DAWN PUTNAM SESSION

Putnam 2009 — Papers A and B

Putnam 120 Archive

Contents

Putnam 2009 — Paper A **5**
2009 A1 5
2009 A2 7
2009 A3 11
2009 A4 13
2009 A5 16
2009 A6 19

Putnam 2009 — Paper B **23**
2009 B1 23
2009 B2 25
2009 B3 28
2009 B4 30
2009 B5 33
2009 B6 36

Putnam 2009 — Paper A

2009 A1

2009 A1. Let f be a real-valued function on the plane such that for every square $ABCD$ in the plane, $f(A) + f(B) + f(C) + f(D) = 0$. Does it follow that $f(P) = 0$ for all points P in the plane?

Fields: plane geometry, linear algebra of functional equations **Position:** A1 — the opener **Answer:** Yes, it follows: $f \equiv 0$. Six instances of the given relation, built from a square centred at P and its four quarter-squares, combine to force $4f(P) = 0$.

Layer 1 THE WHAT — surface analysis

The natural attempt. Fix the simplest square you can write down: the unit square with $A = (0,0)$, $B = (1,0)$, $C = (1,1)$, $D = (0,1)$. This gives one equation, $f(0,0) + f(1,0) + f(1,1) + f(0,1) = 0$. Since the hypothesis holds for every square, the obvious next move is to march this same square across the plane and start subtracting. Take the neighbouring square $B' = (1,0), C' = (2,0), D' = (2,1), A' = (1,1)$:

$$f(1,0) + f(2,0) + f(2,1) + f(1,1) = 0.$$

Subtracting the first equation from this one cancels $f(1,0) + f(1,1)$ and gives $f(2,0) + f(2,1) = -f(0,0) - f(0,1)$. Continue the march and a whole grid of relations accumulates: values at $(m,0), (m,1)$ for every integer m get tied to values at $(0,0), (0,1)$ with alternating signs, and rotating the marching square 90° ties in the y-direction the same way. This is honest progress – it is easy to believe that enough marching, in enough directions, at enough scales, eventually pins every lattice value to 0.

But it stalls exactly where the problem is actually asking a question: at an arbitrary point P, not a lattice point. None of these translated unit squares has a vertex at a generic $P = (x_0, y_0)$ unless x_0, y_0 happen to be integers. To reach $f(P)$ for a general P, the whole grid construction has to be re-anchored – rescaled and re-centred – at P from scratch, and even then it is not obvious in advance that the marching relations telescope down to $f(P) = 0$ rather than to some undetermined constant. The approach is not wrong; it is expensive, and it never explains why $f(P) = 0$ has to be true for a point chosen with no relation to any convenient grid.

The model chain. Let P be an arbitrary point in the plane. Build a square $ABCD$ of side 2 centred at P, say (after translating so P is the origin, then translating back) $A = (0,0)$, $B = (2,0)$, $C = (2,2)$, $D = (0,2)$, so $P = (1,1)$. Let E, F, G, H be the midpoints of AB, BC, CD, DA: $E = (1,0)$, $F = (2,1)$, $G = (1,2)$, $H = (0,1)$.

Two facts, checked directly from coordinates: $EFGH$ is itself a square (side $\sqrt{2}$, centre P, rotated 45° from $ABCD$), and each of $AEPH$, $BFPE$, $CGPF$, $DHPG$ is a unit square. This gives

six admissible instances of the hypothesis:

$$\begin{aligned}
0 &= f(A) + f(B) + f(C) + f(D)\\
0 &= f(E) + f(F) + f(G) + f(H)\\
0 &= f(A) + f(E) + f(P) + f(H)\\
0 &= f(B) + f(F) + f(P) + f(E)\\
0 &= f(C) + f(G) + f(P) + f(F)\\
0 &= f(D) + f(H) + f(P) + f(G).
\end{aligned}$$

Add the last four equations. Each of A, B, C, D appears once, each of E, F, G, H appears exactly twice, and $f(P)$ appears four times:

$$0 = \big(f(A)+f(B)+f(C)+f(D)\big) + 2\big(f(E)+f(F)+f(G)+f(H)\big) + 4f(P).$$

The first bracket is 0 by the first equation, the second bracket is 0 by the second equation, so $0 = 4f(P)$, i.e. $f(P) = 0$. Since P was an arbitrary point, $f \equiv 0$ on the whole plane.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both approaches start by writing down a single square and its equation – that much is forced by the problem. They split at what to do with the freedom the hypothesis grants: "for every square $ABCD$." The natural attempt spends that freedom translating one fixed, convenient square around the plane. The model chain spends it choosing squares whose vertices are deliberately placed relative to P itself.

> *The hypothesis holds for every square, of any size, position, and rotation – so instead of marching one fixed square and hoping the point I care about eventually lands on a vertex, I should ask: what square (or squares) can I draw so that P itself is a vertex from the very first equation?*

> *A single square with P as a vertex gives one equation containing $f(P)$, but three other unknown values alongside it. I need several such squares, sharing P as a common vertex, arranged so that when I add their equations, every value except $f(P)$ cancels – and a square's own centre, together with its edge midpoints, cuts it into four smaller squares that share exactly that centre as a common vertex, which is precisely the shape I need.*

Layer 3 THE WHY

Why does centring the construction at P occur to the model solver and not the marching solver? Because the object being asked about, $f(P)$, is a value at a *point*, and the hypothesis only ever produces an equation containing $f(P)$ when P is chosen to be a *vertex* of one of the squares used. A translating grid of fixed unit squares puts vertices at lattice points, which is a choice made before P is known and has no reason to include P. The moment the target of the proof is a value at a named point, the right question is not "which square is simplest to write down" but "which square, or family of squares, has that point as a vertex" – and since *every* square is admissible, that freedom should be used to serve the goal, not convenience.

The second load-bearing fact is where the four small squares $AEPH, BFPE, CGPF, DHPG$ come from, and this is not an ad hoc construction: it follows from what makes $ABCD$ a square in the first place. A square's diagonals bisect each other at the centre and are equal and perpendicular; consequently the segment from a vertex to the centre, and the segment from that vertex to each adjacent edge-midpoint, are themselves the two perpendicular half-diagonals of a smaller square – this is forced by the same equal-sides-and-right-angles definition that makes $ABCD$ a square, applied at half scale. So decomposing $ABCD$ into four congruent squares

sharing the centre P is not a trick pulled in from outside the problem; it is the structure a square already has, read off its definition.

Once that decomposition exists, the reason summing the four small-square equations works is a counting fact, not a geometric one: P lies on all four small squares (multiplicity 4), each original vertex A, B, C, D lies on exactly one small square (multiplicity 1), and each midpoint E, F, G, H lies on exactly two adjacent small squares (multiplicity 2), because each midpoint sits on the shared edge between two of the four small squares. Those multiplicities are exactly the coefficients $1, 2, 4$ that appear when the sum is taken, and they are exactly what is needed to reduce the sum to a multiple of $f(P)$ once the two outer relations ($ABCD$ itself, and $EFGH$) are subtracted out. The proof is really an accounting of how many small squares each point of the big square touches.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the phrase "for every square $ABCD$ in the plane." On seeing a universally-quantified hypothesis of that strength, the thing to write down immediately is: I get to choose the square, so I should choose it to serve whatever I am trying to prove, not take the first one that comes to hand.

2. That cue suggests linear algebra on the values of f as formal unknowns: pick several squares sharing a common vertex or a nested structure around the target point, and add or subtract their equations until only the target unknown survives.

3. Hesitation sets in right after the first square is written down, when marching it across the plane feels like guaranteed, concrete progress – each new equation is easy to produce. What reinforces the switch away from marching is asking directly: does any equation in this family actually contain $f(P)$ for the specific point P I need? If the answer is no, and stays no as the construction is currently set up, that is the signal to re-anchor the whole construction at P instead of continuing to march.

4. The trap is using the outer square $ABCD$ and the midpoint square $EFGH$ alone and stopping, because both of those equations are true and easy to write down – but neither one contains $f(P)$ at all, so summing just them gives $0 = 0$, no information. The genuinely useful equations are the four small squares that actually touch P; the two big ones are only needed afterward, to cancel the vertices and midpoints out of the sum.

5. Recognise this shape faster by treating any hypothesis of the form "$R(\text{vertices of a figure}) = 0$ for every admissible figure" as an invitation to pick the figure, or a small family of figures, with a distinguished point of the figure (centre, midpoint, vertex) equal to the exact point the question is about.

6. Reusable rule: given a linear relation that holds for every instance of a geometric family, build a small nested set of instances that all pass through your target point, then add or subtract the instances so that every value except the target one is cancelled by a matching multiplicity – count how many times each point of the construction is covered, and let that count tell you which linear combination to take.

▷ **Cue:** a linear relation is asserted to hold for every square/figure through arbitrary points in the plane, and the goal is the value at one specific point P $\longrightarrow$ **Move:** construct a square (or nested family of squares) with P as a vertex or centre, so that at least one instance of the relation contains $f(P)$ directly

▷ **Cue:** a figure decomposes into smaller congruent copies of itself sharing a common point (e.g. a square's centre and edge midpoints cut it into four smaller squares) $\longrightarrow$ **Move:** use that decomposition to write several instances of the hypothesis sharing the common point, then sum them and subtract the outer instances to isolate the value at the common point

2009 A2

2009 A2. Functions f, g, h are differentiable on some open interval around 0 and satisfy the equations and initial conditions

$$f' = 2f^2gh + \frac{1}{gh}, \quad f(0) = 1,$$

$$g' = fg^2h + \frac{4}{fh}, \quad g(0) = 1,$$

$$h' = 3fgh^2 + \frac{1}{fg}, \quad h(0) = 1.$$

Find an explicit formula for $f(x)$, valid in some open interval around 0.

Fields: ordinary differential equations, algebraic manipulation of coupled systems **Position:** A2 **Answer:** $f(x) = 2^{-1/12}\left(\dfrac{\sin(6x+\pi/4)}{\cos^2(6x+\pi/4)}\right)^{1/6}$, obtained by isolating the invariant combination $k = fgh$, which satisfies $k' = 6k^2 + 6$ and solves to $k = \tan(6x + \pi/4)$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The system is symmetric-looking but not symmetric: the coefficients $2, 1, 3$ on the f^2gh, fg^2h, fgh^2 terms differ, and so do the constants $1, 4, 1$ on the reciprocal terms. A first instinct to test is whether $f = g = h$ solves the system, since all three initial conditions equal 1. If $f = g = h$, the three equations collapse to $f' = 2f^3 + 1/f^2$, $f' = f^3 + 4/f^2$, $f' = 3f^3 + 1/f^2$ respectively – three different differential equations for the same function, inconsistent unless $2f^3 + 1/f^2 = f^3 + 4/f^2 = 3f^3 + 1/f^2$ identically, which forces $f^3 = 3/f^2$ and $f^3 = 0$ simultaneously near $x = 0$, a contradiction. So f, g, h are genuinely different functions, and the symmetric ansatz is dead within a line.

With that ruled out, an honest next attempt is a Taylor expansion around 0. Write $f = 1 + a_1x + a_2x^2 + \cdots$, and likewise for g, h with coefficients b_i, c_i. Substituting into $f' = 2f^2gh + 1/(gh)$ and matching the constant term: $f'(0) = a_1 = 2 \cdot 1 \cdot 1 \cdot 1 + 1/(1 \cdot 1) = 3$. The same computation on the other two equations gives $g'(0) = b_1 = 1 + 4 = 5$ and $h'(0) = c_1 = 3 + 1 = 4$. This is genuine, correct information, and it can be pushed to second order and beyond with more algebra at each step. But "explicit formula for $f(x)$" means a closed form, not a power series with a recognisable-only-in-hindsight coefficient pattern – and nothing in the constant-term matching above suggests that $6x + \pi/4$ or a tangent function is about to appear. The series route produces correct numbers without producing the reason for the eventual closed form; pushed far enough it would in principle reveal the Taylor series of $2^{-1/12}(\sin(6x + \pi/4)/\cos^2(6x + \pi/4))^{1/6}$, but recognising that series as this function, rather than as a growing list of fractions, is not something the expansion itself points toward.

The model chain. Multiply the first equation by gh, the second by fh, the third by fg:

$$f'gh = 2f^2g^2h^2 + 1, \qquad g'fh = f^2g^2h^2 + 4, \qquad h'fg = 3f^2g^2h^2 + 1.$$

Add all three. The left side is $f'gh + g'fh + h'fg$, which is exactly the product rule expansion of $(fgh)' = f'gh + fg'h + fgh'$ (the three terms on the left are the same three terms, just grouped as $f'gh$, $fh \cdot g'$, $fg \cdot h'$). The right side is $6f^2g^2h^2 + 6$. Writing $k(x) = f(x)g(x)h(x)$, this is

$$k' = 6k^2 + 6, \qquad k(0) = 1 \cdot 1 \cdot 1 = 1.$$

This is separable: $dk/(k^2 + 1) = 6\,dx$, so $\arctan k = 6x + C$. At $x = 0$, $\arctan 1 = \pi/4 = C$. Hence

$$k(x) = \tan\left(6x + \frac{\pi}{4}\right).$$

Now return to the original f-equation and use $gh = k/f$:

$$f' = 2f^2 \cdot \frac{k}{f} + \frac{1}{k/f} = 2fk + \frac{f}{k} = f\left(2k + \frac{1}{k}\right).$$

So $f'/f = 2k(x) + 1/k(x)$ with $k(x) = \tan(6x + \pi/4)$ now a fully explicit function of x; this is a single first-order equation in f alone. Writing $u = 6x + \pi/4$, $du = 6\,dx$:

$$\ln f = \int (2\tan u + \cot u)\,\frac{du}{6} = \frac{-2\ln|\cos u| + \ln|\sin u|}{6} + C_1 = -\frac{1}{3}\ln|\cos u| + \frac{1}{6}\ln|\sin u| + C_1.$$

At $x = 0$, $u = \pi/4$, $\cos u = \sin u = \frac{\sqrt{2}}{2} = 2^{-1/2}$, and $f(0) = 1$ forces $\ln 1 = 0 = (-\frac{1}{3} + \frac{1}{6})\ln(2^{-1/2}) + C_1 = \frac{1}{12}\ln 2 + C_1$, so $C_1 = -\frac{1}{12}\ln 2$. Exponentiating,

$$f(x) = 2^{-1/12}\left(\frac{\sin(6x + \pi/4)}{\cos^2(6x + \pi/4)}\right)^{1/6}.$$

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree that the coefficients $2, 1, 3$ break any hope of a symmetric ansatz, and both correctly rule that out fast. They split at what to do next: the natural attempt keeps working *coordinate by coordinate* – expanding f (and implicitly g, h) in isolation, term by term. The model chain instead asks whether the *three equations together*, not any one of them, hide a single new unknown.

> *Each equation's awkward reciprocal term – $1/(gh)$, $4/(fh)$, $1/(fg)$ – is exactly a reciprocal of the two variables missing from that equation's polynomial term. If I multiply each equation by precisely those missing variables, every reciprocal term becomes a plain constant, and every polynomial term becomes a multiple of $(fgh)^2$ – so all three equations, cleared this way, are secretly equations about the single quantity $k = fgh$.*

> *The three cleared left-hand sides, $f'gh$, $g'fh$, $h'fg$, are not just similar to each other – they are literally the three summands of $(fgh)'$ by the product rule. That is not a coincidence to notice after the fact; it is the reason to sum the three equations in the first place, rather than try to solve any one of them alone.*

> *The problem only asks for f, and I now know $k = fgh$ explicitly – but g, h never appear in the f-equation except as the product gh, and $gh = k/f$ is already known once k is known. So I don't need to solve for g and h individually; substituting $gh = k/f$ turns the f-equation into a single, separable first-order equation in f alone.*

Layer 3 THE WHY

Why does "multiply by the missing variables" occur to the model solver rather than the direct one? Because the exponents in the problem are not arbitrary: in the first equation, f^2gh multiplied by gh gives $f^2g^2h^2 = (fgh)^2$, and $1/(gh)$ multiplied by gh gives exactly 1 – the same cancellation happens in the other two equations by design. This is visible only if the reciprocal term is read as "one over the product of the two variables *not* appearing (to a power) in the polynomial term" rather than as an isolated fraction to be dealt with separately. Once that pattern is read off the exponents, clearing denominators this way is not a guess – it is the unique choice of multiplier that turns every right-hand side into a function of fgh alone, because the problem's coefficients were built so that it would.

The second fact – that summing the three cleared equations reconstructs $(fgh)'$ – is the product rule, read backwards. The definition of the derivative of a product of three functions is $d(fgh)/dx = f'gh + fg'h + fgh'$: a sum of exactly three terms, each obtained by differentiating one factor and leaving the other two untouched. The three cleared left-hand sides are exactly those three terms, just each attached to its own original equation. Seeing this is not pattern-spotting; it is recognising that the problem has arranged three separate equations to be, in disguise, the three pieces of one product-rule expansion, and that recognition is what makes summing them the forced move rather than one option among several.

The final step – solving for f once k is known, without ever solving for g or h – follows from a simple but easy-to-miss structural fact: in the original f-equation, g and h never occur separately, only as the product gh. A quantity that only ever appears as a product in the equation you care about does not need to be known individually; only the product needs to be known, and the product is k/f, already available. This is the same principle, one level down, as the reduction from three unknowns to one: look at exactly which combinations of the unknowns actually appear in the equation you are solving, and solve for no more than that.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the shape of the reciprocal terms: $1/(gh)$, $4/(fh)$, $1/(fg)$, each missing precisely the variable that appears (squared) in its own equation's polynomial term. On seeing it, write down: multiplying by the missing variables should turn every right-hand side into a function of the full product fgh alone – check this before doing anything else.

2. That cue suggests reducing a coupled nonlinear system in three unknowns to a single separable ODE in the invariant combination $k = fgh$, solving that ODE in closed form, then substituting back to isolate the one unknown actually asked for.

3. Hesitation sets in right after the symmetric ansatz $f = g = h$ fails: the instinct is to fall back on expanding each function separately in a power series, because that always produces *some* correct numbers with no risk of a wrong guess. What reinforces the switch away from series work is noticing that the three reciprocal terms are missing exactly the right variables to combine into $(fgh)^2$ – once that pattern is seen, clearing denominators is guaranteed to simplify the system, whereas the series route is guaranteed only to produce numbers, with no guarantee they will ever be recognised as a closed form.

4. The trap is treating each of the three equations as its own problem to solve in isolation – differentiating, expanding, or eliminating variables equation by equation – rather than asking what combination of all three equations might decouple. A coupled system's real structure often lives in the combination of the equations, not in any one of them read alone.

5. Recognise this shape faster by treating any system where the i-th equation's derivative term is a monomial in all the unknowns, paired with a reciprocal of the product of the *other* unknowns, as a signal that the product of all the unknowns is the true single degree of freedom – check that before attempting to solve for any one unknown directly.

6. Reusable rule: in a coupled system of n ODEs where each equation mixes a monomial in all n unknowns with a reciprocal of the product of the other $n-1$, multiply each equation by its missing factors to expose the common invariant (typically the full product), sum using the product-rule structure this creates, solve the resulting single-variable equation, then substitute the known invariant back into whichever original equation contains only the specific unknown you need – never solve for more unknowns than the target equation actually contains.

▷ **Cue:** n coupled ODEs, each with a term that is the reciprocal of the product of the $n-1$ *other* unknowns $\longrightarrow$ **Move:** multiply equation i by its missing unknowns to clear denominators, and check whether the cleared right-hand sides become functions of the full product of all unknowns

▷ **Cue:** after clearing, each cleared left-hand side is exactly one term of d/dx(product of all unknowns) by the product rule $\longrightarrow$ **Move:** sum all cleared equations to obtain a single separable (or

otherwise tractable) ODE for that product

▷ **Cue:** a target unknown's own equation only ever involves the other unknowns through a fixed combination of them (e.g. a product gh) $\longrightarrow$ **Move:** solve only for that combination, not for the unknowns individually, and substitute it back to isolate the target

2009 A3

2009 A3. Let d_n be the determinant of the $n \times n$ matrix whose entries, from left to right and then from top to bottom, are $\cos 1, \cos 2, \ldots, \cos n^2$. (For example,

$$d_3 = \begin{vmatrix} \cos 1 & \cos 2 & \cos 3 \\ \cos 4 & \cos 5 & \cos 6 \\ \cos 7 & \cos 8 & \cos 9 \end{vmatrix}.$$

The argument of cos is always in radians, not degrees.) Evaluate $\lim_{n\to\infty} d_n$.

Fields: linear algebra, trigonometric identities **Position:** A3 **Answer:** The limit is 0 – in fact $d_n = 0$ exactly for every $n \geq 3$, because the addition formula for cosine writes the whole matrix as a sum of two rank-1 matrices, so its rank never exceeds 2.

Layer 1 THE WHAT — surface analysis

The natural attempt. The entries are transcendental and there is no algebraic pattern visible along a single row or column, so the honest first move is to compute. $d_1 = \cos 1 \approx 0.540$, not zero. $d_2 = \cos 1 \cos 4 - \cos 2 \cos 3 \approx (0.540)(-0.654) - (-0.416)(-0.990) \approx -0.353 - 0.412 = -0.765$, also nonzero, so there is no reason yet to expect the answer is 0. Only at $n = 3$ does something happen: computing d_3 numerically (nine cosines, a 3×3 determinant expansion) gives a value indistinguishable from 0 to many decimal places. That is suggestive but not a proof – transcendental quantities can produce values that look like zero under finite-precision arithmetic without being exactly zero.

Believing the pattern "$d_n = 0$ for $n \geq 3$" and trying to prove it directly invites cofactor expansion or row reduction on the literal matrix of cosines. This is where the natural attempt stalls: expanding a 3×3, let alone an $n \times n$, determinant of cosines of consecutive integers by brute force produces a sum of many products of cosines with no shared structure to cancel against; the addition and product-to-sum identities needed to simplify such an expansion have to be invented on the fly, for every pair of terms, with no guarantee that the mess collapses for general n. Proving $d_n = 0$ this way, one determinant expansion at a time, for every $n \geq 3$ simultaneously, is not a realistic route – the approach that got you *to* the conjecture (compute and look) is not the approach that will *prove* it.

The model chain. Index the matrix by row j and column k, $1 \leq j, k \leq n$, so the (j,k) entry is $\cos((j-1)n + k)$ (row j starts right after row $j-1$ ends, at position $(j-1)n$, and runs through $k = 1, \ldots, n$). Apply the angle-addition formula to split the row-dependence from the column-dependence:

$$\cos\big((j-1)n + k\big) = \cos\big((j-1)n\big)\cos k - \sin\big((j-1)n\big)\sin k.$$

Define column vectors $u, w \in \mathbb{R}^n$ by $u_j = \cos((j-1)n)$, $w_j = \sin((j-1)n)$, and row vectors $v, z \in \mathbb{R}^n$ by $v_k = \cos k$, $z_k = \sin k$. Then the whole matrix is

$$M = u\,v^{\mathsf{T}} - w\,z^{\mathsf{T}},$$

a sum of two rank-1 matrices, since $u_j v_k$ is the (j,k) entry of the outer product uv^{T} and likewise for $w_j z_k$. The rank of a sum of two matrices never exceeds the sum of their ranks, so $\operatorname{rank}(M) \leq \operatorname{rank}(uv^{\mathsf{T}}) + \operatorname{rank}(wz^{\mathsf{T}}) \leq 1 + 1 = 2$.

For $n \geq 3$, M is an $n \times n$ matrix with $\text{rank}(M) \leq 2 < n$. A square matrix with rank strictly less than its size has linearly dependent columns, and a matrix with linearly dependent columns has determinant 0 (by multilinearity and the alternating property of the determinant: if column k_0 is a linear combination of the others, subtracting that combination from column k_0 turns it into the zero column without changing the determinant, and a determinant with a zero column is 0). Hence $d_n = 0$ for every $n \geq 3$, and in particular $\lim_{n\to\infty} d_n = 0$.

This also explains, rather than just confirms, the official column trick: adding column 3 to column 1 of the matrix and applying $\cos x + \cos y = 2\cos\frac{x+y}{2}\cos\frac{x-y}{2}$ produces a first column proportional to the second. That proportionality is not a lucky accident to be discovered by trial; it is guaranteed in advance by $\text{rank}(M) \leq 2$, which forces *any* three columns of M (for $n \geq 3$) to be linearly dependent. The addition-formula argument is what makes that dependency exist; the column trick is one particular way of exhibiting it.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start the same way: write the matrix out, index its entries by row and column, and notice $\cos((j-1)n+k)$ is a cosine of a sum. They split immediately after that observation. The natural attempt treats the sum inside the cosine as just an argument to be computed, and proceeds to direct determinant expansion. The model chain treats the sum inside the cosine as an instruction: *this entry is a value of* $\cos(a+b)$*, and* $\cos(a+b)$ *always factors.*

> *The entry in row j, column k is* $\cos\big($*(something depending only on j) + (something depending only on k)*$\big)$ – and* $\cos(a+b)$ *always expands, by the addition formula, into a sum of two products, each product being one function of a times one function of b. That means every entry of this matrix, no matter how large n is, is built from only two row-only functions and two column-only functions – so the matrix itself can never need more than two independent directions to describe it.*

> *I don't need to find a specific linear dependency among the columns by trial-and-error row reduction. Rank is a property of the whole matrix, not of any one combination I happen to try, and if the addition formula already bounds the rank at 2, then a dependency among any three columns is guaranteed to exist before I go looking for it – so I should establish the rank bound first, and only then, if I want an explicit witness, go find one.*

Layer 3 THE WHY

Why does the addition formula occur to the model solver as a statement about *rank* rather than as an isolated simplification tool? Because $\cos(a+b) = \cos a\cos b - \sin a\sin b$ says something about the function of two variables $(a,b) \mapsto \cos(a+b)$: it lies in the two-dimensional space spanned by the two *separable* functions $\cos a\cos b$ and $\sin a\sin b$ – functions that are each a product of a function of a alone and a function of b alone. A matrix whose (j,k) entry is a separable function of (j,k) is, by definition, a rank-1 matrix (an outer product); a matrix whose entries are a sum of two separable functions is a sum of two rank-1 matrices, hence rank at most 2. This is not a fact about cosines specifically – it is what the addition formula *is*, read as a statement about the function $\cos(a+b)$ rather than as an identity for simplifying a single number. Once the matrix's rank is recognised as controlled by the number of terms in the addition formula, checking that bound against n is the natural next step, because rank versus size is exactly what determines whether a square matrix's determinant must vanish.

The second fact – rank $r < n$ forces determinant 0 for an $n \times n$ matrix – is not an extra theorem imported from outside; it follows directly from what a determinant *is*. The determinant is an alternating multilinear function of the columns, and alternating means it vanishes whenever two columns coincide, which (combined with multilinearity) forces it to vanish whenever one column is a linear combination of the others, since that combination can be subtracted off

without changing the determinant's value, producing an explicit zero column. Rank $r < n$ is precisely the statement that such a dependency exists among the n columns. So the whole proof is two structural facts stacked directly on their definitions: the addition formula bounds the number of independent row/column directions in M, and the definition of determinant as an alternating multilinear form converts any such shortfall in independent directions directly into a zero determinant.

This is also why the official column-addition trick works without needing to be guessed: for $n \geq 3$, three columns of a rank-≤ 2 matrix are automatically dependent, so *some* integer combination of any three columns must vanish. Finding that the specific combination "column 1 plus column 3" happens to be proportional to column 2 is not luck; it is the addition formula for $\cos((j-1)n+1) + \cos((j-1)n+3) = 2\cos((j-1)n+2)\cos 1$ exhibiting, entry by entry, the exact dependency the rank bound already promised.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that the matrix entry in row j, column k is cos of an expression that splits additively into a row-only part and a column-only part. On seeing it, write down: $\cos(a+b)$ always factors via the addition formula, so this matrix's rank is bounded by the number of terms that formula produces – check that bound against n before touching row or column operations.

2. That cue suggests expressing the matrix as a sum of rank-1 outer-product matrices via the addition formula, bounding its rank, and comparing that bound to n to decide whether the determinant must vanish.

3. Hesitation sets in right after the numerical computation suggests $d_3 \approx 0$: the pull is to try to prove it by direct row or column manipulation on the explicit matrix, because that feels like it is working with the actual object rather than an abstract bound. What reinforces the switch to the rank argument is realising that a manipulation found by trial for $n = 3$ gives no guarantee it generalises to every $n \geq 3$ at once, whereas a rank bound derived from the addition formula applies to every n simultaneously, in one argument.

4. The trap is trusting a numerical near-zero as if it were exact. A determinant of cosines of consecutive integers can be extremely small without an algebraic reason, purely from near-cancellation of transcendental quantities – the computation is a reason to suspect $d_n = 0$, never a substitute for the rank argument that proves it exactly.

5. Recognise this shape faster by treating any matrix whose (j,k) entry is $\varphi(\text{row-expression} + \text{column-expression})$, for φ a trig function, exponential, or anything else with a known addition formula, as a matrix whose rank is bounded before a single entry is computed – read the rank off the number of terms in φ's addition formula (2 for cos or sin; 1 for $e^{a+b} = e^a e^b$).

6. Reusable rule: for an $n \times n$ matrix with (j,k) entry $\varphi(\text{row}_j + \text{col}_k)$, expand φ via its addition formula to write the matrix as a sum of (at most) that-many rank-1 outer products; if this bound is less than n, the determinant is 0 automatically, with no need for row reduction, cofactor expansion, or numerical computation.

▷ **Cue:** a matrix entry is a trig or exponential function evaluated at (row-index expression) + (column-index expression) $\longrightarrow$ **Move:** expand via the function's addition formula to write the matrix as a sum of rank-1 outer products, and compare the resulting rank bound to the matrix's size

▷ **Cue:** an $n \times n$ matrix has rank $r < n$ $\longrightarrow$ **Move:** conclude its determinant is exactly 0, via the alternating-multilinear definition of the determinant, without further computation

2009 A4

2009 A4. Let S be a set of rational numbers such that

(a) $0 \in S$;

(b) If $x \in S$ then $x+1 \in S$ and $x-1 \in S$; and

(c) If $x \in S$ and $x \notin \{0,1\}$, then $\frac{1}{x(x-1)} \in S$.

Must S contain all rational numbers?

Fields: number theory (quadratic residues); invariants of closure operations **Position:** A4 — mid-paper. The answer is a single counterexample set; the difficulty is deciding to look for an invariant instead of trying to generate $\mathbb{Q}$. **Answer:** No. The set $S = \mathbb{Q} \setminus \{n + \frac{2}{5} : n \in \mathbb{Z}\}$ satisfies (a), (b), (c).

Layer 1 THE WHAT — surface analysis

The natural attempt. You start generating. From 0 and (b) you get every integer. Then (c) applied to an integer $n \geq 2$ gives $\frac{1}{n(n-1)}$: so $\frac{1}{2}, \frac{1}{6}, \frac{1}{12}, \ldots$ are in S, and by (b) so is every $k + \frac{1}{n(n-1)}$. Apply (c) to $x = \frac{1}{2}$: $x(x-1) = -\frac{1}{4}$, so $-4 \in S$ (already known). Apply it to $x = \frac{3}{2}$: $x(x-1) = \frac{3}{4}$, so $\frac{4}{3} \in S$, hence $\frac{1}{3} \in S$. To $x = -\frac{1}{2}$: $x(x-1) = \frac{3}{4}$ again. To $x = \frac{1}{3}$: $x(x-1) = -\frac{2}{9}$, so $-\frac{9}{2}$. To $x = \frac{4}{3}$: $\frac{4}{9}$, so $\frac{9}{4}$ and hence $\frac{1}{4}$. Thirds, quarters, sixths arrive quickly; the instinct says "yes, everything comes out eventually," and you begin hunting for a general recipe that produces $\frac{m}{n}$ from things with smaller denominator.

Then fifths. Try to produce $\frac{2}{5}$. You need x with $\frac{1}{x(x-1)} \equiv \frac{2}{5} \pmod 1$. Try $x = \frac{p}{q}$ with $p(p-q) = \pm 5$: $x = \frac{5}{6}$ gives $\frac{36}{-5}$, which is $-8 + \frac{4}{5}$; $x = \frac{5}{4}$ gives $\frac{16}{5} = 3 + \frac{1}{5}$; $x = \frac{1}{6}$ gives $-\frac{36}{5} \equiv \frac{4}{5}$; $x = -\frac{1}{4}$ gives $\frac{16}{5} \equiv \frac{1}{5}$. You get $\frac{1}{5}$ and $\frac{4}{5}$ every time, never $\frac{2}{5}$ or $\frac{3}{5}$. This is the wall, and it is also the answer, if you stop to read it.

The model chain. *Step 1: what can (c) produce?* Write $x = \frac{p}{q}$ in lowest terms with $q \geq 1$, $x \notin \{0,1\}$, so $p \neq 0$ and $p \neq q$. Then

$$\frac{1}{x(x-1)} = \frac{q^2}{p(p-q)}.$$

Since $\gcd(p,q) = 1$ we also have $\gcd(p-q,q) = \gcd(p,q) = 1$, so q^2 is coprime to $p(p-q)$ and this fraction is already in lowest terms. Its reduced denominator is $|p(p-q)|$ and its reduced numerator is $\pm q^2$ (sign of $p(p-q)$).

Step 2: the invariant of (b). If $r = \frac{m}{n}$ in lowest terms with $n \geq 1$, then $r \pm 1 = \frac{m \pm n}{n}$ is still in lowest terms with the same denominator and numerator congruent to $m \pmod n$. So the pair $(n,\ m \bmod n)$ — equivalently, the class of r in $\mathbb{Q}/\mathbb{Z}$ — is unchanged by (b).

Step 3: a forbidden class. By Step 1, every output of (c) has the form $\frac{m}{n}$ in lowest terms with $m \equiv \pm q^2 \pmod n$: that is, m or $-m$ is a square modulo n. Modulo 5 the squares are $0, 1, 4$. The class $\frac{2}{5} + \mathbb{Z}$ consists of $\frac{m}{5}$ with $m \equiv 2 \pmod 5$, and neither 2 nor $-2 \equiv 3$ is a square mod 5. So (c) never outputs a number in $\frac{2}{5} + \mathbb{Z}$, whatever its input.

Step 4: the counterexample. Let $S = \mathbb{Q} \setminus (\frac{2}{5} + \mathbb{Z})$. (a): $0 \notin \frac{2}{5} + \mathbb{Z}$. (b): $\frac{2}{5} + \mathbb{Z}$ is closed under ± 1, so its complement is too. (c): by Step 3, $\frac{1}{x(x-1)} \notin \frac{2}{5} + \mathbb{Z}$ for every rational $x \notin \{0,1\}$. So S satisfies all three conditions and omits $\frac{2}{5}$. The answer is no.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains compute $\frac{1}{x(x-1)}$ for several rational x. They split at (†), the moment the fifths keep coming out as $\frac{1}{5}$ and $\frac{4}{5}$. The natural attempt treats this as bad luck with the choice of x and keeps trying inputs. The model treats it as data: something about the operation is refusing $\frac{2}{5}$, so find that something and prove it for all x at once.

> *The question is "must S be everything?", and my rules only say what S must contain. If the answer were yes I'd need a generation recipe for every rational; if no, I need one set closed under the rules that misses something. Let me work out which of those is cheaper before I grind.*

> *Every time I aim at $\frac{2}{5}$ I get $\frac{1}{5}$ or $\frac{4}{5}$. The output is $\frac{q^2}{p(p-q)}$ — its numerator is a square. Mod 5 the squares are 1 and 4. That's why. Shifting by integers doesn't change the numerator mod the denominator, so $\frac{2}{5}$ is never reachable.*

Layer 3 THE WHY

Why is looking for an invariant the natural move, not a flash of insight? Look at the logical shape of the question. Conditions (a)–(c) are closure conditions: they say "S contains the smallest set generated from 0 by these operations." Asking whether S must be all of $\mathbb{Q}$ is asking whether that generated set is all of $\mathbb{Q}$. There are exactly two ways to answer such a question. Either you exhibit a generation route to every rational, or you find a property that 0 has and that both operations preserve, but that some rational lacks. The second is an invariant, by definition. So an invariant is not a trick; it is one of only two possible shapes of the answer, and the phrase "Must S contain…?" is the standing invitation to test the "no" shape.

Why does the invariant live in $\mathbb{Q}/\mathbb{Z}$? Because operation (b) is translation by 1. Anything invariant under (b) must depend only on x modulo 1 — this is forced, not chosen. So the only question is which classes in $\mathbb{Q}/\mathbb{Z}$ operation (c) can hit.

Why look at the numerator mod the denominator? Because the formula $\frac{q^2}{p(p-q)}$ is already in lowest terms, and that gives two clean facts: the output's denominator is a product $p(p-q)$ and its numerator is a square up to sign. A class $\frac{m}{n}+\mathbb{Z}$ is determined by m mod n. The only arithmetic constraint on $\pm q^2$ modulo n is that it is a square. So the classes (c) can reach are among those where $\pm m$ is a quadratic residue mod n. Then the search for a forbidden class is a finite table: mod 3, squares are $0, 1$, and ± 1, ± 2 always include 1; mod 4, squares are $0, 1$, and ± 1, ± 3 include 1; mod 5, squares are $0, 1, 4$, and $\pm 2 = \{2, 3\}$ misses them. Five is the first denominator that works, which is why the answer involves fifths.

Kedlaya's solution states the set $\mathbb{Q} \setminus \{n + \frac{2}{5}\}$ and then checks it by a case analysis on $p \mid 5$ and $(p-q) \mid 5$. Here is where the $\frac{2}{5}$ comes from: it is the smallest class whose numerator is, up to sign, a quadratic non-residue of its denominator. The Remark after his solution states exactly this generalisation. The quadratic-residue argument in Step 3 replaces the eight-case check with one line, and it is the version to remember.

One more observation worth keeping: the natural attempt was not wasted. The first three cases of fifths you computed already contained the complete answer. The failure was treating a pattern in outputs as noise instead of as a hypothesis.

Caveat. Verified independently: a breadth-first closure from 0 under (b) and (c), restricted to rationals of height at most 400 (over 5000 elements), produces fifths only in the classes $\frac{1}{5}, \frac{4}{5}$; and $\frac{1}{x(x-1)}$ was checked to avoid $\frac{2}{5}+\mathbb{Z}$ and $\frac{3}{5}+\mathbb{Z}$ for every $x = \frac{p}{q}$ with $|p| < 300$, $q < 300$. Note that $\mathbb{Q} \setminus (\{\frac{2}{5}, \frac{3}{5}\} + \mathbb{Z})$ is also a valid counterexample.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the question form "Must S contain all…?" together with conditions that only say what S contains. Write down at once: *either a recipe for every rational, or a property preserved by (b) and (c) that 0 has.* Then write the output of (c) in lowest terms: $\frac{q^2}{p(p-q)}$.

2. An invariant. Translation-invariance under (b) forces it to be a property of x mod 1, and the square numerator of (c)'s output points to quadratic residues.

3. Hesitation comes after thirds and quarters are generated easily and the "yes" feeling is strong. It is reinforced away by computing a few targeted outputs with small denominators and reading them: when a target class never appears, stop and explain why.

4. The trap is momentum: the early successes (integers, halves, thirds, quarters) make "yes" feel inevitable, and you spend the exam building a generation recipe that cannot exist. A second trap is forgetting to check that the fraction $\frac{q^2}{p(p-q)}$ is in lowest terms; without $\gcd(q, p-q) = 1$ the residue argument is unsupported.

5. Slogan: *a closure question with a "must" is an invariant question until proved otherwise; compute a few outputs aimed at one target, and if the target keeps being missed, the miss is the invariant.*

6. Decision rule: **when asked whether a set generated by rules must be everything, first identify what each rule preserves (translation by 1 preserves $x \bmod 1$; a formula with a square numerator preserves quadratic-residue structure), then search the smallest cases for a class no rule can reach.**

▷ **Cue:** "Must S contain all…?" where S is only required to be closed under some operations $\longrightarrow$ **Move:** look for an invariant: a property of 0 (the seed) preserved by every operation, and a target lacking it

▷ **Cue:** an operation $x \mapsto x \pm 1$ among the closure rules $\longrightarrow$ **Move:** any invariant must depend only on x mod 1, i.e. on the class in $\mathbb{Q}/\mathbb{Z}$, i.e. on (denominator, numerator mod denominator)

▷ **Cue:** a rational expression whose value, in lowest terms, has a perfect square in the numerator $\longrightarrow$ **Move:** its numerator mod its denominator is $\pm$ a quadratic residue; choose a modulus where some class is a non-residue up to sign

▷ **Cue:** repeated attempts to hit a target always land in the same few residue classes $\longrightarrow$ **Move:** stop varying the input; the pattern is the invariant, and the proof is to show it for all inputs at once

2009 A5

2009 A5. Is there a finite abelian group G such that the product of the orders of all its elements is 2^{2009}?

Fields: group theory (finite abelian groups); congruences **Position:** A5 — the harder half of the paper, but short once the right modulus is chosen. The work is choosing to reduce mod 4 rather than computing the exponent exactly. **Answer:** No such group exists.

Layer 1 THE WHAT — surface analysis

The natural attempt. You reduce to 2-groups first: if $|G|$ had an odd prime factor p, Cauchy's theorem gives an element of order p, and the product of orders would not be a power of 2. So G is an abelian 2-group, and by the structure theorem

$$G \cong \mathbb{Z}/2^{a_1} \times \cdots \times \mathbb{Z}/2^{a_e}, \qquad a_1 \geq \cdots \geq a_e \geq 1.$$

Write the product as $2^{k(G)}$, where $k(G) = \sum_{g \in G} \log_2 \operatorname{ord}(g)$. The natural plan is to compute $k(G)$ exactly as a function of $(a_1, \ldots, a_e)$ and then show it never equals 2009. You compute small cases. $\mathbb{Z}/2$: $k = 1$. $\mathbb{Z}/4$: orders $1, 4, 2, 4$, so $k = 2 + 1 + 2 = 5$. $(\mathbb{Z}/2)^2$: three elements of order 2, $k = 3$. $\mathbb{Z}/8$: $k = 1 + 2 \cdot 2 + 3 \cdot 4 = 17$. $\mathbb{Z}/2 \times \mathbb{Z}/4$: four elements of order 4, three of order 2, so $k = 11$.

The general formula needs the number of elements of order dividing 2^m, which is 2^{s_m} with $s_m = \sum_i \min(a_i, m)$, giving

$$k(G) = \sum_{m \geq 1} m \left(2^{s_m} - 2^{s_{m-1}}\right).$$

Now what? This is an exact expression in the a_i and it can be made to equal many things. Showing that 2009 is never hit by solving this equation directly means a search over partitions. You can bound it — every non-identity element contributes at least 1, so $|G| \le 2010$, so $|G| \le 2^{10}$ and there are finitely many cases (the partitions of $0, 1, \ldots, 10$) — but that is a table of 138 groups, not an exam proof.

The model chain. *Step 1: reduce to 2-groups.* As above: an odd prime $p \mid |G|$ gives, by Cauchy's theorem, an element of order p, and the product would be divisible by p. So G is an abelian 2-group. If G is trivial the product is $1 = 2^0$. Otherwise G has rank $e \ge 1$ in the decomposition above.

Step 2: group elements by the cyclic subgroup they generate. Say $g \sim g'$ if $\langle g \rangle = \langle g' \rangle$. An element of order 2^h generates a cyclic group $C \cong \mathbb{Z}/2^h$, and the elements generating C are its generators, of which there are $\varphi(2^h) = 2^{h-1}$. Each has order 2^h. So the class contributes

$$h \cdot 2^{h-1}$$

to $k(G)$. For $h = 1$ this is 1. For $h \ge 2$ it is divisible by 4: $h = 2$ gives 4, and for $h \ge 3$ the factor 2^{h-1} is already divisible by 4.

Step 3: reduce mod 4.

$$k(G) \equiv \#\{g \in G : \operatorname{ord}(g) = 2\} \pmod 4.$$

The elements with $2g = 0$ are the tuples whose i-th coordinate is 0 or $2^{a_i - 1}$, so there are 2^e of them, and $2^e - 1$ have order exactly 2. Since $2009 \equiv 1 \pmod 4$,

$$2^e - 1 \equiv 1 \pmod 4 \implies 2^e \equiv 2 \pmod 4 \implies e = 1.$$

Step 4: the cyclic case. $G \cong \mathbb{Z}/2^n$ has exactly one cyclic subgroup of each order 2^h, $1 \le h \le n$, so by Step 2

$$k(\mathbb{Z}/2^n) = \sum_{h=1}^{n} h\, 2^{h-1} = (n-1)2^n + 1,$$

(induction: adding $n \cdot 2^{n-1}$ to $(n-2)2^{n-1} + 1$ gives $(2n-2)2^{n-1} + 1$). This is strictly increasing in n, with $k = 1793$ at $n = 8$ and $k = 4097$ at $n = 9$. It skips 2009. No group works.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains reduce to abelian 2-groups and both know the orders of elements are powers of 2. They split at (†): "I need to show $k(G) \ne 2009$." The natural attempt tries to compute $k(G)$ exactly and hit the target. The model asks for the weakest piece of information about $k(G)$ that could rule out 2009 — a congruence — and finds that mod 4 it depends on one number, the count of involutions.

> *I don't need $k(G)$. I need $k(G) \ne 2009$. The cheapest way to show two numbers differ is to find a modulus where they differ. The year is* 1 mod 4 *and* 1 mod 8 *— small moduli first.*

> *Elements that generate the same cyclic subgroup have the same order, and there are 2^{h-1} of them. So each class contributes $h \cdot 2^{h-1}$, which is* 0 mod 4 *unless $h = 1$. Mod* 4*, only the elements of order* 2 *count, and there are $2^e - 1$ of those. That forces the group to be cyclic.*

Layer 3 THE WHY

Why reach for a congruence? Look at what is being asked: not "compute the product" but "can it equal this specific large number?". Whenever a Putnam problem uses the year as a target, the year was chosen because it is ordinary — nothing about 2009 is special except its residues.

So the problem designer's certificate for "no" is almost always a congruence or a size gap, and the solver's first move should be to ask which residue of 2009 the structure controls. Here $2009 = 7^2 \cdot 41$ and $2009 \equiv 1 \pmod 8$; the relevant fact turns out to be $2009 \equiv 1 \pmod 4$.

Why does the grouping by cyclic subgroups come to mind? Because it is the one partition of G on which the quantity being summed, $\log_2 \operatorname{ord}(g)$, is constant. The order of g is the size of $\langle g \rangle$; so the natural classes are "elements with the same $\langle g \rangle$". Counting generators of a cyclic group is Euler's φ, and $\varphi(2^h) = 2^{h-1}$. Once that is on the page the multiples of 4 appear without any search. The summand becomes $h \cdot 2^{h-1}$ — a power of 2 times h, which is highly divisible except at the bottom level. Congruences modulo small powers of 2 are then controlled entirely by the bottom level: the elements of order 2.

Why is the number of involutions $2^e - 1$? The elements killed by 2 form a subgroup that is a vector space over $\mathbb{F}_2$, one dimension per cyclic factor. That is a count every group theorist knows, and it is exactly why "rank" is the invariant that falls out.

Kedlaya's main solution writes $k(G) = \sum_i i(2^{s_i} - 2^{s_{i-1}})$ and then states that $k(G) + 1$ is divisible by 2^{s_1}. Here is where that comes from: $\sum_i i(2^{s_i} - 2^{s_{i-1}}) = n\,2^{s_n} - \sum_{i=0}^{n-1} 2^{s_i}$ by summation by parts, where n is the largest a_i; every term except $2^{s_0} = 1$ is divisible by 2^{s_1}, so $k(G) \equiv -1 \pmod{2^{s_1}}$. Since $s_1 = e$, the congruence says $k \equiv -1 \pmod{2^e}$, and $2009 \equiv -1$ fails modulo 4 (as $2010 \equiv 2$), so $e \leq 1$. The Remark after his solution is the grouping argument above. The two are the same fact; the grouping version is the one a solver generates, because it starts from "which elements have the same order?" rather than from a formula.

Caveat. Verified independently: $k(G)$ was computed from the formula $\sum_m m(2^{s_m} - 2^{s_{m-1}})$ for every abelian 2-group of order at most 2^{11} (all partitions of $0, \ldots, 11$), cross-checked by direct enumeration on $\mathbb{Z}/2 \times \mathbb{Z}/4 \times \mathbb{Z}/4$ ($k = 55$). The congruence $k(G) \equiv 2^e - 1 \pmod 4$ held in every case and no group gave 2009. Since $|G| \leq 2010$ is forced, this search is exhaustive.

Layer $+\alpha$ BACKTESTING

1. The cue is the target 2^{2009}: a specific, ordinary exponent. Write down: *I only need $k(G) \neq 2009$; find a modulus.* Also write $2009 \equiv 1 \pmod 4$ and $\pmod 8$.

2. Partition the group into classes on which the summand is constant (elements generating the same cyclic subgroup), count each class with Euler's φ, and reduce modulo a small power of 2.

3. Hesitation occurs once the exact formula for $k(G)$ is written: it looks like progress and invites solving $k(G) = 2009$. Reinforce away with the rule that an exact formula is only useful if it can be inverted; if not, reduce it modulo something.

4. The trap is exhaustive computation: the bound $|G| \leq 2010$ makes a finite search legal but long. A second trap is forgetting the reduction to 2-groups, or forgetting the trivial group ($k = 0$). A third is miscounting involutions as 2^e instead of $2^e - 1$.

5. Slogan: *a year as a target means a congruence or a gap; group the sum so each class contributes a highly divisible amount, and the congruence is decided by the classes that do not.*

6. Decision rule: **to show a structured quantity cannot equal a given number, do not compute it; find the smallest modulus at which the structure pins it down, and check the target's residue.** Companion rule: to sum a function of $\operatorname{ord}(g)$ over a group, group elements by $\langle g \rangle$ and count with φ.

▷ **Cue:** a competition asks whether a quantity can equal a number built from the year, such as 2^{2009} ⟶ **Move:** look for a congruence (or a gap between consecutive attainable values) rather than solving exactly; write down the year's residues mod 3, 4, 8

▷ **Cue:** a sum over group elements of a function of the element's order ⟶ **Move:** group elements by the cyclic subgroup they generate; a class of order n has $\varphi(n)$ elements

▷ **Cue:** a finite abelian p-group and a question mod p or mod p^2 ⟶ **Move:** count elements killed by p: they form $(\mathbb{Z}/p)^e$ with e the number of cyclic factors

▷ **Cue:** a product of orders (or any multiplicative statistic) required to be a power of a prime p
⟶ **Move:** rule out other primes with Cauchy's theorem, reducing to a p-group

2009 A6

2009 A6. Let $f : [0,1]^2 \to \mathbb{R}$ be a continuous function on the closed unit square such that $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ exist and are continuous on the interior $(0,1)^2$. Let $a = \int_0^1 f(0,y)\,dy$, $b = \int_0^1 f(1,y)\,dy$, $c = \int_0^1 f(x,0)\,dx$, $d = \int_0^1 f(x,1)\,dx$. Prove or disprove: There must be a point (x_0,y_0) in $(0,1)^2$ such that

$$\frac{\partial f}{\partial x}(x_0,y_0) = b-a \quad \text{and} \quad \frac{\partial f}{\partial y}(x_0,y_0) = d-c.$$

Fields: multivariable calculus; mean value theorem; counterexample construction **Position:** A6 — the last problem of the morning, and hard mainly because it invites a proof. Once you suspect it is false, a counterexample falls out of a two-line design. **Answer:** Disprove it: $f(x,y) = (1+y)\sin(2\pi x)$ has $b-a = d-c = 0$, but its gradient never vanishes in $(0,1)^2$ (in fact $|\nabla f| \geq 1$ everywhere).

Layer 1 THE WHAT — surface analysis

The natural attempt. This looks like the Mean Value Theorem in two variables, so you try to prove it. Write f_x, f_y for the partial derivatives.

First component. Fix $y \in (0,1)$. The function $x \mapsto f(x,y)$ is continuous on $[0,1]$ and differentiable on $(0,1)$, so the one-variable MVT gives $\xi(y) \in (0,1)$ with

$$f_x(\xi(y),y) = f(1,y) - f(0,y) =: h(y).$$

Now h is continuous on $[0,1]$ with $\int_0^1 h = b-a$. If h is constant, every y has $h(y) = b-a$. If not, h takes values strictly above and strictly below its average at two points of $[0,1]$, which may be moved slightly into $(0,1)$ by continuity, and the intermediate value theorem between them gives $y^* \in (0,1)$ with $h(y^*) = b-a$. So the point $(\xi(y^*), y^*)$ satisfies the first equation.

Second component. The same argument on vertical lines gives a point satisfying the second equation.

Gluing. Now you need one point that does both. You try to argue that the set $\{f_x = b-a\}$ is a curve crossing the square from bottom to top and $\{f_y = d-c\}$ a curve crossing from left to right, so they must meet. But nothing forces either set to cross the square: the first argument produced one point on each horizontal line only for the special height y^*, not a curve. You try a Rolle-style argument instead: subtract a linear function so the target is $(0,0)$ and look for an interior extremum. But the hypotheses are only about *averages* along the edges, not about boundary values, so the maximum can sit on the boundary. Here is the wall. Two separate existence proofs, and no mechanism to make the two points the same point.

The model chain. *Step 1: reduce the target to* $(0,0)$. For $L(x,y) = \alpha x + \beta y$ the four edge integrals are $a_L = \beta/2$, $b_L = \alpha + \beta/2$, $c_L = \alpha/2$, $d_L = \alpha/2 + \beta$, so $b_L - a_L = \alpha$ and $d_L - c_L = \beta$, which are exactly L_x and L_y. Hence $g = f - (b-a)x - (d-c)y$ satisfies the hypotheses with its own $b-a = d-c = 0$, and f has a point with $\nabla f = (b-a, d-c)$ iff g has an interior critical point. The claim is therefore equivalent to: *if the opposite-edge averages agree, f has a critical point in* $(0,1)^2$.

Step 2: suspect it is false. By Fubini and Step 1's computation, $(b-a, d-c)$ is the average of ∇f over the square (when that average exists). In one variable, "the average is attained" follows from the intermediate value theorem. In the plane the average of a vector field lies in the convex hull of its values, and nothing forces a connected set in $\mathbb{R}^2$ to contain the points of its convex hull. So look for f whose gradient wraps around $(0,0)$ without touching it.

Step 3: design by separation of variables. Try $f(x,y) = p(y)\,g(x)$. Then

$$f_x = p(y)\,g'(x), \qquad f_y = p'(y)\,g(x).$$

Choose $p(y) = 1+y$, which is never zero and has $p' \equiv 1$. Then $\nabla f = 0$ requires $g(x) = 0$ and $g'(x) = 0$ together. So it is enough that g has only simple zeros. The edge conditions: $a = g(0)\int_0^1 p = \frac{3}{2}g(0)$ and $b = \frac{3}{2}g(1)$, so $a = b$ iff $g(0) = g(1)$; $c = p(0)\int_0^1 g$ and $d = p(1)\int_0^1 g = 2\int_0^1 g$, so $c = d$ iff $\int_0^1 g = 0$.

Step 4: the example. $g(x) = \sin(2\pi x)$ has $g(0) = g(1) = 0$, $\int_0^1 g = 0$, and its only zero in $(0,1)$ is $x = \frac{1}{2}$, where $g'(\frac{1}{2}) = 2\pi\cos\pi = -2\pi \neq 0$. So

$$f(x,y) = (1+y)\sin(2\pi x)$$

is smooth, has $a = b = c = d = 0$, and

$$|\nabla f|^2 = 4\pi^2(1+y)^2\cos^2(2\pi x) + \sin^2(2\pi x) \;\geq\; \cos^2(2\pi x) + \sin^2(2\pi x) = 1.$$

There is no point with $\nabla f = (b-a, d-c) = (0,0)$. The statement is false.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains get as far as (†): "each equation separately has a solution." At (†) the natural attempt keeps trying to glue the two points, because the problem looks like a theorem it half-remembers. The model stops and asks what the one-variable MVT actually used, sees that it used the intermediate value theorem, and asks whether the intermediate value theorem exists for maps into $\mathbb{R}^2$. It does not, and the model switches from proving to building.

> *Both halves are true separately and I cannot make them hold at the same point. "Prove or disprove" on an A6 with a claim that looks like a textbook theorem — if it were a theorem, it would be in the textbook. Let me assume it is false and see what a counterexample must look like.*

> *After subtracting a linear function, the target is $(0,0)$: I need a function with equal opposite-edge averages and no critical point. The gradient is a map into the plane, and the plane minus a point is still connected, so the gradient can go round the origin without ever hitting it.*

> *I want $f_x = 0$ and $f_y = 0$ never to happen together. If $f = (1+y)g(x)$ then $f_y = g(x)$ and $f_x = (1+y)g'(x)$, so I only need g and g' never to vanish together, plus $g(0) = g(1)$ and $\int g = 0$. A sine wave over one period.*

Layer 3 THE WHY

Why is the MVT true in one variable at all? Strip it to the axioms. $\frac{f(1)-f(0)}{1} = \int_0^1 f'$ is an average of the values of f'; an average lies between the minimum and maximum; and the intermediate value theorem says a continuous function on an interval hits every value between two it takes. The last step is the whole content: it is the statement that the image of a connected set in $\mathbb{R}$ is an interval, and in $\mathbb{R}$ connected sets are exactly the convex ones.

In the plane the first two steps survive and the third does not. Fubini still gives $(b-a, d-c) = \iint \nabla f$, the average of the gradient, and an average still lies in the convex hull of the values. The image of the square under ∇f is still connected. But in $\mathbb{R}^2$ connected no longer means convex: an annulus is connected and misses its own centre. So the claim fails exactly when the gradient's image is annulus-like around its own average. That is not a trick; it is the precise point at which one-dimensional topology stops being true, and the problem is built to test whether you know which axiom the MVT uses.

Why subtract a linear function? Because the target $(b-a, d-c)$ is a constant, and constants in a gradient come from linear functions; Step 1 checks that a linear function's edge data reproduces its own gradient, so subtracting it changes both sides by the same amount. After that the problem is the two-dimensional Rolle's theorem with "equal boundary values" weakened to "equal boundary averages". Rolle in two dimensions (constant on the boundary forces an interior extremum) is true; the averaged version is the thing being tested, and averages cannot pin down where the maximum is.

Why a product $p(y)g(x)$? Because it splits the gradient into two factors each depending on one variable, so "both partials vanish" becomes a condition you can read off. Choosing $p' \equiv 1$ makes f_y vanish only where g does, and $p > 0$ makes f_x vanish only where g' does. The condition "no common zero of g and g'" is the definition of a function with simple zeros — the most generic behaviour a function can have.

Kedlaya's solution uses $f(x,y) = 3(1+y)(2x-1)^2 - y$. Here is where that comes from. It is the same design with $g(x) = (2x-1)^2 - \frac{1}{3}$ (zero integral, $g(0) = g(1)$, simple zeros), scaled by 3 and shifted by a constant, which does not change the gradient: $3(1+y)(2x-1)^2 - y = 3(1+y)g(x) + 1$. The factor $1+y$ rather than y is deliberate. Merryfield's $y\sin(2\pi x)$ in the Remark also works, but its gradient $(2\pi y\cos 2\pi x,\ \sin 2\pi x)$ tends to $(0,0)$ as $(x,y) \to (\frac{1}{2}, 0)$ — a boundary point, so the example is legal, but only because the problem asks for an interior point. With $1+y$ the gradient stays at distance ≥ 1 from the target on the closed square, and no reader can object that the counterexample is fragile.

Caveat. Verified independently: the edge integrals of $(1+y)\sin(2\pi x)$ were computed symbolically (all zero) and $|\nabla f|$ was evaluated on a 2001×2001 grid of the closed square (minimum 1.000, matching the bound above). Kedlaya's example and $y\big((2x-1)^2 - \frac{1}{3}\big)$ were checked the same way; the latter, like $y\sin 2\pi x$, has gradient approaching 0 only at a boundary point.

Layer $+\alpha$ BACKTESTING

1. The cue is "Prove or disprove" attached to a statement that generalises a one-variable theorem (the MVT) to the plane. Write down: *which step of the one-variable proof is dimension-specific?* and *subtract $(b-a)x + (d-c)y$ to make the target zero.*

2. Reduce to the target $(0,0)$, recognise an averaged Rolle's theorem, and construct a counterexample by separation of variables $f = p(y)g(x)$ so that the two partials vanish on disjoint sets.

3. Hesitation occurs after proving each component separately: that feels like 90% of a proof. Reinforce away with the rule that two existence statements never combine into a simultaneous one without a structural reason, and that "prove or disprove" on a hard slot is a genuine coin with a disprove side.

4. The trap is the phrase "Mean Value Theorem" in your head: it makes the claim feel like a known theorem. A second trap is choosing a counterexample whose gradient approaches the target at the boundary without noticing that you are relying on the word "interior". A third is forgetting that the partials need only be continuous on the open square, so Fubini on ∇f needs care in a proof — irrelevant for a smooth counterexample.

5. Slogan: *one-variable theorems that rest on the intermediate value theorem fail for $\mathbb{R}^2$-valued maps, because the plane minus a point is connected; the counterexample makes the values wind round the target.*

6. Decision rule: **before proving a multivariable version of a one-variable theorem, locate the step that uses the order or connectedness of $\mathbb{R}$; if the values now live in $\mathbb{R}^2$, try to break it by making the values go round the target.** Companion rule: to make two partial derivatives never vanish together, use a product $p(y)g(x)$ with $p > 0$, $p' \neq 0$ and g having only simple zeros.

- ▷ **Cue:** "prove or disprove" on a statement that looks like a theorem you almost remember, especially a one-variable theorem transplanted to several variables $\longrightarrow$ **Move:** find the step of the known proof that is dimension-specific; if there is one, build a counterexample before attempting a proof
- ▷ **Cue:** a condition "some point has $\nabla f = (u, v)$" with (u, v) constant $\longrightarrow$ **Move:** subtract $ux + vy$ and look for a critical point instead
- ▷ **Cue:** average of a vector-valued function claimed to be attained $\longrightarrow$ **Move:** averages lie in the convex hull; the intermediate value theorem only gives the full convex hull in $\mathbb{R}$, so in $\mathbb{R}^2$ look for values that encircle the average
- ▷ **Cue:** need a function with no critical point but with prescribed integral data on the edges of a rectangle $\longrightarrow$ **Move:** try $f = p(y)g(x)$; translate each edge condition into a condition on g or p and require simple zeros
- ▷ **Cue:** a counterexample whose bad behaviour occurs only in the limit at the boundary $\longrightarrow$ **Move:** add a constant inside a factor (e.g. $y \to 1 + y$) to push it away, so the example does not lean on the word "interior"

Putnam 2009 — Paper B

2009 B1

2009 B1. Show that every positive rational number can be written as a quotient of products of factorials of (not necessarily distinct) primes. For example,

$$\frac{10}{9} = \frac{2! \cdot 5!}{3! \cdot 3! \cdot 3!}.$$

Fields: number theory (combinatorics of factorisations) **Position:** B1 – opener, disarmingly concrete for a Putnam B1 **Answer:** Yes: strong induction on the largest prime dividing the numerator or denominator (in lowest terms), using $p! = p \cdot (p-1)!$ to peel primes off one at a time.

Layer 1 THE WHAT — surface analysis

The natural attempt. The worked example is a gift and a trap at once: $10/9 = 2! \cdot 5!/(3! \cdot 3! \cdot 3!)$ shows the target is achievable but says nothing about how to reach it in general, so the honest first move is to look for the single reusable fact that makes primes and factorials talk to each other. That fact arrives almost immediately: $p!/(p-1)! = p$ for any prime p. This is exactly the right instinct – it says every prime is itself a quotient of two factorials – and since every positive rational is, by the Fundamental Theorem of Arithmetic, a product of prime powers $\prod p_i^{e_i}$ (positive and negative e_i allowed), the natural attempt reaches for $p_i = p_i!/(p_i-1)!$ and tries to assemble the whole rational as a product of these ratios. It gets as far as writing

$$\frac{a}{b} = \prod_i \left(\frac{p_i!}{(p_i-1)!} \right)^{e_i}$$

and then stalls, because the problem does not allow $(p_i-1)!$ in the answer – only factorials of primes are permitted, and $p_i - 1$ is prime only in the single case $p_i = 3$. For $p_i = 5$, $(p_i-1)! = 4!$, and 4 is not prime; the natural attempt has produced a formula that is arithmetically correct but built from a forbidden material. At this point the honest student either abandons $p!/(p-1)! = p$ as the wrong lever entirely, or starts hunting for some other identity expressing p purely in terms of factorials of primes – a search that, tried directly, does not terminate, because no single equation of that form exists for every prime simultaneously.

The model chain. Write the target rational in lowest terms as a/b, and induct on p, the largest prime dividing a or b (taking $p = 1$, i.e. no primes, in the base case $a = b = 1$). Base case: $1/1 = 2!/2!$, a quotient of prime factorials. Inductive step: suppose the claim holds for every positive rational whose largest prime factor (in lowest terms) is strictly less than p. Write $a/b = p^k a'/b'$ where $k \in \mathbb{Z}$ is the exact power of p dividing a/b and a', b' are positive integers whose largest prime factor is strictly less than p. Then

$$\frac{a}{b} = p^k \cdot \frac{a'}{b'} = \left(\frac{p!}{(p-1)!} \right)^k \cdot \frac{a'}{b'} = (p!)^k \cdot \frac{a'}{(p-1)!^k \, b'}.$$

The fraction $a'/((p-1)!^k b')$ is a positive rational whose numerator and denominator, after cancellation, involve only primes strictly less than p (since a', b', and $(p-1)!$ all do). By the induction hypothesis it is a quotient of products of prime factorials; multiplying by $(p!)^k$, itself such a quotient, gives a/b as one too. This completes the induction.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths arrive at the identity $p!/(p-1)! = p$ and both use it. They split on what to do about the leftover factor $(p-1)!$ that this identity introduces. The natural attempt treats $(p-1)!$ as the final answer for that slot – something that must itself already be a factorial of a prime, checks whether it is, and gives up when it usually isn't. The model treats $(p-1)!$ as unfinished business: not a value to be judged pass/fail against the target form, but a smaller instance of the very same problem, to be handed to the induction hypothesis.

> $(p-1)!$ *does not need to be a factorial of a prime on its own – it only needs to be expressible as a quotient of products of prime factorials, which is exactly what I am trying to prove for every positive rational, and* $(p-1)!$*'s own largest prime factor is smaller than* p*, so I am allowed to assume this for it already.*

> *The quantity to induct on is not "does this integer happen to be a prime factorial" but "what is the largest prime dividing the numerator or denominator" – a well-ordered measurement that strictly decreases every time I peel off the top prime, so the recursion is guaranteed to terminate even though it may take several rounds to reach* $1/1$.

Layer 3 THE WHY

The identity $p! = p \cdot (p-1)!$ is not a trick borrowed for this problem – it is the defining recursive equation of the factorial function, the one fact about $p!$ that is true by construction rather than by computation. So the impulse to isolate p as $p!/(p-1)!$ is not clever; it is simply asking what a factorial *is*, one layer down. The obstruction the natural attempt meets – that $(p-1)!$ is not itself a factorial of a prime – looks like a wall only if "factorial of a prime" is treated as the atomic, irreducible unit that every piece of the construction must resolve to immediately. But the problem never asks for the primes to bottom out in one step; it asks for a quotient of *products* of prime factorials, an arbitrarily deep object. The right unit to induct on, therefore, is not "is this a prime factorial" but "what is the largest prime dividing it," because that quantity is a positive integer (or the degenerate case of none) and strictly decreases under $a/b \mapsto a'/((p-1)!^k b')$ – the standard shape of a well-founded induction. Kedlaya's solution states the substitution $a/b = (p!)^k a'/((p-1)!^k b')$ and the induction on the largest prime factor in a single sentence, without pausing on why $(p-1)!$ is allowed to remain unresolved; the reason is that the class of "positive rationals expressible as quotients of prime factorials" is exactly characterised by an induction on largest-prime-factor, and once that characterisation is seen, $(p-1)!$ is not a loose end at all – it is precisely an instance one step down the very induction being run. The deeper point: the primes generate every rational multiplicatively (unique factorisation), and $p!$ is the cheapest object built entirely from primes $\leq p$ that contains exactly one factor of p to be cancelled against $(p-1)!$ – so $p!/(p-1)! = p$ is not an accident to be exploited once, but the single move that has to be iterated.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the appearance of $p! = p \cdot (p-1)!$ as soon as the goal ("build a prime from factorials") is stated. On seeing it, write down immediately that this is a recursive identity, not a closed formula – it produces p at the cost of introducing $(p-1)!$, and that cost has to be paid off, not ignored.

2. That cue suggests induction: peel the largest prime off a rational using $p!/(p-1)!$, and treat the resulting smaller rational as a fresh instance of the same problem rather than a final answer to be checked against the target form directly.
3. Hesitation occurs exactly when $(p-1)!$ turns out not to be a prime's factorial (e.g. $p=5$ gives 4!). Reinforce it away by asking: does this leftover object have a smaller "largest prime factor" than the one I started with? If yes, the induction hypothesis is licensed to finish the job, and there is nothing further to verify by hand.
4. The trap is assuming a generating set (here, factorials of primes) must combine to produce the target in one non-recursive step, i.e. that every intermediate quantity produced along the way must itself already lie in the allowed final form.
5. Recognisable faster next time by treating any "construct X purely from a restricted generating set" problem as a candidate for induction on a well-ordered measure of size of the object being built (here, the largest prime index), rather than searching for a single closed-form identity that works for every case simultaneously.
6. Reusable rule: when a problem asks for objects to be built entirely from a graded generating family (primes, prime powers, prime factorials, etc.), look for an operation that removes the extremal generator while landing on a strictly smaller instance of the same class, then induct on that extremal generator's size – never demand that the operation terminate in a single application.

▷ **Cue:** a target object must be expressed purely in terms of a restricted family of generators (e.g. factorials of primes only) $\longrightarrow$ **Move:** find an operation that isolates the largest/extremal generator while producing a strictly smaller instance of the same problem, then induct on that extremal generator's size rather than searching for a one-step closed formula

▷ **Cue:** an identity like $p!/(p-1)! = p$ introduces a leftover term that is not itself in the allowed final form $\longrightarrow$ **Move:** do not reject the identity; check whether the leftover term is a smaller instance of the same problem and can be discharged by the induction hypothesis

2009 B2

> **2009 B2.** A game involves jumping to the right on the real number line. If a and b are real numbers and $b > a$, the cost of jumping from a to b is $b^3 - ab^2$. For what real numbers c can one travel from 0 to 1 in a finite number of jumps with total cost exactly c?

Fields: analysis (Riemann sums, intermediate value theorem) **Position:** B2 – second problem, and the paper's first calculus-flavoured trap **Answer:** Exactly the real numbers c with $1/3 < c \leq 1$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The cost formula $b^3 - ab^2 = b^2(b-a)$ invites an optimisation instinct: treat the number of jumps as fixed and small, and minimise. With two jumps $0 \to x \to 1$ the total cost is $x^3 + (1-x)$, since the jump $0 \to x$ costs $x^3 - 0 \cdot x^2 = x^3$ and the jump $x \to 1$ costs $1^3 - x \cdot 1^2 = 1 - x$. Setting $f(x) = x^3 - x + 1$ and differentiating, $f'(x) = 3x^2 - 1 = 0$ at $x = 1/\sqrt{3}$, giving a minimum cost of $f(1/\sqrt{3}) = 1 - \frac{2}{3\sqrt{3}} = 1 - \frac{2\sqrt{3}}{9} \approx 0.615$. This is a completely correct computation, and it is tempting to read it as the answer to the problem: two jumps cannot cost less than about 0.615, so – the natural attempt reasons – perhaps costs below that threshold, or at least well below $1/2$, are simply unreachable, and the true infimum should be found by optimising over three jumps, then four, watching the minimum creep down, and trying to spot where it stabilises. Carried out, this programme does produce decreasing minima (three jumps beats two, four beats three), but it never says why, and it invites the false belief that the interesting content of the problem is a calculus-of-several-variables optimisation with a fixed, small number of jumps – rather than what *happens as the number of jumps is allowed to grow without bound*, which is where the real content lives.

The model chain. For any sequence $0 = x_0 < x_1 < \cdots < x_m = 1$, the total cost is

$$\sum_{i=1}^{m} x_i^2 (x_i - x_{i-1}).$$

Lower bound. For $t \in (x_{i-1}, x_i)$, $t < x_i$ so $t^2 < x_i^2$, hence $x_i^2(x_i - x_{i-1}) > \int_{x_{i-1}}^{x_i} t^2\,dt$ strictly (the interval has positive length). Summing over i, the total cost is strictly greater than $\int_0^1 t^2\,dt = 1/3$ for *every* finite jump sequence. Taking $x_i = i/m$ for $m \to \infty$ gives cost $\frac{(m+1)(2m+1)}{6m^2} \to \frac{1}{3}$, so $1/3$ is approached but never attained: every $c > 1/3$ arbitrarily close to $1/3$ is approximable, though this alone does not yet give exact achievability.

Upper bound. The single jump $0 \to 1$ costs 1; no sequence with more than one jump can cost more than 1, since inserting an intermediate point y strictly decreases cost (shown below), so $c \leq 1$ for every sequence, with equality exactly for the trivial one-jump sequence.

Achieving every $c \in (1/3, 1]$. Fix a partition $0 = x_0 < \cdots < x_m = 1$ with cost c_m, and for $2 \leq i \leq m$ let c_i be the cost of the sub-sequence $0, x_i, x_{i+1}, \ldots, x_m$ (dropping the earlier points). For $0 < y \leq x_{i-1}$, replacing the initial segment with $0, y, x_i, \ldots$ changes the cost from c_i to $c_i - y(x_i^2 - y^2)$, which is strictly less than c_i and tends to c_i as $y \to 0^+$. By continuity in y, every value in $[c_{i-1}, c_i)$ is achieved as y ranges over $(0, x_{i-1}]$, for each $i = 2, \ldots, m$, together with $c_m = 1$ itself (the trivial jump $0 \to 1$). Chaining these intervals over $i = 2, \ldots, m$ sweeps out $[c_1, 1]$ continuously with no gaps. Since c_1 (the cost of the single leftover point x_1, i.e. of the two-jump sequence $0, x_1, 1$) can be pushed arbitrarily close to $1/3$ by choosing m large and $x_1 = 1/m$ small, the union over all m of these swept intervals covers all of $(1/3, 1]$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths write down the cost as $\sum x_i^2(x_i - x_{i-1})$. They split on what kind of object that sum *is*. The natural attempt reads it as a function to be minimised over finitely many free variables $x_1, \ldots, x_{m-1}$ for a fixed, small m. The model reads the same sum as an upper Riemann sum for $\int_0^1 t^2\,dt$, and immediately asks what happens to a Riemann sum as the mesh shrinks – a question about a limit over an unbounded family of m, not an optimisation over one fixed m.

> *This sum has exactly the shape of an upper Riemann sum for t^2 on $[0,1]$ – x_i^2 is the supremum of t^2 on $[x_{i-1}, x_i]$, multiplied by the interval width – so its infimum over all finite partitions is not found by calculus on a fixed number of jumps; it is $\int_0^1 t^2\,dt = 1/3$, approached only as the number of jumps goes to infinity.*

> *Since the problem allows any finite number of jumps, not a fixed number, I am not looking for a minimum that is attained – I am looking for an infimum that a genuinely unbounded family of sequences approaches, and I need to separately show every value strictly above that infimum, up to the maximum 1, is actually hit exactly, not merely approximated.*

Layer 3 THE WHY

The cost function $b^2(b-a)$ was engineered to look like a strange cubic optimisation problem, but its real identity is given away by matching it against the definition of the Riemann integral: a partition's upper sum for $f(t) = t^2$ is exactly $\sum(\sup_{[x_{i-1},x_i]} f) \cdot (x_i - x_{i-1})$, and $\sup_{[x_{i-1},x_i]} t^2 = x_i^2$ because t^2 is increasing on $[0,1]$. So $b^2(b-a)$ is not an arbitrary cost rule to be attacked with Lagrange multipliers; it is, term by term, the definition of an upper Riemann sum, and the entire lower-bound half of the problem is simply the statement that t^2 is Riemann integrable on $[0,1]$ with integral $1/3$ – upper sums always exceed the integral, and can be pushed arbitrarily close to it by refining the partition, which is the definition of integrability itself, not a special

property of this problem. The natural attempt's error is not a computational one – the two-jump minimisation is correct as far as it goes – it is a category error: treating a question about the infimum over an unbounded family (all finite partitions, of every length) as if it were a question about a single, small, fixed-dimensional optimisation. Once the sum is recognised as an upper Riemann sum, the value $1/3$ is not found, it is recalled – it is simply $\int_0^1 t^2\,dt$, a quantity that requires no fresh computation once the identification is made. The achievability half then follows from the same instinct pursued one step further: since cost varies continuously as jump points move continuously, and one can interpolate continuously between "coarse partition, cost near 1" and "fine partition, cost near $1/3$", the intermediate value theorem – applied along a path of partitions, not along the real line directly – fills in the entire interval. Kedlaya and Ng's solution states the Riemann-sum identification in a single line ("the cost of jumping along a particular sequence is an upper Riemann sum of the function t^2"); the reason that identification is available at all is that the problem's cost rule was constructed, digit for digit, to match $\sup \cdot \Delta x$ for t^2 – once that match is seen, both the bound and its sharpness stop needing separate justification and become restatements of what a Riemann sum is.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the shape $b^2(b-a)$ itself, appearing as a sum over consecutive jumps $\sum x_i^2(x_i - x_{i-1})$. On seeing a sum of (value at right endpoint) $\times$ (interval width), write down immediately: this is an upper Riemann sum for some function – identify the function before doing anything else.

2. That cue suggests the method: bound the sum by comparing it to the corresponding integral (Riemann sums squeeze the integral from above or below), rather than differentiating the sum as a function of finitely many jump positions.

3. Hesitation occurs right after computing the two-jump minimum ≈ 0.615 by calculus, which feels like real progress. Reinforce it away by asking explicitly: is the number of jumps in this problem fixed, or is "finite" doing the work of "arbitrarily many"? If the latter, a fixed-m optimum is not the quantity the problem is asking for.

4. The trap is treating "finite number of jumps" as a small, bounded number to optimise over classically, when the problem's real freedom is in letting that number grow without bound – the calculus-optimisation instinct anchors attention on the wrong variable.

5. Recognisable faster next time by checking, for any sum of the form $\sum g(x_i)(x_i - x_{i-1})$ over a partition, whether g evaluated at the partition endpoints is exactly sup or inf of some natural function on each subinterval – if so, the sum is a Riemann sum in disguise and its extreme behaviour is governed by integrability, not finite-dimensional calculus.

6. Reusable rule: a cost or score defined additively over the steps of a discrete process, of the form (value at the new point) $\times$ (step size), should be tested against the Riemann-sum template before any optimisation is attempted; the achievable range is then (integral, initial trivial value] or similar, with the integral as an unattained bound approached by refining the partition and the gap filled by continuously interpolating between coarse and fine partitions via the intermediate value theorem.

▷ **Cue:** a cost or score is defined as a sum over consecutive partition points, of the form $g(x_i) \cdot (x_i - x_{i-1}) \longrightarrow$ **Move:** recognise it as an upper or lower Riemann sum for some natural function g; bound it by comparing to $\int g$, using that Riemann sums approach (but for strictly monotonic g, never exactly equal) the integral as the partition refines

▷ **Cue:** a finite-optimisation instinct on a small, fixed number of steps produces a minimum that looks final $\longrightarrow$ **Move:** check whether the problem actually bounds the number of steps, or only requires it to be finite; if unbounded, the true extremal behaviour is a limit over the number of steps, not a calculus optimum at any fixed step count

2009 B3

2009 B3. Call a subset S of $\{1, 2, \ldots, n\}$ *mediocre* if it has the following property: Whenever a and b are elements of S whose average is an integer, that average is also an element of S. Let $A(n)$ be the number of mediocre subsets of $\{1, 2, \ldots, n\}$. [For instance, every subset of $\{1, 2, 3\}$ except $\{1, 3\}$ is mediocre, so $A(3) = 7$.] Find all positive integers n such that $A(n+2) - 2A(n+1) + A(n) = 1$.

Fields: combinatorics, number theory (binary representations) **Position:** B3 – the hard pivot of paper B, notorious for disguising itself as pure counting **Answer:** $n = 2^k - 1$ for some positive integer k.

Layer 1 THE WHAT — surface analysis

The natural attempt. The worked example – $A(3) = 7$, with $\{1, 3\}$ the lone exception – reads like an invitation to find a closed form for $A(n)$ directly, and this is a defensible first move: a mediocre set is, on reflection, exactly an arithmetic progression with odd common difference (an easy fact to conjecture from small cases and confirm), so $A(n)$ can be computed by counting arithmetic progressions with odd common difference inside $\{1, \ldots, n\}$, plus the empty set. This is entirely correct and gives an explicit, if slightly fiddly, formula for $A(n)$. The natural next step is to plug that formula into $A(n+2) - 2A(n+1) + A(n)$ and grind through the algebra of a second finite difference, hunting for when the result equals 1. This works in principle, but it is expensive: the closed form for $A(n)$ involves counting APs by common difference and length separately, and taking a second difference of a two-parameter counting sum invites exactly the kind of case-heavy bookkeeping that eats a competition-length solution without ever explaining, structurally, why $2^k - 1$ should be special. Pursued to the end it can be made to work, but the student who commits to it has chosen to fight the algebra of a difference operator on an explicit formula, rather than asking what a second difference of a counting function *means* combinatorially before computing anything.

The model chain. The key structural fact, needed only once and never again: mediocrity is an intrinsic property of a finite set of positive integers – it makes no reference to an ambient universe $\{1, \ldots, n\}$, since if $a, b \in S$ have integer average m, then m automatically satisfies $m \leq \max(a, b) \leq n$. Consequently, for any N, the mediocre subsets of $\{1, \ldots, N\}$ that avoid N are *exactly* the mediocre subsets of $\{1, \ldots, N-1\}$, as sets – no bijection needed, literal equality of families.

Split the mediocre subsets of $\{1, \ldots, n+1\}$ containing 1 by whether they contain $n+1$:

$$\{\text{mediocre } \subseteq \{1, \ldots, n+1\} \text{ containing } 1\} = \{\text{mediocre } \subseteq \{1, \ldots, n\} \text{ containing } 1\} \sqcup \{\text{mediocre } \subseteq \{1, \ldots, n+$$

Combined with the shift bijection $S \mapsto S + 1$ between mediocre subsets of $\{1, \ldots, n\}$ and of $\{2, \ldots, n+1\}$ (translation preserves the averaging condition), $A(n+1) - A(n)$ equals the number of mediocre subsets of $\{1, \ldots, n+1\}$ containing 1. Applying the splitting identity above with $n+1$ in place of n:

$$A(n+2) - 2A(n+1) + A(n) = \big(A(n+2) - A(n+1)\big) - \big(A(n+1) - A(n)\big) = \#\{\text{mediocre } S \subseteq \{1, \ldots, n+2\}$$

So the problem reduces to: for which $N = n + 2$ is the *full interval* $\{1, \ldots, N\}$ the *only* mediocre subset containing both endpoints 1 and N?

If $N - 1 = 2^a b$ with b odd and $b > 1$, the set $\{1 + mb : 0 \leq m \leq 2^a\}$ is a second, proper mediocre subset containing both endpoints: the average of $1 + m_1 b$ and $1 + m_2 b$ is an integer iff $m_1 + m_2$ is even, in which case it equals $1 + \frac{m_1 + m_2}{2} b$, which lies in the set. So whenever $N - 1$ is not a power of 2, the count exceeds 1.

If $N - 1 = 2^k$, induct on k: any mediocre $S \subseteq \{1, \ldots, 2^k + 1\}$ containing 1 and $2^k + 1$ must contain their average $2^{k-1} + 1$ (an integer). By the induction hypothesis applied to $\{1, \ldots, 2^{k-1} +$

1}, S must contain every integer in that range; likewise, translating $\{2^{k-1}+1,\ldots,2^k+1\}$ down by 2^{k-1} reduces to the same statement, forcing S to contain every integer up to 2^k+1. So $S=\{1,\ldots,2^k+1\}$ is forced, and the count is exactly 1. Hence the equation holds iff $n+1=2^k$, i.e. $n=2^k-1$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths eventually confront the same quantity: how the family of mediocre subsets changes as n increases by one. They split on whether that change is studied *after* computing $A(n)$ in closed form, or *before*. The natural attempt computes $A(n)$ explicitly first (via the odd-common-difference characterisation) and only then takes a second finite difference algebraically. The model treats the second difference operator itself as the object of study, and asks what it counts combinatorially – *before* any formula for $A(n)$ is written down.

> *Mediocrity doesn't reference the ambient set $\{1,\ldots,n\}$ at all – any forced average of two elements of S is automatically $\leq \max(S)$ – so it's a property of the finite set S itself. That means "mediocre subsets of $\{1,\ldots,n+1\}$ avoiding $n+1$" isn't merely in bijection with "mediocre subsets of $\{1,\ldots,n\}$" – it* is *that family, literally. This is what lets me split by containment of the top element without any hereditary-property worries.*

> *$A(n+1)-A(n)$ counts mediocre subsets of $\{1,\ldots,n+1\}$ containing the bottom element 1 (via the shift bijection); applying that same idea one level up and splitting by whether the top element $n+2$ is also present tells me the* second *difference is not some abstract algebraic residue – it is exactly the number of mediocre subsets containing* both *endpoints 1 and $n+2$. A second difference of a counting function, here, literally counts something.*

Layer 3 THE WHY

The reason the second-difference identity collapses so cleanly is that mediocrity is defined without any reference to an ambient bound – it is a property of a finite set of positive integers, full stop, since the average of two elements never exceeds the larger of the two. This single observation is what licenses treating "avoids the top element" as literal set-equality across different values of n, rather than merely an isomorphism that needs checking; and it is what makes a first difference count "contains the bottom element" and a second difference count "contains both endpoints" – each application of the split-by-an-element trick removes one layer of the difference operator and adds one forced-containment condition. Once the problem is reduced to "for which N is $\{1,\ldots,N\}$ the only mediocre subset containing 1 and N", the answer is forced by a fact about binary representations: containing both endpoints forces containing their midpoint whenever it is an integer, and the midpoint of 1 and N is an integer precisely when $N-1$ is even. Iterating this bisection is exactly the process of repeatedly halving $N-1$, and it only ever bottoms out cleanly – forcing every single integer in between, with no freedom left over – when $N-1$ is a pure power of 2; any odd factor $b>1$ surviving in $N-1=2^a b$ instead permits an arithmetic-progression-shaped "skip" set that satisfies the endpoint-forcing condition without ever being forced to fill in the intermediate integers not congruent to 1 (mod b). So the answer $n=2^k-1$ is not an arbitrary-looking numeric pattern to be pattern-matched from small cases; it is the exact condition under which repeated integer-bisection of an interval has no room to stop short of the full interval. Kedlaya and Ng's solution states the reduction to "mediocre subsets containing the endpoints" and the bisection induction without dwelling on why mediocrity transfers so freely between $\{1,\ldots,n\}$ and $\{1,\ldots,n+1\}$; the reason is the ambient-independence fact above, and it is what makes the whole telescoping argument rigorous rather than merely suggestive.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the structure of the target equation itself: $A(n+2) - 2A(n+1) + A(n)$ is a second finite difference, not an arbitrary combination. On seeing it, write down immediately: a first difference of a counting function often counts "objects containing a marked new element" – ask what a *second* difference should count, combinatorially, before reaching for a closed form.

2. That cue suggests the method: express each difference as a containment-splitting identity (does the new element appear or not), checking first whether the underlying property (mediocrity) is intrinsic to the set or depends on the ambient universe, since that determines whether the split is a clean set-equality or requires a separate bijection argument.

3. Hesitation occurs right at the point of choosing between grinding out a closed form for $A(n)$ (the odd-AP characterisation) versus working with the difference operator abstractly. Reinforce it away by noting that the target equation only ever needs to equal 1 – an exact closed form for $A(n)$ is far more information than the question requires, so the question is a signal to work with differences directly.

4. The trap is the AP characterisation itself: it is correct, elegant, and immediately achievable from the hint set $A(3) = 7$, and it tempts a full closed-form computation that buries the clean endpoint-forcing structure under case-heavy algebra on common differences and lengths.

5. Recognisable faster next time by treating any problem asking for a finite difference (first, second, or higher) of a counting function $A(n)$ as a request to interpret the difference combinatorially – via inclusion/exclusion on a single marked new element – before ever attempting to find $A(n)$ itself in closed form.

6. Reusable rule: a k-th finite difference of a set-counting function, applied to a property that is intrinsic to the set itself (independent of the ambient bound), counts exactly the objects forced to contain k specific marked elements (the k most recently added extremes); reduce the problem to characterising when that constrained family is forced down to a single object, rather than differencing an explicit closed-form count.

▷ **Cue:** a problem asks for values of n where a finite difference (e.g. $A(n+2) - 2A(n+1) + A(n)$) of a counting function equals a small fixed number $\longrightarrow$ **Move:** interpret the difference combinatorially by splitting on containment of the newest marked element(s) at each level, rather than computing a closed form for the counting function and differencing it algebraically; check first whether the defining property is intrinsic to the set (ambient-independent), which licenses literal set-equality in the splits

▷ **Cue:** a set containing two elements whose average is forced to lie in the set, repeated on the resulting midpoints $\longrightarrow$ **Move:** recognise this as interval bisection in disguise; the set is forced to be the full interval exactly when the endpoint gap is a pure power of 2, since only then does repeated halving never leave an odd remainder that permits a smaller arithmetic-progression alternative

2009 B4

2009 B4. Say that a polynomial with real coefficients in two variables, x, y, is *balanced* if the average value of the polynomial on each circle centered at the origin is 0. The balanced polynomials of degree at most 2009 form a vector space V over $\mathbb{R}$. Find the dimension of V.

Fields: linear algebra (dimension counting, kernels of linear functionals); integration over circles; homogeneous decomposition **Position:** B4 — mid-paper, but in practice one of the more approachable problems on B: the difficulty is entirely in counting the constraints correctly **Answer:** $\dim V = \binom{2011}{2} - 1005 = 2021055 - 1005 = 2020050$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The space W of all polynomials of degree at most 2009 has the monomials $x^a y^b$, $a+b \le 2009$, as a basis; there are $1+2+\cdots+2010 = \frac{2010\cdot 2011}{2} = 2021055$ of them. "Balanced" is a list of linear conditions, so $\dim V = 2021055 - (\text{number of independent conditions})$. Everyone gets that far.

Next the student computes the average of a monomial on the circle of radius r, parametrised by $(r\cos\theta, r\sin\theta)$:

$$\frac{1}{2\pi}\int_0^{2\pi} (r\cos\theta)^a (r\sin\theta)^b \, d\theta = \frac{r^{a+b}}{2\pi}\int_0^{2\pi} \cos^a\theta \, \sin^b\theta \, d\theta.$$

The integral $I(a,b) = \int_0^{2\pi} \cos^a\theta \sin^b\theta \, d\theta$ vanishes when a or b is odd (substitute $\theta \mapsto \pi - \theta$ or $\theta \mapsto -\theta$) and is positive when both are even. So far, correct.

Here is where the attempt goes wrong. The student reasons: "the monomials with a, b both even have nonzero average, the others have zero average, so each even–even monomial must have zero coefficient." That counts one condition per even–even monomial: for $a+b = 2k$ there are $k+1$ such monomials, giving $\sum_{k=0}^{1004}(k+1) = \frac{1005\cdot 1006}{2} = 505515$ conditions and the answer 1515540. It is wrong. The polynomial $x^2 - y^2$ has two even–even monomials with nonzero coefficients and is balanced: its average on the circle of radius r is $\frac{r^2}{2\pi}(I(2,0) - I(0,2)) = \frac{r^2}{2\pi}(\pi - \pi) = 0$. The averages of different monomials can *cancel*. The student has confused "each monomial contributes nothing" with "the sum contributes nothing" — a sum of nonzero numbers can be zero.

The opposite error is also common: "balanced means one integral vanishes, so there is one condition and $\dim V = 2021055 - 1$." That forgets that the problem asks for *each* circle, i.e. for every $r > 0$.

The model chain. **Step 1: split by degree.** Every $P \in W$ is uniquely $P = \sum_{i=0}^{2009} P_i$ with P_i homogeneous of degree i. Let H_i be the space of homogeneous polynomials of degree i; $\dim H_i = i+1$ (basis $x^i, x^{i-1}y, \ldots, y^i$), and $W = H_0 \oplus H_1 \oplus \cdots \oplus H_{2009}$.

Step 2: the average as a polynomial in r. Define the linear functional $\lambda_i : H_i \to \mathbb{R}$, $\lambda_i(Q) = \int_0^{2\pi} Q(\cos\theta, \sin\theta)\, d\theta$. Homogeneity gives $P_i(r\cos\theta, r\sin\theta) = r^i P_i(\cos\theta, \sin\theta)$, so the average of P on the circle of radius r is

$$A_P(r) = \frac{1}{2\pi}\sum_{i=0}^{2009} r^i \, \lambda_i(P_i).$$

P is balanced iff $A_P(r) = 0$ for all $r > 0$. A polynomial in r with infinitely many zeros is the zero polynomial, so this holds iff $\lambda_i(P_i) = 0$ for every i. Therefore

$$V = \ker\lambda_0 \oplus \ker\lambda_1 \oplus \cdots \oplus \ker\lambda_{2009}.$$

Step 3: odd degrees. If i is odd, $P_i(-x,-y) = -P_i(x,y)$. The substitution $\theta \mapsto \theta + \pi$ sends $(\cos\theta, \sin\theta)$ to its negative and preserves $[0, 2\pi]$ up to periodicity, so $\lambda_i(P_i) = -\lambda_i(P_i)$, i.e. $\lambda_i = 0$ and $\ker\lambda_i = H_i$.

Step 4: even degrees. If $i = 2k$, λ_i is not the zero functional: $Q = (x^2+y^2)^k \in H_i$ has $Q(\cos\theta, \sin\theta) = 1$, so $\lambda_i(Q) = 2\pi \neq 0$. A nonzero linear functional on a finite-dimensional space is onto $\mathbb{R}$, so by rank–nullity its kernel has dimension $\dim H_i - 1$.

Step 5: count. There are 1005 even i in $\{0, \ldots, 2009\}$ (namely $0, 2, \ldots, 2008$). Hence

$$\dim V = \sum_{i=0}^{2009}(i+1) - 1005 = 2021055 - 1005 = 2020050. \qquad \square$$

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains compute the monomial averages $\frac{r^{a+b}}{2\pi} I(a,b)$ correctly. They split at the question "how many conditions is that?"

(N) One condition per monomial with nonzero average — conditions indexed by *monomials*.

(M) One condition per power of r — conditions indexed by *degrees*, because the only thing the problem lets you vary is r, and r sees a monomial only through its total degree.

> *A condition is an equation, not a monomial. The equation here is "average on the circle of radius r is 0", one for each r. So write down that average for a general P and see what it is as a function of r. It's a polynomial in r, and the coefficient of r^i only involves the degree-i part. So the conditions are "the coefficient of r^i vanishes", one per i — at most 2010 of them, not one per monomial.*

> *For each i, is that condition real or empty? Odd i: antipodal points cancel, so it's empty. Even i: $(x^2+y^2)^{i/2}$ is identically 1 on the unit circle, so the functional is nonzero and costs exactly one dimension.*

- ▷ **Cue:** "the average (or integral) over each circle / each sphere / each level set is zero" $\longrightarrow$ **Move:** parametrise the family by its size r, pull r out using homogeneity, and read off one linear condition per power of r
- ▷ **Cue:** a space cut out by linear conditions and a request for its dimension $\longrightarrow$ **Move:** count independent conditions, not terms; a nonzero linear functional on a finite-dimensional space removes exactly one dimension
- ▷ **Cue:** an integral of an odd function over a set symmetric under $v \mapsto -v$ $\longrightarrow$ **Move:** it is zero; use this to kill whole families of conditions at once
- ▷ **Cue:** need to show a linear functional is not identically zero $\longrightarrow$ **Move:** exhibit one input on which it is visibly nonzero, preferably one that is constant on the domain of integration, such as $(x^2+y^2)^k$ on a circle

Layer 3 THE WHY

Why must the conditions be indexed by degree? Go back to the definition of a linear condition on a vector space: it is a linear functional set equal to zero. The problem hands you a family of functionals, $P \mapsto A_P(r)$, one per $r > 0$ — infinitely many. The question is the dimension of their joint span. Monomials never enter that question; they are a basis of the domain, not of the space of functionals. The student who counts monomials is counting the wrong space.

Then why does r sort things by degree? Because scaling the circle is the same as scaling the variables, and scaling $(x,y) \mapsto (rx,ry)$ multiplies a degree-i homogeneous polynomial by exactly r^i. That is the definition of homogeneous of degree i. So the functional "average on the circle of radius r" is $\sum_i r^i \lambda_i$, and the family $\{A(r)\}_{r>0}$ spans the same space as $\{\lambda_i\}$: the Vandermonde fact that a polynomial vanishing at infinitely many points is zero. That step is what lets the infinitely many conditions collapse to at most 2010. It is not a trick; it is the question "what does the problem let me vary, and what does that variable see?"

Kedlaya's solution opens with "Any polynomial ... can be written uniquely as a sum $\sum P_i$ in which P_i is homogeneous of degree i." Here is where that comes from: the balanced condition is a statement about all dilations of one circle, and homogeneous components are exactly the eigenvectors of dilation. Whenever a condition is imposed for every scale, decompose into pieces that scale cleanly.

Now why do the odd degrees drop out while each even degree costs one? Odd: the circle is symmetric under $v \mapsto -v$ and an odd-degree homogeneous polynomial changes sign under

it, so the integral is forced to be zero whatever the coefficients. Even: one sample is enough to show nonzeroness, and the sample to pick is one that is constant on the circle, because then the integral needs no computing — and $x^2 + y^2 = 1$ *is* the circle. You are using the equation of the domain as a test function. That is the circle example from the Blueprint in a different costume: the circle is defined by $x^2 + y^2 = r^2$, so $x^2 + y^2$ is the object of interest.

Notice what you never had to do: evaluate $I(2k, 0)$, describe the kernel of λ_{2k} explicitly, or find a basis of V. Dimension counting only needs "nonzero functional, codimension one." Computing more than the question asks is the zombie maths this problem punishes.

Caveat. Verified by exact computer algebra: for every degree bound $d \le 8$ the dimension of the balanced space, computed as (number of monomials) minus the rank of the matrix of r^i-coefficients of the circle averages, equals $\binom{d+2}{2} - (\lfloor d/2 \rfloor + 1)$. At $d = 2009$ this is $2021055 - 1005 = 2020050$, matching the official answer.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "on *each* circle centered at the origin": a family of conditions indexed by the radius. On seeing it, write $A_P(r)$ for a general P and ask what kind of function of r it is.
2. The cue suggests splitting P into homogeneous parts, so that $A_P(r) = \frac{1}{2\pi} \sum r^i \lambda_i(P_i)$, and then using "a polynomial zero for all $r > 0$ has zero coefficients" to turn one condition per radius into one condition per degree.
3. Hesitation occurs after computing that each even–even monomial has a nonzero average: it looks as if every such monomial is forbidden. It is reinforced away by testing one balanced example, $x^2 - y^2$, which shows that averages cancel between monomials of the same degree.
4. The trap is counting monomials instead of equations, giving 505515 conditions. The opposite trap is imposing the unit circle only, giving one condition. Both come from not writing down what one condition actually is.
5. Recognise it faster by asking of any dimension question, before touching a basis: *what exactly are the equations, how are they indexed, and which of them are identically satisfied?*
6. Reusable decision rule: *when a condition is imposed for every value of a scaling parameter, decompose the object into pieces that scale homogeneously; the conditions separate piece by piece, one per power of the parameter. Then each piece costs zero or one dimension according as its functional is zero or not — test zero by symmetry, test nonzero with one visibly positive input.*

2009 B5

2009 B5. Let $f : (1, \infty) \to \mathbb{R}$ be a differentiable function such that

$$f'(x) = \frac{x^2 - f(x)^2}{x^2(f(x)^2 + 1)} \qquad \text{for all } x > 1.$$

Prove that $\lim_{x \to \infty} f(x) = \infty$.

Fields: real analysis; qualitative theory of differential equations; monotonicity and comparison **Position:** B5 — genuinely hard in the arena, because the equation cannot be solved and the natural sign analysis branches into cases; one change of variable removes all of them **Answer:** Proof. In fact $f(x) \sim (3x)^{1/3}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. First reflex: solve the equation. It is not separable (x and f are tangled in the numerator), not linear, not Bernoulli, not homogeneous. Substitutions $f = xv$ or $f = \tan u$

make it worse. After ten minutes the student abandons solving and does what they were taught for autonomous equations: read the sign of f'.

The sign of f' is the sign of $x^2 - f^2$: f increases where $|f| < x$ and decreases where $|f| > x$. So the plan is: "show that eventually $-x < f(x) < x$; then f is increasing, so it has a limit $L \in (-\infty, \infty]$; if L were finite then $f' \to \frac{1}{1+L^2} > 0$, which forces $f \to \infty$, contradiction."

The last two steps are fine. The first is where the cost is. Two dangers:

- f could start above the line $y = x$. There it decreases, but x increases, so the gap $f - x$ shrinks at rate at least 1 — unless f stays above x forever, in which case $f \geq x \to \infty$ and we are done anyway. So the student splits: *Case A*, $f(x) \geq x$ for all x; *Case B*, $f(x_0) < x_0$ for some x_0. In Case B one must prove f never recrosses $y = x$: at a first recrossing point $x_1 > x_0$ we would have $f(x_1) = x_1$, hence $f'(x_1) = 0 < 1$, so $x - f$ has derivative 1 at x_1 and was negative immediately before x_1 — contradicting minimality. That is correct, but it is a delicate first-crossing argument that many students botch (the "least" crossing needs a closed set and continuity).
- f could be very negative, below $-x$. There it decreases too. The student needs a lower bound: from the equation, $f' = \frac{1}{1+f^2} - \frac{f^2}{x^2(1+f^2)} > -\frac{1}{x^2}$, so $f(x) > f(x_0) - \frac{1}{x_0}$ for $x > x_0$. Now f is bounded below while $-x \to -\infty$, so eventually $f > -x$.

With both dangers handled, $|f| < x$ eventually and the plan closes. This is exactly the official first solution. It works; it costs a case split, a first-crossing argument, and a separate lower bound, and in the arena every one of those is a place to lose a mark.

The model chain. **Step 0: rewrite.** Multiply the equation by $f(x)^2 + 1$:

$$\big(1 + f(x)^2\big) f'(x) = 1 - \frac{f(x)^2}{x^2}. \tag{$*$}$$

Put $\varphi(t) = t + \frac{1}{3}t^3$ and $g(x) = \varphi(f(x))$. By the chain rule $g'(x) = (1 + f(x)^2) f'(x)$, so

$$g'(x) = 1 - \frac{f(x)^2}{x^2} \qquad (x > 1).$$

Note that φ is continuous and strictly increasing ($\varphi'(t) = 1 + t^2 > 0$) with $\varphi(t) \to \pm\infty$ as $t \to \pm\infty$.

Step 1: f is bounded below. From the original equation, $f' = \frac{1}{1+f^2} - \frac{f^2}{x^2(1+f^2)} > -\frac{f^2}{x^2(1+f^2)} \geq -\frac{1}{x^2}$. So $u(x) = f(x) - \frac{1}{x}$ has $u' = f' + \frac{1}{x^2} > 0$ and is increasing. For $x \geq 2$: $f(x) > u(x) \geq u(2) = f(2) - \frac{1}{2} =: m$.

Step 2: g grows at most linearly. $g' \leq 1$, so $g(x) - x$ is nonincreasing and $g(x) \leq g(2) + x - 2$ for $x \geq 2$.

Step 3: $f(x)^2/x^2 \to 0$. Let $x \geq 2$. If $f(x) < 0$ then $m < f(x) < 0$ and $f(x)^2 < m^2$. If $f(x) \geq 0$ then, using $f(x) > m$,

$$\tfrac{1}{3} f(x)^3 = g(x) - f(x) < g(2) + x - 2 - m = x + K, \qquad K := g(2) - 2 - m,$$

so $f(x)^2 < \big(3(x+K)\big)^{2/3}$ once $x + K > 0$. Either way, for large x, $f(x)^2 \leq \max\{m^2, (3(x+K))^{2/3}\}$, and dividing by x^2 gives $f(x)^2/x^2 \to 0$.

Step 4: $g \to \infty$. By Step 3, $g'(x) = 1 - f(x)^2/x^2 \to 1$. Choose $X \geq 2$ with $g' \geq \frac{1}{2}$ on $[X, \infty)$; then $g(x) \geq g(X) + \frac{1}{2}(x - X) \to \infty$.

Step 5: $f \to \infty$. Given M, choose X' with $g(x) > \varphi(M)$ for $x > X'$. If $f(x) \leq M$ for such x then $g(x) = \varphi(f(x)) \leq \varphi(M)$, a contradiction; so $f(x) > M$ for all $x > X'$. Hence $\lim_{x\to\infty} f(x) = \infty$. □

Bonus. Steps 2 and 4 give $g(x)/x \to 1$, and $g = f + \frac{1}{3}f^3$ with $f \to \infty$ gives $f(x)^3/(3x) \to 1$: $f(x) \sim (3x)^{1/3}$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start from the same equation and both need a lower bound for f. They split at the first thing done with the right-hand side.

(N) Read off the *sign* of f': positive iff $|f| < x$. The plane is cut by the lines $y = \pm x$ into regions, and the argument must track which region f is in — hence cases.

(M) Clear the denominator that depends only on f: $(1 + f^2)f' = 1 - f^2/x^2$. The left side is the derivative of $f + \frac{1}{3}f^3$, and the right side is ≤ 1 everywhere and $\to 1$ once $f = o(x)$. No regions, no cases.

> *The awkward bit is the $f^2 + 1$ downstairs — it's only a function of f. If I multiply it across, the left side becomes $(1 + f^2)f'$, and that's the derivative of $f + \frac{1}{3}f^3$. Then the right side is $1 - f^2/x^2$, which is at most 1 and should be almost 1. So $f + \frac{1}{3}f^3$ behaves like x.*

> *For large x the f^2/x^2 term ought to be negligible, so the equation is roughly $f' = \frac{1}{1+f^2}$, i.e. $f + \frac{1}{3}f^3 \approx x$, i.e. $f \approx (3x)^{1/3}$. So what I really have to prove is only that f is small compared with x — and $f + \frac{1}{3}f^3 \leq x + C$ gives me that for free, once f is bounded below.*

▷ **Cue:** an ODE $f' = \dfrac{N(x,f)}{D(f)}$ where the denominator depends on f alone $\longrightarrow$ **Move:** multiply by $D(f)$; the left side is $\frac{d}{dx}\Phi(f)$ with $\Phi' = D$; analyse $\Phi(f)$ instead of f

▷ **Cue:** an unsolvable ODE and a question about behaviour at infinity $\longrightarrow$ **Move:** guess the asymptotics by dropping the term that should be negligible, solve the simplified equation, then prove only that the dropped term really is negligible

▷ **Cue:** you need $f \to \infty$ but you control $\varphi(f)$ for an increasing unbounded φ $\longrightarrow$ **Move:** $\varphi(f) \to \infty$ is enough: $f \leq M$ would force $\varphi(f) \leq \varphi(M)$

▷ **Cue:** a derivative bounded below by an integrable function, e.g. $f' > -1/x^2$ $\longrightarrow$ **Move:** add an antiderivative to make it monotone ($f - 1/x$ increasing) and read off a lower bound for f

Layer 3 THE WHY

Why is clearing the f-only denominator the natural first move and not a lucky one? Go back to what the chain rule says: $\frac{d}{dx}\Phi(f(x)) = \Phi'(f(x))f'(x)$. Any expression of the form (function of f) $\times f'$ is an exact derivative. That is the whole theory of separable equations; for $f' = a(x)/D(f)$ you would write $D(f)f' = a(x)$ and integrate both sides. Here the equation is *almost* separable: the numerator $1 - f^2/x^2$ still mixes x and f, but the denominator is pure f. So you do the half of the separation that is available. What you get, $\Phi(f)' = 1 - f^2/x^2$, is no longer an equation you can solve, but it is one you can *bound*, because the right side is squeezed between something and 1. That is the general principle: when you cannot separate, separate as much as you can and bound the rest.

Why is the right-hand side then so tractable? Because the problem's own formula says so if you rewrite it with the x^2 distributed: $\frac{x^2-f^2}{x^2(1+f^2)} = \frac{1}{1+f^2}\left(1 - \frac{f^2}{x^2}\right)$. The factor $1 - f^2/x^2$ is the comparison of f with x — exactly the comparison the natural attempt tracked with cases. In (M) it is not a case distinction but a number that tends to 1 as soon as f grows slower than x. And the transformed equation proves that growth bound itself: $\frac{1}{3}f^3 \leq x + C$. The equation certifies its own approximation. That is why no case split is needed.

Elkies's third official solution begins "Consider the function $g(x) = f(x) + \frac{1}{3}f(x)^3$." Here is where that comes from: it is the antiderivative of the denominator $1 + f^2$. Kedlaya and Ng's first solution instead reads signs and pays in cases; Zara's second solution introduces $f \pm x$ to track the lines $y = \pm x$ — the case structure made into functions. The three solutions are three answers to "what do I do with the denominator?": ignore it (sign analysis), or integrate it.

And the answer $f \sim (3x)^{1/3}$, which the problem does not ask for, is the check that tells you in advance the proof is aiming at the right target: f tends to infinity, but slowly — slower than x, which is why the region $|f| < x$ is where everything happens.

The lower bound (Step 1) is needed in both chains and deserves its own why. The equation says $f' \geq -f^2/(x^2(1+f^2))$, and $f^2/(1+f^2) < 1$ however large f is. That saturating fraction is the key: whatever f does, it cannot fall faster than $1/x^2$, whose total over $[2, \infty)$ is finite. So f can lose at most $\frac{1}{2}$ from $x = 2$ onward.

Caveat. Verified numerically as a sanity check (not part of the proof): integrating the equation from $x = 1.5$ with $f(1.5) \in \{-50, -3, 0, 1.5, 5, 50, 300\}$ up to $x = 10^5$, every solution eventually increases and approaches $(3x)^{1/3}$ (e.g. $f(10^5) \approx 66.93$ against $(3 \cdot 10^5)^{1/3} \approx 66.94$ for the moderate starting values), and the solution starting at -50 never falls below $f(2) - \frac{1}{2}$. The proof above is independent of this check.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the factor $f(x)^2 + 1$ in the denominator: a function of f alone. On seeing it, write $(1 + f^2)f' = 1 - f^2/x^2$ and recognise the left side as $\frac{d}{dx}(f + \frac{1}{3}f^3)$.

2. The cue suggests partial separation: transfer the analysis from f to $g = \varphi(f)$ with $\varphi' = 1 + t^2$, bound g' from above by 1 and show it tends to 1.

3. Hesitation occurs at "but f might be bigger than x, or very negative". It is reinforced away by noting that $g \leq x + C$ already caps f at about $(3x)^{1/3}$, and $f' > -1/x^2$ already floors f; neither needs a case split.

4. The trap is region-tracking: splitting the plane by $y = \pm x$ and proving f eventually stays in $|f| < x$. It can be done, but it invites an unrigorous "first crossing" argument. A secondary trap is forgetting the lower bound entirely and assuming f increasing from the start.

5. Recognise it faster by asking of any differential equation: *which factors depend on the unknown alone?* Every such factor can be moved to the f' side and integrated, even when the equation as a whole cannot be solved.

6. Reusable decision rule: *for the long-run behaviour of an unsolvable equation, first guess the asymptotics by dropping the term you expect to be negligible; then change variable to the quantity that the simplified equation makes exactly linear (here $f + \frac{1}{3}f^3$), and prove only that the dropped term is negligible.* The same rule handles recurrences such as $a_{n+1} = a_n + 1/a_n^2$ (look at a_n^3) and $a_{n+1} = \sin a_n$ (look at $1/a_n^2$).

2009 B6

2009 B6. Prove that for every positive integer n, there is a sequence of integers $a_0, a_1, \ldots, a_{2009}$ with $a_0 = 0$ and $a_{2009} = n$ such that each term after a_0 is either an earlier term plus 2^k for some nonnegative integer k, or of the form $b \bmod c$ for some earlier positive terms b and c. [Here $b \bmod c$ denotes the remainder when b is divided by c, so $0 \leq (b \bmod c) < c$.]

Fields: number theory (congruences, binomial theorem); constructions **Position:** B6 — the last problem, hard because the space of constructions is huge and the budget of steps is fixed while n is not; once the right question is asked, the construction has eight lines **Answer:** Construction. For odd n and $x = 2^h \geq n$: $0,\ 1,\ 2x+1,\ (x+1)^2,\ x^n+1,\ n(x+1),\ x,\ n$, then repeat n; for $n = 2^e m$ with m odd, multiply every term by 2^e.

Layer 1 THE WHAT — surface analysis

The natural attempt. The student reads the two operations and starts with the one that is easy to control: "earlier term plus 2^k". Starting from $a_0 = 0$, write n in binary, $n = 2^{k_1} + 2^{k_2} + \cdots + 2^{k_s}$, and add one power at a time: $a_1 = 2^{k_1}$, $a_2 = a_1 + 2^{k_2}$, …, $a_s = n$. Pad with copies of n (padding is legal: build some $c = 2^K > n$ as $a_0 + 2^K$ along the way, and then $n = n \bmod c$ as often as needed).

The wall: s, the number of 1s in the binary expansion of n, is unbounded. With 2009 steps this handles $n < 2^{2008}$ or so, not every n. The problem says *every* positive integer, with a *fixed* length. So additions alone cannot work, and the mod operation has to do almost all of the work.

The student then experiments with mod on numbers that additions make cheaply — numbers with few binary 1s, such as 2^a, $2^a + 1$, $2^a + 2^b + 1$. For instance $2^a \bmod (2^b + 1)$: since $2^b \equiv -1$, this is $2^{a \bmod b}$ or $2^b + 1 - 2^{a \bmod b}$ depending on the parity of $\lfloor a/b \rfloor$. Both are again sparse in binary. The student may push on and find that $2^a \bmod 3^k$ runs through every residue coprime to 3 (because 2 is a primitive root mod 3^k), which leads to the official second solution — but then one must manufacture 3^{2h} as a remainder, needs the lifting fact about primitive roots, and must fix up n divisible by 3. It is a valid route, and a heavy one. Most students stall between "binary has too many ones" and "I need number theory I cannot recall under time pressure".

The model chain. **Step 1: reduce to odd n.** Suppose $a_0, \ldots, a_{2009}$ is a valid sequence ending at m, and $e \geq 0$. Then $2^e a_0, \ldots, 2^e a_{2009}$ is valid and ends at $2^e m$: $a_i = a_j + 2^k$ becomes $2^e a_i = 2^e a_j + 2^{k+e}$; and if $b = qc + r$ with $0 \leq r < c$ then $2^e b = q(2^e c) + 2^e r$ with $0 \leq 2^e r < 2^e c$, so $(2^e b) \bmod (2^e c) = 2^e (b \bmod c)$, and $2^e b, 2^e c$ are positive exactly when b, c are. Writing $n = 2^e m$ with m odd, it suffices to treat odd n.

Step 2: the key congruence. Let n be odd, choose $h \geq 1$ with $x := 2^h \geq n$ (so in fact $x > n$, since x is even and n odd). Put $y = x + 1$. By the binomial theorem,

$$x^n = (y-1)^n = \sum_{j=0}^{n} \binom{n}{j} y^j (-1)^{n-j} \equiv (-1)^n + n\,y\,(-1)^{n-1} = -1 + ny \pmod{y^2},$$

using that n is odd. Hence $x^n + 1 \equiv n(x+1) \pmod{(x+1)^2}$. Since $0 < n(x+1) < (x+1)^2$ (because $n < x + 1$), this says exactly

$$(x^n + 1) \bmod (x+1)^2 = n(x+1).$$

And $n(x+1) = nx + n$ with $0 \leq n < x$, so $n(x+1) \bmod x = n$.

Step 3: the sequence. Every step is checked against the rules:

$$\begin{aligned}
a_0 &= 0, \\
a_1 &= 1 = a_0 + 2^0, \\
a_2 &= 2x + 1 = a_1 + 2^{h+1}, \\
a_3 &= (x+1)^2 = a_2 + 2^{2h}, \\
a_4 &= x^n + 1 = a_1 + 2^{hn}, \\
a_5 &= n(x+1) = a_4 \bmod a_3 \quad (a_4, a_3 > 0), \\
a_6 &= x = a_0 + 2^h, \\
a_7 &= n = a_5 \bmod a_6 \quad (a_5, a_6 > 0), \\
a_i &= n = a_7 \bmod a_3 \quad \text{for } 8 \leq i \leq 2009 \quad (n < a_3).
\end{aligned}$$

So $a_{2009} = n$. Combined with Step 1 this handles every positive integer. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree that additions alone reach only numbers with few binary 1s, and that n must come out of a single mod. They split at the question asked next.

(N) "Which remainders can I get from sparse numbers?" — an open-ended search through residues, which leads to discrete logarithms and primitive roots.

(M) "Where can an *arbitrary* n appear cheaply?" — answer: as an exponent. x^n with $x = 2^h$ costs one addition however large n is. So the question becomes: which operation converts an exponent into a coefficient? Answer: the binomial theorem, read modulo the square of the thing being expanded.

> *I have a fixed number of steps and n can be anything, so n can't be built up digit by digit. The only place an arbitrary n fits into one step is an exponent:* 2^{hn} *is one addition. So I need to get n down from the exponent.* $(1+y)^n = 1 + ny + \cdots$ *— the n comes down as the coefficient of y, and everything after it is a multiple of* y^2*. So reduce* x^n *modulo* $(x+1)^2$*, with* $y = x + 1$*.*

> *Now I have* $n(x+1)$*, not n. To strip the factor* $x+1$ *I reduce mod x:* $n(x+1) = nx + n$ *leaves n as long as* $n < x$*. And x,* x^n*,* $(x+1)^2 = x^2 + 2x + 1$ *are all a couple of additions from* 0 *because x is a power of* 2*.*

- ▷ **Cue:** a fixed number of operations must produce an arbitrary integer n ⟶ **Move:** put n where it costs nothing — in an exponent — and find an operation that brings it down
- ▷ **Cue:** need to extract n from a^n ⟶ **Move:** expand $a = (y-1)$ or $(1+y)$ and reduce modulo y^2: $(1+y)^n \equiv 1 + ny \pmod{y^2}$
- ▷ **Cue:** a statement about all positive integers, with operations that commute with multiplication by 2 ⟶ **Move:** reduce to odd n by scaling the whole construction by 2^e
- ▷ **Cue:** a remainder that is correct modulo M but must equal a specific number ⟶ **Move:** choose the free parameter large enough that the target already lies in $[0, M)$

Layer 3 THE WHY

Why must n go into an exponent? Count. After t additions starting from 0, with no mod used, the terms have at most t binary 1s. A single mod can produce a number that is not sparse, but the inputs to that mod must themselves be built. Look at the rules for a slot that accepts an unbounded integer in a single step: there is exactly one, the exponent k in 2^k. (The mod rule takes only earlier terms as inputs, so it cannot inject new information; it can only rearrange what the additions supplied.) That slot takes any nonnegative integer at no cost. So the first design decision is forced by the rules as written: n enters as $2^{\text{something involving } n}$. Nothing had to be spotted — it is the only door in the room.

Why the binomial theorem? Because once n is in an exponent, the question "how do I get n out?" has one textbook answer that uses only arithmetic, not logarithms: the first-order term of $(1+y)^n$ is ny. That is the same fact as the derivative of t^n being nt^{n-1} — differentiation is exactly the operation that brings an exponent down, and reducing modulo y^2 is the arithmetic way of keeping only the first-order term. So the move is "take the derivative, arithmetically". A student who knows $\frac{d}{dt}t^n = nt^{n-1}$ already owns this idea; the missing step is recognising that "mod y^2" is how you differentiate with integers.

Why $x + 1$ and a power of 2 for x? The expansion point y must be something whose powers are reachable: $x = 2^h$ makes $x^n = 2^{hn}$ free, and $y = x + 1$ makes $x = y - 1$, so x^n is $(y-1)^n$. The modulus $y^2 = x^2 + 2x + 1$ has three binary 1s: three additions. Every choice is dictated by "keep everything sparse in binary".

Why odd n? Because $(y-1)^n$ carries the sign $(-1)^{n-1}$ on the linear term; for odd n that sign is $+$ and the constant is -1, which the "$+1$" in $a_4 = x^n + 1$ cancels. For even n you could fiddle with signs, but the scaling symmetry in Step 1 is cheaper: both operations commute with multiplication by 2, so the powers of 2 in n come for free. Recognising a symmetry of the rules is always cheaper than a second case.

Kedlaya and Ng's first solution (credited to Yufei Zhao) begins "we may reduce to the case where n is odd" and sets $x = 2^h \geq n$ with no comment. Here is where both come from: oddness makes the sign of the linear term come out right, and $x \geq n$ is the size condition that makes a congruence into an equality of remainders.

Finally, the number 2009. It is the year, not a constraint: the construction uses eight terms. Its only role is to tell you the length is *fixed*, which is the fact that kills binary expansion and forces the exponent idea. The official remarks record a six-step variant (Lenstra, van Luijk, Della Torre) using Euler's theorem; the eight-step one above needs nothing beyond the binomial theorem.

Caveat. Verified by computer: for every n from 1 to 300 the sequence above (with the 2^e scaling for even n, padded to length 2010) was generated and each term a_i, $i \geq 1$, was checked to be either an earlier term plus a power of 2 or b mod c for earlier positive terms b, c; all passed, with $a_{2009} = n$. The general proof above does not depend on this check.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the pairing "every positive integer n" with a *fixed* length 2009. On seeing it, write: "binary expansion fails; n must enter in $O(1)$ steps; where can an arbitrary integer enter for free?"

2. The cue suggests putting n in an exponent (2^{hn}) and extracting it with a congruence — the binomial theorem modulo a square: $(y-1)^n \equiv (-1)^n + (-1)^{n-1}ny \pmod{y^2}$.

3. Hesitation occurs at "but reduction mod $(x+1)^2$ gives $n(x+1)$, not n", and at "what about even n?". It is reinforced away by one more reduction (mod x) and by the observation that the rules are invariant under multiplying everything by 2.

4. The trap is treating 2009 as a resource to spend (binary expansion, digit-by-digit building) rather than as the statement that the length is bounded. A secondary trap is diving into primitive roots and discrete logarithms, which works but needs facts one may not be able to prove in the room.

5. Recognise it faster by asking of any bounded-step construction: *which slot in the rules accepts an arbitrary integer at no cost?* Put the unbounded parameter there, then look for the operation that brings it back out.

6. Reusable decision rule: *to extract an exponent n from a power, expand around a nearby point and keep the first-order term — $(1+y)^n \equiv 1 + ny \pmod{y^2}$ is differentiation done in integers.* The same rule shows that p^{k+1} is the exact power of an odd prime p dividing $(1+p)^{p^k} - 1$ (the lifting-the-exponent lemma), and it underlies Hensel lifting and the dual-number trick $\varepsilon^2 = 0$ for computing derivatives algebraically.