DAWN PUTNAM SESSION

Putnam 1991 — Papers A and B

Putnam 120 Archive

Contents

Putnam 1991 — Paper A **5**
1991 A1 5
1991 A2 7
1991 A3 10
1991 A4 12
1991 A5 15
1991 A6 17

Putnam 1991 — Paper B **21**
1991 B1 21
1991 B2 23
1991 B3 26
1991 B4 28
1991 B5 31
1991 B6 34

Putnam 1991 — Paper A

1991 A1

1991 A1. A 2×3 rectangle has vertices as $(0,0), (2,0), (0,3)$, and $(2,3)$. It rotates 90° clockwise about the point $(2,0)$. It then rotates 90° clockwise about the point $(5,0)$, then 90° clockwise about the point $(7,0)$, and finally, 90° clockwise about the point $(10,0)$. (The side originally on the x-axis is now back on the x-axis.) Find the area of the region above the x-axis and below the curve traced out by the point whose initial position is (1,1).

Fields: geometry, areas, rigid motions (rolling polygons) **Position:** A1 — the opener **Answer:** The area is $6 + \dfrac{7\pi}{2}$: the curve is four quarter-circle arcs, and the region splits into four quarter-disc sectors of total area $\frac{\pi}{4}(2+5+5+2) = \frac{7\pi}{2}$ plus eight right triangles of total area 6.

Caveat. No official solution for 1991 is in the archive's sources. The solution below was constructed here and checked two ways: by the sector-plus-triangle decomposition and by the polygon-plus-circular-segment decomposition (both give $6 + 7\pi/2$), and also by a numerical simulation of the four rotations (trapezoid-rule area $16.99557\ldots$ against $6 + 7\pi/2 = 16.99557\ldots$). The region is read, as intended, as the one bounded by the curve, the x-axis and the vertical segments $x = 1$ and $x = 11$ joining the curve's endpoints to the axis.

Layer 1 THE WHAT — surface analysis

The natural attempt. You track the point, arc by arc, and integrate. A clockwise quarter-turn about $C = (c,0)$ sends the displacement (u,v) from C to $(v,-u)$.

Turn 1, about $(2,0)$. $P = (1,1)$ has displacement $(-1,1)$, which goes to $(1,1)$; so P moves to $(3,1)$ along the circle $(x-2)^2 + y^2 = 2$. The rectangle becomes $[2,5] \times [0,2]$.

Turn 2, about $(5,0)$. Displacement $(-2,1) \mapsto (1,2)$, so $P \to (6,2)$ along $(x-5)^2 + y^2 = 5$. Rectangle: $[5,7] \times [0,3]$.

Turn 3, about $(7,0)$. Displacement $(-1,2) \mapsto (2,1)$, so $P \to (9,1)$ along $(x-7)^2 + y^2 = 5$. Rectangle: $[7,10] \times [0,2]$.

Turn 4, about $(10,0)$. Displacement $(-1,1) \mapsto (1,1)$, so $P \to (11,1)$ along $(x-10)^2 + y^2 = 2$. Rectangle: $[10,12] \times [0,3]$, the original translated by 10, as the problem promises.

Then the area is

$$\int_1^3 \sqrt{2-(x-2)^2}\,dx + \int_3^6 \sqrt{5-(x-5)^2}\,dx + \int_6^9 \sqrt{5-(x-7)^2}\,dx + \int_9^{11} \sqrt{2-(x-10)^2}\,dx.$$

This works. But each integral needs $x = c + r\sin\theta$, the $\cos^2$ double-angle step, and then limits like $\arcsin(-2/\sqrt{5})$ and $\arcsin(1/\sqrt{5})$, whose difference you must then recognise as $\pi/2$. Four of those in an exam is where arithmetic slips live, and at the end you still have to be sure the arcsines combine.

The model chain. The tracking above is necessary and cheap; keep it. Then stop integrating. Each piece of the curve is a quarter-circle arc centred at a pivot C on the x-axis, from a point U to a point V, with U left of C and V right of C (the turn is clockwise and the arc stays above the axis). The region under that arc, between the verticals through U and V, is cut by the segments CU and CV into three pieces:

- the sector CUV, a quarter of a disc of radius $r = |CU|$, area $\frac{\pi}{4}r^2$;
- the right triangle with vertices C, U and the foot of U on the axis;
- the right triangle with vertices C, V and the foot of V.

Now read off:

pivot	U	V	r^2	sector	triangles
$(2,0)$	$(1,1)$	$(3,1)$	2	$\pi/2$	$\frac{1}{2}+\frac{1}{2}$
$(5,0)$	$(3,1)$	$(6,2)$	5	$5\pi/4$	$1+1$
$(7,0)$	$(6,2)$	$(9,1)$	5	$5\pi/4$	$1+1$
$(10,0)$	$(9,1)$	$(11,1)$	2	$\pi/2$	$\frac{1}{2}+\frac{1}{2}$

(For the second row: the triangle at U has legs $5-3=2$ and 1, the one at V has legs $6-5=1$ and 2.) The pieces tile the region without overlap, because the curve is traced left to right (each arc sweeps angles strictly inside $(0,\pi)$) and consecutive arcs meet at the shared points $(3,1)$, $(6,2)$, $(9,1)$. Summing,

$$\text{Area} = \frac{\pi}{4}(2+5+5+2) + (1+2+2+1) = \frac{7\pi}{2} + 6.$$

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree on the tracking: pivots, radii, the four positions $(3,1)$, $(6,2)$, $(9,1)$, $(11,1)$. They split at the line "area under a circular arc $= \int \sqrt{r^2-(x-c)^2}\,dx$". The natural attempt writes the integral; the model draws the two radii to the ends of the arc and sees a sector and two triangles.

> *Every arc is a quarter-turn about a point I know. A quarter-turn sweeps a quarter-disc. So the curved part of the area is a sector, $\frac{\pi}{4}r^2$, and everything else is straight lines — I should cut along the radii, not integrate.*

> *The radii squared are $|PV|^2$ for the vertices V of the rectangle, $2,5,5,2$. The π part is $\frac{\pi}{4}$ times the sum of squared distances from the point to the four corners. The leftover must be polygonal, and 6 is the area of the rectangle; I should check whether that is a coincidence.*

Layer 3 THE WHY

Why is cutting along the radii the natural move rather than a trick? Because of what a rotation is. A rotation about C keeps every point at a fixed distance from C — that is the definition — so the trajectory is a circle centred at C, and the centre is the object of interest in exactly the sense of the Blueprint's circle example. Once the centre is the object of interest, the segments from the centre to the ends of the arc are the lines to draw. And the angle is given to you: every turn is 90°, so every sector is a quarter-disc. The integral $\int \sqrt{r^2-(x-c)^2}\,dx$ throws this away; it re-derives, through a trigonometric substitution, the area of a sector you already know, and it pays for that with arcsine bookkeeping. That is zombie maths: the most primal method applied without first asking what the motion already hands you.

The second critical thought is where the reusable content lies. Look at the triangles. Before a turn about a vertex C, the point P is at height h above the side lying on the axis and at horizontal distance a from C; the triangle is $\frac{1}{2}ah$. That side is on the axis for two consecutive half-steps — it lands at the end of one turn and is lifted at the start of the next — and P's foot splits it into lengths a and b with $a+b$ equal to the side. The two triangles belonging to that side total $\frac{1}{2}(a+b)h$, which is the area of the triangle formed by P and that side. (For the bottom side, the two halves are the first triangle of turn 1 and the last triangle of turn 4, one at each end of the cycle.) The four triangles on P and the four sides tile the rectangle. So for any interior point, one full roll of a rectangle gives

$$\text{Area under the trace} \;=\; \text{Area of rectangle} \;+\; \frac{\pi}{4} \sum_{\text{vertices } V} |PV|^2.$$

Here $6+\frac{\pi}{4}(2+2+5+5)$. Nothing about $(1,1)$ or 2×3 was special; the 6 is not a coincidence. For a convex polygon rolled through one full turn the same argument works, with each sector's angle equal to the exterior angle at that vertex.

A cheap cross-check is worth having in the arena: join the points $(1,1),(3,1),(6,2),(9,1),(11,1)$ by chords. The area under the chords is a sum of trapezoids, $2+4.5+4.5+2=13$; each circular segment is a sector minus the isosceles right triangle CUV of area $\frac{1}{2}r^2$, giving $\frac{\pi}{4}\cdot 14-\frac{1}{2}\cdot 14=\frac{7\pi}{2}-7$. Total $6+\frac{7\pi}{2}$ again.

Layer $+\alpha$ BACKTESTING

1. The first cue is "rotates 90° … about the point". Write down, for each turn, the pivot, the displacement of P from it, and its image under $(u,v)\mapsto(v,-u)$.

2. It suggests that the curve is a chain of quarter-circle arcs with known centres, and that the area should be assembled from sectors ($\frac{\pi}{4}r^2$) and polygons, not integrated.

3. Hesitation comes at the first area: the hand reaches for $\int \sqrt{r^2-(x-c)^2}\,dx$. The reinforcement is to draw the two radii from the pivot to the ends of the arc before writing any integral sign.

4. The trap is twofold: the integration route (four trig substitutions with awkward arcsine limits) and, on the tracking side, rotating the wrong way or about the wrong pivot. The problem's own remark that the original bottom side ends on the axis is a free check: the final rectangle must be $[10,12]\times[0,3]$ and P must end at $(11,1)$.

5. Any trajectory generated by rotations about known centres (rolling polygons, linkages, reflected paths) is a union of circular arcs, and areas swept by it are sectors plus polygons.

6. Reusable rule: when a curve is made of circular arcs whose centres and angles are known, never integrate the arc — cut along the radii to the endpoints, count sectors as $\frac{\theta}{2}r^2$, and handle the rest as polygons.

▷ **Cue:** a figure rotates about a given point and you must follow a marked point ⟶ **Move:** write the displacement from the pivot and apply $(u,v)\mapsto(v,-u)$ for a clockwise quarter-turn; check the final position against any invariant the problem states

▷ **Cue:** area under a curve made of circular arcs with known centres ⟶ **Move:** draw radii to the ends of each arc; area = sectors + triangles (or chords' polygon + circular segments)

▷ **Cue:** a polygon rolls through a full turn and an area under a traced point is requested ⟶ **Move:** area = polygon's area $+\sum_V \frac{\theta_V}{2}|PV|^2$ with θ_V the exterior angle at V

1991 A2

1991 A2. Let $\mathbf{A}$ and $\mathbf{B}$ be different $n\times n$ matrices with real entries. If $\mathbf{A}^3=\mathbf{B}^3$ and $\mathbf{A}^2\mathbf{B}=\mathbf{B}^2\mathbf{A}$, can $\mathbf{A}^2+\mathbf{B}^2$ be invertible?

Fields: linear algebra, non-commutative algebra **Position:** A2 — early, short once seen **Answer:** No: $(\mathbf{A}^2+\mathbf{B}^2)(\mathbf{A}-\mathbf{B}) = \mathbf{A}^3-\mathbf{A}^2\mathbf{B}+\mathbf{B}^2\mathbf{A}-\mathbf{B}^3 = \mathbf{0}$ with $\mathbf{A}-\mathbf{B} \neq \mathbf{0}$, so $\mathbf{A}^2+\mathbf{B}^2$ is singular.

Caveat. No official solution for 1991 is in the archive's sources. The one-line identity below was constructed and verified here by direct expansion; the example showing the hypotheses can hold ($\mathbf{A} = \mathbf{0}$, $\mathbf{B}$ a nonzero nilpotent with $\mathbf{B}^3 = \mathbf{0}$) was also checked numerically.

Notation: write A, B for $\mathbf{A}, \mathbf{B}$, I for the identity and 0 for the zero matrix.

Layer 1 THE WHAT — surface analysis

The natural attempt. You see $A^3 = B^3$ and think of numbers: for real a, b, $a^3 = b^3$ forces $a = b$ because $a^3 - b^3 = (a-b)(a^2+ab+b^2)$ and the second factor vanishes only at $a = b = 0$. So you try to copy it:

$$A^3 - B^3 \stackrel{?}{=} (A-B)(A^2+AB+B^2).$$

Expanding the right-hand side gives $A^3 + A^2B + AB^2 - BA^2 - BAB - B^3$. The four middle terms cancel only if A and B commute, and nothing says they do. So you look at the second hypothesis, $A^2B = B^2A$, and try to force commutativity out of it. You can't: it is one relation between two cubic words, and it does not imply $AB = BA$.

Next you try determinants: $\det(A)^3 = \det(B)^3$ gives $\det A = \det B$. True, and useless — it says nothing about $A^2 + B^2$.

Then you try examples, looking for an invertible $A^2 + B^2$. Take $A = 0$: the hypotheses become $B^3 = 0$ and $0 = 0$, so any nonzero B with $B^3 = 0$ works, e.g. the 3×3 shift $B = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$. Then $A^2 + B^2 = B^2$ is singular. Every example you build comes out singular, which suggests the answer is no, but examples are not a proof, and you are now fifteen minutes in with the scalar factorisation in pieces.

The model chain. Answer: no.

The hypotheses give you four cubic words: A^3, B^3, A^2B, B^2A. Look for a product of $A^2 + B^2$ with something of degree one whose expansion contains exactly those words. Multiplying on the right by $A - B$:

$$(A^2+B^2)(A-B) = A^3 - A^2B + B^2A - B^3 = (A^3 - B^3) - (A^2B - B^2A) = 0 - 0 = 0.$$

Now $A \neq B$, so $A - B \neq 0$; choose a column v of $A - B$ that is not the zero vector. The identity above says $(A^2 + B^2)v = 0$ with $v \neq 0$, so $A^2 + B^2$ has a nontrivial kernel and is not invertible. (Equivalently: if $A^2 + B^2$ were invertible, multiply the identity on the left by its inverse to get $A - B = 0$, contradicting $A \neq B$.)

The question is not vacuous: $A = 0$ and B the shift matrix above are different, satisfy both hypotheses, and give a singular $A^2 + B^2$, as the theorem says they must.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start from $A^3 = B^3$ and the scalar instinct "factor the difference of cubes". They split at the choice of which factorisation to attempt. The natural attempt insists on $(A-B)(A^2+AB+B^2)$ — the factorisation of the scalar identity — and dies on commutativity. The model reverses the question: it starts from the object asked about, $A^2 + B^2$, and asks what to multiply it by so that only the given words appear.

> *I am asked about $A^2 + B^2$, not about $A - B$. If it is singular, the witness should be $(A^2 + B^2)X = 0$ for some nonzero X, and the candidate for a nonzero X is $A - B$, since "different" is the only nonzero thing I am given.*

> *The hypotheses put the squares on the left: A^2B and B^2A. So the square factor must be on the left and the linear factor on the right: $(A^2 + B^2)(A - B)$, not $(A - B)(A^2 + B^2)$.*

Layer 3 THE WHY

Why is this the natural move and not a rabbit from a hat? Start from what invertibility means. A square matrix M fails to be invertible exactly when $Mv = 0$ for some $v \neq 0$, equivalently when $MX = 0$ for some nonzero matrix X. So to answer "can M be invertible?" with no, you need a nonzero annihilator. In this problem the only nonzero object guaranteed by the hypotheses is $A - B$ — the word "different" in the statement is there to supply it. So the problem is telling you to compute $(A^2 + B^2)(A - B)$. The phrasing is the cue.

Then the order of the factors is not a choice either. In a non-commutative ring, the words that can appear are fixed by where each letter sits. $(A^2 + B^2)(A - B)$ expands to $A^3 - A^2B + B^2A - B^3$: squares first, linear letter last — exactly the shape of the given $A^2B = B^2A$. The other order, $(A - B)(A^2 + B^2) = A^3 + AB^2 - BA^2 - B^3$, produces AB^2 and BA^2, about which you know nothing. So the hypotheses are shaped to match one product and only one.

The scalar factorisation is the trap because it smuggles in commutativity. $a^3 - b^3 = (a - b)(a^2 + ab + b^2)$ uses $ab = ba$ in the middle terms; for matrices, $(A - B)(A^2 + AB + B^2)$ has six terms, not two. The deeper lesson: with matrices, write every product as a word and ask which words you have been given. The hypotheses are a list of available words; the target is to find a product whose expansion uses only those words.

A final remark on why the problem feels harder than it is: the scalar picture says $a^2 + b^2$ is invertible whenever a and b are real and not both zero, which primes you to expect "yes". The matrix answer is "no", not because $A^2 + B^2$ has some hidden sign problem but because the relations force it to kill a nonzero matrix.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the word "different" together with "can … be invertible?". Write down: "not invertible $\iff$ $MX = 0$ for some $X \neq 0$; the given nonzero matrix is $A - B$."

2. It suggests expanding $(A^2 + B^2)(A - B)$ and checking it against the hypotheses.

3. Hesitation occurs while trying to salvage $(A - B)(A^2 + AB + B^2)$ or trying to deduce $AB = BA$. The reinforcement: list the words the hypotheses mention (A^3, B^3, A^2B, B^2A) and only multiply things whose expansion stays inside that list.

4. The trap is the scalar difference-of-cubes factorisation, which silently assumes commutativity, and the determinant detour, which is true but irrelevant.

5. Whenever a matrix problem gives polynomial relations and asks about invertibility, look for a product (target)$\times$(known nonzero thing) that the relations make zero; the order of factors is dictated by the order of letters in the relations.

6. Reusable rule: to show M is singular, find a nonzero X with $MX = 0$; take X to be the one nonzero object the hypotheses give you, and in a non-commutative setting choose the side of multiplication so that the expansion produces only words the hypotheses control.

▷ **Cue:** "can this matrix be invertible?" and the hypotheses hand you a nonzero matrix (e.g. $A \neq B$) $\longrightarrow$ **Move:** multiply the target by that nonzero matrix and try to show the product is 0

▷ **Cue:** matrix relations such as $A^2B = B^2A$ with the squares on the left $\longrightarrow$ **Move:** put the square factor on the left and the linear factor on the right: expand $(A^2 + B^2)(A - B)$

▷ **Cue:** a scalar identity like $a^3 - b^3 = (a-b)(a^2+ab+b^2)$ is about to be used on matrices
⟶ **Move:** expand as words first; if the cross terms need $AB = BA$, abandon it and list the words the hypotheses control

1991 A3

1991 A3. Find all real polynomials $p(x)$ of degree $n \geq 2$ for which there exist real numbers $r_1 < r_2 < \cdots < r_n$ such that

1. $p(r_i) = 0, \qquad i = 1, 2, \ldots, n$, and
2. $p'\left(\frac{r_i + r_{i+1}}{2}\right) = 0 \qquad i = 1, 2, \ldots, n-1,$

where $p'(x)$ denotes the derivative of $p(x)$.

Fields: algebra, real polynomials, logarithmic derivative **Position:** A3 — middle of the first half
Answer: Exactly the quadratics with two distinct real roots, $p(x) = a(x-r_1)(x-r_2)$ with $a \neq 0$ and $r_1 < r_2$. For $n \geq 3$ none exist: at the midpoint of r_1, r_2 the terms of $p'/p = \sum_j 1/(x-r_j)$ from r_1, r_2 cancel and every other term is negative.

Caveat. No official solution for 1991 is in the archive's sources. The proof below was constructed and checked here line by line; the $n = 3$ spot-check ($x^3 - x$, whose critical points are $\pm 1/\sqrt{3}$, not $\pm\frac{1}{2}$) agrees with it.

Layer 1 THE WHAT — surface analysis

The natural attempt. You start with Rolle's theorem. p has the n distinct zeros $r_1 < \cdots < r_n$, so p' has a zero strictly inside each (r_i, r_{i+1}); that is $n-1$ zeros, and p' has degree $n-1$, so these are all of them. Condition 2 says each of them sits exactly at a midpoint. This is a clean restatement, and it tells you nothing yet about whether such p exist.

So you test. For $n = 2$, $p = a(x-r_1)(x-r_2)$, $p' = a(2x - r_1 - r_2)$, which vanishes at the midpoint. Every quadratic with distinct real roots works.

For $n = 3$ you reach for the most symmetric example, roots $-1, 0, 1$: $p = x^3 - x$, $p' = 3x^2 - 1$, zeros at $\pm 1/\sqrt{3} \approx \pm 0.577$. The midpoints are $\pm\frac{1}{2}$. Fails. So you try to adjust: roots $0, s, t$, $p = x(x-s)(x-t) = x^3 - (s+t)x^2 + stx$, $p' = 3x^2 - 2(s+t)x + st$, and impose $p'(s/2) = 0$ and $p'((s+t)/2) = 0$. The first gives $\frac{3}{4}s^2 - s(s+t) + st = 0$, i.e. $-\frac{1}{4}s^2 = 0$, so $s = 0$, contradicting $0 < s$. So $n = 3$ is impossible — but you got there by a computation that gives no hint of how to do $n = 4$ or general n, where the same approach means solving $n-1$ polynomial equations in n unknowns. The wall is the general case: a pattern seen at $n = 3$ is not a proof for all n.

(Look closely at that $n = 3$ computation, though. The condition at the *first* midpoint alone killed it. That is the clue the model follows.)

The model chain. Since p has degree n and the n distinct numbers $r_1, \ldots, r_n$ are zeros, $p(x) = a\prod_{j=1}^{n}(x - r_j)$ with $a \neq 0$. At any x that is not a root, the product rule gives

$$\frac{p'(x)}{p(x)} = \sum_{j=1}^{n} \frac{1}{x - r_j}. \tag{$*$}$$

Let $m = \frac{r_1 + r_2}{2}$. It lies strictly between r_1 and r_2, so it is not a root, $p(m) \neq 0$, and $p'(m) = 0$ holds if and only if the right side of $(*)$ is 0 at $x = m$. With $d = r_2 - r_1 > 0$,

$$\frac{1}{m - r_1} + \frac{1}{m - r_2} = \frac{2}{d} - \frac{2}{d} = 0, \qquad \text{so} \qquad \sum_{j=1}^{n} \frac{1}{m - r_j} = \sum_{j=3}^{n} \frac{1}{m - r_j}.$$

If $n \geq 3$, every r_j with $j \geq 3$ satisfies $r_j > r_2 > m$, so each term $\frac{1}{m-r_j}$ is negative and the sum is strictly negative. Hence $p'(m) \neq 0$, and condition 2 fails at $i = 1$. No polynomial of degree $n \geq 3$ qualifies.

If $n = 2$, the sum at m is exactly 0, so $p'(m) = 0$; condition 2 has only the one instance $i = 1$. So every $p(x) = a(x - r_1)(x - r_2)$ with $a \neq 0$, $r_1 < r_2$ qualifies — equivalently, every real quadratic with positive discriminant.

Answer: the real quadratics with two distinct real roots, and nothing of higher degree.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree that $p = a\prod(x - r_j)$ and that condition 2 is about where p' vanishes. They split at how to evaluate p' at a midpoint. The natural attempt expands p' as a polynomial in x with coefficients in the r_j and substitutes. The model divides by p, so that $p'(m) = 0$ becomes a sum of n separate reciprocals, one per root, whose signs can be read off.

> *p is given to me as a product over its roots, and I need information about p'. Differentiating a product is ugly; differentiating its logarithm turns it into a sum, one term per root. So I look at $p'/p = \sum 1/(x - r_j)$.*

> *At the midpoint of r_1 and r_2 those two terms cancel exactly, because the midpoint is equidistant from them. Everything else is to the right, so everything else pulls the same way. The sum cannot be zero unless there is nothing else.*

Layer 3 THE WHY

Why the logarithmic derivative, and why is it natural rather than a known trick you had to have seen? Because of the form the hypotheses force on p. The problem hands you all n roots, so p is a product, and the question is about the derivative of that product. The derivative of a product of n factors is a sum of n products of $n - 1$ factors — unreadable. Divide by p and each of those products collapses to a single reciprocal $1/(x - r_j)$. This is the same reason ln is used to differentiate x^x or a long quotient: logarithms turn products into sums, and a sum of simple terms is what you can control term by term. It is the A-level logarithmic differentiation you already know, applied to a polynomial.

Then why look at the *first* midpoint, not a middle one? Because of signs. At a midpoint m_i of r_i, r_{i+1}, the two nearest terms cancel by equidistance — the midpoint is precisely the point where $\frac{1}{x-r_i}$ and $\frac{1}{x-r_{i+1}}$ are equal and opposite. The remaining terms have sign determined by which side of m_i each root lies: roots to the left contribute positive terms, roots to the right negative ones. At an interior midpoint both kinds appear and could balance. At the first midpoint there are no roots to the left of r_1, so every surviving term has the same sign. An extreme position is where a sign argument works; that is why you choose it. (The last midpoint works equally well, by the mirror-image argument: every surviving term is positive.)

There is a physical reading that makes all this feel inevitable. Put a unit charge at each r_j repelling with force 1/distance on the line; the force at x is $\sum_j \frac{1}{x-r_j}$, and the zeros of p' between roots are the equilibrium points. At the midpoint of the first two charges, their forces cancel, and every other charge is on the right pushing leftwards. Equilibrium is impossible unless there are no other charges. With $n = 2$ the midpoint is the equilibrium by symmetry, which is the familiar fact that a parabola's vertex is halfway between its roots.

Finally, Rolle's theorem in the natural attempt was true and correct and led nowhere. It locates the zeros of p' in intervals; the question needed their exact positions, and a location theorem cannot give an exact position. Knowing when a true theorem is the wrong tool is part of the skill.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that all n roots of a degree-n polynomial are named. Write down $p(x) = a\prod_j (x - r_j)$ and, since p' is asked about, $p'/p = \sum_j \frac{1}{x-r_j}$.

2. It suggests evaluating the logarithmic derivative at the midpoints and reading off signs, starting at an extreme midpoint.

3. Hesitation happens after the $n = 2$ and $n = 3$ cases, when the expanded-coefficient approach visibly will not scale. The reinforcement: the $n = 3$ failure came from the first midpoint alone — look at the first midpoint for general n.

4. The trap is expanding p' in coefficients (elementary symmetric functions of the roots) and solving a nonlinear system; and the smaller trap of forgetting that $n = 2$ is a genuine, infinite family of answers, not a degenerate case to be discarded.

5. Any condition on p' (or on critical points) for a polynomial given by its roots should be converted to $\sum 1/(x - r_j) = 0$; symmetric positions give cancellations and extreme positions give one-signed sums.

6. Reusable rule: when a product is given and a derivative is asked about, take the logarithmic derivative to make it a sum; then test the condition at an extreme position, where all surviving terms share a sign.

▷ **Cue:** a polynomial given by all its real roots, with a condition on p' $\longrightarrow$ **Move:** write $p'/p = \sum_j 1/(x - r_j)$ and evaluate the sum instead of expanding p'

▷ **Cue:** a condition must hold at each of several positions (every midpoint, every gap) $\longrightarrow$ **Move:** check it first at an extreme position, where the contributions all have one sign

▷ **Cue:** a point is equidistant from two roots $\longrightarrow$ **Move:** their reciprocal terms $1/(x - r_i)$ cancel exactly; only the other roots remain

1991 A4

1991 A4. Does there exist an infinite sequence of closed discs $D_1, D_2, D_3, \ldots$ in the plane, with centers $c_1, c_2, c_3, \ldots$, respectively, such that

1. the c_i have no limit point in the finite plane,
2. the sum of the areas of the D_i is finite, and
3. every line in the plane intersects at least one of the D_i?

Fields: geometry (plane, lines and discs); analysis (convergent series); construction **Position:** A4 — mid-paper; the difficulty is entirely in deciding the answer, after which the construction is short. **Answer:** Yes. On the circle of radius k about the origin ($k = 1, 2, 3, \ldots$) place $\lceil \pi k^4 \rceil$ equally spaced discs of radius k^{-3}; they cover that circle, every line meets some such circle, and the total area is at most $\sum_k (\pi^2 k^{-2} + \pi k^{-6}) < \infty$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You read three conditions and feel them pulling against each other. Finite total area means the radii r_i tend to 0 fast; no limit point means the discs march off to infinity; and yet every line — including lines far from the origin, in every direction — has to be caught. Your instinct says "no", and you look for an obstruction.

The first obstruction you try is area-based: "lines are everywhere, so the discs must cover a positive proportion of the plane, which needs infinite area". But a line has zero area, and a family of lines can be caught by a set of small area — a single disc of radius 1 catches every line through it. So the area argument does not even get started.

The second obstruction is a projection. Fix a direction, say vertical lines $x = t$. The line $x = t$ meets D_i exactly when t lies in the projection of D_i onto the x-axis, an interval of length $2r_i$. So the intervals of lengths $2r_i$ must cover the whole real line, which forces

$$\sum_i r_i = \infty.$$

You write this down, hoping it contradicts $\sum_i \pi r_i^2 < \infty$. It does not: $r_i = 1/i$ satisfies both. The obstruction you were hunting for does not exist, and you have spent the time finding that out.

The model chain. *Answer: yes.* Let C_k denote the circle of radius k centred at the origin O, for $k = 1, 2, 3, \ldots$. Set

$$r_k = \frac{1}{k^3}, \qquad N_k = \left\lceil \pi k^4 \right\rceil,$$

and on C_k place N_k points $P_{k,1}, \ldots, P_{k,N_k}$ equally spaced, at angles $2\pi t/N_k$. Let the discs be the closed discs of radius r_k centred at these points, listed ring by ring as one sequence.

Each circle is covered. Take a point Q on C_k. Some centre $P_{k,t}$ differs from Q in angle by $\delta \le \pi/N_k$. The chord length is

$$|QP_{k,t}| = 2k \sin(\delta/2) \le k\delta \le \frac{\pi k}{N_k} \le \frac{\pi k}{\pi k^4} = r_k,$$

so Q lies in that disc. Hence $C_k \subseteq \bigcup_t D_{k,t}$.

Condition 3. A line ℓ lies at some finite distance d from O. Choose an integer $k > d$. The foot of the perpendicular from O to ℓ is inside C_k, and ℓ is unbounded, so ℓ meets C_k (indeed in two points). That point of C_k lies in one of the discs, so ℓ meets a disc.

Condition 2. The discs on ring k have total area

$$N_k \, \pi r_k^2 \le (\pi k^4 + 1)\, \pi k^{-6} = \pi^2 k^{-2} + \pi k^{-6},$$

and $\sum_k (\pi^2 k^{-2} + \pi k^{-6})$ converges.

Condition 1. A bounded set lies within some radius R of O, and contains only centres from rings $k \le R$: finitely many rings, each with finitely many centres. So no bounded region contains infinitely many centres, and there is no finite limit point. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains reach the projection inequality $\sum r_i = \infty$. The natural attempt reads it as a failed obstruction and keeps looking for a better one. The model reads it the other way round: the *only* constraint found so far is on $\sum r_i$, while the hypothesis constrains $\sum r_i^2$, and a sequence can make the first diverge while the second converges. That gap is the answer "yes".

> *Catching lines costs length, not area — the projection of a disc has length $2r$. The budget is on r^2. When r is small, r^2 is far smaller than r, so I can buy as much length as I like for as little area as I like. The answer is yes; now I need a shape made of tiny discs that no line can avoid.*

> *Which curve does every line eventually cross? A line at distance d from the origin crosses every circle about the origin of radius bigger than d. So I cover the circles of radius $1, 2, 3, \ldots$ with tiny discs.*

Layer 3 THE WHY

There is no official solution in the archive's sources for 1991; the construction above is my own and every step has been checked, including the chord estimate and the area sum.

Why is "yes" the natural answer rather than a surprise? Go to what "a line meets a disc" means. A line ℓ meets the closed disc of centre c and radius r exactly when the distance from c

to ℓ is at most r. That condition is linear in r. Among all lines of a fixed direction, those meeting the disc form a strip of width $2r$. So the "size" of a disc, as seen by lines, is its width, not its area. The problem's second condition controls area, πr^2; its third condition needs width, $2r$. Two different powers of r are being compared, and for small r the lower power dominates. Once you see that the hypothesis and the requirement are measured in different units, the answer is decided: tiny discs are cheap in the currency you are charged and valuable in the currency you need.

That is the whole mechanism, and it deserves stating as an inequality. On ring k the covering needs total width about $2\pi k$ (the circumference), bought with N_k discs of radius r_k, so $N_k r_k \approx \pi k$. The area is $N_k \pi r_k^2 \approx \pi^2 k\, r_k$. Width is fixed by geometry; area is width times radius. Shrinking the radius sends area to zero at fixed width. Any r_k with $\sum k\, r_k < \infty$ works; $r_k = k^{-3}$ is a convenient choice.

Why circles? Because condition 3 quantifies over *every* line, and the cleanest way to catch all lines is a family of curves each of which is crossed by a large, explicit set of lines. A closed curve surrounding the origin is crossed by every line that passes inside it, and "passes inside a circle of radius k" is the statement "distance from O is less than k" — a condition that every line satisfies for large enough k. The circle is the set of points at distance k from the centre, so distance from the centre is the natural parameter for lines too. (A square of side $2k$ would work equally well; the circle only makes the covering estimate one line.)

And why does condition 1 not bite? Because it forbids accumulation, not number. Each ring is finite and the rings go to infinity, so the centres are locally finite. The condition is there to forbid the cheap cheat of piling infinitely many discs near one point, which would not help anyway.

Finally, the projection argument from the natural attempt is not wasted: it is the necessary condition $\sum r_i = \infty$, and it tells you exactly how the construction must look — divergent total radius, convergent total squared radius. Reading a necessary condition as a design specification, rather than as a failed contradiction, is the move.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the pairing of "sum of the areas … finite" with "every line … intersects". On seeing it, write down what a disc costs (πr^2) and what it buys against lines of one direction (a strip of width $2r$).

2. Compare the two powers of r: the requirement is on $\sum r$, the budget on $\sum r^2$. That suggests many small discs — a construction, not an impossibility proof.

3. Hesitation occurs at the answer itself: condition 1 and condition 2 both sound like smallness conditions, and the instinct says "no". Reinforce it away by testing the one necessary condition you can derive, $\sum r_i = \infty$, against the hypothesis. $r_i = 1/i$ satisfies both; there is no conflict, so stop hunting for one.

4. The trap is spending the time on an impossibility proof because the conditions "feel" incompatible, and the secondary trap is the area heuristic, which confuses how much of the plane is covered with how many lines are caught.

5. Slogan: *lines see width, not area.* Whenever a family of thin objects must be hit by a family of lower-dimensional objects, the relevant size is the one the hitting family sees.

6. Decision rule: in a "does there exist" problem, derive the necessary condition first. If it involves a different quantity from the one the hypothesis bounds, and the two can be decoupled by scaling (here r versus r^2), the answer is yes, and the construction is "make the objects small and numerous".

▷ **Cue:** finite total area but every line must be hit $\longrightarrow$ **Move:** compare width $2r$ (what lines see) with area πr^2 (what is budgeted); many tiny discs win

▷ **Cue:** a condition quantified over every line in the plane $\longrightarrow$ **Move:** use nested closed curves (circles of radius k) about a fixed point; a line at distance d crosses all with $k > d$

▷ **Cue:** need a curve covered by small discs of radius r $\longrightarrow$ **Move:** space centres along the curve at arc-gap at most $2r$

▷ **Cue:** a necessary condition that fails to contradict the hypothesis $\longrightarrow$ **Move:** treat it as the design specification for a construction

1991 A5

1991 A5. Find the maximum value of

$$\int_0^y \sqrt{x^4 + (y - y^2)^2}\, dx$$

for $0 \le y \le 1$.

Fields: analysis (definite integrals with a parameter); inequalities **Position:** A5 — late in the paper, but the difficulty is a choice of route, not depth: the direct route is long, the right one is four lines. **Answer:** $\frac{1}{3}$, attained only at $y = 1$, where the integrand becomes x^2.

Layer 1 THE WHAT — surface analysis

The natural attempt. You see "maximum value … for $0 \le y \le 1$" and the parameter y sits in two places: the upper limit and inside the root. The reflex is calculus in y. Write $c = y - y^2$ and

$$I(y) = \int_0^y \sqrt{x^4 + c^2}\, dx.$$

Differentiate with respect to y: one term from the moving upper limit, one from differentiating under the integral sign, with $c' = 1 - 2y$:

$$I'(y) = \sqrt{y^4 + c^2} + c(1 - 2y) \int_0^y \frac{dx}{\sqrt{x^4 + c^2}}.$$

Now you have to decide the sign of this. The integral left over is elliptic — there is no closed form — so you cannot compute I' and solve $I'(y) = 0$. For $y \le \frac{1}{2}$ both terms are non-negative, fine. For $y > \frac{1}{2}$ the second term is negative and you need a bound on the integral. Bounding $\sqrt{x^4 + c^2} \ge c$ gives $\int_0^y dx/\sqrt{x^4 + c^2} \le y/c$, so the negative term is at least $-(2y - 1)y$, and you must show

$$\sqrt{y^4 + (y - y^2)^2} > y(2y - 1),$$

which after squaring and dividing by y^2 reads $2y^2 - 2y + 1 > 4y^2 - 4y + 1$, i.e. $2y(1 - y) > 0$. True on $(0, 1)$. So I is increasing and the maximum is $I(1) = \int_0^1 x^2\, dx = \frac{1}{3}$.

This works. It is also fifteen minutes of exam: you had to justify differentiating an integral whose limit and integrand both move, handle two sign cases, find a bound for an integral you cannot evaluate, and square an inequality (checking both sides are positive first). Every one of those is a place to lose marks.

The model chain. For $u, v \ge 0$,

$$\sqrt{u^2 + v^2} \le \sqrt{u^2 + 2uv + v^2} = u + v,$$

with equality iff $uv = 0$. Put $u = x^2$ and $v = y - y^2$, both non-negative for $0 \le x \le y \le 1$:

$$I(y) \le \int_0^y \left(x^2 + y - y^2\right) dx = \frac{y^3}{3} + y^2 - y^3 = y^2 - \frac{2}{3}y^3.$$

Let $g(y) = y^2 - \frac{2}{3}y^3$. Then $g'(y) = 2y - 2y^2 = 2y(1 - y) \ge 0$ on $[0, 1]$, so g is non-decreasing and $g(y) \le g(1) = \frac{1}{3}$. Hence $I(y) \le \frac{1}{3}$ for all $y \in [0, 1]$.

At $y = 1$ we have $y - y^2 = 0$, so $I(1) = \int_0^1 \sqrt{x^4}\,dx = \int_0^1 x^2\,dx = \frac{1}{3}$. The bound is attained, and the maximum is $\frac{1}{3}$.

(For $0 < y < 1$ the first inequality is strict, since $uv = x^2(y - y^2) > 0$ for $x \in (0, y]$; so $y = 1$ is the only maximiser. I also checked numerically that I is increasing on $[0, 1]$ with $I(1) = 0.3333\ldots$) $\square$

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start from the same object, $\int_0^y \sqrt{x^4 + (y - y^2)^2}\,dx$. At that line the natural attempt asks "how does this change with y?" and differentiates. The model asks "what is inside the root, and can I get rid of the root?" and bounds the integrand before integrating anything.

> *I can't integrate $\sqrt{x^4 + c^2}$ — nobody can in closed form. So I'm not going to compute this integral; I'm going to bound it. What's the simplest thing above a square root of a sum of two squares? The sum itself: $\sqrt{a^2 + b^2} \le a + b$. And that bound is exact when one of them is zero — and $y - y^2$ is zero at $y = 1$. So the bound is sharp exactly where I'd guess the maximum is.*

> *Check the endpoints before doing anything else: $y = 0$ gives 0, $y = 1$ kills the $(y - y^2)^2$ term and gives $\int_0^1 x^2 = \frac{1}{3}$. My guess is $\frac{1}{3}$. Now I need an upper bound that equals $\frac{1}{3}$ at $y = 1$.*

Layer 3 THE WHY

The MAA volume for 1985–2000 (project doc `Mathematical_Competition_1985_2000_Probl.pdf`) gives both routes: the derivative argument as Solution 1 and the $\sqrt{u^2 + v^2} \le u + v$ bound as Solution 2. I solved and checked both independently before comparing, and both agree with the answer $\frac{1}{3}$. Solution 2 opens with "If $u, v \ge 0$, then $\sqrt{u^2 + v^2} \le u + v$" and gives no reason for choosing that inequality. Here is where it comes from.

Start with what the question asks: a *maximum*. A maximum is proved by two things — a bound valid everywhere, and one point where the bound is attained. Nothing in that definition asks you to know the function exactly. The natural attempt quietly assumes that to maximise I you must know I, or at least I'. You don't. You need an upper bound that is sharp somewhere.

Next, the integrand. $\sqrt{x^4 + (y - y^2)^2}$ is the length of a vector with components x^2 and $y - y^2$ — a hypotenuse. The legs of a right triangle add up to at least the hypotenuse (the triangle inequality: going along two sides is at least as long as the direct side). So $\sqrt{u^2 + v^2} \le u + v$ is not an algebraic trick; it is the statement that a straight path is shortest. It is also the cheapest upper bound available, because it removes the root and leaves a polynomial you can integrate at sight.

Third, why is the bound good enough? A bound is useless if it overshoots everywhere. The triangle inequality is an equality when the triangle collapses — when one leg is zero. Look at the legs: x^2 is zero only at $x = 0$, but $y - y^2$ is zero at $y = 0$ and $y = 1$, for every x. So at $y = 1$ the bound is exact across the whole range of integration. The problem's own phrasing plants this: $y - y^2 = y(1 - y)$ vanishes at both ends of the stated interval $0 \le y \le 1$. An expression that vanishes at the endpoints of the given range is a signal that the extremum sits at an endpoint and that something simplifies there.

Finally, after bounding, the problem is a one-variable polynomial $g(y) = y^2 - \frac{2}{3}y^3$, and the derivative $2y(1 - y)$ is again the quantity $y - y^2$ doubled — non-negative on the interval for the same reason. So the calculus you wanted to do is still done, but on a cubic rather than on an elliptic integral.

Why the derivative route is zombie maths here: it treats the integral as the object and tries to follow it exactly, including through its non-elementary part. Notice that even the derivative route had to use a bound in the end ($\sqrt{x^4 + c^2} \ge c$) to control the integral it could not evaluate.

The bound was always going to appear; the model applies it once, at the start, where it does the most work.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is a square root of a sum of two squares inside an integral you cannot do in closed form, together with a parameter range $0 \le y \le 1$ on which $y - y^2$ vanishes at both ends. Write down $I(0) = 0$ and $I(1) = \frac{1}{3}$ immediately.

2. Maximise by bound-and-attain: find an upper bound for the integrand that is exact at the candidate point, integrate the bound, and maximise the resulting elementary function.

3. Hesitation occurs at the moment you realise the integral has no antiderivative. The reflex is to differentiate in y instead. Reinforce it away with the definition: a maximum needs a bound and a point of attainment, not a formula for the function.

4. The trap is the parameter in the upper limit, which invites differentiation under the integral sign — a correct but long route with two sign cases and an elliptic integral to bound. A second trap is using $\sqrt{u^2+v^2} \le u+v$ without checking that $u, v \ge 0$; here $x^2 \ge 0$ and $y - y^2 \ge 0$ on the given range, and that must be said.

5. Slogan: *if you cannot integrate it, bound it where it is sharp.* Test the endpoints first; the endpoint that kills a term tells you where the bound must be exact.

6. Decision rule: when asked for an extremum of a quantity you cannot compute exactly, guess the extremiser from the special points (endpoints, where a term vanishes), then look for the cheapest inequality that becomes an equality there. Differentiate only the bounded, elementary expression.

▷ **Cue:** $\sqrt{u^2+v^2}$ with $u, v \ge 0$ to be bounded above $\longrightarrow$ **Move:** use $\sqrt{u^2+v^2} \le u+v$, exact when $u = 0$ or $v = 0$

▷ **Cue:** maximise an integral with no closed-form antiderivative $\longrightarrow$ **Move:** bound the integrand by an integrable function that is exact at the candidate maximiser

▷ **Cue:** a factor $y(1-y)$ in a problem restricted to $0 \le y \le 1$ $\longrightarrow$ **Move:** evaluate at $y = 0$ and $y = 1$ first; the extremum likely sits where that factor vanishes

▷ **Cue:** extremum problem ("find the maximum value") $\longrightarrow$ **Move:** prove it as upper bound everywhere plus one point of equality; do not insist on knowing the function

1991 A6

1991 A6. Let $A(n)$ denote the number of sums of positive integers

$$a_1 + a_2 + \cdots + a_r$$

which add up to n with

$$a_1 > a_2 + a_3, a_2 > a_3 + a_4, \ldots,$$
$$a_{r-2} > a_{r-1} + a_r, a_{r-1} > a_r.$$

Let $B(n)$ denote the number of $b_1 + b_2 + \cdots + b_s$ which add up to n, with

1. $b_1 \ge b_2 \ge \cdots \ge b_s$,
2. each b_i is in the sequence $1, 2, 4, \ldots, g_j, \ldots$ defined by $g_1 = 1$, $g_2 = 2$, and $g_j = g_{j-1} + g_{j-2} + 1$, and
3. if $b_1 = g_k$ then every element in $\{1, 2, 4, \ldots, g_k\}$ appears at least once as a b_i.

Prove that $A(n) = B(n)$ for each $n \ge 1$.

(For example, $A(7) = 5$ because the relevant sums are $7, 6+1, 5+2, 4+3, 4+2+1$, and $B(7) = 5$ because the relevant sums are $4+2+1, 2+2+2+1, 2+2+1+1+1, 2+1+1+1+1+1, 1+1+1+1+1+1+1$.)

Fields: combinatorics (partitions, bijections); recurrences (Fibonacci numbers) **Position:** A6 — the last problem of the morning, hard to see and short once seen. **Answer:** Set $a_{r+1} = a_{r+2} = 0$ and $d_i = a_i - a_{i+1} - a_{i+2}$. The A-conditions say exactly $d_1, \dots, d_r \geq 1$, and $\sum_i a_i = \sum_j d_j g_j$. So the A-sum with slacks $(d_1, \dots, d_r)$ corresponds to the B-sum in which g_j occurs exactly d_j times ($j = 1, \dots, r$); this is a bijection.

Layer 1 THE WHAT — surface analysis

The natural attempt. You list small cases and compare. Both sides count $1, 1, 2, 2, 3, 3, 5, 5, 7, 7, 10, 10, 13, 14, \dots$ for $n = 1, 2, \dots$, so the statement is believable, and you look for a recursion. On the A side you try to remove a_1 or a_r; on the B side you try to remove the largest part g_k or one copy of 1. Neither removal preserves the conditions cleanly: deleting a_1 leaves a valid A-sum but of an unknown total, and deleting a copy of g_k from a B-sum may destroy condition 3. You then try generating functions. The B side is fine,

$$\sum_n B(n)x^n = \sum_{k\geq 1} \prod_{j=1}^{k} \frac{x^{g_j}}{1 - x^{g_j}},$$

but the A side, with its overlapping inequalities $a_i > a_{i+1} + a_{i+2}$, has no evident product form. The attempt stalls here: the A-conditions are not independent, and nothing has made them so.

The model chain. *Step 1: uniform conditions.* Given an A-sum $a_1 + \cdots + a_r$, put $a_{r+1} = a_{r+2} = 0$. Then the full list of conditions — the displayed inequalities together with $a_r \geq 1$ — is

$$a_i > a_{i+1} + a_{i+2} \qquad (i = 1, \dots, r):$$

for $i \leq r - 2$ it is given; for $i = r - 1$ it reads $a_{r-1} > a_r$; for $i = r$ it reads $a_r > 0$.

Step 2: slack variables. Define $d_i = a_i - a_{i+1} - a_{i+2}$ for $i = 1, \dots, r$. Step 1 says: the a_i form an A-sum iff every d_i is an integer ≥ 1. Conversely, given integers $d_1, \dots, d_r \geq 1$, the relations $a_{r+1} = a_{r+2} = 0$, $a_i = d_i + a_{i+1} + a_{i+2}$ determine $a_r, a_{r-1}, \dots, a_1$ uniquely, and they are positive integers satisfying Step 1. So A-sums with r terms correspond bijectively to r-tuples of positive integers $(d_1, \dots, d_r)$.

Step 3: solve for a_i. Let $F_0 = 0$, $F_1 = F_2 = 1$, $F_t = F_{t-1} + F_{t-2}$. Claim: $a_i = \sum_{j=i}^{r} F_{j-i+1} d_j$. By downward induction: for $i = r$, $a_r = d_r = F_1 d_r$; for $i = r - 1$, $a_{r-1} = d_{r-1} + d_r$. If the claim holds for $i + 1$ and $i + 2$ then

$$a_i = d_i + \sum_{j\geq i+1} (F_{j-i} + F_{j-i-1}) d_j = \sum_{j\geq i} F_{j-i+1} d_j.$$

Step 4: the total. Summing over i and swapping the order,

$$n = \sum_{i=1}^{r} a_i = \sum_{j=1}^{r} d_j \sum_{i=1}^{j} F_{j-i+1} = \sum_{j=1}^{r} d_j \, (F_1 + \cdots + F_j).$$

Write $S_j = F_1 + \cdots + F_j$. Then $S_1 = 1$, $S_2 = 2$, and $S_j = F_1 + F_2 + \sum_{t=3}^{j} (F_{t-1} + F_{t-2}) = 1 + S_{j-1} + S_{j-2}$ (the first two terms supply $F_1 + F_2 = 1 + F_1$, and the rest regroup as $S_{j-1} - F_1$ and S_{j-2}). This is the recurrence of g_j with the same initial values, so $S_j = g_j$ and

$$n = \sum_{j=1}^{r} d_j \, g_j.$$

Step 5: read off the B-sum. A B-sum is determined by its largest part g_k and the multiplicities $m_1, \ldots, m_k$ of $g_1, \ldots, g_k$; condition 3 says each $m_j \geq 1$, and the total is $\sum_j m_j g_j$. The g_j are distinct (strictly increasing), so these data determine the B-sum and conversely. Map the A-sum with slacks $(d_1, \ldots, d_r)$ to the B-sum with $k = r$, $m_j = d_j$. By Steps 2 and 4 this is a bijection from A-sums of n to B-sums of n, so $A(n) = B(n)$. □

Check with $n = 7$: $4 + 3$ has $d = (4 - 3, 3) = (1, 3)$, giving $1 + 2 + 2 + 2$; $4 + 2 + 1$ has $d = (1, 1, 1)$, giving $4 + 2 + 1$; 7 gives seven 1s.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains look at the inequalities $a_i > a_{i+1} + a_{i+2}$. The natural attempt treats them as constraints to be respected while recursing. The model names the amount by which each inequality holds, $d_i = a_i - a_{i+1} - a_{i+2}$, and uses those slacks as the new coordinates. After that there are no constraints left except $d_i \geq 1$.

> *These inequalities overlap — each a_i appears in three of them. I can't count overlapping constraints. But I can count independent ones: if I call the slack in each inequality d_i, the conditions become "each $d_i \geq 1$", and those don't interact at all.*

> *The last two conditions look different from the rest. If I pad with $a_{r+1} = a_{r+2} = 0$, they become the same inequality, and I get exactly r slacks for r variables.*

> *The rule $g_j = g_{j-1} + g_{j-2} + 1$ is Fibonacci plus one. My a_i are built from the d's by a Fibonacci recursion. The "+1" is a partial sum. These must be the same numbers.*

Layer 3 THE WHY

There is no official solution in the archive's sources for 1991; the bijection above is my own. It has been checked algebraically and by computer for $n \leq 18$, where both counts agree $(A(n) = B(n) = 1, 1, 2, 2, 3, 3, 5, 5, 7, 7, 10, 10, 13, 14, 18, 19, 23, 25)$, and against the example $n = 7$ term by term.

Why slacks? Go to what a partition count is. The B side is a count of tuples of multiplicities $(m_1, \ldots, m_k)$, each constrained independently $(m_j \geq 1)$, weighted by a linear form $\sum m_j g_j$. That is the shape every product-form generating function has. To match it, the A side must also become "independent positive integers, weighted linearly". The inequalities $a_i > a_{i+1} + a_{i+2}$ are linear, and the standard way to turn linear inequalities into independent ones is to introduce the difference as a new variable — exactly as $a_1 > a_2 > \ldots$ becomes "differences ≥ 1" in ordinary partition theory, where strict chains are straightened by subtracting consecutive terms. Here each inequality involves three consecutive terms instead of two, so the difference involves three terms. The move is the same.

Why is the change of variables invertible? Because it is triangular: d_i involves a_i with coefficient 1 and otherwise only later a's. Solving from the bottom (a_r first) needs no division, so integers go to integers in both directions. That is what makes it a bijection rather than an injection.

Why do the g_j appear? The inverse map is the recursion $a_i = d_i + a_{i+1} + a_{i+2}$, a Fibonacci recursion with forcing term d_i. So each d_j contributes to $a_j, a_{j-1}, \ldots, a_1$ with Fibonacci weights $F_1, F_2, \ldots, F_j$, and to the total with their sum. Partial sums of a Fibonacci sequence satisfy the Fibonacci recursion plus a constant — differencing S_j gives back F_j, and the "+1" is the boundary term. The problem's odd definition $g_j = g_{j-1} + g_{j-2} + 1$ is therefore not decoration; it is the fingerprint of summing the inequalities. Once you have the slacks, the g_j are forced, and the setter wrote them into the statement as the cue.

This is also why recursion on n failed. Removing a part changes one a_i and hence three slacks; removing a copy of g_j changes one multiplicity. The two sides have different natural coordinates, and no recursion in n aligns them. The bijection has to be built in the right coordinates first.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the chained inequalities $a_i > a_{i+1} + a_{i+2}$. On seeing them, write $d_i = a_i - a_{i+1} - a_{i+2} \geq 1$, after padding $a_{r+1} = a_{r+2} = 0$ to make the last two conditions of the same form.

2. Slack variables, then express the total n as a linear form in the slacks and compare the coefficients with the parts allowed on the other side.

3. Hesitation occurs at the boundary: the conditions $a_{r-1} > a_r$ and $a_r \geq 1$ look different. Reinforce it away by padding with zeros; the special cases become the general case.

4. The trap is recursion on n (removing a part from each side), which does not respect either set of constraints. A second trap is not recognising $g_j + 1$ as Fibonacci numbers, which hides why the coefficients match.

5. Slogan: *overlapping linear inequalities become independent when you name the slacks.* And: *a recurrence "Fibonacci plus a constant" means partial sums of Fibonacci numbers.*

6. Decision rule: to prove two partition counts equal, rewrite each as a count of independent non-negative (or positive) integer tuples weighted by a linear form; if the weights agree, the identity map on tuples is the bijection.

▷ **Cue:** chained inequalities $a_i > a_{i+1} + a_{i+2}$ (or $a_i > a_{i+1}$) $\longrightarrow$ **Move:** introduce slacks $d_i = a_i - a_{i+1} - a_{i+2} \geq 1$ as new independent variables

▷ **Cue:** boundary conditions that look different from the general inequality $\longrightarrow$ **Move:** pad the sequence with zeros so they become instances of it

▷ **Cue:** a recurrence $g_j = g_{j-1} + g_{j-2} + 1$ $\longrightarrow$ **Move:** recognise $g_j = F_{j+2} - 1 = F_1 + \cdots + F_j$, partial sums of Fibonacci numbers

▷ **Cue:** a partition condition "every part up to the largest appears" $\longrightarrow$ **Move:** count by multiplicities $m_j \geq 1$, weighted by $\sum m_j g_j$

Putnam 1991 — Paper B

1991 B1

1991 B1. For each integer $n \geq 0$, let $S(n) = n - m^2$, where m is the greatest integer with $m^2 \leq n$. Define a sequence $(a_k)_{k=0}^{\infty}$ by $a_0 = A$ and $a_{k+1} = a_k + S(a_k)$ for $k \geq 0$. For what positive integers A is this sequence eventually constant?

Fields: number theory (squares, floor of square root), sequences **Position:** B1 – opening problem of the second session **Answer:** Exactly when A is a perfect square; for every other A the sequence is strictly increasing and never meets a square.

Layer 1 THE WHAT — surface analysis

The natural attempt. You compute. With $A = 2$: $S(2) = 1$, so $a_1 = 3$; $S(3) = 2$, $a_2 = 5$; $S(5) = 1$, $a_3 = 6$; then $8, 12, 15, 21, 26, 27, 29, 33, 41, 46, 56, 63, 77, 90, \ldots$ With $A = 3$ you join the same orbit at 5. With $A = 7$: $7, 10, 11, 13, 17, 18, 20, 24, 32, 39, 42, \ldots$ Nothing ever settles, and with a perfect square $A = m^2$ you have $S(A) = 0$ at once, so the sequence is constant from the start. The conjecture writes itself: eventually constant if and only if A is a square.

Now the proof. Since $S \geq 0$, the sequence never decreases, and it stops moving exactly when it lands on a square. So you need: a non-square never steps onto a square. The instinct is to hunt for an invariant, and the first place a strong student looks is a congruence – squares are 0 or $1 \bmod 4$, $0, 1, 4 \bmod 8$, and so on, so perhaps the orbit of a non-square stays in the non-square residues. It does not. The orbit of 2 above visits $8 \equiv 0$, $12 \equiv 0$, $21 \equiv 1$, $29 \equiv 1 \pmod 4$; modulo 3 it visits $6 \equiv 0$ and $46 \equiv 1$. Every modulus you try, the orbit wanders through the square residues. At this point the student either gives up on invariants and writes "by inspection the sequence never hits a square", which is not a proof, or starts a case analysis on $a_k \bmod$ something that has no reason to close.

The model chain. Step 1: reformulate. Because $a_{k+1} - a_k = S(a_k) \geq 0$, the sequence is non-decreasing, and $a_{k+1} = a_k$ exactly when $S(a_k) = 0$, i.e. when a_k is a perfect square. If some a_K is a square, then $a_{K+1} = a_K$, and by induction every later term equals a_K: eventually constant. Conversely, if the sequence is constant from index K on, then $S(a_K) = a_{K+1} - a_K = 0$, so a_K is a square. Hence

$$(a_k) \text{ is eventually constant} \iff \text{some } a_k \text{ is a perfect square.}$$

If A is a square this happens at $k = 0$.

Step 2: a non-square never steps onto a square. Let a be a positive non-square, and let m be the greatest integer with $m^2 \leq a$, so $m^2 < a < (m+1)^2$ and $S(a) = a - m^2$. Then

$$a + S(a) = 2a - m^2.$$

Suppose $2a - m^2 = n^2$ for some integer $n \geq 0$. From $m^2 < a < (m+1)^2$,

$$m^2 < 2a - m^2 < 2(m+1)^2 - m^2 = m^2 + 4m + 2,$$

so $m^2 < n^2 < m^2 + 4m + 2 < (m+2)^2$. The only possibility is $n = m + 1$. But then

$$2a = m^2 + (m+1)^2 = 2m^2 + 2m + 1,$$

which is odd, while $2a$ is even. Contradiction. So $a + S(a)$ is not a square.

Step 3: conclude. If A is not a square, Step 2 and induction show that no a_k is a square; by Step 1 the sequence is not eventually constant (indeed it is strictly increasing). Therefore the sequence is eventually constant precisely when A is a perfect square.

Check by the forward route. Writing $a = m^2 + r$ with $1 \le r \le 2m$, the next term is $m^2 + 2r$. If $r \le m$ it lies in $(m^2, m^2 + 2m]$, strictly below $(m+1)^2$; if $r \ge m + 1$ it equals $(m+1)^2 + (2r - 2m - 1)$ with $2r - 2m - 1$ odd and between 1 and $2m - 1$, so strictly between $(m+1)^2$ and $(m+2)^2$. The same conclusion, and it shows where the oddness lives: the overshoot past $(m+1)^2$ is always odd. I have also checked the statement by direct simulation for every $A < 3000$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree on the reformulation "eventually constant $\iff$ the orbit hits a square". They split at the next line. The natural attempt looks for a global invariant of the whole orbit (a residue class). The model looks at a single step and asks, locally, "if the next term were a square, which square could it be?" – then pins it between m^2 and $(m+2)^2$.

> *S is defined by the largest square below a. So the step only knows about the window $m^2 < a < (m+1)^2$. I should not look for something the whole orbit preserves; I should ask whether one step out of that window can land on a square, and if so which one.*

> *The next term is $2a - m^2$, and a is less than $(m+1)^2$, so the next term is less than $m^2 + 4m + 2$. That is below $(m+2)^2$, so the only square it could be is $(m+1)^2$ – one candidate, and I can test it.*

Layer 3 THE WHY

Why is the local window the natural object, rather than a residue class? Read the definition of S again: m is *the greatest integer with* $m^2 \le n$. Everything the recursion does is measured from m^2. A function defined through "the largest square below" has no reason to respect congruences, because the reference point m^2 jumps whenever a crosses a square, and the jump sizes $2m + 1$ grow. So a mod-q invariant is fighting the definition. The definition itself is telling you the natural coordinate: the pair (m, r) with $a = m^2 + r$. Once you are in those coordinates, the recursion is $r \mapsto 2r$, measured from m^2 – the step doubles the distance to the square below. Doubling is simple; the only complication is crossing $(m+1)^2$.

Second: why does the argument reduce to one candidate? Because squares are sparse and the step is bounded. The step can at most double the distance above m^2, and the distance is less than $2m + 1$, so the next term is less than $m^2 + 4m + 2$. The gap from m^2 to $(m+2)^2$ is $4m + 4$. A doubled sub-gap cannot jump two whole gaps. So the question "can the next term be a square?" collapses to "can it be $(m+1)^2$?" That is the recurring pattern with the floor of a square root: the gaps $2m + 1$ are almost constant on a local scale, so bounded moves have only one or two possible landing squares.

Third: why parity? Landing exactly on $(m+1)^2$ means $m^2 + 2r = (m+1)^2$, i.e. $2r = 2m + 1$. The left side is even, the right side is odd. That is the whole obstruction, and it comes from the fact that consecutive squares differ by an odd number while the step from m^2 always travels an even distance $2r$. So there *is* an invariant after all – but it is the parity of the distance from m^2, a local quantity, not the residue of a itself. That is why the congruence hunt on a failed: it was looking at the wrong variable.

There is no official solution in the archive's sources for 1991, so the chain above is my own; it has been verified by hand in both directions (backward: which square could the next term be; forward: where exactly does it land) and by computer for $A < 3000$.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "m is the greatest integer with $m^2 \le n$". On seeing it, write $a = m^2 + r$ with $0 \le r \le 2m$ and rewrite the recursion in terms of r: the next term is $m^2 + 2r$.
2. It suggests localisation: work inside one gap between consecutive squares, and ask where one step can land, rather than tracking the whole orbit.
3. Hesitation occurs after the numerical experiments, when you believe the conjecture but cannot see a proof and start trying moduli. What reinforces it away is the observation that the problem only needs *one step* to be safe (induction does the rest), so you should interrogate a single step.
4. The trap is the congruence invariant on a_k: squares have restricted residues, so it looks promising, but the orbit visits every residue and the search burns time. A smaller trap is forgetting the converse in Step 1 – that an eventually constant sequence must actually contain a square.
5. When a recursion is defined through "the largest square (power, multiple) not exceeding n", the right coordinates are the base point and the remainder. Bounded steps then have only one or two possible landing targets, which you test individually.
6. To show an orbit never reaches a set, prove that one step from outside the set cannot enter it; to prove that, bound the step and list the at-most-few targets it could reach, then kill each by a parity or size argument.

▷ **Cue:** a function is defined via "the greatest m with $m^2 \le n$" $\longrightarrow$ **Move:** write $n = m^2 + r$, $0 \le r \le 2m$, and express the recursion in (m, r)
▷ **Cue:** need to show an orbit never enters a set $\longrightarrow$ **Move:** prove the one-step statement "outside stays outside" and finish by induction
▷ **Cue:** a bounded quantity must be a perfect square $\longrightarrow$ **Move:** trap it strictly between m^2 and $(m+2)^2$ so the only candidate is $(m+1)^2$, then test that one
▷ **Cue:** an equation says an even quantity equals a difference of consecutive squares $\longrightarrow$ **Move:** consecutive squares differ by the odd number $2m+1$: parity contradiction
▷ **Cue:** a congruence invariant keeps failing on the terms themselves $\longrightarrow$ **Move:** look for the invariant in the local coordinate (the remainder above a base point), not in the term

1991 B2

1991 B2. Suppose f and g are non-constant, differentiable, real-valued functions defined on $(-\infty, \infty)$. Furthermore, suppose that for each pair of real numbers x and y,

$$f(x+y) = f(x)f(y) - g(x)g(y),$$
$$g(x+y) = f(x)g(y) + g(x)f(y).$$

If $f'(0) = 0$, prove that $(f(x))^2 + (g(x))^2 = 1$ for all x.

Fields: functional equations, real analysis (differentiation), complex numbers **Position:** B2 – second problem of the second session **Answer:** Differentiating in y at $y = 0$ gives $f' = -g'(0)\, g$ and $g' = g'(0)\, f$, so $(f^2 + g^2)' = 0$; the constant equals $f(0)^2 + g(0)^2 = 1$ because $f(0) = 1, g(0) = 0$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You recognise the addition formulae for cos and sin, and you know the target $\cos^2 + \sin^2 = 1$. The algebraic instinct is to reproduce the trigonometric proof: put $y = -x$. That gives

$$f(0) = f(x)f(-x) - g(x)g(-x), \qquad g(0) = f(x)g(-x) + g(x)f(-x).$$

If you knew $f(0) = 1$, $g(0) = 0$, f even and g odd, the first line would read $1 = f(x)^2 + g(x)^2$ and you would be done. So you try to prove evenness and oddness from the equations. Put $x = 0$: $f(y) = f(0)f(y) - g(0)g(y)$ and $g(y) = f(0)g(y) + g(0)f(y)$. From $x = y = 0$ you can extract $f(0)$ and $g(0)$ (see below). But parity will not come. Swapping x and y gives nothing new – the equations are already symmetric. Substituting $y = -x$ gives two equations in four unknowns $f(\pm x), g(\pm x)$. And there is a reason it will not come: the functions $f(x) = e^{ax}\cos bx$, $g(x) = e^{ax}\sin bx$ satisfy both equations for every real a, b, and for $a \neq 0$ the function f is not even and $f^2 + g^2 = e^{2ax} \neq 1$. The functional equations alone do not imply the result. You have not used $f'(0) = 0$, and pure substitution cannot use it. That is the wall.

The model chain. *Step 1: values at* 0. Put $x = y = 0$:

$$f(0) = f(0)^2 - g(0)^2, \qquad g(0) = 2f(0)g(0).$$

If $g(0) \neq 0$, the second equation gives $f(0) = \frac{1}{2}$, and then the first gives $g(0)^2 = \frac{1}{4} - \frac{1}{2} < 0$, impossible. So $g(0) = 0$ and $f(0) = f(0)^2$, i.e. $f(0) \in \{0, 1\}$. If $f(0) = 0$, then putting $y = 0$ in the first equation gives $f(x) = f(x)f(0) - g(x)g(0) = 0$ for all x, so f is constant, contrary to hypothesis. Hence

$$f(0) = 1, \qquad g(0) = 0.$$

Step 2: differentiate in y at $y = 0$. Fix x. Both sides of each equation are differentiable functions of y (the left sides because f, g are differentiable at $x + y$; the right sides are linear combinations of $f(y), g(y)$ with constant coefficients). Differentiating with respect to y and setting $y = 0$, with $k = g'(0)$ and using $f'(0) = 0$:

$$f'(x) = f(x)f'(0) - g(x)g'(0) = -k\,g(x), \qquad g'(x) = f(x)g'(0) + g(x)f'(0) = k\,f(x).$$

Step 3: the sum of squares is constant. Let $h(x) = f(x)^2 + g(x)^2$. Then

$$h'(x) = 2f(x)f'(x) + 2g(x)g'(x) = -2k\,f(x)g(x) + 2k\,g(x)f(x) = 0$$

for every x, so h is constant on $\mathbb{R}$ (mean value theorem). Its value is $h(0) = 1^2 + 0^2 = 1$. Hence $f(x)^2 + g(x)^2 = 1$ for all x. ■

A second way to pin the constant. Squaring and adding the two equations gives the two-squares identity

$$f(x+y)^2 + g(x+y)^2 = \big(f(x)^2 + g(x)^2\big)\big(f(y)^2 + g(y)^2\big),$$

i.e. $h(x + y) = h(x)h(y)$. With $h \equiv c$ this gives $c = c^2$, so $c \in \{0, 1\}$; and $c = 0$ would force $f \equiv g \equiv 0$, which are constant. So $c = 1$ without Step 1.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree on reading the equations as the addition formulae for cos, sin. They split at the first manipulation. The natural attempt substitutes values ($y = -x$, $y = 0$) and hunts for parity. The model differentiates with respect to y at $y = 0$, turning two functional equations into two ordinary differential equations.

> *The hypothesis $f'(0) = 0$ is about a derivative at one point. Substitution can never touch it. The only way a derivative at 0 enters an equation valid for all x is by differentiating in y and then setting $y = 0$ – so that is the move.*

> *I want to show $f^2 + g^2$ is constant, and a quantity is constant when its derivative vanishes. So I need f' and g' in terms of f and g; the addition law gives exactly that once I differentiate at 0.*

> *Before I try to prove this from the equations alone, let me test it on the obvious family.* $e^{ax}\cos bx$, $e^{ax}\sin bx$ *satisfy both equations and give* e^{2ax}. *So the equations cannot be enough, and* $f'(0) = 0$ *is what kills a. That tells me where the proof must use it.*

Layer 3 THE WHY

Why is differentiating in y at $y = 0$ natural rather than a trick? Go to what an addition law is. An equation $F(x+y) = \Phi(F(x), F(y))$ says the value at $x+y$ is determined by values at x and y. Take y tiny: then the value at $x+y$ – the value *slightly to the right of* x – is determined by the value at x and the values near 0. That is precisely a differential equation: the rate of change at x is fixed by the state at x and by the rates at 0. So every differentiable addition law is secretly a first-order ODE whose coefficients are derivatives at 0. The problem hands you one of those derivatives, $f'(0)$, and asks for a global conclusion. The global-from-infinitesimal passage is the derivative at 0 propagated by the addition law. There is nothing else it could be.

Why does $f'(0) = 0$ produce a *conserved* quantity? Package the pair as a complex function $F = f + ig$. The two equations are the real and imaginary parts of

$$F(x+y) = F(x)F(y),$$

because $(f(x)+ig(x))(f(y)+ig(y)) = (f(x)f(y) - g(x)g(y)) + i(f(x)g(y) + g(x)f(y))$. Differentiating in y at 0 gives $F' = \lambda F$ with $\lambda = F'(0) = f'(0) + ig'(0)$, so $F(x) = F(0)e^{\lambda x} = e^{\lambda x}$. Then $|F|^2 = f^2 + g^2 = e^{2\operatorname{Re}\lambda\, x} = e^{2f'(0)x}$. The quantity f^2+g^2 is the squared modulus, and the modulus of $e^{\lambda x}$ is controlled by the real part of λ alone. $f'(0)$ *is* that real part. So the hypothesis is exactly "the exponent is purely imaginary", which is exactly "the motion is rotation, not spiral". In the real computation of Step 3, this is the cancellation $-2kfg + 2kgf = 0$: the rotation part contributes nothing to the radius, and the stretching part has been set to zero.

This also explains why the natural attempt was doomed: parity of f and g holds only when $\operatorname{Re}\lambda = 0$, so any proof of parity must use $f'(0) = 0$, which substitution cannot reach.

Finally, why must Step 1 (or the multiplicativity of h) be there at all? $h' = 0$ gives constancy; the value of the constant still has to come from somewhere. The degenerate solution $f \equiv g \equiv 0$ satisfies the equations with $h \equiv 0$, and the phrase "non-constant" in the statement exists to exclude it. When a statement contains a hypothesis that looks decorative, find the counterexample it excludes – here, the zero solution.

There is no official solution for 1991 in the archive's sources; the proof above is my own and each step has been checked, including the family $e^{ax}\cos bx$, $e^{ax}\sin bx$ as a sanity test of both hypotheses.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the shape $f(x+y) = f(x)f(y) - g(x)g(y)$, $g(x+y) = f(x)g(y) + g(x)f(y)$ – the cosine and sine addition formulae. On seeing it, write $F = f + ig$ and $F(x+y) = F(x)F(y)$.

2. Together with a hypothesis about a derivative at 0, it suggests differentiating the functional equation in y and setting $y = 0$, producing the system $f' = f'(0)f - g'(0)g$, $g' = g'(0)f + f'(0)g$.

3. Hesitation occurs after $y = -x$, when parity looks one step away and will not come. What reinforces it away is the test family $e^{ax}\cos bx$: it shows the equations alone are insufficient, so you must stop substituting and use $f'(0) = 0$.

4. The trap is to imitate the trigonometric proof (which uses parity) and to try to derive parity by substitution. A quieter trap is to forget that "constant" is not "equal to 1": the zero solution must be excluded via "non-constant".

5. Any functional equation of addition type plus differentiability is an ODE in disguise; any hypothesis on a derivative at a single point is a coefficient of that ODE. To show a quantity is constant, differentiate it and substitute the ODE.

6. Before attacking a functional equation, test the conclusion on the general solution family you know; the hypotheses the family violates tell you exactly where each hypothesis must be used in the proof.

▷ **Cue:** a functional equation $F(x+y) = \Phi(F(x), F(y))$ with differentiable F $\longrightarrow$ **Move:** differentiate in y and set $y = 0$ to get an ODE with coefficients $F'(0)$

▷ **Cue:** a hypothesis about a derivative at a single point $\longrightarrow$ **Move:** it can only enter through differentiating a functional identity and evaluating at that point

▷ **Cue:** the pair $f(x)f(y) - g(x)g(y)$, $f(x)g(y) + g(x)f(y)$ $\longrightarrow$ **Move:** write $F = f + ig$; the system is $F(x+y) = F(x)F(y)$ and $f^2 + g^2 = |F|^2$

▷ **Cue:** prove an expression is constant for all x $\longrightarrow$ **Move:** show its derivative is zero, then evaluate at a convenient point

▷ **Cue:** a hypothesis such as "non-constant" or "nonzero" looks decorative $\longrightarrow$ **Move:** find the degenerate solution it excludes; that is where it will be used

1991 B3

1991 B3. Does there exist a real number L such that, if m and n are integers greater than L, then an $m \times n$ rectangle may be expressed as a union of 4×6 and 5×7 rectangles, any two of which intersect at most along their boundaries?

Fields: combinatorics (tilings), elementary number theory (Frobenius / coin problem) **Position:** B3 – third problem of the second session **Answer:** Yes. $L = 197$ works: every $m \times n$ rectangle with $m, n \geq 198$ can be tiled by stacking horizontal strips of heights 20, 28 and 35, each of which tiles at every length $n \geq 30$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You read "does there exist" and consider both answers. The first necessary condition is area: $mn = 24a + 35b$ for some non-negative integers a, b. Since $\gcd(24, 35) = 1$, every integer above $24 \cdot 35 - 24 - 35 = 781$ is of this form, so area never obstructs large rectangles. That makes "yes" plausible but proves nothing: area is necessary, not sufficient. So you look for a colouring obstruction – the standard weapon against tilings – and try chessboard colourings, colourings mod 4 or mod 5 along rows. None bites: a 4×6 tile and a 5×7 tile carry different colour counts under almost every simple colouring, and the rectangle can be large in both directions, so there is nothing to balance against. Then you turn to construction and try to tile a specific $m \times n$ directly: split by m mod 4, m mod 5, and so on, fitting tiles into corners. The case tree has dozens of branches, the pieces do not fit together across branches, and there is no end in sight. The attempt stalls because it is trying to solve a two-dimensional problem in two dimensions at once.

The model chain. Read "4×6 rectangle" as a rectangle of those side lengths in either orientation (the tiles are geometric rectangles; the statement fixes no orientation). The answer is *yes*.

Step 1: columns. A column of width w and height h is filled by a vertical stack of identical tiles of width w. The four available widths are:

- width 6: tiles 4 high, so any height divisible by 4;
- width 4: tiles 6 high, so any height divisible by 6;
- width 7: tiles 5 high, so any height divisible by 5;
- width 5: tiles 7 high, so any height divisible by 7.

Step 2: strips. A strip of height h and length n is tiled by placing columns of height h side by side, so it suffices that n is a non-negative combination of the widths available at height h. Three strips suffice:

- $h = 20$ (divisible by 4 and 5): widths 6 and 7. Every $n \geq 30$ is $6s + 7t$ with $s, t \geq 0$, since the largest integer not so expressible is $6 \cdot 7 - 6 - 7 = 29$.
- $h = 28$ (divisible by 4 and 7): widths 6 and 5. Every $n \geq 20$ works, since $5 \cdot 6 - 5 - 6 = 19$.
- $h = 35$ (divisible by 5 and 7): widths 7 and 5. Every $n \geq 24$ works, since $5 \cdot 7 - 5 - 7 = 23$.

So for every $n \geq 30$, strips of heights 20, 28, 35 and length n can all be tiled.

Step 3: stack the strips. An $m \times n$ rectangle is a stack of strips of length n whose heights add to m. So it suffices that $m = 20a + 28b + 35c$ with $a, b, c \geq 0$. Claim: every $m \geq 198$ has this form. Since $35 \equiv 3 \pmod 4$, the values $35c$ for $c = 0, 1, 2, 3$ cover all residues mod 4; choose $c \in \{0, 1, 2, 3\}$ with $35c \equiv m \pmod 4$. Then $m - 35c \geq 198 - 105 = 93$ and is divisible by 4, hence $m - 35c = 4t$ with $t \geq 24$. Now $20a + 28b = 4(5a + 7b)$, and every $t \geq 24$ is $5a + 7b$ with $a, b \geq 0$, since $5 \cdot 7 - 5 - 7 = 23$. So $m = 20a + 28b + 35c$.

Step 4: conclude. Take $L = 197$. If $m, n > 197$ then $m \geq 198$ and $n \geq 198 \geq 30$, so write $m = 20a + 28b + 35c$, tile a strips of height 20, b of height 28 and c of height 35, each of length n (Step 2), and stack them. Tiles meet only along boundaries. (The bound 197 is sharp for this three-strip method – a computer search confirms 197 is the largest integer not of the form $20a + 28b + 35c$ – but the problem asks only for existence, and no attempt is made at the least possible L.)

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree that area is no obstruction and that the answer is probably "yes". They split at the first construction line. The natural attempt tries to tile a given $m \times n$ by case analysis on residues in both directions at once. The model fixes one dimension: it builds strips of a few fixed heights that tile at *every* large length, and reduces the other dimension to writing m as a sum of those heights.

> *Two free dimensions is too many. If I can make a strip of some fixed height that tiles for every long enough length, then I only have to worry about heights, and heights add. That is a one-dimensional coin problem.*

> *A strip of height h tiles at every large length if two columns of coprime widths both fit height h. Width 6 needs $4 \mid h$, width 7 needs $5 \mid h$ – so $h = 20$. Now I need a few such heights whose gcd is 1.*

> *20 and 42 are both even, so they will never reach odd m. I need an odd strip height: 35, using the two orientations of the 5×7 tile.*

Layer 3 THE WHY

Why is "reduce to strips" natural and not a trick? Look at what the conclusion quantifies over: *all* $m, n > L$. You are asked for a construction that works uniformly in two independent parameters. Uniformity in one parameter is a solved problem you already own: the Frobenius (coin) problem says that if $\gcd(p, q) = 1$ then every sufficiently large integer is a non-negative combination of p and q (indeed every integer above $pq - p - q$). That theorem is about *lengths that add*. In a tiling, lengths add exactly when you place blocks of equal height side by side, or blocks of equal width on top of each other. So the moment you ask "where in a tiling do integers add?", the answer is "along a strip", and the construction organises itself: fix a height, make

lengths add along it; then fix the length, make heights add up the stack. The two-dimensional problem is the product of two one-dimensional coin problems. That is the whole idea, and it comes from the definition of what a tiling does to lengths.

Why do the specific heights $20, 28, 35$ appear? Each column has a width and a divisibility condition on its height. For a strip to be flexible in length, it needs two columns whose widths are coprime; the four widths are $4, 5, 6, 7$, and every pair except $\{4, 6\}$ is coprime. The strip height must then be a common multiple of the two corresponding tile heights: pairs $(6, 7) \to 20$, $(6, 5) \to 28$, $(7, 5) \to 35$, $(4, 7) \to 30$, $(4, 5) \to 42$. That is a finite menu read directly off the tile dimensions. From it you need heights with overall gcd 1 (otherwise some residue class of m is never reached). The even ones cannot do it alone, so 35 – the only odd entry – is forced, and 20 with 28 supply a gcd of 4 that 35 breaks. Nothing here is chosen by inspiration; it is the menu, filtered by a gcd condition.

Why does the colouring hunt fail? A colouring argument proves *impossibility*. Before reaching for it, ask whether impossibility is even plausible: the area condition is satisfied for all large mn, and both tiles can be rotated, which destroys most colour asymmetries. When a statement says "for all sufficiently large" and the necessary counting condition holds eventually, the expected answer is "yes" by a uniform construction; spend the time on the construction.

There is no official solution for 1991 in the archive's sources; the construction and the bound $L = 197$ are my own, verified by hand (Step 3 is a complete proof of the representation of every $m \geq 198$) and by a computer check of the numerical semigroup generated by $20, 28, 35$.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "if m and n are integers greater than L" – uniformity for all large parameters. On seeing it, write down the coin-problem fact: if $\gcd(p, q) = 1$, every integer greater than $pq - p - q$ is $ap + bq$ with $a, b \geq 0$.

2. It suggests separating dimensions: build strips of fixed height that tile at every large length, then stack strips so that heights add to m.

3. Hesitation occurs when the first strip heights found (20, 30, 42) are all even and you fear odd m is impossible. What reinforces it away is listing the full menu of strip heights and noticing that 35 uses both orientations of the same tile.

4. The trap is to fight the problem in two dimensions at once – residue case analysis on m and n simultaneously, or a hunt for a colouring obstruction to a statement that is true.

5. Whenever a tiling, packing or sum problem asks for a construction valid for all large parameters, ask where quantities add; turn each dimension into a coin problem with coprime generators.

6. Reduce a two-parameter "all sufficiently large" construction to a one-parameter family of building blocks, each valid for all large values of the first parameter, whose sizes in the second parameter generate a numerical semigroup of gcd 1.

▷ **Cue:** a statement of the form "for all integers $m, n > L$, the object can be built from pieces" $\longrightarrow$ **Move:** fix one dimension; build strips that work for all large lengths; stack them

▷ **Cue:** need every large integer as a non-negative combination $\longrightarrow$ **Move:** find generators with gcd 1; the coin problem gives all integers beyond $pq - p - q$ for two coprime generators

▷ **Cue:** a rectangular tile of size $a \times b$ in a tiling $\longrightarrow$ **Move:** a column of width a can have any height divisible by b; list the width/height-divisibility menu

▷ **Cue:** all available block sizes share a common factor $\longrightarrow$ **Move:** look for a block using a different orientation or tile that breaks the factor

▷ **Cue:** the necessary counting condition (area, sum) holds for all large parameters $\longrightarrow$ **Move:** expect a construction, not an obstruction; do not start with colourings

1991 B4

1991 B4. Suppose p is an odd prime. Prove that

$$\sum_{j=0}^{p} \binom{p}{j}\binom{p+j}{j} \equiv 2^p + 1 \pmod{p^2}.$$

Fields: number theory (binomial coefficients modulo prime powers), combinatorics (binomial identities) **Position:** B4 – fourth problem of the second session **Answer:** The congruence holds. The two end terms give 1 and $\binom{2p}{p} \equiv 2$; each middle term is congruent to $\binom{p}{j}$ modulo p^2, and those add up to $2^p - 2$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You test $p = 3$: the terms are $1, 3 \cdot 4 = 12, 3 \cdot 10 = 30$ and $1 \cdot 20 = 20$, total $63 = 7 \cdot 9$, and $2^3 + 1 = 9$. So both sides are 0 modulo 9. The check works, but it does not tell you where to go next. The natural move is to hunt for a closed form for the sum $\sum_j \binom{p}{j}\binom{p+j}{j}$ with p replaced by a general n: these are the central Delannoy numbers $1, 3, 13, 63, 321, \ldots$, and they have no closed form. You try expanding $\binom{p+j}{j}$ with Vandermonde as $\sum_i \binom{p}{i}\binom{j}{i}$, obtain a double sum, swap the order, and get a new double sum with no obvious collapse. Twenty minutes on, you have an identity-shaped mess and you have not used the one thing the problem gave you: that p is prime and you only need the answer modulo p^2. This is zombie maths – exact algebra on a quantity you are only asked about approximately.

The model chain. Write $T_j = \binom{p}{j}\binom{p+j}{j}$ for $0 \le j \le p$. Split the sum into the two end terms and the middle.

End terms. $T_0 = 1$. $T_p = \binom{2p}{p}$.

Middle terms, $1 \le j \le p-1$. Two facts.

(a) $p \mid \binom{p}{j}$. Indeed $j!\,(p-j)!\,\binom{p}{j} = p!$; the prime p divides the right side but divides neither $j!$ nor $(p-j)!$ (every factor is less than p), so p divides $\binom{p}{j}$.

(b) $\binom{p+j}{j} \equiv 1 \pmod{p}$. Indeed

$$j!\binom{p+j}{j} = (p+1)(p+2)\cdots(p+j) \equiv 1 \cdot 2 \cdots j = j! \pmod{p},$$

and $j!$ is coprime to p because $j < p$, so it can be cancelled.

Hence

$$T_j - \binom{p}{j} = \binom{p}{j}\left(\binom{p+j}{j} - 1\right)$$

is a product of two integers each divisible by p, so $T_j \equiv \binom{p}{j} \pmod{p^2}$.

The last end term. Choosing p objects from $2p$ objects split into two groups of p (Vandermonde) gives

$$\binom{2p}{p} = \sum_{k=0}^{p}\binom{p}{k}\binom{p}{p-k} = \sum_{k=0}^{p}\binom{p}{k}^2 = 2 + \sum_{k=1}^{p-1}\binom{p}{k}^2.$$

By (a) every square in the last sum is divisible by p^2, so $\binom{2p}{p} \equiv 2 \pmod{p^2}$.

Assemble.

$$\sum_{j=0}^{p} T_j \equiv 1 + \sum_{j=1}^{p-1}\binom{p}{j} + 2 = 1 + (2^p - 2) + 2 = 2^p + 1 \pmod{p^2}. \qquad \blacksquare$$

(Checked numerically for every prime up to 23. The argument never uses that p is odd; for $p = 2$ the sum is $13 \equiv 5 = 2^2 + 1 \pmod 4$ as well.)

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths agree up to writing the sum down. They split at the first line: the natural attempt asks "what *is* this sum?"; the model asks "what is each term *modulo* p^2?" and never evaluates the sum at all.

> *I don't need the sum, I need it mod p^2. The factor $\binom{p}{j}$ already carries one p for every middle j, so I only need the other factor $\binom{p+j}{j}$ modulo p – and modulo p, adding p to something is the same as adding nothing.*

> *The right-hand side is $2^p + 1$. 2^p is $\sum \binom{p}{j}$, so the target is telling me each term should collapse to $\binom{p}{j}$, with one extra 1 left over. The ends are where the extra must come from: $1 + \binom{2p}{p}$ should be $1 + 2$.*

> *Now $\binom{2p}{p}$ mod p^2. Squares of $\binom{p}{k}$ are divisible by p^2, so if I can write $\binom{2p}{p}$ as a sum of such squares I'm done. Vandermonde does exactly that.*

Layer 3 THE WHY

Start from what a congruence modulo p^2 is. It is a statement about the remainder on division by p^2, which means any summand that is a multiple of p^2 is invisible. So the first question on seeing "$\pmod{p^2}$" is not "what is the sum?" but "which pieces are invisible, and which pieces are visible only through their residue mod p?" A product ab with $p \mid a$ depends, modulo p^2, only on a and on b mod p: if $b \equiv b' \pmod{p}$ then $ab - ab' = a(b - b')$ carries two factors of p. That single observation is the whole engine, and it is forced by the definition of the modulus, not discovered by inspiration.

Why is $\binom{p}{j}$ divisible by p? Because $\binom{p}{j} = p!/(j!(p-j)!)$ and the only factor p in the numerator has nothing to cancel against. This is the same fact that gives $(a+b)^p \equiv a^p + b^p \pmod{p}$, and the problem is built on it: the left factor of every middle term is a "free" p.

Why is $\binom{p+j}{j} \equiv 1$? Write it as a product $\prod_{i=1}^{j} \frac{p+i}{i}$. Modulo p each factor is $\frac{i}{i} = 1$. Shifting the top by p is invisible mod p, so $\binom{p+j}{j} \equiv \binom{j}{j} = 1$. (This is the simplest case of Lucas's theorem, but you do not need the theorem; the product formula is enough.) The one place that shortcut breaks is $j = p$, where the product contains $\frac{2p}{p}$ and dividing by p is illegal mod p. That is exactly why $j = p$ has to be treated separately, and why the model splits off the end terms before doing anything else.

Why Vandermonde for $\binom{2p}{p}$? Because the previous step has trained you to want factors of $\binom{p}{k}$, and the question is how to see $\binom{2p}{p}$ in terms of $\binom{p}{\cdot}$. The set of size $2p$ is two sets of size p glued together; choosing p from the union is choosing k from one half and $p-k$ from the other. That turns $\binom{2p}{p}$ into a sum of products $\binom{p}{k}\binom{p}{p-k}$, and every middle product has *two* free factors of p. The only survivors are $k = 0$ and $k = p$, each contributing 1.

Kedlaya's first solution packages all of this as a generating function: the sum is the coefficient of x^p in $\sum_j \binom{p}{j}(1+x)^{p+j} = (2+x)^p(1+x)^p$, which equals $\sum_k \binom{p}{k}^2 2^k$, and only $k = 0, p$ survive mod p^2. Here is where that comes from: $\binom{p+j}{j} = \binom{p+j}{p}$ is a coefficient of $(1+x)^{p+j}$, and $\sum_j \binom{p}{j} y^j = (1+y)^p$ is the binomial theorem with $y = 1 + x$. It is the same idea – every surviving product has two $\binom{p}{\cdot}$ factors – achieved in one algebraic stroke rather than term by term. The second official solution is the term-by-term argument above, with $\binom{2p}{p} \equiv 2$ proved by pairing i with i^{-1} modulo p.

Layer $+\alpha$ BACKTESTING

1. The cue is "$\pmod{p^2}$" next to a sum of products containing $\binom{p}{j}$. On seeing it, write: "$p \mid \binom{p}{j}$ for $0 < j < p$, so I need the other factor only mod p."

2. Term-by-term reduction: split off the ends, reduce the middle factor mod p, and reserve the ends for separate treatment.

3. Hesitation comes at $j = p$: the product $\prod(p+i)/i$ now contains $2p/p$ and the mod-p shortcut fails. The reinforcement is to expect this: whenever a shortcut divides by i for $i \le j$, it breaks the moment j reaches p. Isolate that term and ask which known congruence covers it ($\binom{2p}{p} \equiv 2$).

4. The trap is the closed form. The central Delannoy numbers have none, and Vandermonde-expanding $\binom{p+j}{j}$ produces a double sum that is correct and useless. A lesser trap: the hypothesis "odd" suggests the proof must use oddness (pairing j with $p-j$, say). It does not – the statement is also true for $p = 2$ – and chasing a role for it wastes time.

5. The right-hand side $2^p + 1$ is a signal: 2^p is the row sum of $\binom{p}{j}$. When the target is a recognisable sum plus a small correction, conjecture that the terms reduce individually to that sum's terms and look for the correction in the boundary terms.

6. When asked for a sum modulo m, reduce the terms, not the sum. If some factor already carries part of the modulus, reduce the co-factor only modulo what remains. Treat separately every index at which your reduction divides by the prime.

▷ **Cue:** a congruence mod p^2 for a sum of products, one factor being $\binom{p}{j}$ ⟶ **Move:** use $p \mid \binom{p}{j}$ for $0 < j < p$ and reduce the other factor only mod p

▷ **Cue:** $\binom{p+j}{j}$ or any binomial whose top is shifted by p, with $j < p$ ⟶ **Move:** write it as $\prod_{i \le j}(p+i)/i$ and read it mod p as 1

▷ **Cue:** $\binom{2p}{p}$ modulo p^2 ⟶ **Move:** Vandermonde: $\binom{2p}{p} = \sum_k \binom{p}{k}^2$, and all middle squares vanish mod p^2, giving 2

▷ **Cue:** the target of a congruence is a recognisable sum (such as $2^p = \sum \binom{p}{j}$) plus a constant ⟶ **Move:** guess that each term reduces to the matching term of that sum and find the constant at the boundary indices

▷ **Cue:** a sum with no closed form, asked about only modulo something ⟶ **Move:** stop evaluating; reduce term by term

1991 B5

1991 B5. Let p be an odd prime and let $\mathbb{F}_p$ denote (the field of) integers modulo p. How many elements are in the set

$$\{x^2 : x \in \mathbb{F}_p\} \cap \{y^2 + 1 : y \in \mathbb{F}_p\}?$$

Fields: number theory (finite fields, quadratic residues), counting **Position:** B5 – fifth problem of the second session **Answer:** $\lceil p/4 \rceil = \lfloor (p+3)/4 \rfloor$: that is, $\frac{p+3}{4}$ if $p \equiv 1 \pmod 4$ and $\frac{p+1}{4}$ if $p \equiv 3 \pmod 4$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You rename the set first: an element a lies in the intersection exactly when a is a square and $a-1$ is a square (squares here include 0). So you are counting *consecutive squares* $a-1, a$ in $\mathbb{F}_p$. Then you tabulate. For $p = 3$ the squares are $\{0,1\}$ and the shifted squares $\{1,2\}$, intersection $\{1\}$: one element. For $p = 5$: $\{0,1,4\} \cap \{1,2,0\} = \{0,1\}$, two. For $p = 7$: $\{0,1,2,4\} \cap \{1,2,3,5\} = \{1,2\}$, two. For $p = 11$: squares $\{0,1,3,4,5,9\}$, shifted $\{1,2,4,5,6,10\}$, intersection $\{1,4,5\}$, three. For $p = 13$: squares $\{0,1,3,4,9,10,12\}$, shifted $\{1,2,4,5,10,11,0\}$, intersection $\{0,1,4,10\}$, four.

The data say "about $p/4$", and you have a heuristic that explains it: roughly half the elements are squares, and if "a is a square" and "$a-1$ is a square" behaved independently, a quarter of all a would qualify. The heuristic is also exactly what you cannot prove. Whether a is a square and whether $a-1$ is a square are not independent events in any sense you can make rigorous at A-level; they are two facts about the same number. If you know Legendre symbols you push harder: with $\chi(a) = \left(\frac{a}{p}\right)$, the number of square roots of a is $1+\chi(a)$, so

$$\#\{(x,y)\} \text{ with } x^2 = a,\ y^2 = a-1 \ = \ (1+\chi(a))(1+\chi(a-1)),$$

and summing over a gives $p + \sum_a \chi(a) + \sum_a \chi(a-1) + \sum_a \chi(a)\chi(a-1) = p+0+0+\sum_a \chi(a)\chi(a-1)$. Now the whole problem sits in the correlation sum $\sum_a \chi(a)\chi(a-1)$, which is exactly the "independence" you could not justify, in symbolic form. Many students stop here. (It can be finished: for $a \neq 0$, $\chi(a)\chi(a-1) = \chi(a^2)\chi(1-a^{-1}) = \chi(1-a^{-1})$, and as a runs over the non-zero elements, $1-a^{-1}$ runs over every element except 1, so the sum is $0-\chi(1) = -1$. That is Kedlaya's Solution 2. But finding that substitution is the same insight as the model chain below, arriving late and in heavier notation.)

The model chain. Let $S = \{(x,y) \in \mathbb{F}_p^2 : x^2 - y^2 = 1\}$, and let $\varphi : S \to \mathbb{F}_p$, $\varphi(x,y) = x^2$. An element a is in the intersection if and only if $a = x^2$ and $a = y^2 + 1$ for some x, y, that is, if and only if $a \in \varphi(S)$. So the answer is $|\varphi(S)|$.

Step 1: count S. Factor: $x^2 - y^2 = (x+y)(x-y)$. Put $u = x+y$, $v = x-y$. Since p is odd, 2 is invertible and $x = (u+v)/2$, $y = (u-v)/2$, so $(x,y) \mapsto (u,v)$ is a bijection $\mathbb{F}_p^2 \to \mathbb{F}_p^2$. It carries S onto $\{(u,v) : uv = 1\}$. For each $u \neq 0$ there is exactly one v, namely u^{-1}, and for $u = 0$ there is none. Hence $|S| = p-1$.

Step 2: fibre sizes. Fix $a \in \varphi(S)$ and one point (x_0, y_0) over it. The fibre $\varphi^{-1}(a)$ is $\{(x,y) : x^2 = a,\ y^2 = a-1\}$. In a field, $x^2 = x_0^2$ means $(x-x_0)(x+x_0) = 0$, so $x = \pm x_0$; likewise $y = \pm y_0$. Since p is odd, $x_0 \neq -x_0$ unless $x_0 = 0$, and similarly for y_0. So:

- if $a \neq 0$ and $a \neq 1$: $x_0 \neq 0$, $y_0 \neq 0$, and the fibre is $\{(\pm x_0, \pm y_0)\}$, of size 4;
- $a = 1$: $y_0 = 0$, $x_0 = \pm 1$, fibre of size 2; and $1 = 1^2 = 0^2 + 1$ is always in $\varphi(S)$;
- $a = 0$: $x_0 = 0$, $y_0^2 = -1$, fibre of size 2; this a is in $\varphi(S)$ if and only if -1 is a square mod p.

Let $c = 1$ if -1 is a square mod p and $c = 0$ otherwise, and let k be the number of $a \in \varphi(S)$ other than $0, 1$.

Step 3: add the fibres. S is the disjoint union of its fibres, so

$$p - 1 = |S| = 4k + 2 + 2c, \qquad |\varphi(S)| = k + 1 + c = \frac{p-1-2-2c}{4} + 1 + c = \frac{p+1+2c}{4}.$$

Step 4: determine c. $|\varphi(S)|$ is an integer, so $p + 1 + 2c \equiv 0 \pmod 4$. If $p \equiv 1 \pmod 4$ this forces $c = 1$; if $p \equiv 3 \pmod 4$ it forces $c = 0$. (This is itself a proof that -1 is a square mod p exactly when $p \equiv 1 \pmod 4$; no separate theorem is needed.) Therefore

$$|\varphi(S)| = \begin{cases} \dfrac{p+3}{4}, & p \equiv 1 \pmod 4, \\ \dfrac{p+1}{4}, & p \equiv 3 \pmod 4, \end{cases}$$

which is $\lceil p/4 \rceil$. This matches the table: $1, 2, 2, 3, 4$ for $p = 3, 5, 7, 11, 13$, and a computer check confirms it for every odd prime below 200.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree on the translation "a is a square and $a-1$ is a square". They split at the next line. The natural attempt keeps a as the unknown and asks, for each a,

whether two conditions hold – which needs the joint behaviour of two square-tests. The model changes the unknown: it counts the *witnesses* (x, y) with $x^2 = y^2 + 1$, a single equation in two variables, and recovers the count of a by dividing out the fibre size.

> *I can't count the a directly, because each a has to pass two tests at once. But each good a comes with its witnesses x and y, and those satisfy one equation, $x^2 = y^2 + 1$. Count the pairs, then divide by how many pairs each a has.*

> *$x^2 - y^2 = 1$ is a difference of two squares. It factors: $(x + y)(x - y) = 1$. Once it's a product equal to 1, the first factor is anything non-zero and the second is forced.*

> *Each a gets four witnesses $(\pm x, \pm y)$ – unless x or y is 0. So the edge cases are $a = 0$ and $a = 1$, and I must count them separately rather than let them spoil the division by 4.*

Layer 3 THE WHY

Why is "count witnesses instead of values" natural rather than a trick? Look at how the set is defined: $\{x^2 : x \in_p\}$ and $\{y^2 + 1 : y \in_p\}$ are *images*. An image is the set of values of a map, and a value can be hit many times. Counting the points of an image is hard; counting the points of the domain is easy; the bridge between them is the fibre size. That is not a Putnam idea, it is the definition of an image, together with the one counting principle that every problem of this kind rests on: a set that is partitioned into blocks has size equal to the sum of the block sizes. The intersection of two images is the image of the set of pairs that land on the same value, and "land on the same value" is the equation $x^2 = y^2 + 1$. So the moment you write the set-builder notation honestly, the equation $x^2 = y^2 + 1$ is sitting there. The model chain is what you get by reading the statement literally.

Why does the equation then give up its count so easily? Because $x^2 - y^2$ is the one quadratic form you already know how to factor, and in a field a product equal to a non-zero constant is completely controlled: choose the first factor freely (non-zero), and the second is its inverse. This is the finite-field version of the fact that the hyperbola $x^2 - y^2 = 1$ is parametrised by $x = \frac{r+r^{-1}}{2}$, $y = \frac{r-r^{-1}}{2}$ (think cosh and sinh with $r = e^t$) – one free parameter, no square roots. Kedlaya's remark says the same thing: the conic is "birationally equivalent to the line". Stripped of the vocabulary, that means "a conic through a known point can be swept out by one parameter, so it has about as many points as a line". Here the count is exactly $p - 1$ because the one excluded parameter value is $r = 0$.

Why does the answer then follow from integrality? Because the fibre-size bookkeeping produces a formula $\frac{p+1+2c}{4}$ with one unknown bit c, and a count is an integer. Kedlaya's Solution 1 uses exactly this step, and it looks like sleight of hand: the problem seems to require the theorem "-1 is a square iff $p \equiv 1 \pmod 4$", and instead the argument *proves* that theorem. Here is where that comes from. A counting identity is an equation between integers; any unknown quantity appearing in it with a small coefficient is pinned down mod that coefficient. Whenever a count comes out as "(known) + (unknown small correction)" over a denominator, check whether divisibility fixes the correction before reaching for a theorem.

Finally, why does the natural attempt stall exactly where it does? The Legendre-sum route is correct but it keeps a as the variable, so the dependence between $\chi(a)$ and $\chi(a - 1)$ appears as an opaque sum. The substitution that evaluates it, $\chi(a)\chi(a - 1) = \chi(1 - a^{-1})$, is a change of variable – the same move as passing from a to a parameter on the curve. Both routes need the insight "stop indexing by a"; the model chain has it at line two, the character-sum chain at line ten.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that the set is an intersection of two *images* of squaring-type maps. On seeing it, write down the equation of coincidence: $x^2 = y^2 + 1$, i.e. $x^2 - y^2 = 1$.
2. It suggests counting solutions of the equation (points on a conic), then converting to a count of values through the fibre sizes of $(x, y) \mapsto x^2$.
3. Hesitation occurs at the division by 4: the fibres are not all of size 4. What reinforces it away is asking explicitly "when is $x = -x$ or $y = -y$?" – only at 0 in odd characteristic – so the exceptions are $a = 0$ and $a = 1$, each with a fibre of size 2.
4. The trap is the independence heuristic: "half the numbers are squares, so a quarter of pairs are both squares". It gives the right size and no proof, and the attempt to prove it leads into a character sum that looks harder than the problem. A second trap is forgetting 0 is a square, which loses the $a = 0$ case when $p \equiv 1 \pmod 4$.
5. Whenever you are asked to count the elements of an image, or of an intersection of images, count the pairs that map there and divide by fibre sizes. Whenever an equation is a difference of squares over a field, factor it and parametrise by one factor.
6. Count values via witnesses: replace "how many a have a representation" by "how many representations are there", compute that by a bijection or parametrisation, then divide by the (generic) number of representations per a, handling degenerate fibres separately; if an unknown bit survives, let integrality of the final count decide it.

▷ **Cue:** count the elements of $\{f(x)\} \cap \{g(y)\}$ $\longrightarrow$ **Move:** count solutions of $f(x) = g(y)$, then divide by fibre sizes
▷ **Cue:** an equation $x^2 - y^2 = c$ with $c \neq 0$ over a field of odd characteristic $\longrightarrow$ **Move:** substitute $u = x + y$, $v = x - y$; the solution set is $uv = c$, exactly $p - 1$ points
▷ **Cue:** dividing a count of solutions by the number of sign choices $(\pm x, \pm y)$ $\longrightarrow$ **Move:** isolate the values where $x = 0$ or $y = 0$; those fibres are smaller
▷ **Cue:** a count comes out as $\frac{\text{known}+2c}{4}$ with an unknown $c \in \{0, 1\}$ $\longrightarrow$ **Move:** use integrality to determine c before invoking any theorem
▷ **Cue:** a heuristic "the events are independent, so the probability multiplies" $\longrightarrow$ **Move:** treat it as a guess for the answer only; convert the joint condition into a single equation in more variables

1991 B6

1991 B6. Let a and b be positive numbers. Find the largest number c, in terms of a and b, such that

$$a^x b^{1-x} \leq a\frac{\sinh ux}{\sinh u} + b\frac{\sinh u(1-x)}{\sinh u}$$

for all u with $0 < |u| \leq c$ and for all x, $0 < x < 1$. (Note: $\sinh u = (e^u - e^{-u})/2$.)

Fields: real analysis (inequalities, hyperbolic functions), ordinary differential equations (comparison / maximum principle) **Position:** B6 – last and hardest problem of the second session; in the official tally of the top 213 papers, 2 scored full marks, 30 scored 0 and 173 left it blank **Answer:** $c = |\ln(a/b)|$. The inequality holds for every $x \in (0, 1)$ when $0 < |u| \leq |\ln(a/b)|$, is an identity in x at $|u| = |\ln(a/b)|$, and fails for every $x \in (0, 1)$ when $|u| > |\ln(a/b)|$. (When $a = b$ this gives $c = 0$: no nonzero u works.)

Layer 1 THE WHAT — surface analysis

The natural attempt. You see two sides that agree at the ends: at $x = 0$ both equal b, at $x = 1$ both equal a. So you look at the slope at $x = 0$, since an inequality that is tight at an endpoint is decided, to first order, there. Write $k = \ln(a/b)$, so $a = be^k$ and the left side is be^{kx} with slope bk

at 0. The right side has slope

$$\frac{au\cosh 0 - bu\cosh u}{\sinh u} = \frac{u\,(a - b\cosh u)}{\sinh u}$$

at 0. The inequality near $x = 0$ needs $u(e^k - \cosh u)/\sinh u \geq k$. Try $u = k$ (with $k > 0$): the left side is $k\sinh k/\sinh k = k$. Equality. That is a strong hint that $c = |k|$, and it is the right answer.

Now you try to prove it. Clearing $\sinh u$ gives the two-variable inequality $be^{kx}\sinh u \leq be^k \sinh ux + b\sinh u(1-x)$, which you expand into exponentials: eight terms in $e^{\pm u}$, $e^{\pm ux}$, e^{kx}. You try to factor it and cannot, because k and u are unrelated except at the point you care about. You try Taylor series in u: the right side is $ax + b(1-x) - \frac{u^2}{6}x(1-x)\big[a(1+x) + b(2-x)\big] + O(u^4)$, which confirms that the right side falls as $|u|$ grows from 0 and that the $u \to 0$ limit is weighted AM–GM. But it gives only local information near $u = 0$, and $c = |k|$ need not be small. The endpoint slope gave necessity at best (and even that needs the slope function to be monotone in u, which you have not shown); sufficiency for every interior x is a two-parameter inequality in transcendental functions, and brute force has no handle on it. This is the wall.

The model chain. *Reductions.* sinh is odd, so each fraction $\sinh(ux)/\sinh u$ is even in u; the right side is even in u and it suffices to treat $u > 0$. Put $k = \ln(a/b)$, so $a^x b^{1-x} = be^{kx}$.

Recognise both sides as solutions of one differential equation. For fixed $u > 0$ define on $[0,1]$

$$f(x) = a\frac{\sinh ux}{\sinh u} + b\frac{\sinh u(1-x)}{\sinh u}, \qquad g(x) = be^{kx}.$$

Then $f'' = u^2 f$ and $g'' = k^2 g$, and $f(0) = g(0) = b$, $f(1) = g(1) = a$. Let $h = f - g$, the gap we want to be ≥ 0. Then $h(0) = h(1) = 0$ and

$$h'' - u^2 h = u^2 f - k^2 g - u^2 f + u^2 g = (u^2 - k^2)\,g. \qquad (*)$$

Since $g > 0$, the sign of the right side of $(*)$ is the sign of $u^2 - k^2$, the same for every x.

Case $0 < u \leq |k|$: the inequality holds. Suppose $h(x_1) < 0$ for some $x_1 \in (0,1)$. h is continuous on $[0,1]$ and zero at both ends, so it attains a negative minimum at some interior x_0, where $h''(x_0) \geq 0$. Then $h''(x_0) - u^2 h(x_0) \geq -u^2 h(x_0) > 0$, while $(*)$ gives $(u^2 - k^2)g(x_0) \leq 0$. Contradiction. So $h \geq 0$ on $(0,1)$ for every $u \in (0, |k|]$.

Case $u > |k|$: the inequality fails at every x. Now the right side of $(*)$ is strictly positive. If h were positive somewhere, it would have a positive interior maximum at some x_0 with $h''(x_0) \leq 0$, giving $h''(x_0) - u^2 h(x_0) < 0$, contradiction. So $h \leq 0$. If $h(x_0) = 0$ at an interior point, that point is a maximum, $h''(x_0) \leq 0$ and $h''(x_0) - u^2 h(x_0) \leq 0$, again contradicting $(*)$. Hence $h < 0$ on all of $(0,1)$: the inequality fails for every x.

Conclusion. The inequality holds for all $0 < |u| \leq c$ and all $x \in (0,1)$ exactly when $c \leq |k|$. So the largest c is $|\ln(a/b)|$. When $a = b$, $k = 0$, every $u \neq 0$ lies in the failing case, and the largest c is 0 (the condition is vacuous there). As a check on the boundary case $u = |k| > 0$: $(*)$ says $h'' = k^2 h$ with $h(0) = h(1) = 0$, whose only solution is $h \equiv 0$; directly, $be^k \sinh kx + b\sinh k(1-x) = \frac{b}{2}\big(e^{k+kx} - e^{-k+kx}\big) = be^{kx}\sinh k$, so the two sides are identical.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree up to the observation that the two sides match at $x = 0$ and $x = 1$, and the natural attempt even finds the answer by testing $u = k$. They split at the next line. The natural attempt treats the right side as a formula in two variables and tries to manipulate it (clear denominators, expand, Taylor-expand). The model asks what *kind of function of x* each side is, sees that each is a solution of $y'' = \lambda^2 y$ with the same boundary values, and subtracts to get $(*)$.

> *Both sides equal b at $x = 0$ and a at $x = 1$. So this is a comparison of two curves with the same endpoints. What equation does each curve satisfy?* $\sinh ux$ *and* $\sinh u(1-x)$ *both have second derivative u^2 times themselves, and be^{kx} has second derivative k^2 times itself.*

> *At $u = k$ the two sides solve the same equation with the same boundary values, so they must be the same function. That is why $u = k$ gave equality – it is not a coincidence I should verify, it is uniqueness.*

> *The difference h vanishes at both ends and satisfies $h'' - u^2 h = (u^2 - k^2)g$ with a sign that does not depend on x. A function with fixed-sign $h'' - u^2 h$ and zero boundary values cannot dip the wrong way: look at its extreme point.*

Layer 3 THE WHY

Why is "what differential equation does each side solve?" a natural question and not a rabbit from a hat? Start from the definition. sinh is built from e^u and e^{-u}, and $e^{\pm ux}$ are precisely the two independent solutions of $y'' = u^2 y$. So $\sinh ux$ and $\sinh u(1-x)$ are just two convenient solutions of that equation: the first vanishes at $x = 0$, the second at $x = 1$. Divide each by $\sinh u$ and they become the pair that equals 1 at the other end. So the right-hand side is, by construction, *the solution of $y'' = u^2 y$ with $y(0) = b$ and $y(1) = a$* – the "hyperbolic interpolant" between b and a, the exact analogue of the linear interpolant $ax + b(1-x)$ (which is the solution of $y'' = 0$, the $u \to 0$ limit). The left side $a^x b^{1-x} = be^{kx}$ is the *geometric* interpolant, and it too solves an equation of that family, $y'' = k^2 y$. Once you see that the problem's right side was manufactured as a boundary-value solution, comparing it to another boundary-value solution is the only thing the structure asks you to do. That is the Blueprint's circle: the right side *is* an interpolant, therefore the endpoints and the governing equation are the objects of interest, therefore subtracting and studying h on $[0,1]$ is not a trick.

Why does the extremum argument work, and why should you reach for it? h is zero at both ends. The question "is $h \geq 0$ on $(0,1)$?" is a question about where h can have a negative minimum. At an interior minimum you know the sign of h'' (second-derivative test from A-level) and the sign of h. $(*)$ ties those together with a right side of known sign. That is the one-dimensional maximum principle, and it needs nothing beyond knowing that $h'' \geq 0$ at an interior minimum.

Why is the threshold $|k|$ and why is it sharp? $h'' - u^2 h$ changes sign exactly when u^2 crosses k^2: the governing equation of the right side becomes "more curved" than that of the left. A solution of $y'' = \lambda^2 y$ between fixed positive endpoints sags lower in the middle as λ grows (a large λ makes the solution hug 0 away from the ends). So the right side sinks as u grows, and it passes through the left side exactly when the two equations coincide.

Now the official solution. Kedlaya, Poonen and Vakil's Solution 1 proves that for each fixed x, $u \mapsto \sinh(\alpha u)/\sinh u$ is decreasing for $u > 0$ and $0 < \alpha < 1$, so the gap $F(u)$ decreases in u; it vanishes at $u = \ln(a/b)$, and the answer follows. Their motivation says the substitution $v = r$ (with $v = e^u$, $r = a/b$) needs "some inspiration". Here is where that comes from: $v = r$ is $u = k$, the point where the two defining equations coincide, so uniqueness forces the two sides to agree. Their Solution 2 shows $\ln\big(a\sinh ux + b\sinh u(1-x)\big)$ has second derivative with the sign of $2ab\cosh u - a^2 - b^2$, so it is concave exactly when $u \leq |k|$; concavity of a logarithm is the statement that the function lies above its geometric interpolant. All three proofs are the same fact seen from different sides. The comparison proof above is my own; I verified each step and checked the conclusion numerically on 20,000 random (a, b, u, x).

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that the two sides agree at $x = 0$ (both b) and $x = 1$ (both a). On

seeing it, write down: "compare two functions of x on $[0,1]$ with equal boundary values; set $h = \text{RHS} - \text{LHS}$, $h(0) = h(1) = 0$."

2. It suggests asking what equation each side satisfies in x. $\sinh ux$ and e^{kx} point to $y'' = \lambda^2 y$, and the method is to subtract the equations and use the sign of $h'' - u^2 h$ at an interior extremum (the maximum principle).

3. Hesitation occurs after finding equality at $u = k$: you suspect the answer but do not know how to prove it for all x at once. What removes it is the uniqueness thought: equality at $u = k$ for every x means the two sides are the same boundary-value solution, so the general case is a perturbation of the equation, not of the formula.

4. The trap is expanding into exponentials or Taylor series in u and fighting an eight-term, two-parameter transcendental inequality; the Taylor route only sees u near 0 while the answer $|k|$ can be large. A second trap is forgetting $a = b$, where the answer is $c = 0$, and forgetting that the right side is even in u.

5. Recognise faster by reading the special functions as solutions of their differential equations: sinh, cosh, $e^{\lambda x}$ solve $y'' = \lambda^2 y$; sin, cos solve $y'' = -\lambda^2 y$. A normalised combination such as $\frac{\sinh ux}{\sinh u}$ that is 0 at one end and 1 at the other is a boundary-value interpolant.

6. When an inequality compares two functions that agree at the ends of an interval, find the linear differential operator each one satisfies; apply that operator to the difference and use the sign at an interior extremum. The parameter where the two operators coincide is the threshold.

▷ **Cue:** an expression of the form $A\frac{\sinh ux}{\sinh u} + B\frac{\sinh u(1-x)}{\sinh u}$ $\longrightarrow$ **Move:** recognise the solution of $y'' = u^2 y$ with $y(0) = B$, $y(1) = A$

▷ **Cue:** both sides of an inequality in x agree at the two endpoints of an interval $\longrightarrow$ **Move:** set $h =$ difference, note $h = 0$ at both ends, and study where h could have an interior extremum of the wrong sign

▷ **Cue:** $h(0) = h(1) = 0$ and $h'' - \lambda h$ has a fixed sign on the interval ($\lambda > 0$) $\longrightarrow$ **Move:** maximum principle: at an interior extremum h'' and h have known signs; derive a contradiction

▷ **Cue:** a substitution makes an inequality an identity in the free variable $\longrightarrow$ **Move:** ask which uniqueness theorem forces the identity; the threshold parameter is where the two defining equations coincide

▷ **Cue:** $a^x b^{1-x}$ appears $\longrightarrow$ **Move:** write it as be^{kx} with $k = \ln(a/b)$: a geometric interpolant, solution of $y'' = k^2 y$

▷ **Cue:** "find the largest c such that the inequality holds for all $0 < |u| \le c$" $\longrightarrow$ **Move:** check evenness in u, look for the parameter value giving equality, then prove monotone behaviour on each side of it