noticing, before committing to either, that the inner product ranges over *all* $b = 1, \ldots, 2015$ – a complete residue system – and that "product over a complete residue system" is a much stronger and more specific structure than "a single complex number to simplify"; that specificity is worth pausing on before picking an identity.

4. The trap is the apparent innocence of $1 + e^{i\theta} = 2\cos(\theta/2)e^{i\theta/2}$: it is true and it is the identity most students reach for on autopilot for any $1 + e^{i\theta}$ expression, but applying it here discards the residue-system structure of the exponents ab mod 2015, without which the sum of 4,060,225 log-cosine terms has no accessible closed form.

5. Recognise it faster by treating "product of $(1 + \zeta)$ (or any linear expression in ζ) over ζ ranging across a complete or near-complete set of roots of unity" as a standing cue for the cyclotomic factorisation $z^n - 1 = \prod(z - \zeta^b)$, evaluated at the appropriate point, rather than for any per-term trigonometric simplification.

6. Reusable rule: when a product runs over ω^{ab} for b across a full residue system mod n and $\gcd(a, n) = 1$, replace the specific exponents by "all n-th roots of unity" (multiplication by a unit permutes residues) and evaluate the resulting product as a cyclotomic polynomial identity, $\prod_{b=1}^{n}(z - \zeta^b) = z^n - 1$, at the relevant point z – this converts any function of ω^{ab} that is polynomial in ω^{ab} into a closed form in one step, independent of the specific a.

▷ **Cue:** a product runs over $1 + \omega^{ab}$ (or any polynomial in ω^{ab}) for b across $1, \ldots, n$, with $\gcd(a, n) = 1$ $\longrightarrow$ **Move:** recognise $b \mapsto ab$ mod n permutes residues, replace the product by one over all n-th roots of unity, and evaluate via $z^n - 1 = \prod_{b=1}^{n}(z - \zeta^b)$

▷ **Cue:** a problem asks for $\log_2$ (or $\log_p$) of a product, rather than the product itself $\longrightarrow$ **Move:** expect the answer to be an exact power of 2 (or p); look for a mechanism that produces that power termwise, such as a cyclotomic evaluation at $z = -1$

▷ **Cue:** a sum $\sum_{a=1}^{n} \gcd(a, n)$, or any sum of a function of $\gcd(a, n)$ over $a = 1, \ldots, n$ $\longrightarrow$ **Move:** group by divisor $d \mid n$, use that exactly $\phi(n/d)$ values of a have $\gcd(a, n) = d$, and recognise the resulting sum as a Dirichlet convolution, hence multiplicative – evaluate on prime powers and multiply

2015 A4

2015 A4. For each real number x, let

$$f(x) = \sum_{n \in S_x} \frac{1}{2^n},$$

where S_x is the set of positive integers n for which $\lfloor nx \rfloor$ is even. What is the largest real number L such that $f(x) \geq L$ for all $x \in [0,1)$? (As usual, $\lfloor z \rfloor$ denotes the greatest integer less than or equal to z.)

Fields: real analysis (infimum of a function defined by a series); combinatorics on words (binary expansions, lexicographic order); Beatty-type sequences $\lfloor nx \rfloor$ **Position:** A4 — mid-paper. Short once seen, but the answer has to be guessed before it can be proved, and the proof has two halves (a bound and a near-extremal family). **Answer:** $L = 4/7$. The infimum is approached as $x \to 2/3$ from below and is never attained.

Layer 1 THE WHAT — surface analysis

The natural attempt. You read "largest L such that $f(x) \geq L$ for all x" as "compute inf f", and the obvious way to find an infimum is to evaluate the function. So you tabulate. $x = 0$: every floor is 0, $S_x = \mathbb{N}$, $f = 1$. $x = 1/2$: the floors are $0, 1, 1, 2, 2, 3, 3, \ldots$, even at $n = 1, 4, 5, 8, 9, \ldots$, and $f = \frac{1}{2} + \frac{1}{16} + \frac{1}{32} + \cdots$, which sums to $\frac{1}{2} + \frac{3/32}{1-1/16} = 0.6$. $x = 2/3$: floors $0, 1, 2, 2, 3, 4, 4, 5, 6, \ldots$, even at $n = 1, 3, 4, 6, 7, \ldots$, and $f = \frac{1}{2} + \frac{3/16}{1-1/8} = \frac{5}{7} \approx 0.71$. x near 1: floors $0, 1, 2, 3, \ldots$ for a long stretch, even at $n = 1, 3, 5, \ldots$, and $f \approx 2/3$.

Now the student tries to treat f as a function and minimise it. It is a step function with a jump at every rational p/n — every point where some nx crosses an integer — so derivatives are useless and there is no finite list of candidate points. The tabulation says the minimum is somewhere around 0.57 to 0.6, and the student starts evaluating at $x = 3/5, 5/8, 7/11, \ldots$, each a new infinite series, each a slightly different number. Forty minutes in, there is a table of decimals and no idea which one is the infimum, let alone a proof that nothing lower exists.

The wall is structural: treating $f(x)$ as a real-valued function of a real variable throws away the one thing that makes it computable, which is that $f(x)$ is a binary number whose digits are the indicator of S_x.

The model chain. Step 1: read $f(x)$ as a binary expansion. For $S \subseteq \mathbb{N}$ write $F(S) = \sum_{n \in S} 2^{-n}$. Then $F(S)$ is the number $0.s_1 s_2 s_3 \ldots$ in base 2, where $s_n = 1$ if $n \in S$. Every term 2^{-n} equals the sum of all later terms, $\sum_{m>n} 2^{-m}$. Consequently: if $S \neq S'$ and m is the least element of the symmetric difference, and $m \in S$, then

$$F(S) - F(S') \;\geq\; 2^{-m} - \sum_{k>m} 2^{-k} \;=\; 0,$$

with equality only if S' contains every integer beyond m and S contains none of them. So minimising F is minimising a binary string lexicographically: make the early digits 0.

Step 2: which digits can be 0? Write $s_n(x)$ for the indicator of $n \in S_x$.

- $n = 1$: $\lfloor x \rfloor = 0$ for $x \in [0,1)$, so $s_1 = 1$ always.
- Consecutive floors: since $0 \leq x < 1$, $\lfloor (n+1)x \rfloor - \lfloor nx \rfloor \in \{0, 1\}$.
- Three consecutive zeros are impossible once $s_2 = 0$. If $s_2 = 0$ then $\lfloor 2x \rfloor = 1$, so $x \geq 1/2$. If $\lfloor nx \rfloor, \lfloor (n+1)x \rfloor, \lfloor (n+2)x \rfloor$ were all odd, then, since each step is 0 or 1 and a step of 1 changes parity, both steps are 0, so all three floors are equal; but $(n+2)x - nx = 2x \geq 1$, so the first and last floors differ. Contradiction.

So the lexicographically smallest conceivable string is $1\,0\,0\,1\,0\,0\,1\,0\,0\ldots$: $S = T := \{1, 4, 7, 10, \ldots\}$, with

$$F(T) = \frac{1/2}{1 - 1/8} = \frac{4}{7}.$$

Step 3: the lower bound. Suppose $f(x) < 4/7$ for some $x \in [0, 1)$. Then $S_x \neq T$; let m be the least element of the symmetric difference. By Step 1, $m \in T \setminus S_x$ (otherwise $f(x) \geq 4/7$). Since $1 \in S_x \cap T$, $m = 3j + 1$ with $j \geq 1$, and S_x agrees with T below m. Hence $2 \notin S_x$ and $3j - 1, 3j, 3j + 1 \notin S_x$: three consecutive zeros with $s_2 = 0$, impossible by Step 2. So $f(x) \geq 4/7$ for all x.

In fact the inequality is strict. Equality in Step 1 would need one of S_x, T to be finite. T is infinite, and S_x is infinite too: if $x = 0$ then $S_x = \mathbb{N}$, and if $x > 0$ the floors increase without bound in steps of 0 or 1, so they pass through every even integer. And $S_x = T$ exactly is impossible, since it would force $\lfloor 3kx \rfloor = 2k - 1$ for every k, i.e. $\frac{2k-1}{3k} \leq x < \frac{2}{3}$ for all k, and the left side tends to $2/3$.

Step 4: $4/7$ is sharp. Take $x = \frac{2}{3} - \varepsilon$ with $0 < \varepsilon < \frac{1}{9n}$. For $1 \leq k \leq n$ every multiple $r\varepsilon$ with $r \leq 3n$ lies in $(0, \frac{1}{3})$, so

$$\lfloor (3k-2)x \rfloor = \left\lfloor 2k - 2 + \tfrac{2}{3} - (3k-2)\varepsilon \right\rfloor = 2k - 2, \quad \lfloor (3k-1)x \rfloor = 2k - 1, \quad \lfloor 3kx \rfloor = \left\lfloor 2k - 3k\varepsilon \right\rfloor = 2k - 1.$$

So $S_x \cap \{1, \ldots, 3n\} = \{1, 4, \ldots, 3n - 2\}$, and

$$f(x) \;\leq\; \sum_{k=1}^{n} 2^{-(3k-2)} + \sum_{m>3n} 2^{-m} \;=\; \sum_{k=1}^{n} 2^{-(3k-2)} + 2^{-3n} \;\xrightarrow[n\to\infty]{}\; \frac{4}{7}.$$

Hence no $L > 4/7$ works, and $L = 4/7$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start at the same place: "I need $\inf_x f(x)$." At (†) — deciding what kind of object $f(x)$ is — they split. The natural attempt treats f as a function of x and samples it. The model treats $f(x)$ as a *binary expansion* $0.s_1s_2s_3\ldots$ and asks which digit strings x can produce. After (†) the model never evaluates f at a single point until the extremal string $1001001\ldots$ has already been identified.

> *Every term is $1/2^n$ and each term is a yes/no. That's a binary number, and its digits are the parities of $\lfloor nx \rfloor$. Making a binary number small means making its* first *digits zero — so I don't need to understand f everywhere, I need to know how early and how often the digit string can be forced to 0.*

> *Digit 1 is always 1. To get digit 2 to be 0 I need $x \geq 1/2$, and then the floors move by at least 1 every two steps, so I can never have three zeros in a row. The best I can hope for is $100\,100\,100\ldots$, which is $4/7$. Now check whether some x nearly produces it: I want floors $0, 1, 1, 2, 3, 3, \ldots$, i.e. slope $2/3$, approached from below.*

Layer 3 THE WHY

Why is the binary reading not a trick? Because of what $\sum_{n\in S} 2^{-n}$ *is*. By definition of base-2 notation, $\sum_{n\in S} 2^{-n}$ is the real number whose nth binary digit is 1 exactly when $n \in S$. The problem writes a binary expansion and calls it f. The ratio $1/2$ is not incidental: it is precisely the ratio for which "one term outweighs the entire tail" holds with equality, $2^{-n} = \sum_{m>n} 2^{-m}$. With ratio $1/3$ it would hold strictly and comparison would be even cleaner; with ratio $2/3$ it would fail, and this problem would be a genuinely different (and much harder) optimisation. So the cue is the ratio $1/2$ attached to an indicator. Once seen, the order on values is the dictionary order on digit strings, and minimisation is a greedy question: how many leading zeros can I buy, then how many more?

Why does the greedy question have a short answer? Because $\lfloor nx \rfloor$ for $0 \le x < 1$ is a *staircase with steps of height* 0 *or* 1. That single structural fact is the whole constraint system. The parity string of a staircase flips exactly where the staircase steps up, so a run of r equal parities is a run of r equal floor values, which needs $(r-1)x < 1$. Digit 1 is forced to be 1 (the staircase starts at 0). The first zero we can buy, at $n = 2$, costs $x \ge 1/2$ — and that very cost forbids runs of three odd floors. The two constraints fight, and 100 repeating is where they balance. Kedlaya's solution introduces $T = \{1, 4, 7, \dots\}$ in its first line without comment; here is where it comes from. It is the lexicographic minimum of the strings the staircase is allowed to produce.

Why 2/3? A staircase with parity pattern even, odd, odd, repeating, climbs 2 units every 3 steps: slope $2/3$. At exactly $2/3$ the third step lands *on* an integer and the floor jumps early (floor of 2 is 2, even), which is why the extremum is approached from below and not attained. That is also why this is an infimum problem phrased as "largest L": the setters knew the minimum does not exist.

Caveat. The strict inequality $f(x) > 4/7$ (non-attainment) is proved in Step 3 above; it is not needed for the answer but it explains the phrasing. I also checked numerically: evaluating f (truncated at $n = 80$) on the grid $x = k/20000$ gives minimum $0.571428\dots$ near $x = 0.6624$, and $f(2/3 - 10^{-5})$ agrees with $4/7$ to sixteen digits.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the summand $1/2^n$ attached to membership in a set. Write down immediately: $f(x) = 0.s_1 s_2 s_3 \dots$ in binary, with $s_n = 1$ iff $\lfloor nx \rfloor$ is even, and $2^{-n} = \sum_{m>n} 2^{-m}$.

2. Lexicographic (greedy) minimisation of a digit string subject to constraints, followed by a near-extremal construction to show sharpness.

3. Hesitation comes after guessing $4/7$: "the infimum is not attained, so how do I prove it's the largest L?" Reinforce with the standard two-part template for a sharp bound: (i) prove $f \ge L$ for all x, (ii) exhibit x_n with $f(x_n) \to L$. There is no need for the infimum to be a minimum.

4. Two traps. First, the sampling trap: f is a step function with a jump at every rational, so evaluating it at nice points (like $x = 2/3$ itself, where $f = 5/7$) points you *away* from the answer. Second, a rigour trap in the construction: at $x = 2/3 - \varepsilon$ the floor pattern is only correct up to $n \approx 1/(3\varepsilon)$, so ε must be tied to n, and the tail beyond $3n$ must be bounded crudely (by 2^{-3n}), not ignored.

5. Slogan: *a sum of* 2^{-n} *over a set is a binary number; compare binary numbers by their first differing digit.* More generally, whenever a quantity is a weighted sum with geometrically dominating weights (each weight at least the sum of all later ones), the optimisation is lexicographic, and you should optimise the first coordinate, then the second, and so on.

6. Decision rule: **before optimising a function, ask whether its values are ordered by a simpler combinatorial object (a digit string, a sequence, a partition); if so, optimise the object greedily and only then build a real input that nearly realises it.** Companion rule: $\lfloor nx \rfloor$ with $0 \le x < 1$ is a staircase with steps 0 or 1, and a run of r equal values costs $(r-1)x < 1$.

▷ **Cue:** a sum $\sum_{n \in S} 2^{-n}$ (or any weights where each weight at least equals the sum of all later weights) $\longrightarrow$ **Move:** read it as a binary number; the order on values is lexicographic order on the indicator string, so optimise greedily from the first digit

▷ **Cue:** $\lfloor nx \rfloor$ for $0 \le x < 1$ and a condition on parity or residue $\longrightarrow$ **Move:** consecutive floors differ by 0 or 1; a run of r equal floors forces $(r-1)x < 1$

▷ **Cue:** "find the largest L with $f(x) \ge L$ for all x" and the candidate extremal pattern is only approached $\longrightarrow$ **Move:** prove the bound for all x, then give a family x_n with $f(x_n) \to L$; tie the perturbation ε to n

▷ **Cue:** a periodic target pattern for $\lfloor nx \rfloor$ that climbs a units every b steps $\longrightarrow$ **Move:** try $x = a/b$ perturbed slightly downward or upward, depending on whether the pattern needs

the floor to jump late or early

2015 A5

2015 A5. Let q be an odd positive integer, and let N_q denote the number of integers a such that $0 < a < q/4$ and $\gcd(a, q) = 1$. Show that N_q is odd if and only if q is of the form p^k with k a positive integer and p a prime congruent to 5 or 7 modulo 8.

Fields: elementary number theory: inclusion–exclusion with the Möbius function, parity arguments, characters of $(\mathbb{Z}/8\mathbb{Z})^\times$; (in the alternative route) Gauss's lemma and the quadratic character of -2
Position: A5 — hard but short. The whole difficulty is deciding to count modulo 2 from the start.
Answer: A proof. $N_q \equiv \#\{\text{squarefree } d \mid q : q/d \equiv 5,7 \pmod 8\} \pmod 2$, and that count is odd exactly for $q = p^k$, $p \equiv 5,7 \pmod 8$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You compute. $q = 5$: $a \in \{1\}$, $N_5 = 1$, odd, and $5 \equiv 5$. $q = 7$: $a = 1$, $N_7 = 1$. $q = 9$: $a \in \{1, 2\}$, $N_9 = 2$. $q = 11$: $a \in \{1, 2\}$, $N_{11} = 2$, and $11 \equiv 3$. $q = 13$: $a \in \{1, 2, 3\}$, $N_{13} = 3$. $q = 15$: $a \in \{1, 2\}$, $N_{15} = 2$ — even, although $15 \equiv 7 \pmod 8$, so the "prime power" clause is doing real work. $q = 25$: $a \in \{1, 2, 3, 4, 6\}$, $N_{25} = 5$, odd.

The table confirms the statement but does not suggest a mechanism. The next instinct, for a prime $q = p$, is the exact count: every a with $0 < a < p/4$ is coprime to p, so $N_p = \lfloor p/4 \rfloor$. That is odd iff $p \equiv 5,7 \pmod 8$. Good — the prime case is one line. Then $q = p^k$: remove the multiples of p, $N_{p^k} = \lfloor p^k/4 \rfloor - \lfloor p^{k-1}/4 \rfloor$, and you check the parities by cases on p mod 8 and k. Still manageable. Then $q = pr$ with two primes: $N_{pr} = \lfloor pr/4 \rfloor - \lfloor r/4 \rfloor - \lfloor p/4 \rfloor + \lfloor 1/4 \rfloor$, and now there are sixteen residue combinations of (p, r) mod 8 to check, and for three primes sixty-four, and with exponents the tree keeps branching. The student who ploughs on is doing inclusion–exclusion as an *exact* count, carrying integers when only their parities matter, and runs out of time in the general case.

The model chain. *Step 1: inclusion–exclusion.* For $d \mid q$, the number of multiples of d in $(0, q/4)$ is $\lfloor q/(4d) \rfloor$ (as q is odd, $q/4$ is never an integer, so the open and closed counts agree, and $\lfloor \lfloor q/4 \rfloor / d \rfloor = \lfloor q/(4d) \rfloor$ in any case). By Möbius inversion over the divisors of q,

$$N_q = \sum_{d \mid q} \mu(d) \left\lfloor \frac{q/d}{4} \right\rfloor,$$

where $\mu(d) = (-1)^r$ if d is a product of r distinct primes and $\mu(d) = 0$ otherwise.

Step 2: reduce modulo 2. Since $-1 \equiv 1 \pmod 2$, $\mu(d) \equiv 1$ for squarefree d and 0 otherwise:

$$N_q \equiv \sum_{\substack{d \mid q \\ d \text{ squarefree}}} \left\lfloor \frac{q/d}{4} \right\rfloor \pmod 2.$$

Step 3: the parity of $\lfloor m/4 \rfloor$ for odd m. Write $m = 8t + s$ with $s \in \{1, 3, 5, 7\}$. Then $\lfloor m/4 \rfloor = 2t + \lfloor s/4 \rfloor$, which is even for $s = 1, 3$ and odd for $s = 5, 7$. Define $\chi(m) = +1$ if $m \equiv 1, 3 \pmod 8$ and $\chi(m) = -1$ if $m \equiv 5, 7 \pmod 8$. So

$$N_q \equiv \#\{d \mid q \text{ squarefree} : \chi(q/d) = -1\} \pmod 2.$$

Step 4: χ is multiplicative. The set $\{1, 3\}$ is a subgroup of the multiplicative group $\{1, 3, 5, 7\}$ modulo 8 (as $3^2 = 9 \equiv 1$), of index 2; χ is the homomorphism to $\{\pm 1\}$ with that kernel. Hence $\chi(mn) = \chi(m)\chi(n)$ for odd m, n, and since $\chi(d)^2 = 1$, $\chi(q/d) = \chi(q)\chi(d)$.

Step 5: count. Let $p_1, \ldots, p_r$ be the distinct primes dividing q. Squarefree divisors d correspond to subsets $I \subseteq \{1, \ldots, r\}$ via $d = \prod_{i \in I} p_i$, and $\chi(d) = \prod_{i \in I} \chi(p_i)$. We want the number of I with $\chi(d) = -\chi(q)$.

- $r = 0$ ($q = 1$): the only d is 1, $\chi(1) = 1$, the count is 0. Even.
- All $\chi(p_i) = +1$: then every $\chi(d) = +1$ and $\chi(q) = +1$, so the count is 0. Even.
- Some $\chi(p_j) = -1$: toggling $j \in I$ flips the sign of $\chi(d)$, a bijection between the subsets with $\chi(d) = +1$ and those with $\chi(d) = -1$. Each class has 2^{r-1} members, so the count is 2^{r-1}, odd iff $r = 1$.

So N_q is odd iff $r = 1$ and $\chi(p_1) = -1$: that is, iff $q = p^k$, $k \geq 1$, $p \equiv 5, 7 \pmod 8$. ■

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains write inclusion–exclusion, $N_q = \sum_{d|q} \mu(d) \lfloor q/(4d) \rfloor$. At (†) — the next line — they split. The natural attempt evaluates the sum as integers, case by case on residues of primes. The model reads the question again ("N_q is odd") and reduces the sum modulo 2 before evaluating anything, which kills the signs of μ and turns each term into a bit depending only on q/d mod 8.

> *The question only asks whether N_q is odd. So I should never compute N_q. I should compute it mod 2 from the very first line — and mod 2 the Möbius signs vanish, so the whole sum is just "how many squarefree cofactors q/d have $\lfloor (q/d)/4 \rfloor$ odd".*

> *$\lfloor m/4 \rfloor$ is odd exactly when $m \equiv 5, 7 \pmod 8$ — that's where the "5 or 7 mod 8" in the statement comes from. And $\{1, 3\}$ is closed under multiplication mod 8, so this is a sign character. Counting subsets by the sign of a product of ± 1's: one factor equal to -1 splits them exactly in half.*

Layer 3 THE WHY

Why reduce modulo 2? Because the problem's phrasing tells you nothing else will be asked. "Show that N_q is odd if and only if…" is a statement about N_q mod 2, and reduction mod 2 is a ring homomorphism: it commutes with sums and products, so you may apply it at any stage of a computation — and the earlier you apply it, the more information you are permitted to throw away. Here the first thing it destroys is the sign of μ, which is the entire source of the combinatorial mess in the natural attempt. It is the same reason you work mod 9 to check digit sums: you ask only the question you were asked.

Why must $\lfloor m/4 \rfloor$ then be examined mod 8? Because replacing m by $m + 8$ adds exactly 2 to $\lfloor m/4 \rfloor$, so the *parity* of $\lfloor m/4 \rfloor$ is periodic in m with period $2 \cdot 4 = 8$. That is a general fact — the parity of $\lfloor m/k \rfloor$ depends on m mod $2k$ — and it explains, from the definitions alone, why the statement is about residues mod 8. The statement itself advertises this: "5 or 7 modulo 8" is the upper half of the odd residues, exactly the ones with $\lfloor s/4 \rfloor = 1$.

Why is multiplicativity the next thing to check? Because after Step 3 you are counting divisors by a property of the *cofactor* q/d, and divisors of q are products of primes. The only properties of a product that you can read off its factors are multiplicative ones. So the question "is χ a homomorphism?" is forced. It is: $\{1, 3\}$ is a subgroup of $(\mathbb{Z}/8)^\times$. Readers who know quadratic reciprocity will recognise $\chi(p) = \left(\frac{-2}{p}\right)$, the Legendre symbol of -2, which is $+1$ exactly for $p \equiv 1, 3 \pmod 8$. That is not a coincidence, and it is the content of the official second solution.

Kedlaya's first solution handles Step 5 by pairing divisors c, pc when some prime is 1 or 3 mod 8, and by quadruples when two primes are 5 or 7; it is correct but reads as a case split. Here is where it comes from: it is the single fact that $\sum_I \prod_{i \in I} \chi(p_i) = \prod_i (1 + \chi(p_i))$, which vanishes as soon as one $\chi(p_i) = -1$. A factor $(1 - 1)$ in a product is the algebraic form of "toggle p_j in and out".

The official second solution goes the other way: it multiplies the even residues $2, 4, \ldots, q-1$ coprime to q, reduces them into $(-q/2, q/2)$, and derives $(-1)^{N_q} \equiv (-2)^{\varphi(q)/2} \pmod q$ — Gauss's lemma for -2. Where does that come from? From asking "what natural product has N_q sign changes?" If you already know Gauss's lemma, the phrase "number of $a < q/4$" is its fingerprint, since halving the even residues above $q/2$ lands exactly in $(q/4, q/2)$. The two solutions together prove that -2 is a square mod p iff $p \equiv 1, 3 \pmod 8$, as the official remark notes.

Caveat. Verified by direct computation: for every odd $q < 4000$, N_q is odd exactly when q is a prime power p^k with $p \equiv 5, 7 \pmod 8$, and χ was checked to be multiplicative on odd residues. The proof above is complete and does not depend on the computation.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the conclusion "N_q is odd if and only if". Write down: *I will compute N_q mod 2 only.* The second cue is "$\gcd(a, q) = 1$" on an interval: write $N_q = \sum_{d|q} \mu(d) \lfloor q/(4d) \rfloor$.

2. Inclusion–exclusion, then reduction mod 2 (which turns μ into the indicator of squarefree), then counting subsets of primes by a multiplicative sign.

3. Hesitation occurs at $q = 15$, where $15 \equiv 7 \pmod 8$ but N_{15} is even: the student wonders whether the residue of q is the right invariant. It is not; the residues of the *cofactors* q/d are. Reinforce by writing the parity sum out for $q = 15$: cofactors $15, 5, 3, 1$ have $\chi = -1, -1, +1, +1$, two minus signs, even.

4. The trap is exact computation: running inclusion–exclusion over integers and splitting by residues of each prime, which branches exponentially in the number of prime factors. A smaller trap: forgetting $q = 1$, where $N_1 = 0$ and the prime-power clause needs $k \geq 1$.

5. Slogan: *if only the parity (or any residue) is asked, reduce at the first line.* And: the parity of $\lfloor m/k \rfloor$ is a function of m mod $2k$; a sign that depends on a residue class and is closed under multiplication is a character, and sums of characters over subsets factor as $\prod(1 + \chi(p))$.

6. Decision rule: **match the modulus of your computation to the modulus of the question, and apply it before the first simplification, not after.** Companion rule: when counting divisors by a property of a product, check first whether the property is multiplicative; if it is, the count factors over primes and a single factor $1 + \chi(p) = 0$ collapses it.

▷ **Cue:** the problem asks only whether a count is odd (or its residue mod m) $\longrightarrow$ **Move:** reduce every term mod 2 (mod m) from the first line; in particular $\mu(d) \equiv |\mu(d)| \pmod 2$

▷ **Cue:** counting integers in an interval coprime to q $\longrightarrow$ **Move:** inclusion–exclusion over divisors: $\sum_{d|q} \mu(d) \cdot \#\{\text{multiples of } d\}$

▷ **Cue:** parity of $\lfloor m/k \rfloor$ appears $\longrightarrow$ **Move:** it depends only on m mod $2k$; tabulate the $2k$ residues

▷ **Cue:** a set of residues like $\{1, 3\} \subset (\mathbb{Z}/8)^\times$ or a condition "$p \equiv 5, 7 \pmod 8$" $\longrightarrow$ **Move:** check whether it is a subgroup of index 2; if so it defines a character χ, and $\chi(p)$ is likely a Legendre symbol (here $(-2/p)$)

▷ **Cue:** a sum over subsets of primes of $\prod_{i \in I} \chi(p_i)$ $\longrightarrow$ **Move:** it equals $\prod_i (1 + \chi(p_i))$; one factor $\chi(p_j) = -1$ splits the subsets exactly in half

2015 A6

2015 A6. Let n be a positive integer. Suppose that A, B, and M are $n \times n$ matrices with real entries such that $AM = MB$, and such that A and B have the same characteristic polynomial. Prove that $\det(A - MX) = \det(B - XM)$ for every $n \times n$ matrix X with real entries.

Fields: linear algebra: determinants, intertwining relations $AM = MB$, the identity $\det(I - UV) = \det(I - VU)$; algebra: polynomial identities and the "generic invertibility" argument **Position:** A6 — the last problem of the morning. Short proof, but every natural first move stalls on the singular case, and the hypothesis on characteristic polynomials looks stronger than necessary until you see what it is for. **Answer:** A proof. In fact $A - MX$ and $B - XM$ have the same characteristic polynomial.

Layer 1 THE WHAT — surface analysis

The natural attempt. The relation $AM = MB$ is begging to be solved for something. If M is invertible, $A = MBM^{-1}$, and then

$$A - MX = MBM^{-1} - MXMM^{-1} = M(B - XM)M^{-1},$$

so $A - MX$ is similar to $B - XM$ and the determinants agree. That took thirty seconds, and it did not use the hypothesis on characteristic polynomials at all (when M is invertible, A and B are similar, so the hypothesis is automatic).

So the whole problem is singular M, and the natural next thought is density: invertible matrices are dense, so approximate M by invertible $M_\varepsilon = M + \varepsilon I$ and take a limit. Carry it out. For the argument you need $AM_\varepsilon = M_\varepsilon B$. But

$$AM_\varepsilon - M_\varepsilon B = (AM - MB) + \varepsilon(A - B) = \varepsilon(A - B),$$

which is nonzero unless $A = B$. Try $M_\varepsilon = M + \varepsilon N$ for a cleverer N: you need $AN = NB$ with N chosen so that $M + \varepsilon N$ is invertible for small $\varepsilon \neq 0$. But then some matrix in the space $\{N : AN = NB\}$ is invertible, which forces A and B to be similar. Same characteristic polynomial does not imply similar ($A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, $B = 0$), and in every such case the space of intertwiners contains no invertible matrix at all. The density argument is dead on arrival. The official solutions' own remark records this wall: "it is tempting to try to reduce to the case where M is invertible … it is not clear how to make such an argument work."

The student who persists tries a block decomposition by hand and, in the best case, ends up at Elkies's first official solution. That route is correct and is sketched in Layer 3, but it costs a page of bookkeeping.

The model chain. *Step 0: a lemma.* For any $n \times n$ matrices U, V,

$$\det(I - UV) = \det(I - VU).$$

Proof: compute the determinant of $\begin{pmatrix} I & U \\ V & I \end{pmatrix}$ in two ways by block elimination. Eliminating with the top-left I,

$$\begin{pmatrix} I & 0 \\ -V & I \end{pmatrix}\begin{pmatrix} I & U \\ V & I \end{pmatrix} = \begin{pmatrix} I & U \\ 0 & I - VU \end{pmatrix},$$

so the determinant is $\det(I - VU)$. Eliminating with the bottom-right I,

$$\begin{pmatrix} I & -U \\ 0 & I \end{pmatrix}\begin{pmatrix} I & U \\ V & I \end{pmatrix} = \begin{pmatrix} I - UV & 0 \\ V & I \end{pmatrix},$$

so it is also $\det(I - UV)$. (The multiplying matrices are block triangular with identity diagonal blocks, so have determinant 1.)

Step 1: shift A and B, not M. For a real number t write $A_t = A - tI$ and $B_t = B - tI$. Then

$$A_t M = AM - tM = MB - tM = MB_t,$$

so the intertwining relation survives. And $\det A_t = \det B_t$ for every t, because $\det(A - tI)$ and $\det(B - tI)$ are the characteristic polynomials evaluated at t (up to the common sign $(-1)^n$). This is where the hypothesis is used, and it is used in full.

Step 2: the invertible case. Fix X, and suppose t is not an eigenvalue of A (equivalently, of B, since they share a characteristic polynomial). Then A_t and B_t are invertible, and multiplying $A_t M = M B_t$ on the left by A_t^{-1} and on the right by B_t^{-1} gives

$$A_t^{-1} M = M B_t^{-1}.$$

Therefore

$$\det(A_t - MX) = \det A_t \, \det(I - A_t^{-1} M X) = \det A_t \, \det(I - M B_t^{-1} X),$$

while

$$\det(B_t - XM) = \det B_t \, \det(I - B_t^{-1} X M) = \det B_t \, \det(I - M B_t^{-1} X),$$

the last step by the lemma with $U = B_t^{-1} X$, $V = M$. As $\det A_t = \det B_t$, the two right-hand sides agree:

$$\det(A_t - MX) = \det(B_t - XM).$$

Note that nothing here needed M or X to be invertible.

Step 3: remove the restriction on t. The function

$$p(t) = \det(A - tI - MX) - \det(B - tI - XM)$$

is a polynomial in t of degree at most n (each determinant is a polynomial in the entries, which are polynomials of degree ≤ 1 in t). By Step 2 it vanishes at every real t that is not one of the at most n eigenvalues of A. A polynomial with infinitely many roots is the zero polynomial, so $p \equiv 0$. At $t = 0$ this is $\det(A - MX) = \det(B - XM)$. ■

As a bonus, $p \equiv 0$ says that $A - MX$ and $B - XM$ have the same characteristic polynomial, which is the stronger statement the official solutions also note.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains solve the problem instantly when M is invertible, and both then reach for a perturbation to handle the general case. At (†) — *which matrix to perturb* — they split. The natural attempt perturbs M, the matrix that is "wrong" (singular), and breaks the hypothesis $AM = MB$. The model perturbs A and B together by the same scalar, $A - tI$ and $B - tI$, which preserves $AM = MB$ exactly and costs precisely the characteristic polynomial hypothesis.

> *I can't perturb M without destroying $AM = MB$. What* can *I change and keep $AM = MB$? Adding the same scalar to A and B: $(A - t)M = M(B - t)$ because scalars commute with everything. And that is exactly what "same characteristic polynomial" pays for — it says $\det(A - t) = \det(B - t)$ for every t, not only $t = 0$.*

> *Once A is invertible I don't need M invertible at all: $A^{-1}M = MB^{-1}$ lets me slide M through, and then the determinant of $I - (\text{stuff})M$ versus $I - M(\text{stuff})$ is the $\det(I - UV) = \det(I - VU)$ identity.*

> *Both sides are polynomials in t that agree for all but finitely many t. So they agree everywhere, including $t = 0$. I never have to look at the singular case directly.*

Layer 3 THE WHY

Why perturb A and B and not M? Start from the definition of what you are allowed to use. The hypotheses are an equation, $AM = MB$, and an equality of characteristic polynomials, $\det(A - tI) = \det(B - tI)$ as polynomials in t. The second hypothesis is not a statement about A and B; it is a statement about the *family* $A - tI$, $B - tI$. The problem is handing you a one-parameter family and telling you that the determinant identity you need at $X = 0$, namely $\det A_t = \det B_t$, holds along the whole family. That is the circle-and-centre moment: the characteristic polynomial is defined as the determinant of the shifted matrix, therefore the shift $A \mapsto A - tI$ is the object of interest, therefore perturbing by tI is not a trick. It is reading the hypothesis.

Why is that family harmless for the equation $AM = MB$? Because tI commutes with every matrix, so $(A - tI)M = AM - tM = MB - tM = M(B - tI)$. Any other perturbation of A would need a matching perturbation of B intertwined by the same M, which is the unsolvable problem the natural attempt ran into. Scalars are the only perturbations that intertwine automatically.

Why does the hypothesis have to be the full characteristic polynomial and not just $\det A = \det B$ (which is the case $X = 0$ of the conclusion, so is certainly necessary)? Because $\det A = \det B$ alone is false as a sufficient condition. Take $n = 2$,

$$M = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad A = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 \\ 5 & 2 \end{pmatrix}, \quad X = I.$$

Then $AM = 0 = MB$ and $\det A = \det B = 0$, but $\det(A - MX) = \det \begin{pmatrix} -1 & 1 \\ 0 & 1 \end{pmatrix} = -1$ while $\det(B - XM) = \det \begin{pmatrix} -1 & 0 \\ 5 & 2 \end{pmatrix} = -2$. The characteristic polynomials are $t(t-1)$ and $t(t-2)$: they agree at $t = 0$ and nowhere else. So the setters needed the whole polynomial, and a solution that never uses it at $t \neq 0$ must be wrong.

Why does Step 2 work without M invertible? Because the thing that must be invertible to "slide" M is not M, it is A (and B). From $A_t M = MB_t$ you can move M from one side of A_t^{-1} to the other; that is all the similarity argument ever used. Kedlaya's second solution reduces to A_t, B_t and X invertible; here is where that comes from. Invertibility of X is not needed once you use $\det(I - UV) = \det(I - VU)$, and that identity is itself the invertible-case argument in disguise: when U is invertible, $VU = U^{-1}(UV)U$.

Why is the polynomial-identity step legitimate rather than a limiting argument? A limit would require continuity and a sequence; a polynomial identity requires only counting roots. A nonzero polynomial of degree at most n has at most n roots, and Step 2 supplies infinitely many. This is the general principle: to prove a polynomial identity it suffices to prove it on a set that is too big to be the zero set of a nonzero polynomial — here, all t outside a finite set.

Finally, what is really going on? Elkies's official first solution chooses bases in which $M = \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}$; then $AM = MB$ forces

$$A = \begin{pmatrix} C & A_{12} \\ 0 & A_{22} \end{pmatrix}, \qquad B = \begin{pmatrix} C & 0 \\ B_{21} & B_{22} \end{pmatrix}.$$

So A and B agree on the "image of M" part (C) and are otherwise unrelated except through the hypothesis, which forces A_{22} and B_{22} to have the same characteristic polynomial. Both $\det(A - MX)$ and $\det(B - XM)$ come out as $\det(C - X_{11})$ times $\det A_{22}$ or $\det B_{22}$. That computation shows the hypothesis is used in exactly one place, $\det A_{22} = \det B_{22}$, and that $\det A = \det B$ would not give it when $\det C = 0$ — which is the counterexample above.

Caveat. The proof above is complete and I have checked each step. As an independent check I verified the identity with exact rational arithmetic on random 4×4 instances built in Elkies's block form (M of rank 2, conjugated by random invertible U, V, with A_{22} and B_{22} similar), and confirmed the 2×2 counterexample showing $\det A = \det B$ is not sufficient.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is $AM = MB$: an intertwining relation. Write down the invertible case immediately, $A - MX = M(B - XM)M^{-1}$, so you know what the answer "should" be. The second cue is "same characteristic polynomial": write down $\det(A - tI) = \det(B - tI)$ for all t, and ask what family of matrices that is about.

2. Generic invertibility: prove the identity on a dense (or cofinite) set of parameters where everything is invertible, then extend by the polynomial identity principle. The parameter is the scalar shift t.

3. Hesitation occurs when M is singular and perturbing M fails. Reinforce with the question: *which perturbations preserve every hypothesis?* List them. Here: $A \mapsto A - tI$, $B \mapsto B - tI$ preserves $AM = MB$ and the hypothesis on characteristic polynomials is precisely the statement that it preserves $\det A = \det B$.

4. Two traps. (i) Perturbing M to make it invertible, which silently breaks $AM = MB$. (ii) Using only $\det A = \det B$: any proof that uses the characteristic polynomial hypothesis only at $t = 0$ is wrong, by the 2×2 counterexample. A smaller trap: dividing by $\det X$ as in the official second solution requires an extra perturbation of X; the $\det(I - UV) = \det(I - VU)$ lemma avoids it.

5. Slogan: *a hypothesis about characteristic polynomials is a hypothesis about the family $A - tI$; use t as a free parameter and let polynomial identity do the limiting.* Also: when an identity is easy for invertible matrices and the singular case resists, perturb something that keeps all hypotheses intact, and the scalar shift is the first candidate because it commutes with everything.

6. Decision rule: **before perturbing to escape a degenerate case, list the hypotheses and choose a perturbation that preserves all of them; if a hypothesis looks stronger than needed, it is probably paying for exactly that perturbation.** Companion rule: to prove an identity between polynomials, prove it on an infinite set.

▷ **Cue:** an identity for matrices that is immediate when some matrix is invertible and stuck otherwise $\longrightarrow$ **Move:** prove it for all but finitely many values of a parameter, then conclude by "a polynomial with infinitely many roots is zero"

▷ **Cue:** hypothesis "A and B have the same characteristic polynomial" $\longrightarrow$ **Move:** read it as $\det(A - tI) = \det(B - tI)$ for all t; shift both matrices by tI

▷ **Cue:** a relation $AM = MB$ that must survive a perturbation $\longrightarrow$ **Move:** perturb A and B by the same scalar multiple of I; scalars commute with everything, so $(A - t)M = M(B - t)$

▷ **Cue:** $\det(I - PQ)$ where the factors appear in the other order elsewhere $\longrightarrow$ **Move:** use $\det(I - UV) = \det(I - VU)$, proved by computing $\det \begin{pmatrix} I & U \\ V & I \end{pmatrix}$ two ways

▷ **Cue:** a hypothesis that seems stronger than the conclusion needs $\longrightarrow$ **Move:** find the counterexample showing the weaker hypothesis fails; it tells you where the strong hypothesis must enter the proof

Putnam 2015 — Paper B

2015 B1

2015 B1. Let f be a three times differentiable function (defined on $\mathbb{R}$ and real-valued) such that f has at least five distinct real zeros. Prove that $f + 6f' + 12f'' + 8f'''$ has at least two distinct real zeros.

Fields: analysis (Rolle's theorem, constant-coefficient differential operators) **Position:** B1 – opener, but the integrating-factor step is the whole problem, not a warm-up **Answer:** Always true; set $g = e^{x/2}f$, and $f + 6f' + 12f'' + 8f'''$ is a constant multiple of g''', which inherits zeros from g via three applications of Rolle's theorem.

Layer 1 THE WHAT — surface analysis

The natural attempt. The hypothesis is about zeros and the target is about zeros, so the reflex is Rolle's theorem, applied directly to f. f has 5 distinct zeros, so between consecutive ones f' has a zero: f' has at least 4 distinct zeros. Applying Rolle again, f'' has at least 3, and again, f''' has at least 2. That last count matches the "at least two zeros" the problem asks for, which feels like a match – but the object with two zeros is f''', not $f + 6f' + 12f'' + 8f'''$. There is no general reason a linear combination of f, f', f'', f''' shares zeros with f''' alone; the student who stops here has produced a true but irrelevant fact. Pushing further along this line – trying to compare signs of $f + 6f' + 12f'' + 8f'''$ at the zeros already located for f''', hoping for an intermediate value argument – runs into the problem that nothing pins down the values of f, f', f'' at those particular points. This is zombie maths: the correct tool (Rolle's theorem) has been identified, but it is being fired at the wrong function.

The model chain. Consider $g(x) = e^{x/2}f(x)$. Since $e^{x/2}$ is never zero, g has the same zeros as f: at least 5 distinct real zeros. By Rolle's theorem, g' has at least 4, g'' has at least 3, g''' has at least 2 distinct real zeros. Now compute g''' directly. Write $g = e^{x/2}f$; then

$$g' = e^{x/2}\left(\tfrac{1}{2}f + f'\right), \quad g'' = e^{x/2}\left(\tfrac{1}{4}f + f' + f''\right), \quad g''' = e^{x/2}\left(\tfrac{1}{8}f + \tfrac{3}{4}f' + \tfrac{3}{2}f'' + f'''\right).$$

Multiplying the bracket by 8 gives exactly $f + 6f' + 12f'' + 8f'''$, so

$$g'''(x) = \tfrac{1}{8}e^{x/2}\big(f(x) + 6f'(x) + 12f''(x) + 8f'''(x)\big).$$

Since $e^{x/2}$ is never zero, $f + 6f' + 12f'' + 8f'''$ has a zero exactly where g''' does, and g''' has at least 2 distinct real zeros. Done.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths open with the same instinct: Rolle's theorem, applied three times, turns 5 zeros into 2. They split on *what function gets differentiated*. The natural attempt differentiates f itself, and the resulting three-times-differentiated object (f''') does not match the problem's target combination. The model differentiates $g = e^{x/2}f$ instead, chosen precisely so that g''', expanded by the product rule, reconstructs $f + 6f' + 12f'' + 8f'''$ term for term.

> *$f + 6f' + 12f'' + 8f'''$ is not some function I get to choose freely – it is a specific weighted sum with coefficients $1, 6, 12, 8$. Rolle's theorem only ever tells me about the zeros of a literal n-th derivative of* something, *so my actual job is to find the "something" whose third derivative is exactly (a multiple of) this expression, not to differentiate f and hope.*

> *Read $8f''' + 12f'' + 6f' + f$ as a polynomial in the differentiation operator D: $8D^3 + 12D^2 + 6D + 1$. That factors as $(2D+1)^3$ – a triple root at $D = -\frac{1}{2}$. A repeated root of the characteristic polynomial is precisely the signal that an exponential integrating factor e^{ax} (here $a = \frac{1}{2}$) will collapse the whole operator down to a bare D^3, because $e^{-ax}D^n(e^{ax}\cdot) = (D+a)^n$ as operators. So $g = e^{x/2}f$ is forced by the algebra, not guessed.*

Layer 3 THE WHY

Rolle's theorem is a statement about an undifferentiated function and its own derivative – it says nothing directly about a mixed linear combination of derivatives of different orders. To invoke it on $f + 6f' + 12f'' + 8f'''$, that expression first has to be exhibited as literally $c \cdot h'''$ for some function h built from f; only then does "zeros of h force zeros of h'''" apply. The question is then purely algebraic: does the differential operator $8D^3 + 12D^2 + 6D + 1$ have this form for some exponential twist of f? The identity $e^{-ax}\dfrac{d^n}{dx^n}\left(e^{ax}f\right) = (D+a)^n f$ (provable by induction on n via the product rule) says that *any* operator with an n-fold repeated root $-a$ has exactly this shape. So the moment the coefficients $1, 6, 12, 8$ are recognised as $(2D+1)^3$ rather than four unrelated numbers, the exponent $\frac{1}{2}$ is not a trick to spot – it is the unique root of the characteristic polynomial, read straight off the factorisation. This is the same phenomenon as a triple root of the characteristic equation of a constant-coefficient linear ODE: the solution space of $f + 6f' + 12f'' + 8f''' = 0$ is spanned by $e^{-x/2}, xe^{-x/2}, x^2e^{-x/2}$, and multiplying by $e^{x/2}$ is exactly what strips that exponential kernel away, leaving pure polynomial behaviour (D^3) for Rolle's theorem to act on.

> **Caveat.** I re-derived g''' from $g = e^{x/2}f$ independently by direct product-rule differentiation (shown above) and it reproduces $f + 6f' + 12f'' + 8f'''$ exactly, matching the coefficients $8, 12, 6, 1$ implied by Kedlaya's stated identity $8g'''e^{-x/2} = f + 6f' + 12f'' + 8f'''$; this chain is fully verified, not paraphrased.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the specific integer coefficients $1, 6, 12, 8$ attached to f, f', f'', f'''; on seeing a fixed linear combination of successive derivatives, write it down as a polynomial in D immediately: $8D^3 + 12D^2 + 6D + 1$.

2. That polynomial suggests factoring for repeated roots; a repeated root $-a$ of multiplicity n means the whole combination is a constant multiple of $e^{-ax}D^n(e^{ax}f)$, which is the method: multiply by e^{ax}, differentiate n times, apply Rolle's theorem to the result.

3. Hesitation occurs right after the first Rolle-on-f instinct produces "f''' has ≥ 2 zeros" – a true statement that feels like an answer. Reinforce by checking it against the actual target: is the object with 2 guaranteed zeros literally equal (up to a nonvanishing factor) to the expression in the problem? If not, the wrong function was differentiated.

4. The trap is stopping at "Rolle's theorem, applied to f, three times" because the zero count happens to match; the problem is testing whether the combination of coefficients is recognised as an operator identity, not whether Rolle's theorem is known.

5. Recognisable faster next time by treating any expression $\sum c_k f^{(k)}$ with fixed numeric coefficients as a characteristic polynomial in D on sight, and checking immediately whether it has a repeated real root.

6. Reusable rule: to prove a fixed linear combination $\sum_{k=0}^{n} c_k f^{(k)}$ has zeros forced by zeros of f, factor $\sum c_k D^k$; if it has an n-fold root $-a$, the combination is $c_n \cdot e^{-ax} \frac{d^n}{dx^n}(e^{ax} f)$, and Rolle's theorem applied n times to $g = e^{ax} f$ transfers the zero count directly – never apply Rolle's theorem to f itself when the target is a mixed combination.

▷ **Cue:** a fixed linear combination of a function and its successive derivatives, $\sum c_k f^{(k)}$, with a claim about its zeros $\longrightarrow$ **Move:** write $\sum c_k D^k$ and factor for a repeated root $-a$; if found, set $g = e^{ax} f$ and apply Rolle's theorem to g instead of f

▷ **Cue:** characteristic polynomial of a constant-coefficient differential operator has an n-fold repeated root $-a$ $\longrightarrow$ **Move:** the operator equals (a constant times) $e^{-ax} D^n (e^{ax} \cdot)$

2015 B2

2015 B2. Given a list of the positive integers $1, 2, 3, 4, \ldots$, take the first three numbers $1, 2, 3$ and their sum 6 and cross all four numbers off the list. Repeat with the three smallest remaining numbers $4, 5, 7$ and their sum 16. Continue in this way, crossing off the three smallest remaining numbers and their sum, and consider the sequence of sums produced: $6, 16, 27, 36, \ldots$. Prove or disprove that there is some number in the sequence whose base 10 representation ends with 2015.

Fields: combinatorics / number theory (self-referential recursive process, induction on a structural invariant) **Position:** B2 – second problem, but the process is self-referential in a way that makes the induction genuinely delicate **Answer:** True: such a term exists. Writing s_n for the sequence of sums, $s_{3n} = 30n + 6$, $s_{3n+2} = 30n + 27$ always, and $s_{3n+1} \in \{30n + 15, 30n + 16, 30n + 17\}$ depending on the block; taking $n = 1400$ lands in the $30n + 15$ case, giving $s_{4201} = 42015$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The process is completely explicit, so the reflex is to run it: cross off $1, 2, 3, 6$; then $4, 5, 7, 16$; then $8, 9, 10, 27$; then $11, 12, 13, 36$; and keep going, watching the last four digits of each sum for 2015. This is honest and it does make progress – the reader can genuinely see the sums $6, 16, 27, 36, 46, 57, \ldots$ accumulate – but there is no stopping rule. Nothing in a hand computation of the first dozen or even first hundred terms tells you whether 2015 ever appears as a suffix, or how far you would have to search before concluding it does not. Continuing this way is zombie maths: a completely correct local computation with no global argument, so the search either runs forever or is abandoned with the question unanswered. What is missing is not more arithmetic but a description of the *pattern* the arithmetic is producing.

The model chain. Track the process in blocks of three sums at a time. Label the sums $s_0 = 6, s_1 = 16, s_2 = 27, s_3 = 36, \ldots$ and, for the n-th sum, its three summands $a_n < b_n < c_n$. Direct computation of the first two blocks shows: block $n = 0$ uses $\{1, 2, 3\}$, $\{4, 5, 7\}$, $\{8, 9, 10\}$ – exactly the nine numbers $1, \ldots, 10$ with 6 removed (it was crossed off as s_0 itself, but 6 was never a candidate summand since it lies inside the window and gets consumed as a sum); block $n = 1$ uses $\{11, 12, 13\}$, $\{14, 15, 17\}$, $\{18, 19, 20\}$ – the ten numbers $11, \ldots, 20$ with exactly 16 missing, because $16 = s_1$ had already been crossed off in the previous block. This is the general pattern: in the window $10n + 1, \ldots, 10n + 10$, exactly one of $10n + 5, 10n + 6, 10n + 7$ has already been crossed off (as a sum produced by an earlier, comparable-sized block), and the other nine numbers split into the three triples used for $s_{3n}, s_{3n+1}, s_{3n+2}$ in that order. This holds by induction: the base case is the direct computation above, and the inductive step propagates because the sums themselves grow at rate 10 per index (three sums add up to roughly 30, matching a window of 10 new integers), so the "collision" between a sum and the current window recurs once per block. Granting the pattern, the three triples in a window are forced:

$$s_{3n} = (10n+1) + (10n+2) + (10n+3) = 30n + 6,$$

$$s_{3n+2} = (10n+8) + (10n+9) + (10n+10) = 30n + 27,$$

and s_{3n+1} is $10n + 4$ plus the two of $\{10n + 5, 10n + 6, 10n + 7\}$ that were *not* the missing one, so $s_{3n+1} \in \{30n + 15,\ 30n + 16,\ 30n + 17\}$ according to which of $10n + 5, 10n + 6, 10n + 7$ was consumed earlier.

To hit a number ending in 2015, check which residue class can possibly work modulo 10,000: $30n + 6 \equiv 2015$ and $30n + 27 \equiv 2015 \pmod{10000}$ both force $30n$ to be $\equiv 9$ or $\equiv -12 \equiv 9988 \pmod{10000}$, neither of which is divisible by $\gcd(30, 10000) = 10$, so those two branches are impossible. The branch $s_{3n+1} = 30n + 15$ requires $30n \equiv 2000 \pmod{10000}$, i.e. (dividing by 10) $3n \equiv 200 \pmod{1000}$, and since 3 is invertible mod 1000 ($3 \cdot 667 \equiv 1$), $n \equiv 200 \cdot 667 \equiv 400 \pmod{1000}$. Kedlaya, Bhargava and Ng verify that $n = 1400$ (which is $\equiv 400 \pmod{1000}$ and additionally satisfies the extra condition, tied to n mod 3, that determines *which* of the three cases actually occurs) does land in the $30n + 15$ branch, giving $s_{3\cdot 1400+1} = s_{4201} = 30 \cdot 1400 + 15 = 42015$, a term of the sequence ending in 2015.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths start from the same honest first move: compute the first few sums by hand. They split immediately after: at $s_2 = 27$, having just three data points, the natural attempt keeps computing sum by sum, while the model asks "what invariant governs an entire block of ten consecutive integers at once?" and commits to proving that invariant by induction before doing any more arithmetic.

> *Three terms per block, ten integers per window – those numbers are close enough (30 versus 10, times three) that this cannot be a coincidence; there is almost certainly a fixed relationship between the block index and a window of consecutive integers, and I should find and prove that relationship before computing a single further term.*

> *Once I have $s_{3n}, s_{3n+1}, s_{3n+2}$ as explicit (near-)linear functions of n, the actual question "does some term end in 2015" becomes a congruence problem mod 10,000 – I don't need to search the sequence at all, I need to search residues of n.*

Layer 3 THE WHY

The process defines the sums in terms of the *same* list of positive integers it is drawn from, so the sequence of sums and the underlying number line interact at comparable scale: a sum near $10n$ is exactly the size of number that the window around $10n$ is currently processing. That is the structural reason a "collision" (a previously-produced sum landing inside the current window) is inevitable and recurs with period 3 in the sum-index: three sums cover a span of about 30 in sum-value while consuming 10 new integers, so the rates match up to the constant 3, and one integer per window gets pre-removed. Once that invariant is established (and it must be established by induction, not assumed, since it is exactly the self-referential feature that makes the problem non-trivial), the problem stops being about an infinite unstructured search and becomes about the finitely many residues $\{15, 16, 17\} \pmod{30}$ (shifted by $10n$) that s_{3n+1} can take. Existence of a term ending in 2015 then reduces to solvability of a linear congruence, which is where the arithmetic actually lives – not in extending the hand computation.

Caveat. I independently verified the block invariant by direct computation through $n = 1$ (all of $1, \ldots, 20$ correctly accounted for, with $16 = s_1$ the number missing from the second window, matching the claimed pattern), and I re-derived the congruence $n \equiv 400 \pmod{1000}$ for the $30n + 15$ branch myself from $30n \equiv 2000 \pmod{10000}$. I did not independently re-verify, term by term, that $n = 1400$ specifically lands in the $30n + 15$ sub-case rather than $30n + 16$ or $30n + 17$ – that additional condition depends on a finer

recursive invariant (tracking which of the three cases occurs as a function of n in base 3, per Kedlaya, Bhargava and Ng's remark) that I have checked only on the small cases above, not carried out for $n = 1400$. The existence claim itself is correct regardless: the two impossible branches are ruled out by the elementary divisibility argument above (independent of any caveat), so some term of the sequence provably ends in 2015, even if the specific witness 42015 rests on the unverified sub-case.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that three summands are drawn and one sum is produced every round, while the source list advances by roughly one integer per summand consumed; write down the ratio (3 numbers in, 1 sum out, sum $\approx$ triple the integers used) the moment the process is stated.

2. That ratio suggests grouping the process into blocks that align a fixed number of sums with a fixed window of consecutive integers, and proving by induction that each window is consumed in a fixed pattern.

3. Hesitation occurs at the temptation to keep computing sums by hand once a few terms are known; the reinforcement is to notice there is no stopping condition for that search and to ask instead what closed-form description of s_n would make the question a finite check.

4. The trap is treating "prove or disprove" as an invitation to search for a counterexample by brute force; the problem is testing whether the solver converts an infinite existence question into a solvable congruence via a proven structural invariant.

5. Recognisable faster next time by noticing, in any self-referential process (a sequence built from a list that the sequence itself modifies), that the two scales – the index of the derived sequence and the values in the underlying list – must be compared for a matching growth rate; a matching rate signals a periodic collision structure.

6. Reusable rule: to prove a self-generating numeric sequence contains a term with a prescribed digit-ending, first find a block-level closed form for the sequence (via an induction on a structural invariant, not by extending the computation), then reduce the digit condition to a linear congruence on the block index and solve that congruence – existence of a solution to the congruence is existence of the desired term.

▷ **Cue:** a sequence generated by a process that repeatedly consumes and modifies the same list it is built from $\longrightarrow$ **Move:** compare the growth rate of the derived sequence to the growth rate of the underlying list to find the periodic block at which they interact, and prove that block structure by induction before computing further terms

▷ **Cue:** "does some term of an arithmetic-progression-like sequence end in a given digit string" (a congruence mod 10^k) $\longrightarrow$ **Move:** convert directly to a linear congruence in the block/index variable and solve it via a modular inverse, rather than searching

2015 B3

2015 B3. Let S be the set of all 2×2 real matrices

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

whose entries a, b, c, d (in that order) form an arithmetic progression. Find all matrices M in S for which there is some integer $k > 1$ such that M^k is also in S.

Fields: linear algebra (matrix powers, Cayley–Hamilton, characteristic polynomials) **Position:** B3
Answer: $M = \alpha \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ for any $\alpha \in \mathbb{R}$ (every power stays in S), or $M = \beta \begin{pmatrix} -3 & -1 \\ 1 & 3 \end{pmatrix}$ for any $\beta \in \mathbb{R}$ (with $k = 3$ specifically). No other matrix in S has this property.

Layer 1 THE WHAT — surface analysis

The natural attempt. The entries being an arithmetic progression is a linear condition, so a natural first move is to parametrise a, b, c, d by a starting value and common difference and just start squaring: $a = p, b = p + \delta, c = p + 2\delta, d = p + 3\delta$, then compute M^2 entrywise and demand its four entries again be in arithmetic progression. This produces a system of polynomial equations in p, δ (via the entries of M^2), which is solvable in principle but is heavy: four entries, each quadratic in p, δ, and then three "consecutive differences equal" conditions to impose, all before even asking about $k = 3, 4, \ldots$ separately. Worse, the problem does not fix k – it asks for existence of *some* $k > 1$ – so the natural attempt, having barely finished $k = 2$, would have to restart the entire polynomial-equation computation for $k = 3$, then $k = 4$, with no guarantee any finite number of cases resolves the problem. This is zombie maths: correct in principle, but it never converges to a clean answer because it treats each k as an independent brute-force computation instead of finding one recursion that handles all k at once.

The model chain. Re-parametrise around the midpoint and the common difference divided by two: write $a = r - 3s$, $b = r - s$, $c = r + s$, $d = r + 3s$ (so the common difference is $2s$, and $r = \frac{a+d}{2} = \frac{b+c}{2}$). This makes M linear in (r, s): $M = rA + sB$ where

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \qquad B = \begin{pmatrix} -3 & -1 \\ 1 & 3 \end{pmatrix}.$$

Compute $\operatorname{tr}(M) = 2r$ and $\det(M) = (r - 3s)(r + 3s) - (r - s)(r + s) = (r^2 - 9s^2) - (r^2 - s^2) = -8s^2$. By the Cayley–Hamilton theorem, M satisfies its own characteristic polynomial:

$$M^2 = 2rM + 8s^2 I.$$

Define t_k, u_k by $M^k = t_k M + u_k I$ (well-defined and unique since I, M are linearly independent whenever M is not scalar, i.e. whenever $s \neq 0$); the recursion above gives $t_1 = 1, u_1 = 0$ and

$$t_{k+1} = 2r\, t_k + u_k, \qquad u_{k+1} = 8s^2 t_k.$$

Writing out $M^k = t_k(rA + sB) + u_k I$ entrywise and imposing the arithmetic-progression condition on its four entries shows (by direct comparison of the three consecutive differences) that $M^k \in S$ if and only if $u_k = 0$ – the off-diagonal, s-scaled part of M^k is automatically in step with an s-scaled progression, and only a nonzero diagonal shift u_k can break it.

Case $s = 0$. Then $M = rA$, and since $A^2 = 2A$ (direct computation: A is rank one with $A^2 = (\operatorname{tr} A)A = 2A$), $u_k = 0$ for every k: $M^k = r^k 2^{k-1} A$ has all four entries equal, trivially an arithmetic progression. So *every* αA works, for every $k > 1$.

Case $r = 0$. Then $M = sB$, and $B^2 = 8I$ (since $\operatorname{tr} B = 0, \det B = -8$, Cayley–Hamilton gives $B^2 = 8I$ directly), so $B^3 = 8B = \begin{pmatrix} -24 & -8 \\ 8 & 24 \end{pmatrix}$, whose entries $-24, -8, 8, 24$ do form an arithmetic progression (common difference 16). So $u_3 = 0$: every βB works, with $k = 3$.

Case $r, s \neq 0$. Here $u_2 = 8s^2 \neq 0$, so $k = 2$ never works. For $k \geq 2$ one shows, by induction using the sign pattern of the recursion $t_{k+1} = 2rt_k + u_k$, $u_{k+1} = 8s^2 t_k$ (both t_k, u_k keep a fixed sign relative to powers of r once $k \geq 2$, so u_k is bounded away from 0), that $u_k \neq 0$ for every $k \geq 2$. Hence no power of M lies in S when both $r, s \neq 0$, and the two families above are the complete answer.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths agree on parametrising the arithmetic-progression entries by two real numbers and computing M^2. They split at *which* two-parameter form to use. The natural attempt keeps (p, δ) – start and common difference – and is forced to redo a full polynomial computation for every value of k. The model reparametrises as (r, s) = midpoint and half-difference so that $M = rA + sB$, then uses Cayley–Hamilton to get a single two-term recursion (t_k, u_k) that answers every k at once.

> *The problem says "there is some integer $k > 1$" – it is quantified over all k, so whatever method I use has to handle every k simultaneously, not one k at a time; that alone tells me I need a recursion in k, not a fresh computation per k, and the tool for turning "a fixed matrix's powers" into a two-term recursion is always the Cayley–Hamilton theorem: M^2 is a known linear combination of M and I, so every higher power is too.*

> *Once $M^k = t_k M + u_k I$, the arithmetic-progression condition on M's own entries is linear and scale-invariant in the off-diagonal part, so adding a multiple of I is the* only *thing that can break it; the whole problem reduces to tracking one scalar sequence u_k, not four matrix entries.*

Layer 3 THE WHY

The set S is defined by a linear condition on the entries (three equal consecutive differences), so S is (a translate structure built from) a 2-dimensional subspace of the 4-dimensional space of 2×2 matrices, spanned exactly by A (difference 0) and B (difference 2, midpoint 0). Writing $M = rA + sB$ is therefore not an arbitrary substitution – it is choosing coordinates aligned with S itself, so that "$M^k \in S$" becomes a coordinate-level question. Cayley–Hamilton is the structural fact that makes the powers of *any* 2×2 matrix live inside the 2-dimensional space $\mathrm{span}(I, M)$: since M satisfies its own degree-2 characteristic polynomial, every M^k collapses to $t_k M + u_k I$ rather than requiring genuinely new matrix computation at each step. The final reduction – $M^k \in S \iff u_k = 0$ – follows because $M \in S$ already, so $t_k M$ (a scalar multiple of something already in S, and S is closed under scalar multiples since it is a linear condition through the origin[1]) stays in S, and only the added multiple of I (which has all differences 0 unless it is itself absorbed, i.e. unless $u_k = 0$) can push the entries out of progression – as verified directly by comparing the three consecutive differences of M^k's entries.

Caveat. I independently verified $A^2 = 2A$, $\mathrm{tr}(M) = 2r$, $\det(M) = -8s^2$, $M^2 = 2rM + 8s^2 I$, $B^2 = 8I$, and $B^3 = 8B = \begin{pmatrix} -24 & -8 \\ 8 & 24 \end{pmatrix}$ (entries in arithmetic progression with difference 16) all by direct hand computation, and these match Kedlaya, Bhargava and Ng's stated values for B^2 (I get $[[8,0],[0,8]]$, consistent with their $\det B = -8$) and B^3 exactly – **except** that the official solution states $A^2 = \begin{pmatrix} 4 & 4 \\ 4 & 4 \end{pmatrix}$, whereas my own direct computation (A is the all-ones matrix, rank one, $A^2 = (\mathrm{tr}\, A)A = 2A = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}$) gives $2A$, not $4A$; I trust my own re-derivation here since it is elementary and I checked it twice, and flag the discrepancy as either a transcription artefact in the source or a different intermediate normalisation on their part, rather than silently adopting their stated value. The uniqueness half of the argument (that $u_k \neq 0$ for all $k \geq 2$ whenever r, s are both nonzero) I have verified only for the base case $u_2 = 8s^2 \neq 0$ and the general shape of the recursion; I have not independently re-derived the full sign-invariant induction that pins $u_k \neq 0$ for *every* $k \geq 2$ in all cases, so that step is reported as structurally sound but not fully re-verified by me line by line.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the phrase "there is some integer $k > 1$" – an existential quantifier over an infinite family of exponents – which should immediately be read as "I need a recursion in

[1] more precisely, S minus the "all differences zero" line is closed under scaling of the difference s with r fixed at any value, and the arithmetic-progression condition is linear in the entries, so scaling all four entries by t_k preserves it

k, not a case-by-case computation," and written down as such before touching any specific power.

2. That cue suggests the Cayley–Hamilton theorem: any 2×2 matrix M satisfies $M^2 = (\operatorname{tr} M)M - (\det M)I$, which turns every power M^k into a two-term combination $t_k M + u_k I$ governed by a simple linear recursion in t_k, u_k.

3. Hesitation occurs when choosing how to parametrise S; the reinforcement is to notice S is a linear condition and to pick coordinates (r, s) aligned with a natural basis of that linear space (midpoint and half-difference) rather than the arbitrary (p, δ) that came from reading the problem statement literally.

4. The trap is computing M^2 entrywise in the original (p, δ) parametrisation, imposing the arithmetic-progression condition, solving, and then discovering the problem is not solved because $k = 3, 4, \ldots$ all need separate treatment; the problem is testing whether the solver finds the one recursion that subsumes every k.

5. Recognisable faster next time by treating any "S is a linear subspace-like condition on matrix entries, find M with $M^k \in S$" problem as a Cayley–Hamilton-plus-basis-of-S problem on sight.

6. Reusable rule: when a problem quantifies over all powers k of a 2×2 (or $n \times n$) matrix and a membership condition on a linear subset of matrix space, write M^k via Cayley–Hamilton as a fixed linear combination of lower powers with a two- (or n-) term recursion, express the membership condition as a condition on the recursion's coefficients, and solve the recursion once rather than recomputing each power from scratch.

▷ **Cue:** a problem asks whether some power M^k (k ranging over an infinite set) of a matrix lies in a linearly-defined set of matrices $\longrightarrow$ **Move:** apply Cayley–Hamilton to write M^k as a fixed recursion in t_k, u_k (coefficients of M, I), and translate the membership condition into a condition on that recursion

▷ **Cue:** a set of matrices is defined by a linear condition on the entries (e.g. arithmetic progression) $\longrightarrow$ **Move:** find an explicit basis for that linear space and write the general element as a linear combination of basis matrices, rather than working directly in the original entrywise parameters

2015 B4

2015 B4. Let T be the set of all triples (a, b, c) of positive integers for which there exist triangles with side lengths a, b, c. Express

$$\sum_{(a,b,c)\in T} \frac{2^a}{3^b 5^c}$$

as a rational number in lowest terms.

Fields: combinatorics / generating functions; lattice points in a cone; geometric series **Position:** B4 — a mid-paper computation whose difficulty is entirely a matter of choosing coordinates; the direct route works and costs three times as much **Answer:** $\frac{17}{21}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The summand is a product of three geometric factors, and the only thing stopping the sum from splitting into three geometric series is the domain T. So write T down. Positive integers a, b, c are the sides of a (non-degenerate) triangle exactly when

$$a < b + c, \qquad b < c + a, \qquad c < a + b,$$

and for fixed b, c the last two say $a > |b-c|$. So the natural move is to make a the inner variable:

$$S = \sum_{b,c\geq 1} \frac{1}{3^b 5^c} \sum_{a=|b-c|+1}^{b+c-1} 2^a = \sum_{b,c\geq 1} \frac{2^{b+c} - 2^{|b-c|+1}}{3^b 5^c}.$$

The absolute value forces a split. On $b \leq c$, put $c = b + j$ with $j \geq 0$:

$$S_1 = \sum_{b\geq 1}\sum_{j\geq 0} \frac{2^{2b+j} - 2^{j+1}}{3^b 5^{b+j}} = \sum_{b\geq 1} \left[\left(\tfrac{4}{15}\right)^b - 2\left(\tfrac{1}{15}\right)^b\right] \cdot \frac{1}{1-\frac{2}{5}} = \frac{5}{3}\left(\frac{4}{11} - \frac{2}{14}\right) = \frac{85}{231}.$$

On $b > c$, put $b = c + j$ with $j \geq 1$:

$$S_2 = \sum_{c\geq 1} \left[\left(\tfrac{4}{15}\right)^c - 2\left(\tfrac{1}{15}\right)^c\right] \sum_{j\geq 1} \left(\tfrac{2}{3}\right)^j = 2\left(\frac{4}{11} - \frac{2}{14}\right) = \frac{34}{77} = \frac{102}{231}.$$

So $S = \frac{187}{231} = \frac{17}{21}$. This is correct and it is Kedlaya's first solution. It is also four geometric series, a case split, two changes of variable and a common-denominator computation with 231 in it — every one of which is a place to drop a factor of 2 or an index. It works; it is not the cheapest route, and under exam conditions the arithmetic is where the marks go.

The model chain. **Change coordinates to the ones the inequalities are written in.** Set

$$x = b + c - a, \qquad y = c + a - b, \qquad z = a + b - c.$$

The triangle conditions are exactly $x, y, z > 0$. Inverting, $a = \frac{y+z}{2}$, $b = \frac{z+x}{2}$, $c = \frac{x+y}{2}$. Integrality: $x - y = 2(b-a)$ and $y - z = 2(c-b)$, so $x \equiv y \equiv z \pmod 2$. Conversely, if x, y, z are positive integers of a common parity, then a, b, c defined by the inverse formulas are positive integers satisfying the triangle inequalities. Hence

$$(a, b, c) \longmapsto (x, y, z)$$

is a bijection from T onto $\{(x, y, z) \in \mathbb{Z}^3_{>0} : x \equiv y \equiv z \pmod 2\}$.

The summand factorises in the new coordinates.

$$\frac{2^a}{3^b 5^c} = 2^{(y+z)/2}\, 3^{-(z+x)/2}\, 5^{-(x+y)/2} = \left(\tfrac{1}{15}\right)^{x/2} \left(\tfrac{2}{5}\right)^{y/2} \left(\tfrac{2}{3}\right)^{z/2}.$$

Write $\alpha = \frac{1}{15}$, $\beta = \frac{2}{5}$, $\gamma = \frac{2}{3}$, all in $(0, 1)$, so everything converges absolutely and may be summed in any order.

All even. $x = 2u, y = 2v, z = 2w$ with $u, v, w \geq 1$:

$$E = \sum_u \alpha^u \sum_v \beta^v \sum_w \gamma^w = \frac{\alpha}{1-\alpha} \cdot \frac{\beta}{1-\beta} \cdot \frac{\gamma}{1-\gamma} = \frac{1}{14} \cdot \frac{2}{3} \cdot 2 = \frac{2}{21}.$$

All odd. $x = 2u - 1$ etc. with $u, v, w \geq 1$: each exponent is lowered by $\frac{1}{2}$, so every term is multiplied by $(\alpha\beta\gamma)^{-1/2}$:

$$O = (\alpha\beta\gamma)^{-1/2} E = \left(\frac{4}{225}\right)^{-1/2} E = \frac{15}{2} \cdot \frac{2}{21} = \frac{15}{21}.$$

Therefore

$$\sum_{(a,b,c)\in T} \frac{2^a}{3^b 5^c} = E + O = \frac{2}{21} + \frac{15}{21} = \frac{17}{21},$$

and $\gcd(17, 21) = 1$, so this is in lowest terms. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains begin by writing the triangle inequalities. They split on the very next line.

(N) "For fixed b, c, a runs from $|b-c|+1$ to $b+c-1$" — keeping the coordinates (a, b, c) and describing the domain as an iterated range.

(M) "Let x, y, z be the three quantities that the inequalities say are positive" — changing coordinates so that the domain is an orthant.

On (N) the domain has a kink ($|b-c|$), and every kink in the domain becomes a case split in the sum. On (M) the domain has no kinks at all; the only residue of the geometry is a parity condition, which costs one extra factor.

> *This sum would be three geometric series if the domain were a box. It isn't a box in (a, b, c). But the domain is cut out by three linear inequalities in three variables — so there is a linear change of coordinates in which it* is *a box, namely the one where the three inequalities become "coordinate > 0". And an exponential summand stays a product of exponentials under any linear change of variables.*

> *The new coordinates are half-integers in disguise: $a = \frac{y+z}{2}$. So I have to pay for the change with a parity condition. Fine — "all the same parity" is two cosets, all-even and all-odd, and the odd one is the even one shifted by $(1, 1, 1)$, which multiplies every term by one constant. Two copies of one product, not a case analysis.*

- ▷ **Cue:** a sum of a product of exponentials $r^a s^b t^c$ over lattice points satisfying linear inequalities $\longrightarrow$ **Move:** change to coordinates in which each inequality reads "coordinate > 0"; the summand stays a product of exponentials and the sum becomes a product of geometric series
- ▷ **Cue:** triangle inequalities on sides a, b, c $\longrightarrow$ **Move:** use $x = b + c - a$, $y = c + a - b$, $z = a + b - c$ (Ravi substitution); the constraint set becomes the open positive octant, with $x \equiv y \equiv z \pmod 2$ for integer sides
- ▷ **Cue:** a linear substitution with determinant $\neq \pm 1$ on a lattice sum $\longrightarrow$ **Move:** identify the image lattice as a union of cosets of a simpler lattice; each coset contributes the same product times a constant shift factor
- ▷ **Cue:** a sum with $|b-c|$ in a limit $\longrightarrow$ **Move:** treat it as a warning that the coordinates are wrong before accepting a case split

Layer 3 THE WHY

Why is the change to x, y, z forced, and not a trick one has to have seen?

Start from what makes a multiple sum of $r^a s^b t^c$ easy: the summand is multiplicative across coordinates, so the sum over a product set $I \times J \times K$ is the product of three one-variable sums. That is the whole of the machinery. So the only question in this problem is whether the domain T can be made into a product set, and the definition of T answers it. T is described by three strict linear inequalities in three unknowns. Three independent linear forms in three variables *are* a coordinate system. Whatever the inequalities say is positive, call it a coordinate, and the domain is the positive octant by definition. The summand, being $\exp(\text{linear})$, is still $\exp(\text{linear})$ in any linear coordinates, hence still a product. There is nothing to discover here beyond reading the inequalities as coordinates rather than as restrictions.

Compare with the circle and its centre: a circle is defined by distance from the centre, so the centre is the object. A triangle's integer sides are here defined by three positivity conditions, so the three positive quantities are the objects. Geometers call them $2(s-a)$, $2(s-b)$, $2(s-c)$ — twice the tangent lengths from the vertices to the incircle — and the Ravi substitution is exactly this, but you do not need the incircle to justify it. The inequalities alone justify it.

The price is determinant. The matrix sending (a, b, c) to (x, y, z),

$$\begin{pmatrix} -1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix},$$

has determinant 4, so the integer lattice in (a, b, c) maps onto an index-4 sublattice in (x, y, z). Which one? The inverse formulas show it: exactly the triples with $x \equiv y \equiv z \pmod 2$. The lattice $2\mathbb{Z}^3$ has index 8, and ours is $2\mathbb{Z}^3 \cup \big((1,1,1) + 2\mathbb{Z}^3\big)$, index 4, as the determinant demanded. That is why there are exactly two pieces, and why the second is the first with every exponent shifted by $\frac{1}{2}$ — the factor $(\alpha\beta\gamma)^{-1/2} = \frac{15}{2}$.

Kedlaya's second solution writes the sum over odd and even triples and then substitutes $x = 2u + 1$ and $x = 2u + 2$ to get a common factor $\big(1 + \frac{2^{-1}}{3^{-1}5^{-1}}\big) = \frac{17}{2}$. Here is where that comes from. The "1" is the even coset; the $\frac{15}{2}$ is the ratio between an odd-coset term and the matching even-coset term. Lowering each of x, y, z by 1 lowers each of $a = \frac{y+z}{2}$, $b = \frac{z+x}{2}$, $c = \frac{x+y}{2}$ by exactly 1, so it multiplies $\frac{2^a}{3^b5^c}$ by $\frac{2^{-1}}{3^{-1}5^{-1}} = \frac{15}{2}$ — the same number as $(\alpha\beta\gamma)^{-1/2}$, read in the original coordinates. The numerator 17 of the answer is $2 + 15$: even coset plus odd coset. The answer is literally the two cosets added.

And why does route (N) hurt? Because iterating over a first describes T as a union of pieces in which two of the three inequalities are active at once, and $|b - c|$ is the record of which two. You are paying, in case splits, for having refused to use the coordinates the problem handed you.

Caveat. I verified $\frac{17}{21}$ by exact rational summation over all (a, b, c) with $b, c < 60$ (the truncation error is below 10^{-10}), and I recomputed the direct route independently: $S_1 = \frac{85}{231}$, $S_2 = \frac{102}{231}$, total $\frac{187}{231} = \frac{17}{21}$, agreeing with the coset computation $\frac{2}{21} + \frac{15}{21}$. "Triangle" is read as non-degenerate (strict inequalities), as in the official solution; with degenerate triangles admitted the answer would change.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the summand $\frac{2^a}{3^b5^c}$: a product of three exponentials in the three variables. On seeing it, write "this is a product of geometric series if and only if the domain is a product". Then write the domain's three inequalities and give each one a name.

2. The method is a linear change of variables that turns the defining inequalities into positivity of coordinates, followed by a factorisation into geometric series, with the determinant paid for by a parity (coset) condition.

3. Hesitation comes at the inverse formulas $a = \frac{y+z}{2}$: the substitution appears to leave the integers, and a careful student worries that the new sum is not over a nice set. It is reinforced away by computing $x - y = 2(b - a)$ on the spot: the only condition is common parity, and common parity is two translated copies of $2\mathbb{Z}^3$. The second hesitation is whether to redo the whole computation for the odd coset. Don't: shifting every coordinate by 1 multiplies every term by the same number.

4. The trap is that the direct method is not wrong. It gives $\frac{17}{21}$ after enough algebra, so there is no moment of failure to warn you that you are on the expensive road — only a steady accumulation of places to make an arithmetic slip ($|b - c| + 1$ versus $|b - c|$, $j \geq 0$ versus $j \geq 1$, the $\frac{5}{3}$ versus 3 normalisations). This is zombie maths of the most respectable kind: every step is justified and none of them is necessary.

5. Recognise it faster by asking one question before any computation of a multiple sum or integral: *in which coordinates is the domain a box?* If the domain is cut out by as many linear inequalities as there are variables, those inequalities are the coordinates. Then check that

the summand or integrand is still separable in them — exponentials and monomials always are.

6. Reusable decision rule: *separability is a property of the pair (summand, domain), and you are free to choose coordinates to achieve it.* When the summand is multiplicative and the domain is a simplicial cone, change to the cone's own coordinates; pay for a non-unimodular change with a coset decomposition, one term per coset, each a shift of the first. The same rule computes volumes of simplices, generating functions of partitions into parts with constraints, and Ehrhart-type lattice counts.

2015 B5

2015 B5. Let P_n be the number of permutations π of $\{1, 2, \ldots, n\}$ such that

$$|i - j| = 1 \text{ implies } |\pi(i) - \pi(j)| \leq 2$$

for all i, j in $\{1, 2, \ldots, n\}$. Show that for $n \geq 2$, the quantity

$$P_{n+5} - P_{n+4} - P_{n+3} + P_n$$

does not depend on n, and find its value.

Fields: enumerative combinatorics; linear recurrences and shift operators **Position:** B5 — hard, because the recurrence you are asked to verify is not the one the structure produces; the work is to find the natural recurrence and see the given one as a consequence **Answer:** The quantity equals 4 for every $n \geq 2$.

Layer 1 THE WHAT — surface analysis

Throughout, write a permutation as the word $\pi(1)\pi(2)\cdots\pi(n)$. The condition says that consecutive letters differ by at most 2: the word is a Hamiltonian path, traversed in a direction, in the graph on $\{1, \ldots, n\}$ joining values that differ by 1 or 2.

The natural attempt. Compute. $P_1, \ldots, P_{10}$ are $1, 2, 6, 12, 20, 34, 56, 88, 136, 208$, and the combination $P_{n+5} - P_{n+4} - P_{n+3} + P_n$ at $n = 2, 3, 4, 5$ is $4, 4, 4, 4$. (At $n = 1$ it is $P_6 - P_5 - P_4 + P_1 = 34 - 20 - 12 + 1 = 3$. Keep that 3; it matters later.) So the answer is 4. Now prove it.

The standard attack on a permutation-count recurrence is to insert or delete the largest element: take a valid word of length $n + 1$, delete $n + 1$, and ask what is left. The letter $n + 1$ has neighbours among $\{n - 1, n\}$ only, so it sits at an end of the word next to n or $n - 1$, or in the interior between $n - 1$ and n. Deleting it from the interior leaves $n - 1, n$ adjacent — fine — but deleting it from an end leaves a word whose end-letter was constrained, and when $n + 1$ sits between $n - 1$ and n you must know whether $n - 1$ and n are consecutive in the smaller word. Each answer depends on a feature of the smaller word that you now have to track, and tracking it requires a further feature one step down. This is what Kedlaya's solution does: it introduces five classes U_n, V_n, W_n, T_n, S_n, derives five coupled recurrences by case analysis on where $n + 1$ can go, and then pushes P_{n+5} down through ten lines of substitutions until $P_{n+4} + P_{n+3} - P_n + 4$ appears.

That is a proof. Its cost is that the five classes are chosen with the target in view, the case analysis for each has to be exhaustive and is easy to get wrong, and — worst — nothing in it tells you *why* the coefficients $1, -1, -1, 0, 0, 1$ are the ones that appear. If the problem had asked for $P_{n+5} - P_{n+4} - P_{n+3} + P_{n+1}$ you would not know until the last line whether it was constant.

The model chain. Work from the small end, where the structure is rigid, and let the recurrence be whatever it is.

Symmetry. For $n \geq 2$, reversing the word is a fixed-point-free involution that swaps "1 before 2" with "2 before 1". So $P_n = 2P'_n$, where P'_n counts valid words with 1 before 2. Split $P'_n = Q_n + R_n$: Q_n counts valid words beginning with 1, and R_n counts valid words with 1 before 2 and 1 not first.

Lemma (paths from 1 to 2). For each $n \geq 2$ there is exactly one valid word of length n that begins with 1 and ends with 2. *Proof.* $n = 2$: 12. $n = 3$: 132. For $n \geq 4$: the second letter is 3 (not 2, which is last). The letter before the final 2 is a neighbour of 2 other than 1, so 3 or 4; it is not 3, since 3 is second and the word would then be 132 with $n = 3$. So the word is $1\,3 \cdots 4\,2$, and deleting 1 and 2 leaves a valid word on $\{3, \ldots, n\}$ from 3 to 4; conversely each such word extends. Subtracting 2 from each letter, induction gives exactly one. (It is $1\,3\,5 \cdots 6\,4\,2$.) □

Recurrence for Q. Let $n \geq 4$ and take a valid word starting with 1. The second letter is 2 or 3.

- $1\,2 \cdots$: deleting 1 leaves a valid word on $\{2, \ldots, n\}$ starting at 2, and conversely. Q_{n-1} words.
- $1\,3\,2 \cdots$: the letter after 2 is a neighbour of 2 not yet used, so it is 4. Deleting $1, 3, 2$ leaves a valid word on $\{4, \ldots, n\}$ starting at 4; conversely prefixing 1 3 2 works because $|2 - 4| = 2$. Q_{n-3} words.
- 1 3 with 2 not third: 2's neighbours in the word lie in $\{1, 3, 4\}$; 1 is first and followed by 3, and 3 is not followed by 2, so 2 has only one available neighbour and must be the last letter. By the Lemma there is exactly 1 such word.

Hence

$$Q_n = Q_{n-1} + Q_{n-3} + 1 \qquad (n \geq 4). \tag{1}$$

Recurrence for R. In a word counted by R_n, the letter 1 is interior, so its two neighbours are exactly 2 and 3; since 1 precedes 2, the word contains 3 1 2.

- 2 is last: the word ends $\cdots 3\,1\,2$. Delete 1, 2 and reverse: a valid word on $\{3, \ldots, n\}$ starting at 3. Conversely every such word, reversed and followed by 1 2, is valid. Q_{n-2} words.
- 2 is not last: its successor is an unused neighbour of 2, hence 4, so the word contains 3 1 2 4. Delete 1 (leaving 3 2 4, still valid), subtract 1 from every letter (giving 2 1 3 on $\{1, \ldots, n-1\}$) and reverse (giving 3 1 2): a word counted by R_{n-1}. Each step is reversible — starting from a word of R_{n-1}, reverse, add 1, and insert 1 between the 3 and the 2, which is valid because $|3 - 1| = 2$ — so this is a bijection. R_{n-1} words.

Hence

$$R_n = R_{n-1} + Q_{n-2} \qquad (n \geq 3). \tag{2}$$

Shift operators. Let E be the shift $(Ex)_n = x_{n+1}$ and $L = E^3 - E^2 - 1$, so $(Lx)_n = x_{n+3} - x_{n+2} - x_n$. The target is

$$P_{n+5} - P_{n+4} - P_{n+3} + P_n = \big((E^5 - E^4 - E^3 + 1)P\big)_n, \qquad E^5 - E^4 - E^3 + 1 = (E^2 - 1)(E^3 - E^2 - 1).$$

So the target is $(LP)_{n+2} - (LP)_n$.

By (1), $(LQ)_m = 1$ for all $m \geq 1$. By (2), $R_{k+1} - R_k = Q_{k-1}$ for $k \geq 2$, so for $m \geq 2$

$$(LR)_{m+1} - (LR)_m = \big(Q_{m+2} - Q_{m+1} - Q_{m-1}\big) = (LQ)_{m-1} = 1.$$

Therefore, for $n \geq 2$,

$$(LP')_{n+2} - (LP')_n = \underbrace{(LQ)_{n+2} - (LQ)_n}_{0} + \underbrace{(LR)_{n+2} - (LR)_n}_{2} = 2,$$

and since $P_m = 2P'_m$ for every index $m \geq 2$ that occurs,

$$P_{n+5} - P_{n+4} - P_{n+3} + P_n = 2 \cdot 2 = 4 \qquad (n \geq 2). \qquad \square$$

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree that the answer is 4 and that the proof must be a recursive decomposition. They split at the choice of which element to decompose on.

(N) "Delete the largest element $n+1$." The largest element's position depends on how $n-1$ and n sit in the smaller word, which forces auxiliary classes, which force more classes.

(M) "Look at where 1 is and what comes after it." The start of a word beginning with 1 has only three possible shapes, 1 2, 1 3 2 4, and the unique zigzag, and they remove 1 and 3 letters respectively.

And a second split, after the counts are in hand: (N) substitutes recurrences into the target until it collapses; (M) factors the target polynomial first, and reads off which recurrence it annihilates.

> *Deleting $n+1$ and deleting 1 are the same thing by the symmetry $k \mapsto n+1-k$, so the difference between them is not the element, it is the anchor. If I also fix that 1 is the* first letter, *the word has a rigid start: after 1 comes 2 or 3, and if 3, then 2 is trapped — it has only 4 left as a neighbour. Trapped elements are what make a count recursive, so I should anchor at an endpoint where something gets trapped.*

> *The steps in my recurrence are 1 and 3: 1 2 costs one letter and 1 3 2 costs three. So the natural characteristic polynomial is $x^3 - x^2 - 1$. Does it divide $x^5 - x^4 - x^3 + 1$? $(x^3 - x^2 - 1)(x^2 - 1) = x^5 - x^4 - x^3 + 1$. Yes. The problem's strange combination is $(E^2 - 1)$ applied to my recurrence — the extra factor is there to kill whatever constants and linear terms the inhomogeneous parts leave behind.*

> *The problem says $n \geq 2$, not $n \geq 1$. Something must fail at $n = 1$. It is $P_1 = 1$: the one-letter word is its own reversal, so $P_1 \neq 2P'_1$. That tells me the proof must go through the reversal symmetry, and that is a check on my decomposition, not an afterthought.*

- ▷ **Cue:** a permutation condition on consecutive entries, $|\pi(i) - \pi(i+1)| \leq k$ $\longrightarrow$ **Move:** read the word as a Hamiltonian path in the graph joining values at distance at most k; count paths by anchoring at an endpoint
- ▷ **Cue:** a counting problem where a small element placed early leaves a neighbour with only one unused neighbour $\longrightarrow$ **Move:** that trapped element forces a fixed block; each forced block is one term of the recurrence, and its length is the shift
- ▷ **Cue:** "show that $\sum c_i P_{n+i}$ does not depend on n" $\longrightarrow$ **Move:** factor $\sum c_i x^i$; one factor will be the characteristic polynomial of the natural recurrence, and the remaining factor ($x - 1$, $x^2 - 1$, ...) kills the inhomogeneous terms
- ▷ **Cue:** a hypothesis "for $n \geq 2$" on an identity that looks uniform $\longrightarrow$ **Move:** find the case where the identity fails; it identifies the symmetry or base case the proof must use
- ▷ **Cue:** a count invariant under reversal of the word $\longrightarrow$ **Move:** halve it: count only words with 1 before 2, and double at the end

Layer 3 THE WHY

Why should anyone anchor at 1 and at the start, and why should anyone factor the polynomial? Both have reasons that sit in the definitions.

The anchor. A recursive count needs an operation that removes a fixed number of letters and leaves an object of the same kind. In the path graph with edges of length 1 and 2, the vertices 1 and n have degree 2, the vertices 2 and $n-1$ have degree 3, and every other vertex has degree 4. The low-degree vertices are at the ends of the value range. Low degree means few choices, and few choices is what makes a case analysis short and exhaustive. So the ends of the value range are forced. Among positions, the ends of the word are where a vertex uses only one of its edges, so anchoring 1 at the first position leaves exactly two choices for the next letter, and the "trapping" of 2 after 1 3 is an immediate degree count: 2 has neighbours $1, 3, 4$, two of which are already used up. Every line of the Q-recurrence is this degree count. Deleting $n+1$ is the mirror image of deleting 1, but without fixing its position the choices do not collapse — that is exactly why route (N) needs five classes.

The factorisation. A linear recurrence with constant coefficients is the statement that a polynomial in the shift E annihilates the sequence. Two facts about shift operators do all the work here: polynomials in E commute and multiply like polynomials; and $(E-1)$ lowers the degree of a polynomial sequence by one, so $(E-1)$ kills constants and $(E-1)^2$ kills linear sequences. The problem hands you $x^5 - x^4 - x^3 + 1$. The phrase "does not depend on n" means "is annihilated by $E-1$". So the claim is that $(E-1)(E^5 - E^4 - E^3 + 1)$ annihilates P — which means P is built from a homogeneous part with some characteristic polynomial dividing $x^5 - x^4 - x^3 + 1$ and a polynomial-in-n part of low degree. The structure gave $x^3 - x^2 - 1$ from the step lengths 1 and 3, plus a constant from the lone zigzag, plus one more level of summation from R. Put together: LQ is constant, LR is linear, so LP' is linear, and (E^2-1) applied to a linear sequence $\ell(n)$ gives $\ell(n+2) - \ell(n)$, twice the slope. The constant $LQ = 1$ is the lone zigzag; R is a running sum of Q, so LR grows by $LQ = 1$ per step, which is the slope. Then (E^2-1) turns slope 1 into 2, the reversal doubles it, and $2 \cdot 2 \cdot 1 = 4$.

Kedlaya's solution introduces U_n, V_n, W_n, T_n, S_n without saying why those five. Here is where they come from: they are the states of a finite automaton that reads the word near the largest element, and the number of states is what the choice "delete the maximum, position unknown" costs. His ten-line reduction is the automaton's transfer matrix being eliminated by hand. The Mahlburg remark in the official solutions is the anchored version given here; the operator factorisation is what makes it finish in three lines instead of ten.

Finally, the $n \geq 2$: at $n = 1$ the target is $P_6 - P_5 - P_4 + P_1 = 34 - 20 - 12 + 1 = 3$, not 4, because $P_1 = 1 = 2P_1' - 1$. The reversal involution has a fixed point only when the word has length 1. The hypothesis in the problem is the reversal symmetry announcing itself.

Caveat. Every recurrence above was checked by exhaustive enumeration for $n \leq 13$: $Q_n - Q_{n-1} - Q_{n-3} = 1$ for $4 \leq n \leq 13$; $R_n = R_{n-1} + Q_{n-2}$ for $3 \leq n \leq 13$; $P_n = 2P_n'$ for $2 \leq n \leq 13$; $(LP')_n = n+1$ for $1 \leq n \leq 10$; and $P_{n+5} - P_{n+4} - P_{n+3} + P_n = 4$ for $2 \leq n \leq 10$, with the value 3 at $n = 1$. The proofs given are complete; the enumeration is corroboration, not the argument.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the condition itself, $|i-j| = 1 \Rightarrow |\pi(i) - \pi(j)| \leq 2$: consecutive letters differ by at most 2. Write down "Hamiltonian path in the graph with edges of length 1 and 2" and list the degrees: $2, 3, 4, \ldots, 4, 3, 2$. The second cue is the target polynomial; write down its factorisation $(x^2-1)(x^3 - x^2 - 1)$ before computing anything.

2. Anchor at a lowest-degree vertex in an end position, and derive a recurrence by forced-block analysis; then use shift-operator algebra to relate the natural recurrence to the requested one.

3. Hesitation occurs at the moment the natural recurrence turns out to be third order with a constant term, while the target is fifth order with no constant: it looks as if you have found the wrong recurrence. It is reinforced away by dividing polynomials: $x^3 - x^2 - 1$ divides

the target, and the quotient $x^2 - 1$ has 1 as a root, which is exactly what absorbs a constant. A second hesitation is at R_n: the bijection with R_{n-1} (delete, shift, reverse) feels suspicious. Check it on a small case — $n = 4$: $R_4 = 2$, the words 3 1 2 4 and 4 3 1 2 — and it holds.

4. Two traps. First, the insert-the-maximum reflex, which produces a correct but opaque five-state system in which the target identity appears only after long substitution. Second, $n = 1$: if you verify the identity at $n = 1$ you get 3 and may conclude that your enumeration is wrong. It isn't; the problem excluded $n = 1$ because $P_1 \neq 2P_1'$.

5. Faster next time: whenever a problem asks you to prove that a specific linear combination of a sequence is constant, factor the characteristic polynomial of the combination before touching the combinatorics. The factors tell you which recurrence to look for, and the factor $x - 1$ (or $x^2 - 1$) tells you how many layers of inhomogeneity to expect.

6. Reusable decision rule: *a given recurrence is usually a consequence of a simpler natural one; find the natural one from the structure, then divide.* On the structural side, decompose at the most constrained element in the most constrained position, where constraints propagate as forced blocks. On the algebraic side, treat the recurrence as a polynomial in the shift operator, and remember that $(E - 1)^k$ annihilates polynomials of degree below k. This handles tiling counts, lattice-path counts and constrained-permutation counts alike.

2015 B6

2015 B6. For each positive integer k, let $A(k)$ be the number of odd divisors of k in the interval $[1, \sqrt{2k})$. Evaluate

$$\sum_{k=1}^{\infty} (-1)^{k-1} \frac{A(k)}{k}.$$

Fields: analysis (conditionally convergent series, integral representations); elementary number theory (divisor sums) **Position:** B6 — the last problem, hard in two separate ways: the rewriting is short once seen, and the analysis needed to make it legitimate is where most attempts silently fail **Answer:** $\dfrac{\pi^2}{16}$.

Layer 1 THE WHAT — surface analysis

Notation: $T_\ell = \sum_{m \geq \ell} \frac{(-1)^{m-1}}{m}$ denotes the tail of the alternating harmonic series from $m = \ell$; it converges by the alternating series test, and $|T_\ell| \leq \frac{1}{\ell}$ (an alternating tail with decreasing terms is bounded by its first term).

The natural attempt. First contact: tabulate. $A(1) = 1$, $A(2) = 1$, $A(3) = 1$ ($3 \not< \sqrt{6}$), $A(4) = 1$, $A(5) = 1$, $A(6) = 2$ (1, 3 with $3 < \sqrt{12}$), … and the partial sums wander round 0.6 with no recognisable limit. A divisor-counting function in a series is the standard signal to swap the order of summation — count pairs rather than integers — so write $k = dm$ with d odd and sum over d first:

$$\sum_{k} \frac{(-1)^{k-1} A(k)}{k} \overset{?}{=} \sum_{d \text{ odd}} \frac{1}{d} \sum_{m:\, d < \sqrt{2dm}} \frac{(-1)^{dm-1}}{m}.$$

This is the right idea, but a strong student now stalls in one of two places. One is technical: the original series does not converge absolutely ($A(k) \geq 1$, since $d = 1$ always counts, so $\sum A(k)/k$ diverges like the harmonic series), and its terms are not monotone, so the swap marked $\overset{?}{=}$ is unjustified and the alternating series test does not even show the left side converges. The other is computational: after simplifying the inner condition, the inner sum is a tail T_ℓ of the alternating harmonic series, which has no closed form, and you are left with $\sum_\ell \frac{T_\ell}{2\ell - 1}$ — a sum of non-elementary quantities. Expanding $T_\ell = \ln 2 - \sum_{m<\ell} \frac{(-1)^{m-1}}{m}$ turns it into a double sum of the same difficulty. That is the wall: the series is in a form where neither summation can be carried out first.

The model chain. **Step 1: rewrite $A(k)$ as a count of pairs.** An odd divisor d of k is $d = 2\ell - 1$ with $\ell \geq 1$, and $k = (2\ell-1)m$ for a positive integer m. The condition $d < \sqrt{2k}$ is $d^2 < 2dm$, i.e. $2\ell - 1 < 2m$, i.e. $m \geq \ell$. Since d is odd, $k \equiv m \pmod 2$, so $(-1)^{k-1} = (-1)^{m-1}$. Hence each term $(-1)^{k-1}A(k)/k$ is the finite sum of $\frac{(-1)^{m-1}}{(2\ell-1)m}$ over the pairs (ℓ, m) with $m \geq \ell$ and $(2\ell-1)m = k$.

Step 2: the series converges, to the rearranged sum. Let

$$S = \sum_{\ell \geq 1} \frac{T_\ell}{2\ell - 1}.$$

This converges absolutely, since $\left|\frac{T_\ell}{2\ell-1}\right| \leq \frac{1}{\ell(2\ell-1)}$. Fix n. The partial sum $\Sigma_n = \sum_{k \leq n}(-1)^{k-1}A(k)/k$ is a finite sum over pairs with $m \geq \ell$ and $(2\ell-1)m \leq n$; such pairs exist for a given ℓ exactly when $\ell(2\ell-1) \leq n$, and then m runs from ℓ to $M_\ell = \lfloor n/(2\ell-1) \rfloor$. So

$$S_n - \Sigma_n = \sum_{\ell(2\ell-1) \leq n} \frac{1}{2\ell-1} \sum_{m > M_\ell} \frac{(-1)^{m-1}}{m}, \qquad S_n := \sum_{\ell(2\ell-1) \leq n} \frac{T_\ell}{2\ell-1}.$$

Each inner tail is at most $\frac{1}{M_\ell+1} < \frac{2\ell-1}{n}$ in absolute value, so each ℓ contributes less than $\frac{1}{n}$. Since $\ell(2\ell-1) \geq \ell^2$, at most $\sqrt{n}$ values of ℓ occur. Thus $|S_n - \Sigma_n| < \frac{1}{\sqrt{n}} \to 0$, and $S_n \to S$ because S converges. So the given series converges, and its value is S.

Step 3: an integral formula for the tail. For $0 \leq t \leq 1$ and $M \geq \ell$,

$$\sum_{m=\ell}^{M} (-1)^{m-1} t^{m-1} = \frac{(-t)^{\ell-1} - (-t)^M}{1+t}.$$

Integrate over $[0,1]$. The term with $(-t)^M$ has absolute value at most $\int_0^1 t^M\,dt = \frac{1}{M+1} \to 0$, so

$$T_\ell = \int_0^1 \frac{(-t)^{\ell-1}}{1+t}\,dt = \int_0^1 \frac{2(-1)^{\ell-1}u^{2\ell-1}}{1+u^2}\,du \qquad (t = u^2).$$

Step 4: sum over ℓ under the integral. With $a_L(u) = \sum_{\ell=1}^{L} \frac{(-1)^{\ell-1}u^{2\ell-1}}{2\ell-1}$,

$$\sum_{\ell=1}^{L} \frac{T_\ell}{2\ell-1} = \int_0^1 a_L(u)\,\frac{2\,du}{1+u^2}.$$

Integrating the finite geometric identity $\frac{1}{1+s^2} = \sum_{\ell=1}^{L}(-1)^{\ell-1}s^{2\ell-2} + \frac{(-1)^L s^{2L}}{1+s^2}$ from 0 to u gives $|\arctan u - a_L(u)| \leq \frac{u^{2L+1}}{2L+1} \leq \frac{1}{2L+1}$ on $[0,1]$. Hence the right side differs from $\int_0^1 \arctan u\,\frac{2\,du}{1+u^2}$ by at most $\frac{\pi/2}{2L+1} \to 0$, and

$$S = \int_0^1 \frac{2\arctan u}{1+u^2}\,du = \Big[(\arctan u)^2\Big]_0^1 = \left(\frac{\pi}{4}\right)^2 = \frac{\pi^2}{16}. \qquad \square$$

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains reach the same rewritten form,

$$\sum_{\ell \geq 1} \frac{1}{2\ell-1} \sum_{m \geq \ell} \frac{(-1)^{m-1}}{m}. \tag{$*$}$$

They split at the next line. The natural attempt tries to *evaluate* an inner or outer sum in $(*)$ as a number and fails, since neither has a closed form while the other index is free. The model writes both reciprocals as integrals of powers, $\frac{(-1)^{m-1}}{m} = \int_0^1 (-t)^{m-1}dt$, so that the sums become geometric series inside an integral — and geometric series always have closed forms. A second, independent split is at the rearrangement itself: the natural attempt writes $\stackrel{?}{=}$ and moves on; the model proves the partial sums converge to $(*)$ before touching it.

> *$A(k)$ counts divisors in a range, so I should count pairs (d, m) with $dm = k$ instead. The range is $d < \sqrt{2k}$ — a square root of k — and with $k = dm$ that is $d^2 < 2dm$, i.e. $d < 2m$: the square root was there to make the condition linear in the pair. The odd d is there to make $(-1)^k = (-1)^m$. So the problem's odd-looking definitions are exactly what makes the pair sum clean.*

> *This series is not absolutely convergent — $d = 1$ always counts, so $A(k) \geq 1$ — so I am not allowed to reorder it. Before I compute anything I must show the partial sums in k approach the partial sums of my rearrangement. The discrepancy is a handful of alternating tails, each short by $1/n$, and there are only about $\sqrt{n}$ of them.*

> *The inner sum is a tail of $1 - \frac{1}{2} + \frac{1}{3} - \cdots$ and has no closed form as a number. It does have one as an integral: $\frac{1}{m} = \int_0^1 t^{m-1}\,dt$ turns every sum of $\pm\frac{1}{m}$ into an integral of a geometric series. And $\frac{1}{2\ell-1}$ is also $\int u^{2\ell-2}$ — that is* arctan *waiting to happen.*

- ▷ **Cue:** a series involving a divisor-counting function $\sum_{d|k,\ d\in I(k)}$ ⟶ **Move:** write $k = dm$ and sum over pairs (d, m); a threshold like $d < \sqrt{ck}$ becomes the linear condition $d < cm$
- ▷ **Cue:** a series with non-negative $A(k)$ bounded below and a sign $(-1)^k$ ⟶ **Move:** it is conditionally convergent at best; prove the partial sums match the rearranged form before rearranging
- ▷ **Cue:** a sum of $\pm\frac{1}{m}$ or $\frac{1}{2\ell-1}$ with no closed form ⟶ **Move:** replace $\frac{1}{m}$ by $\int_0^1 t^{m-1}\,dt$; the sum becomes an integral of a geometric series
- ▷ **Cue:** an answer of the form $\pi^2/16 = (\pi/4)^2$ ⟶ **Move:** expect the square of the Leibniz series, and an integral of the shape $\int 2ff' = f^2$ with $f = \arctan$
- ▷ **Cue:** interchanging an infinite sum and an integral ⟶ **Move:** use the finite geometric identity with an explicit remainder, and bound the remainder integral; this avoids convergence theorems entirely

Layer 3 THE WHY

There are three moves, and each is forced by something on the page.

The pair rewriting. A divisor count is by definition a count of factorisations $k = dm$. Summing a divisor count over k is therefore summing over the set of pairs (d, m), sorted by product — the pairs are the real index set and k is a label on them. So the question is only how the restriction on d looks in pair coordinates, and here the problem's choices announce themselves. $\sqrt{2k}$ is the threshold for which $d < \sqrt{2k} \iff d < 2m$: a square-root threshold on a divisor is always a linear threshold on the pair, since $d \leq \sqrt{ck}$ compares d with k/d. And "odd" is the condition under which $(-1)^k = (-1)^m$, so the sign lives on one coordinate. The problem was written in pair coordinates and translated into k; the rewriting translates it back.

The convergence argument. It is forced by a one-line check that every competitor should run: is $\sum |a_k|$ finite? Here $|a_k| = A(k)/k \geq 1/k$, so no. Riemann's rearrangement theorem says a conditionally convergent series can be rearranged to any value, so the sum over k and the sum $(*)$ over pairs are *a priori* different objects. The comparison has to be made at finite n, and the natural thing to compare is what each version has included by the time k reaches n. The sum over $k \leq n$ includes, for each ℓ, an initial segment of the inner alternating series; $(*)$ includes the whole of it. The difference is a set of alternating tails, and an alternating tail is at most its first term. The proof in Step 2 is nothing but this accounting. Kedlaya's solution states the bound "$\sqrt{n}$ summands, each at most $\frac{1}{n}$" in one line; the reason it works is that the threshold $d < \sqrt{2k}$ caps the number of ℓ at about $\sqrt{n}$ — the same threshold that made the pairs clean also makes the error small.

The integral representation. Kedlaya writes $\frac{(-1)^{m-1}}{m} = (-1)^{m-1}\int_0^1 t^{m-1}\,dt$ without comment. Here is where it comes from. The obstruction in $(*)$ is that $\frac{1}{m}$ does not factor as a function of m in any way that geometric summation can use; t^{m-1} does. The identity $\frac{1}{m} = \int_0^1 t^{m-1}\,dt$ is the exchange of a reciprocal (hostile to summation) for a power (the one thing summation handles perfectly), paid for with an integral you can postpone. After the exchange, $\sum_{m\geq\ell}(-t)^{m-1} = \frac{(-t)^{\ell-1}}{1+t}$ — the tail of a geometric series is a single term, unlike the tail of the harmonic series. Then the outer factor $\frac{1}{2\ell-1}$ with sign $(-1)^{\ell-1}$ and the power $u^{2\ell-1}$ assemble into the Taylor series of $\arctan u$, which is exactly the function whose derivative $\frac{1}{1+u^2}$ is already sitting in the integrand. The final integral $\int_0^1 2\arctan u\cdot\arctan' u\,du$ is the chain rule for $(\arctan u)^2$. Its value is $(\arctan 1)^2 = (1-\frac{1}{3}+\frac{1}{5}-\cdots)^2$: the answer is the square of the Leibniz series, which is what a "triangular" double sum over $\ell \leq m$ tends to produce — $\int_0^1 2ff'$ is the continuous version of summing a product over the half-square $\ell \leq m$.

Why is this hard even so? Because each move is standard and the problem requires all three, and because Steps 2 and 4 are where a correct-looking solution becomes an incorrect one. Putnam graders on B6 read for exactly those two justifications. This is a problem where the answer can be guessed from the heuristic calculation and the marks are in the analysis.

Caveat. Numerically, the partial sum Σ_n at $n = 2\times10^6$ is 0.6168317, against $\pi^2/16 =$ 0.6168503; the difference 1.9×10^{-5} is well inside the proved bound $|S_n - \Sigma_n| < 1/\sqrt{n} \approx 7\times10^{-4}$. Steps 3 and 4 use only finite geometric identities with integrated remainders, in place of the official solution's appeal to monotone convergence, so the argument is self-contained at first-year analysis level; the tail bound $|T_\ell| \leq \frac{1}{\ell}$ and the alternating tail bound in Step 2 are the standard alternating-series estimate.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "the number of odd divisors of k in $[1,\sqrt{2k})$". On seeing it, write $k = (2\ell-1)m$ and translate every condition: $d < \sqrt{2k} \iff m \geq \ell$, and $(-1)^{k-1} = (-1)^{m-1}$. The second cue is the absence of absolute convergence; write "$A(k) \geq 1$, so conditionally convergent at best" on the page before any rearrangement.

2. Swap to a double sum over divisor pairs, justified by a truncation comparison; then evaluate by integral representation of $\frac{1}{m}$, turning the inner sum into a geometric tail and the outer sum into the arctangent series.

3. Hesitation occurs at $(*)$: both sums look non-evaluable and the solver goes looking for a closed form for T_ℓ, or for a telescoping identity. It is reinforced away by the rule that a sum of reciprocals with no closed form should be converted to an integral of powers — after which the tail is one term. The second hesitation is over rigour: how much justification do the interchanges need? Enough to bound an explicit remainder; the finite geometric identity supplies the remainder for free.

4. The trap is to compute correctly and prove nothing: rearrange a conditionally convergent series, interchange sum and integral, and arrive at $\pi^2/16$ with no argument that the original series converges, let alone to that value. The answer is right and the solution scores poorly. The secondary trap is to try to evaluate $\sum_\ell T_\ell/(2\ell-1)$ by writing $T_\ell = \ln 2 - (\text{partial sum})$, which trades one hard double sum for another.

5. Recognise faster by three reflexes: a divisor function in a series means "sum over pairs"; a sign and a non-summable modulus means "prove convergence of the rearrangement at finite n"; a reciprocal $\frac{1}{m}$ that obstructs summation means "$\int_0^1 t^{m-1}\,dt$". And the answer $\pi^2/16$ in hindsight is $(\arctan 1)^2$, a signal to look for arctan and its derivative together.

6. Reusable decision rule: *when a sum will not evaluate, change what is being summed, not how hard you try.* Replace reciprocals by integrals of powers, so that sums become geometric and

integrals do the remaining work; and whenever you reorder or interchange, pay at finite truncation with an explicit error term rather than invoking a theorem you have not checked the hypotheses of. The same pair of habits evaluates $\sum \frac{1}{n^2}$-type constants, Dirichlet series of divisor functions, and most "evaluate this infinite sum" problems whose answer involves π or $\ln 2$.