DAWN PUTNAM SESSION

Putnam 1998 — Papers A and B

Putnam 120 Archive

Contents

Putnam 1998 — Paper A **5**
1998 A1 5
1998 A2 7
1998 A3 9
1998 A4 11
1998 A5 14
1998 A6 17

Putnam 1998 — Paper B **21**
1998 B1 21
1998 B2 23
1998 B3 25
1998 B4 28
1998 B5 30
1998 B6 33

Putnam 1998 — Paper A

1998 A1

1998 A1. A right circular cone has base of radius 1 and height 3. A cube is inscribed in the cone so that one face of the cube is contained in the base of the cone. What is the side-length of the cube?

Fields: solid geometry, similar triangles **Position:** A1 — the opener **Answer:** The side length is $s = \dfrac{9\sqrt{2}-6}{7}$, forced by requiring the cube's *top-face corners* — not its top-edge midpoints — to sit on the cone's lateral surface.

Layer 1 THE WHAT — surface analysis

The natural attempt. The instinct is to slice. You draw the axial cross-section that looks most natural to draw: the plane through the cone's axis and the midpoints of two opposite edges of the cube's top face. In that plane the cone's cross-section is the isosceles triangle you expect, base 2, height 3; and the cube's cross-section is a rectangle of width s (the cube's own side) sitting on the base, rising to height s.

You inscribe this rectangle in the triangle by similar triangles: at height s, the triangle's half-width is $1 - s/3$, so the rectangle's half-width $s/2$ must equal that,

$$\frac{s}{2} = 1 - \frac{s}{3} \implies \frac{s}{2} + \frac{s}{3} = 1 \implies \frac{5s}{6} = 1 \implies s = \frac{6}{5}.$$

This is a complete, clean derivation, and every step is correct *for the rectangle you drew*. The number $6/5 = 1.2$ even feels plausible: a cube of side just over 1 tucked into a cone of height 3, base radius 1.

It is wrong. Check it: if $s = 6/5$, the top face of the cube is a square of side $6/5$ centred on the axis at height $6/5$. Its corners sit at horizontal distance $\frac{s\sqrt{2}}{2} = \frac{6\sqrt{2}}{10} = \frac{3\sqrt{2}}{5} \approx 0.849$ from the axis. The cone's radius at height $6/5$ is $1 - \frac{6/5}{3} = 1 - \frac{2}{5} = \frac{3}{5} = 0.6$. The corners, at 0.849, are already outside the cone at 0.6 — the cube you just constructed pokes through the cone's wall. The cross-section through edge-midpoints was inscribed correctly in its own plane; it simply wasn't the constraint that matters.

The model chain. The cube touches the cone's lateral surface where it is furthest from the axis, and for a square cross-section at a given height, that is at the *corners*, not the edge midpoints. So take the axial cross-section through the axis and two *opposite vertices* of the cube's bottom (and top) face. In this plane, the cone's cross-section is the same isosceles triangle, base 2, height 3. The cube's cross-section is now a rectangle of width equal to the square's diagonal, $s\sqrt{2}$, and height s, sitting on the base.

Inscribe this rectangle by similar triangles: at height s, the triangle's half-width is $1 - s/3$, and the rectangle's half-width is $\frac{s\sqrt{2}}{2}$, so

$$\frac{s\sqrt{2}}{2} = 1 - \frac{s}{3}.$$

Multiply through by 6:

$$3s\sqrt{2} = 6 - 2s \implies s(3\sqrt{2}+2) = 6 \implies s = \frac{6}{3\sqrt{2}+2}.$$

Rationalise by multiplying by $\frac{3\sqrt{2}-2}{3\sqrt{2}-2}$, using $(3\sqrt{2})^2 - 2^2 = 18 - 4 = 14$:

$$s = \frac{6(3\sqrt{2}-2)}{14} = \frac{3(3\sqrt{2}-2)}{7} = \frac{9\sqrt{2}-6}{7}.$$

Numerically $s \approx 0.961$, a slightly smaller cube than the wrong 1.2 — exactly what "the true constraint bites earlier" should produce.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains draw an axial cross-section and apply the same similar-triangles machinery to a rectangle inscribed in the same triangle. They split at the moment of *choosing which cross-section to draw*: the natural attempt slices through the edge midpoints, producing a rectangle of half-width $s/2$; the model slices through the diagonal, producing a rectangle of half-width $s\sqrt{2}/2$. Everything after that choice is mechanically identical — correct similar-triangle algebra applied to the wrong (respectively right) rectangle.

> *The cube doesn't touch the cone all the way around its top edge at once; it touches where it's furthest out, and for a square that's the corners, at distance $s\sqrt{2}/2$ from the centre, not the edge midpoints at distance $s/2$. So the cross-section I need is the one that actually passes through the farthest point, and I should ask "which point of this top face is furthest from the axis?" before I pick a slicing plane at all — not slice first and hope the plane I happened to draw was the binding one.*

Layer 3 THE WHY

Why does slicing through the wrong pair of points produce a self-consistent but false answer, rather than an obviously broken one? Because "inscribe a rectangle in a triangle" is a well-posed 2D problem regardless of which rectangle you hand it; the 2D algebra has no way to know it has been fed the cross-section of a square that is not, in 3D, tangent to the cone along its actual widest points. The cone is a surface of revolution: every point on it is equidistant, in the relevant sense, from the axis at a rate fixed by height alone, $r(z) = 1 - z/3$. The binding constraint on an inscribed solid of revolution's complement is always "does the solid stay inside $r \le r(z)$ at every height," and for a square cross-section at height $z = s$, the quantity that must satisfy that inequality with equality (for the largest inscribable cube) is the *maximum* distance from the axis at that height, which is achieved at the corners of the square, distance $s\sqrt{2}/2$, not at the edge midpoints, distance $s/2$. This is the same fact that makes a square peg's binding dimension its diagonal, not its side, when you're asking whether it fits through a round hole: the constraining direction is whichever one is largest, and axial symmetry means it is always the corner. The edge-midpoint cross-section is a legitimate slice of the true 3D configuration — it's just not the slice where the cone's wall is actually reached, so inscribing a rectangle *in that plane* answers a different, easier question than the one asked.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "cube inscribed in a solid of revolution, one face on the base" – on seeing it, write down that the cone's cross-sectional radius is a function of height alone, $r(z) = 1 - z/3$, before choosing any cutting plane.
2. That cue suggests the real question is not "what does some cross-section look like" but "what is the furthest point of the cube from the axis at each height," since that is what the radius function constrains.

3. Hesitation occurs at the moment of picking which axial plane to slice through; edge-midpoints and diagonal both look like reasonable choices and both hand you a clean triangle-inscribed-in-triangle picture. It is reinforced away by asking, before slicing, which points of the square are actually farthest from the centre – the corners, at $s\sqrt{2}/2$ – and slicing through those.

4. The trap is that the wrong cross-section (edge midpoints) is entirely self-consistent within its own plane: the similar-triangles algebra has no internal check that would flag it as answering the wrong question, so a confidently wrong answer emerges with no visible error.

5. Recognise this family faster by treating "square (or any non-circular) cross-section inscribed inside a surface of revolution" as an automatic instruction to identify the cross-section's furthest point from the axis of symmetry first, and slice through that.

6. Reusable rule: when a solid with corners is inscribed inside a solid of revolution, the binding constraint at any height is always the corner furthest from the axis, not an edge or face midpoint – reduce the 3D inscribing problem to a 2D one only after locating that furthest point, never before.

▷ **Cue:** a solid with flat, cornered faces (cube, square, polygon) is inscribed inside a solid of revolution (cone, sphere, cylinder) $\longrightarrow$ **Move:** find the point of the cornered solid furthest from the axis of revolution at the relevant height, and build the cross-section through that point, not through an edge or face midpoint

▷ **Cue:** an axial cross-section through a cube produces a clean rectangle-in-triangle similar-triangles problem $\longrightarrow$ **Move:** check that the plane chosen passes through the cube's actual extreme points before trusting the answer – any axial plane produces an equally clean-looking but potentially wrong picture

1998 A2

1998 A2. Let s be any arc of the unit circle lying entirely in the first quadrant. Let A be the area of the region lying below s and above the x-axis and let B be the area of the region lying to the right of the y-axis and to the left of s. Prove that $A + B$ depends only on the arc length, and not on the position, of s.

Fields: calculus, coordinate geometry **Position:** A2 **Answer:** $A + B$ equals the radian measure of s – i.e. $A + B$ equals the arc length of s, since the circle has radius 1.

Layer 1 THE WHAT — surface analysis

The natural attempt. The problem hands you two regions and asks about their sum, so the natural first move is to compute each one on its own. Parametrise s by angle, $\theta \in [\theta_1, \theta_2] \subset [0, \pi/2]$, so a point of s is $(\cos\theta, \sin\theta)$. Region A sits under the arc, above the x-axis, between the verticals dropped from the arc's endpoints; that is the ordinary area-under-a-curve integral, run in x from $\cos\theta_2$ to $\cos\theta_1$ (with $\theta_1 < \theta_2$, so $\cos\theta_1 > \cos\theta_2$):

$$A = \int_{\theta_1}^{\theta_2} \sin\theta \cdot \sin\theta \, d\theta = \int_{\theta_1}^{\theta_2} \sin^2\theta \, d\theta = \left[\frac{\theta}{2} - \frac{\sin 2\theta}{4}\right]_{\theta_1}^{\theta_2}.$$

Region B is the mirror computation in y:

$$B = \int_{\theta_1}^{\theta_2} \cos^2\theta \, d\theta = \left[\frac{\theta}{2} + \frac{\sin 2\theta}{4}\right]_{\theta_1}^{\theta_2}.$$

This works. Add them and the $\sin 2\theta$ terms cancel:

$$A + B = \left[\frac{\theta_2}{2} - \frac{\sin 2\theta_2}{4}\right] - \left[\frac{\theta_1}{2} - \frac{\sin 2\theta_1}{4}\right] + \left[\frac{\theta_2}{2} + \frac{\sin 2\theta_2}{4}\right] - \left[\frac{\theta_1}{2} + \frac{\sin 2\theta_1}{4}\right] = \theta_2 - \theta_1.$$

It's a complete, correct proof. But notice what happened: two separate double-angle expansions, four bracket evaluations, and a cancellation you only discover *after* doing all the work of computing each half-angle formula in full. Nothing told you in advance that the position-dependent terms would cancel; you had to grind out both integrals to find out.

The model chain. Don't compute A and B separately – compute $A + B$ directly, as a single line integral. Region A's area, run as θ decreases from θ_2 to θ_1 (so x increases), is $-\int y\,dx$; region B's area is $\int x\,dy$. So

$$A + B = \int_{\theta_1}^{\theta_2} (x\,dy - y\,dx).$$

On the unit circle, $x = \cos\theta$, $y = \sin\theta$, $dx = -\sin\theta\,d\theta$, $dy = \cos\theta\,d\theta$, so

$$x\,dy - y\,dx = \cos\theta \cdot \cos\theta\,d\theta - \sin\theta \cdot (-\sin\theta\,d\theta) = (\cos^2\theta + \sin^2\theta)\,d\theta = d\theta.$$

Therefore

$$A + B = \int_{\theta_1}^{\theta_2} d\theta = \theta_2 - \theta_1,$$

the radian measure – and, since the circle has radius 1, the arc length – of s, with no reference to where θ_1 and θ_2 individually sit. The Pythagorean identity kills the position dependence in one line instead of two double-angle expansions that happen to cancel.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains parametrise by θ and both end up integrating a combination of $\sin^2\theta$ and $\cos^2\theta$-type quantities. They split at the very first decision: whether to write down A and B as two separate integrals to be evaluated and then added, or to write down the single quantity $A + B$ as one integral, $\int (x\,dy - y\,dx)$, before evaluating anything.

> *I don't actually need A and B individually – the problem only ever asks about $A + B$. So I should form the combined integrand $x\,dy - y\,dx$ before I do any calculus, not after, because whatever cancellation is going to happen is far more likely to be visible in the combined expression than in two separate half-angle expansions that I'd have to fully expand first to notice they cancel.*

Layer 3 THE WHY

Why does forming the sum first, rather than adding two finished answers, make the position-dependence vanish so cheaply? Because $x\,dy - y\,dx$ is not an arbitrary combination – on the unit circle it equals $x^2\,d\theta + y^2\,d\theta = (x^2 + y^2)\,d\theta$, and $x^2 + y^2 = 1$ is exactly the equation that defines the circle. The quantity $A + B$ was engineered by the problem (region below the arc, plus region to the left of the arc) to be precisely the pairing that turns into $x\,dy - y\,dx$: one region contributes $-y\,dx$, the other $+x\,dy$, and together these are the two halves of the identity that collapses to $d\theta$ only because every point of s satisfies the unit circle's defining equation. If s lay on some other curve, $x^2 + y^2$ would not be constant, and $A + B$ would genuinely depend on position. The reason the answer must reduce to the radian measure of the arc, rather than to something else, is that $d\theta = x\,dy - y\,dx$ is the two-dimensional statement of "the constraint $x^2 + y^2 = 1$ is what makes θ, and only θ, matter" – computing A and B separately buries this fact inside two Pythagorean-identity-flavoured integrals ($\sin^2 + \cos^2$ reappearing as $\frac{1}{2} - \frac{1}{2}\cos 2\theta$ and $\frac{1}{2} + \frac{1}{2}\cos 2\theta$) and only exposes it after both have been fully expanded and added.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that the problem asks about a *sum* of two areas, $A + B$, not about either one alone. On seeing "prove that a sum of two quantities is position-independent," write down the sum as a single expression before computing either summand.

2. That cue suggests representing both areas as line integrals in the same variable (θ) and combining them into one integrand, $x\,dy - y\,dx$, rather than evaluating two separate definite integrals and adding the results afterward.
3. Hesitation occurs at the decision of how to set up the integrals at all – separate-and-add feels like the default because A and B are defined as two distinct regions. It is reinforced away by noticing the problem never actually needs A and B individually, only their sum, which licenses combining first.
4. The trap is that the separate-integrals route is not wrong, only expensive: it produces two correct half-angle expansions that happen to cancel, so a solver who grinds through it gets full marks but spends the exam discovering by brute force a cancellation that combining first would have revealed for free.
5. Recognise this family faster by treating "prove a sum of two quantities is independent of some parameter" as a standing instruction to try forming the sum symbolically first and simplifying that, before ever substituting in and integrating either piece.
6. Reusable rule: when a curve satisfies a defining constraint (here $x^2 + y^2 = 1$) and a problem asks about a sum of two areas or quantities built from that curve, form the combined expression $x\,dy - y\,dx$ (or the analogous combination) and substitute the constraint into it directly – the constraint is usually exactly what makes the combined quantity collapse, and evaluating summands separately hides that mechanism until the end.

▷ **Cue:** a problem asks to prove a sum of two areas/quantities attached to a curve is independent of position, only of some other parameter (arc length, etc.) $\longrightarrow$ **Move:** form the combined expression as a single integral before evaluating either piece separately; look for the curve's defining equation to appear directly inside the combined integrand

▷ **Cue:** a curve satisfies $x^2 + y^2 = \text{const}$ (unit circle or any circle) $\longrightarrow$ **Move:** expect $x\,dy - y\,dx$ to reduce to (radius)$^2\,d\theta$, i.e. to angular measure, whenever that combination appears

1998 A3

1998 A3. Let f be a real function on the real line with continuous third derivative. Prove that there exists a point a such that

$$f(a) \cdot f'(a) \cdot f''(a) \cdot f'''(a) \geq 0.$$

Fields: real analysis (monotonicity, convexity, contradiction) **Position:** A3 **Answer:** There always exists such an a; the proof is by contradiction, normalising signs via $f \mapsto -f$ and $f(x) \mapsto f(-x)$, then showing a global strict-negativity assumption forces every one of f, f', f'', f''' to be eventually forced positive, which is self-contradictory.

Layer 1 THE WHAT — surface analysis

The natural attempt. The instinct on seeing four functions multiplied together with an inequality to prove is to look for a relation between consecutive ones – Rolle's theorem, or the Mean Value Theorem applied to f'' and f''', say. You might try: pick any a with $f''(a) = 0$ if one exists (then you're done, the product is 0). If not, f'' never vanishes, so by the Intermediate Value Theorem it has constant sign, say $f'' > 0$ everywhere. Then you reach for Taylor's theorem with remainder around some point, trying to pin down the sign of $f(a)f'(a)f''(a)f'''(a)$ at a specific, cleverly-chosen a – perhaps a point where f' is minimised, or where f''' takes some extremal-looking value.

This line of attack is not unreasonable, but it stalls. There is no single point a that the problem hands you for free – no critical point of f, no root, nothing guaranteed to exist by the hypotheses alone. Trying to manufacture one directly, via Taylor expansions or MVT applied between two named points, produces sprawling case-work (what if f''' is positive at one end

and negative at the other? what if f'' has a root but f doesn't?) without ever closing off all the possibilities. The natural attempt is trying to *find a*; the problem is actually asking you to show one *must* exist, and those are different tasks.

The model chain. Argue by contradiction: suppose $f(x)f'(x)f''(x)f'''(x) < 0$ for every real x (if the product is ever ≥ 0 somewhere, take that as a and stop). In particular none of f, f', f'', f''' vanishes anywhere – if one did, the product would be 0 there, contradicting strict negativity – so by continuity and the Intermediate Value Theorem, each of the four functions has *constant sign* on all of $\mathbb{R}$.

Now normalise. Replacing $f(x)$ by $-f(x)$ flips the sign of all four functions simultaneously, so the sign of their product (four factors, $(-1)^4 = 1$) is unchanged; use this freedom to assume $f''(x) > 0$ for all x. Replacing $f(x)$ by $f(-x)$ flips the sign of f' and f''' only (the odd derivatives) while leaving f and f'''s signs alone, and reflects the product pointwise without changing its sign anywhere; use this freedom to assume $f'''(x) > 0$ for all x too.

With $f'' > 0$ everywhere, f' is strictly increasing; with $f''' > 0$ everywhere, f' is strictly convex, so for every x and every $a > 0$,

$$f'(x + a) > f'(x) + a\, f''(x).$$

Fix x; as $a \to \infty$ the right side $\to +\infty$ since $f''(x) > 0$ is a fixed positive number. So $f'(x + a) \to +\infty$, meaning f' is unbounded above. Since f' has constant sign on $\mathbb{R}$ and is unbounded above, that sign must be positive: $f'(x) > 0$ for all x.

Run the same argument one derivative down: $f' > 0$ everywhere means f is strictly increasing; convexity is not needed again, but the identical unboundedness trick applies using $f'(x) > 0$ fixed:

$$f(x + a) > f(x) + a\, f'(x) \xrightarrow[a\to\infty]{} +\infty,$$

so f is unbounded above, and since f has constant sign, that sign must be positive: $f(x) > 0$ for all x.

So $f > 0$, $f' > 0$, $f'' > 0$, $f''' > 0$ everywhere, and their product is strictly positive – contradicting the assumption that it was strictly negative everywhere. Hence the assumption fails: there exists a with $f(a)f'(a)f''(a)f'''(a) \geq 0$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains open with "handle the case where one factor vanishes trivially, then reduce to constant signs via IVT." They split at what to do with four functions of constant, but a priori unrelated, sign. The natural attempt tries to *locate* a specific point a where the signs cooperate. The model chain instead assumes, for contradiction, that they never cooperate anywhere, and shows that assumption is globally untenable – it never needs to name a point at all until the very end, and even then only implicitly, via the contradiction.

> *I don't need to find the point a where the product is non-negative – I need to show that "the product is negative* everywhere*" cannot happen. That's a much stronger, more rigid statement to attack: if it's negative everywhere, then every one of the four functions is nonzero everywhere, hence has one fixed sign for all of* $\mathbb{R}$ *by the Intermediate Value Theorem – and a function with one fixed sign forever, combined with a growth estimate from convexity, is something I can actually pin down.*

Layer 3 THE WHY

Why does assuming global negativity, rather than searching for a witness point, unlock the argument? Because "negative everywhere" is a statement about *all* of $\mathbb{R}$ at once, and a function that is nonzero and continuous on all of $\mathbb{R}$ is forced by the Intermediate Value Theorem to have one sign forever – that is the exact mechanism converting a pointwise hypothesis (product < 0 at a) into a global one (each factor has fixed sign on $\mathbb{R}$), and global hypotheses are what convexity

and monotonicity estimates are built to use. A single witness point carries no structure to exploit; a function of fixed sign for all x does, because you can now ask what happens as $x \to +\infty$, and $f'' > 0$ everywhere plus $f''' > 0$ everywhere is precisely the pair of hypotheses that force f' to blow up in the positive direction (convexity gives a linear-in-a lower bound whose slope is the fixed positive number $f''(x)$, and a linear lower bound with positive slope is unbounded). The normalisation step ($f \mapsto -f$, $f(x) \mapsto f(-x)$) is not decoration: it is what makes "the four functions have some fixed signs" collapse to a single case (f, f', f'', f''' all positive, WLOG) instead of sixteen, because both substitutions preserve the sign of the four-fold product exactly (one flips all four signs, $(-1)^4 = 1$; the other flips only the two odd-order derivatives, which leaves the pointwise sign of the whole product exactly where it was, just reflected in x). Once the case count is down to one, the chain "$f'' > 0$ and $f''' > 0$ force $f' \to +\infty$, hence (having fixed sign) $f' > 0$ everywhere, hence by the same mechanism $f > 0$ everywhere" is not four separate miracles but one estimate applied twice, because convexity-implies-unbounded and monotone-implies-unbounded are the same growth argument one derivative apart.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "prove there exists a point where an inequality holds" phrased as the negation of "the opposite strict inequality holds everywhere." On seeing it, write down the contradiction hypothesis (product < 0 for all x) immediately, rather than trying to construct a witness point directly.

2. That cue suggests, once the contradiction hypothesis is in hand, checking what "for all x" plus continuity and non-vanishing forces via the Intermediate Value Theorem – namely, constant sign for each factor on all of $\mathbb{R}$.

3. Hesitation occurs right after establishing constant signs, at the question "now what?" – four functions of fixed but a priori independent sign look like they need case analysis. This is reinforced away by noticing the substitutions $f \mapsto -f$ and $f(x) \mapsto f(-x)$ each preserve the sign of the four-fold product exactly, collapsing sixteen sign patterns to one representative case.

4. The trap is trying to find a explicitly via Taylor's theorem or the Mean Value Theorem applied between two named points, which produces open-ended case-work with no guaranteed termination, because the hypotheses give no explicit point to start from.

5. Recognise this family faster by treating "prove $\exists\, a$ such that (inequality)" as an instruction to consider its negation as a global hypothesis first, and to ask immediately what constant-sign consequences the Intermediate Value Theorem extracts from that negation.

6. Reusable rule: to prove an inequality holds *somewhere*, assume the strict opposite holds *everywhere*, use continuity plus non-vanishing to upgrade that to constant sign via the Intermediate Value Theorem, then exploit whatever growth structure (convexity, monotonicity, boundedness) the problem supplies to show a function of fixed sign cannot actually stay that way – searching for the witness point directly is usually the harder task.

▷ **Cue:** prove there exists a point where a (possibly strict) inequality holds, with no explicit candidate point available $\longrightarrow$ **Move:** assume the opposite strict inequality holds everywhere and derive a contradiction, rather than trying to construct the witness point directly

▷ **Cue:** a continuous, nowhere-vanishing function on all of $\mathbb{R}$ $\longrightarrow$ **Move:** apply the Intermediate Value Theorem to conclude it has one fixed sign on all of $\mathbb{R}$, then use growth or boundedness hypotheses to constrain that sign

▷ **Cue:** a product of several functions and their derivatives, with a symmetry under $f \mapsto -f$ or $f(x) \mapsto f(-x)$ $\longrightarrow$ **Move:** use the sign-preserving substitution to collapse many sign cases to one representative case before doing any further analysis

1998 A4

1998 A4. Let $A_1 = 0$ and $A_2 = 1$. For $n > 2$, the number A_n is defined by concatenating the decimal expansions of A_{n-1} and A_{n-2} from left to right. For example $A_3 = A_2A_1 = 10$, $A_4 = A_3A_2 = 101$, $A_5 = A_4A_3 = 10110$, and so forth. Determine all n such that 11 divides A_n.

Fields: number theory (divisibility, modular arithmetic); recurrences (Fibonacci) **Position:** A4 — middle of the paper, but mostly bookkeeping once concatenation is translated into arithmetic. **Answer:** $11 \mid A_n$ exactly when $n \equiv 1 \pmod 6$, i.e. $n = 1, 7, 13, 19, \ldots$

Layer 1 THE WHAT — surface analysis

The natural attempt. You write out the first few terms, because the problem hands you four of them and invites more:

$$A_1 = 0,\ A_2 = 1,\ A_3 = 10,\ A_4 = 101,\ A_5 = 10110,\ A_6 = 10110101,\ A_7 = 1011010110110,\ \ldots$$

Then you test each for divisibility by 11 with the alternating digit sum (units digit $+$, tens digit $-$, and so on). $A_3 = 10$: $0 - 1 = -1$, no. $A_4 = 101$: $1 - 0 + 1 = 2$, no. $A_5 = 10110$: $0 - 1 + 1 - 0 + 1 = 1$, no. A_6: eight digits, alternating sum 1, no. A_7: thirteen digits, and after some care the alternating sum is 0 — yes. A_8 has 21 digits, A_9 has 34, A_{13} has 233.

The attempt works for the first half-dozen terms and then drowns. The strings grow like Fibonacci numbers, so writing A_{13} out is a page of digits, and you still would not have a *proof* that the pattern "$n = 1, 7, \ldots$" continues. Worse, the alternating sum of A_n is not obviously related to the alternating sums of A_{n-1} and A_{n-2}, because the sign attached to each digit depends on its position from the right, and gluing A_{n-2} onto the right-hand end shifts every digit of A_{n-1} by $\text{len}(A_{n-2})$ places. That shift is the whole problem, and the digit-by-digit approach never names it.

The model chain. *Step 1: concatenation as arithmetic.* Let d_n be the number of decimal digits of A_n. Writing the digits of A_{n-1} and then those of A_{n-2} means

$$A_n = A_{n-1} \cdot 10^{d_{n-2}} + A_{n-2} \qquad (n \geq 3).$$

(Every A_n with $n \geq 2$ begins with the digit 1, so there are no leading-zero issues; $A_1 = 0$ is the single digit "0", so $d_1 = 1$.)

Step 2: the lengths. Concatenation adds lengths: $d_n = d_{n-1} + d_{n-2}$ with $d_1 = d_2 = 1$. So $d_n = F_n$, the Fibonacci numbers $1, 1, 2, 3, 5, 8, \ldots$

Step 3: reduce mod 11. Since $10 \equiv -1 \pmod{11}$,

$$A_n \equiv (-1)^{F_{n-2}} A_{n-1} + A_{n-2} \pmod{11}.$$

Only the parity of F_{n-2} matters. Mod 2 the Fibonacci numbers run $1, 1, 0, 1, 1, 0, \ldots$ with period 3 (from $F_1 \equiv F_2 \equiv 1$: $F_3 \equiv 0$, $F_4 \equiv 1$, $F_5 \equiv 1$, and the pair $(1, 1)$ recurs).

Step 4: iterate. Working mod 11:

$$\begin{aligned} A_3 &\equiv -A_2 + A_1 = -1, & A_4 &\equiv -A_3 + A_2 = 2, & A_5 &\equiv +A_4 + A_3 = 1, \\ A_6 &\equiv -A_5 + A_4 = 1, & A_7 &\equiv -A_6 + A_5 = 0, & A_8 &\equiv +A_7 + A_6 = 1. \end{aligned}$$

(Signs: F_1, F_2 odd give $-$; F_3 even gives $+$; F_4, F_5 odd give $-$; F_6 even gives $+$.) Check against the given values: $10 \equiv -1$, $101 = 99 + 2 \equiv 2$, $10110 = 11 \cdot 919 + 1 \equiv 1$. Consistent.

Step 5: periodicity. $(A_7, A_8) \equiv (0, 1) = (A_1, A_2)$. The sign used to produce A_{n+6} is $(-1)^{F_{n+4}}$, and $F_{n+4} \equiv F_{n-2} \pmod 2$ because $3 \mid 6$. So the sequence from index 7 onwards obeys exactly the same recurrence, with the same initial pair, as the sequence from index 1; by induction $A_{n+6} \equiv A_n \pmod{11}$ for all $n \geq 1$.

Step 6: read off. Over one period, $A_1, \ldots, A_6 \equiv 0, 1, -1, 2, 1, 1$. Only A_1 vanishes. Hence $11 \mid A_n$ if and only if $n \equiv 1 \pmod 6$. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start by looking at the concatenation. The natural attempt keeps the numbers as *strings* and applies a digit test. The model converts the string operation into the formula $A_n = A_{n-1} \cdot 10^{d_{n-2}} + A_{n-2}$ before doing anything else. The split is the very first line: do you treat "concatenate" as a typographical operation, or as multiplication by a power of 10 followed by addition?

> *Concatenating means shifting A_{n-1} left by however many digits A_{n-2} has, then adding A_{n-2}. So $A_n = A_{n-1} \cdot 10^{len(A_{n-2})} + A_{n-2}$. Now I have a recurrence, and I need to know the lengths.*

> *Mod 11, a power of 10 is ± 1. So I don't need the lengths, only whether they're even or odd — and the lengths are Fibonacci, whose parities go odd, odd, even forever.*

> *The state is the pair (A_{n-1}, A_{n-2}) mod 11 together with where I am in the parity cycle of period 3. If the pair returns to $(0, 1)$ at an index that is 1 more than a multiple of 3, everything repeats.*

Layer 3 THE WHY

Why is the arithmetic translation the natural move and not a trick? Go to the definition of a decimal expansion. The integer whose digits are the string XY is, by the definition of place value, $X \cdot 10^{|Y|} + Y$, where $|Y|$ is the number of digits of Y. That is not a technique; it is what the words "decimal expansion" mean. The problem statement says "concatenating the decimal expansions" — it is telling you the operation lives in base 10, and any divisibility question about a base-10 construction should be pulled back into integers before it is attacked.

Once you have $A_{n-1} \cdot 10^{d_{n-2}} + A_{n-2}$, the choice of modulus becomes transparent. The problem asks about 11, and $11 = 10 + 1$. The one fact about 11 that interacts with base 10 is $10 \equiv -1$. That fact is the *reason* the alternating digit-sum test works in the first place: $\sum a_k 10^k \equiv \sum a_k (-1)^k$. So the natural attempt was already using $10 \equiv -1$; it was using it digit by digit, which is the expensive way. Applying it to the block $10^{d_{n-2}}$ as a whole is the same fact used once instead of F_n times.

Next: why must you track lengths, and why is the Fibonacci recurrence for them immediate? Because concatenation adds lengths, and the recursion for A_n uses the previous two terms, so the recursion for d_n is the same shape: $d_n = d_{n-1} + d_{n-2}$. Any sequence built by gluing the previous two objects has Fibonacci sizes. You are not "noticing" Fibonacci numbers; you are applying the length function to the defining equation.

Finally, periodicity. Kedlaya's solution says "$A_7 \equiv 0$ and $A_8 \equiv 1 \ldots$ it follows that $A_n \equiv A_{n+6}$." Here is what the word "follows" is hiding. A second-order recurrence whose coefficients depend on n is not periodic merely because two consecutive values repeat; the *coefficients* must also be in phase. The coefficient at step n is $(-1)^{F_{n-2}}$, which has period 3. So the full state of the system is (pair of residues, position mod 3). There are at most $11^2 \cdot 3$ states, so eventual periodicity is guaranteed by pigeonhole; the computation shows the state $((0, 1),\ n \equiv 1 \bmod 3)$ recurs at $n = 7$, giving period 6. Write that sentence in a solution — it is where a grader checks that you know why the repetition propagates.

One boundary point worth stating: $A_1 = 0$ is divisible by 11, so $n = 1$ is included in the answer, as it is in the official one ("$n = 6k + 1$ for some nonnegative integer k"). The residues above were also checked by computing A_n in full for $n \leq 25$: the multiples of 11 are exactly $n = 1, 7, 13, 19, 25$.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the word "concatenating". On seeing it, write down $A_n = A_{n-1} \cdot 10^{d_{n-2}} + A_{n-2}$ with d_k the digit count.

2. Reduce the recurrence modulo the given divisor; since $10 \equiv -1 \pmod{11}$, the recurrence becomes $A_n \equiv \pm A_{n-1} + A_{n-2}$ with the sign set by the parity of a length. Then iterate a finite-state system to a repeat.

3. Hesitation occurs at "but I don't know the lengths". Reinforce it away by asking what the modulus needs: only $10^d \bmod 11$, i.e. $d \bmod 2$. A second hesitation is at "the pair repeated, so am I done?": check the coefficient cycle is also in phase.

4. The trap is the digit-sum test. It is correct and feels like progress, and it cannot produce a proof for all n because it never exposes the recurrence. A secondary trap: forgetting $n = 1$, since $A_1 = 0$ is divisible by everything.

5. Slogan: *a construction described in terms of digits is a construction in terms of powers of the base — translate before you test.* And: *when the modulus is $b \pm 1$ for base b, only the parity (or the length itself mod the order of b) of each block matters.*

6. Decision rule: **for any recursively defined sequence asked about modulo m, identify the full finite state (values mod m plus any periodic coefficients), and compute until the state repeats; the period of the answer is the period of the state, not of the values alone.**

▷ **Cue:** a number built by concatenating decimal strings $\longrightarrow$ **Move:** write $\overline{XY} = X \cdot 10^{\mathrm{len}(Y)} + Y$

▷ **Cue:** divisibility by 11 (or by $b+1$ in base b) $\longrightarrow$ **Move:** use $10 \equiv -1$, so $10^d \equiv (-1)^d$; only parities of lengths matter

▷ **Cue:** objects built by gluing the previous two objects $\longrightarrow$ **Move:** their sizes satisfy the Fibonacci recurrence

▷ **Cue:** Fibonacci numbers mod 2 $\longrightarrow$ **Move:** period 3: odd, odd, even

▷ **Cue:** a recurrence mod m with n-dependent coefficients of period p $\longrightarrow$ **Move:** the state is (values, $n \bmod p$); periodicity needs both to recur

1998 A5

1998 A5. Let $\mathcal{F}$ be a finite collection of open discs in $\mathbb{R}^2$ whose union contains a set $E \subseteq \mathbb{R}^2$. Show that there is a pairwise disjoint subcollection $D_1, \ldots, D_n$ in $\mathcal{F}$ such that

$$E \subseteq \cup_{j=1}^{n} 3D_j.$$

Here, if D is the disc of radius r and center P, then $3D$ is the disc of radius $3r$ and center P.

Fields: geometry; analysis (covering lemmas); combinatorics (greedy algorithms) **Position:** A5 — late in the paper, and it is the finite Vitali covering lemma; hard if you have never met it, a few lines if you find the right greedy order. **Answer:** Choose discs greedily in order of decreasing radius, always taking the largest disc disjoint from those already chosen; the enlarged discs $3D_j$ cover every disc of $\mathcal{F}$, hence E.

Layer 1 THE WHAT — surface analysis

The natural attempt. The phrase "pairwise disjoint subcollection" makes you think of a maximal disjoint family, so you build one in the most obvious way: go through $\mathcal{F}$ in whatever order it is listed, keep a disc if it misses everything kept so far, and throw it away otherwise. The kept discs are pairwise disjoint and every discarded disc meets a kept one.

Now you try to prove the covering. Take $x \in E$; it lies in some $D \in \mathcal{F}$ with centre Q and radius r. If D was kept, fine. If not, D meets some kept disc D_j, centre P_j, radius s. Then $|Q - P_j| < r + s$, so

$$|x - P_j| \le |x - Q| + |Q - P_j| < r + (r + s) = 2r + s.$$

You want this to be at most $3s$, which needs $r \le s$. And here is the wall: in an arbitrary order nothing makes $r \le s$. A tiny disc listed first can be kept, and a huge disc overlapping it is then discarded; the huge disc's far side is at distance roughly $2r$ from the tiny centre, far outside $3D_j$. Concretely: a disc of radius ε at the origin listed first, then a disc of radius 1 centred at $(1, 0)$. The second is thrown away, and the point $(1.9, 0)$ is in it but not in the disc of radius 3ε about the origin. The estimate is right; the order is wrong.

The model chain. *Step 1: the greedy choice.* Since $\mathcal{F}$ is finite, we may run the following procedure. Let D_1 be a disc in $\mathcal{F}$ of largest radius. Having chosen $D_1, \dots, D_k$, let $\mathcal{C}_k$ be the set of discs in $\mathcal{F}$ disjoint from all of $D_1, \dots, D_k$. If $\mathcal{C}_k$ is empty, stop and set $n = k$; otherwise let D_{k+1} be a disc in $\mathcal{C}_k$ of largest radius. Each step removes at least one disc from consideration, so the procedure terminates. (If $\mathcal{F}$ is empty then $E = \varnothing$ and the empty subcollection works.) By construction $D_1, \dots, D_n$ are pairwise disjoint.

Step 2: every disc of $\mathcal{F}$ meets a chosen disc no smaller than itself. Let $D \in \mathcal{F}$ have centre Q and radius r. If D is one of the D_j, it meets itself. Otherwise, since the procedure stopped, $\mathcal{C}_n$ is empty, so D meets at least one chosen disc; let j be the *smallest* index with $D \cap D_j \neq \varnothing$. Then D is disjoint from $D_1, \dots, D_{j-1}$, i.e. $D \in \mathcal{C}_{j-1}$ (with $\mathcal{C}_0 = \mathcal{F}$). At step j the procedure picked D_j as a largest disc of $\mathcal{C}_{j-1}$, so its radius s satisfies $s \ge r$.

Step 3: the triangle inequality. Let P be the centre of that D_j. Two open discs meet exactly when the distance between centres is less than the sum of radii, so $|Q - P| < r + s$. For any $x \in D$,

$$|x - P| \le |x - Q| + |Q - P| < r + (r + s) = 2r + s \le 3s.$$

So $D \subseteq 3D_j$.

Step 4: conclude. Every $x \in E$ lies in some $D \in \mathcal{F}$, and $D \subseteq 3D_j$ for some j. Hence $E \subseteq \bigcup_{j=1}^{n} 3D_j$. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains build a maximal disjoint subfamily and both reach the estimate $|x - P_j| < 2r + s$. At that line the natural attempt discovers it needs $r \le s$ and has no way to get it. The model had arranged it in advance, by choosing the processing order: largest first. The split is at the moment of choosing an order in which to consider the discs — before any estimate is written.

> *I'll get $|x - P| < 2r + s$ from the triangle inequality whatever I do. To make that at most $3s$ I need the disc I'm covering to be no bigger than the chosen disc it bumped into. So I should choose the big discs first — then anything that gets rejected was rejected by something at least its own size.*

> *"Rejected by" needs a definite culprit. Take the* first *chosen disc it meets: up to that moment it was still eligible, so the one chosen then was at least as large.*

Layer 3 THE WHY

Why is "largest first" the natural order and not a trick? Start from what the conclusion asks. You must cover a discarded disc D by an enlargement of a *kept* disc D_j. The only information linking D to D_j is that they intersect. From "they intersect" the triangle inequality gives exactly one thing: the far edge of D is within $2r + s$ of P_j. That quantity involves r, the radius of the disc you are trying to cover, which the problem gives you no control over. Enlarging D_j by a fixed factor can only swallow $2r + s$ if r is controlled by s. So the entire proof reduces to one requirement:

every discarded disc meets a kept disc at least as large as itself.

Read that requirement as a design specification for the algorithm and the algorithm writes itself. If you process discs from largest to smallest, then whenever a disc is thrown out, the kept disc responsible was processed earlier, hence is at least as large. Nothing was guessed: the order is the unique ordering that turns "meets" into "meets something bigger".

Where does the 3 come from? It is $2 + 1$: the 2 is the diameter of D in units of r (you have to reach its far side), and the 1 is the radius of D_j. With $r \leq s$, $2r + s \leq 3s$. So the constant is not arbitrary — and it is sharp for this statement. Take two discs of radius 1 whose centres are $2 - \varepsilon$ apart. They overlap, so any disjoint subcollection contains at most one of them, and E (their union) contains points of the other disc at distances approaching $3 - \varepsilon$ from the chosen centre. For any constant $c < 3$, choosing $\varepsilon < 3 - c$ defeats cD.

Kedlaya's solution states the greedy rule and says a disc is skipped "only" because it meets a chosen disc of radius $s \geq r$. That word "only" is carrying Step 2 of the model chain, and it is exactly where students lose marks: the skipped disc may meet *several* chosen discs, some of them smaller than itself (chosen later, from a different region). The claim is true for the *first* chosen disc it meets, and the proof must say so. Write "let j be the least index with $D \cap D_j \neq \varnothing$".

Finally, where finiteness is used: in Step 1, to guarantee that a largest disc exists at each stage and that the process ends. For an infinite family with bounded radii the maximum need not be attained; the standard repair (the infinite Vitali lemma) chooses a disc of radius more than half the supremum, and the constant becomes 5. That is a useful fact to know when this lemma turns up again in measure theory, but it is not needed here.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "pairwise disjoint subcollection" together with an enlargement factor $3D$. On seeing it, write the triangle-inequality estimate for a disc D meeting a chosen disc D_j: $|x - P_j| < 2r + s$, and note what is needed to make it $\leq 3s$.

2. A greedy algorithm whose order is dictated by what the final estimate needs: largest radius first.

3. Hesitation occurs at "which chosen disc is responsible for rejecting D?" Reinforce it away by taking the first one in the order of choice; at that stage D was still a candidate, so the chosen disc is at least as large.

4. The trap is building the disjoint family in an arbitrary order: it produces a maximal disjoint family, which looks like the whole job, and then the covering estimate fails for a large disc rejected by a small one. A secondary trap is asserting "it met a disc at least as large" without specifying which disc.

5. Slogan: *write the final estimate first; the inequality it needs tells you the order in which a greedy algorithm must run.*

6. Decision rule: **when a greedy construction must control each rejected object by the one that rejected it, process objects in decreasing order of the quantity that must be dominated, and attribute each rejection to the first chosen object responsible.**

- ▷ **Cue:** choose a disjoint subfamily of discs (or intervals, balls) whose enlargements cover everything $\longrightarrow$ **Move:** greedy by decreasing radius; each rejected disc meets an earlier, larger chosen disc
- ▷ **Cue:** two open discs of radii r, s intersect $\longrightarrow$ **Move:** distance between centres $< r + s$; far side of the first is within $2r + s$ of the second centre
- ▷ **Cue:** "the object was rejected because of some chosen object" with several candidates $\longrightarrow$ **Move:** pick the first chosen object in the order of the algorithm; at that step the rejected object was still eligible
- ▷ **Cue:** a covering constant like 3 $\longrightarrow$ **Move:** read it as $2r + s$ with $r \leq s$; test sharpness with two equal overlapping discs

1998 A6

1998 A6. Let A, B, C denote distinct points with integer coordinates in $\mathbb{R}^2$. Prove that if

$$(|AB| + |BC|)^2 < 8 \cdot [ABC] + 1$$

then A, B, C are three vertices of a square. Here $|XY|$ is the length of segment XY and $[ABC]$ is the area of triangle ABC.

Fields: geometry (lattice points, triangle inequalities); number theory (integrality) **Position:** A6 — nominally the hardest A problem, but the difficulty is entirely in seeing the integrality squeeze; once seen, the proof is four lines. **Answer:** The hypothesis forces $|AB| = |BC|$ and $\angle ABC = 90^\circ$, so A, B, C are consecutive vertices of a square (and conversely every such lattice triple satisfies the hypothesis).

Layer 1 THE WHAT — surface analysis

The natural attempt. You put coordinates on it: B at the origin after translation, $A = (p, q)$, $C = (r, t)$ with integer entries. Then $|AB| = \sqrt{p^2+q^2}$, $|BC| = \sqrt{r^2+t^2}$, $[ABC] = \frac{1}{2}|pt - qr|$, and the hypothesis reads

$$p^2 + q^2 + r^2 + t^2 + 2\sqrt{(p^2+q^2)(r^2+t^2)} < 4|pt - qr| + 1.$$

You remember Lagrange's identity $(p^2+q^2)(r^2+t^2) = (pr+qt)^2 + (pt-qr)^2$, so the square root is at least $|pt - qr|$, and you get

$$p^2 + q^2 + r^2 + t^2 < 2|pt - qr| + 1.$$

Now $p^2 + q^2 + r^2 + t^2 - 2|pt - qr|$ is $(p \mp t)^2 + (q \pm r)^2$ for the appropriate sign, a non-negative integer less than 1, hence 0. So $t = \pm p$, $r = \mp q$, meaning C is A rotated by 90° about B. Done.

This works, and it is worth seeing that it works. But look at what it cost: coordinates, Lagrange's identity, and a completing-the-square step with a sign case split. The one decisive move — a non-negative integer less than 1 is 0 — arrived at the very end, buried. If you had not remembered Lagrange's identity you would have been stuck with a square root of a product of sums of squares, and the attempt would have stalled at the first display.

The model chain. Write $a = |AB|$, $c = |BC|$ and $K = [ABC]$, and let $\beta = \angle ABC$.

Two integrality facts. (i) a^2 and c^2 are integers, by Pythagoras applied to integer coordinate differences. (ii) $2K$ is an integer: with B at the origin, $A = (p, q)$, $C = (r, t)$, $K = \frac{1}{2}|pt - qr|$.

Two inequalities. (iii) $a^2 + c^2 \geq 2ac$, since $(a - c)^2 \geq 0$. (iv) $ac \geq 2K$, since $K = \frac{1}{2}ac \sin\beta \leq \frac{1}{2}ac$.

The squeeze. Expand the hypothesis: $(a + c)^2 = a^2 + c^2 + 2ac$. Using (iv), $2ac \geq 4K$, and using (iii) and (iv), $a^2 + c^2 \geq 2ac \geq 4K$. Hence

$$8K \;\leq\; a^2 + c^2 + 4K \;\leq\; a^2 + c^2 + 2ac \;=\; (a+c)^2 \;<\; 8K + 1.$$

The quantity $N = a^2 + c^2 + 4K$ is an integer by (i) and (ii), and so is $8K$. An integer N with $8K \leq N < 8K + 1$ equals $8K$. Therefore $a^2 + c^2 = 4K$.

Equality cases. Now $4K \leq 2ac \leq a^2 + c^2 = 4K$, so both inequalities are equalities: $2ac = a^2 + c^2$ gives $a = c$, and $2ac = 4K$ gives $\sin\beta = 1$, i.e. $\beta = 90^\circ$. (Since A, B, C are distinct, $a, c > 0$, so $K = \frac{1}{2}a^2 > 0$; the triangle is non-degenerate.) Thus BA and BC are perpendicular segments of equal length from B, and A, B, C are three consecutive vertices of the square with fourth vertex $A + C - B$. □

Converse (for orientation). If $a = c$ and $\beta = 90^\circ$ then $K = a^2/2$ and $(a + c)^2 = 4a^2 = 8K < 8K + 1$. So the hypothesis characterises exactly these triples.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains reach an inequality between $(a+c)^2$ and $8K$. The natural attempt then tries to control the irrational term $2ac$ algebraically in coordinates. The model looks at the "$+1$" and asks which quantities in sight are integers. The split is at the line

$$(a+c)^2 < 8K+1 :$$

do you treat it as an inequality between real numbers, or as "something is an integer strictly less than $8K+1$"?

> *Why "$+1$"? A $+1$ on the right of a strict inequality only matters if the left is an integer or close to one — it's saying "at most $8K$". What here is an integer? a^2, c^2, $2K$. The only thing that isn't is the cross term $2ac$.*

> *I can't make $2ac$ an integer, but I can replace it by something smaller that is. $ac \geq ac\sin\beta = 2K$. So $(a+c)^2 \geq a^2+c^2+4K$, which is an integer, and it's trapped in $[8K, 8K+1)$.*

> *If the chain of inequalities collapses to equality, every link is tight — so read off the equality cases: $a=c$ from AM–GM, $\sin\beta = 1$ from the area formula. That's a right isosceles triangle, which is exactly "three vertices of a square".*

Layer 3 THE WHY

Why is the integrality squeeze the natural move and not a rabbit from a hat? Go back to the phrasing. The problem is set on integer lattice points, and the bound contains a bare "$+1$". A lattice hypothesis in a problem about lengths and areas is there for one reason: it makes certain quantities integers, and integers cannot lie strictly between n and $n+1$. The structural fact at the bottom is the discreteness of $\mathbb{Z}$: if $N \in \mathbb{Z}$ and $n \leq N < n+1$ for an integer n, then $N = n$. Any time a problem places a strict "$<$ integer $+1$" next to integer-valued data, it is offering you exactly this.

So the question becomes: which quantities are integers? Squared lengths, by Pythagoras on integer differences. Twice the area, by the determinant (shoelace) formula. Lengths themselves are not, and neither is the product ac. That tells you precisely what to do with $(a+c)^2 = a^2+c^2+2ac$: keep a^2+c^2 (integer), and replace the offending $2ac$ by an integer-valued lower bound. The obvious lower bound for ac that involves area is $ac \geq ac\sin\beta = 2K$ — the area formula $K = \frac{1}{2}ac\sin\beta$ is the only formula that connects a product of two sides to the area. So the replacement $2ac \to 4K$ is forced by the requirement "integer, and comparable to ac".

That produces the upper half of the squeeze. The lower half, $8K \leq a^2+c^2+4K$, is the same two inequalities used again: $a^2+c^2 \geq 2ac \geq 4K$. You need both halves because an integer between two reals is only pinned when the interval has length at most 1 and has an integer endpoint; $8K$ is that endpoint.

Why are AM–GM and $\sin\beta \leq 1$ the right inequalities? Because the conclusion is "square", i.e. equal sides and a right angle, and each of these two inequalities has exactly one of those conditions as its equality case. A problem whose conclusion is a rigid shape is almost always proved by squeezing a chain of inequalities to equality; the shape is the conjunction of the equality cases. Choose inequalities whose equality cases are the features of the shape.

Kedlaya's solution writes the chain $8[ABC] \leq |AB|^2+|BC|^2+4[ABC] \leq \cdots$ without saying where the middle term came from. The middle term is $(a+c)^2$ with its one non-integer piece, $2ac$, replaced by the largest integer-valued quantity guaranteed to lie below it. That is the whole idea, and it is what the solution omits.

Comparing with the natural attempt: the coordinate proof is the same argument in disguise. Lagrange's identity $(p^2+q^2)(r^2+t^2) = (pr+qt)^2+(pt-qr)^2$ is $a^2c^2 = (ac\cos\beta)^2+(ac\sin\beta)^2$,

and dropping the first square is $ac \geq ac|\sin\beta|$. The coordinate route works, but it makes you rediscover the area formula in algebraic form before the integrality observation can fire.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the "$+1$" in a strict inequality with lattice-point data. On seeing it, write down the list of integer-valued quantities: $|AB|^2$, $|BC|^2$, $2[ABC]$.

2. An integrality squeeze: bound the given expression below by an integer that is at least $8K$, so that it is trapped in $[8K, 8K+1)$, then read off equality cases.

3. Hesitation occurs at the cross term $2|AB||BC|$, which is not an integer. Reinforce it away with the question "what integer lies below it?" and the answer $4K$, from $K = \frac{1}{2}ac\sin\beta$.

4. The trap is working in coordinates and fighting the square root $\sqrt{(p^2+q^2)(r^2+t^2)}$. It can be done (via Lagrange's identity), but a student who does not know that identity stalls. A second trap: forgetting that an inequality chain forced to equality forces *every* link to equality — that is where the square comes from.

5. Slogan: *a "$+1$" next to lattice data means "an integer is squeezed"; list the integers, and replace each non-integer term by an integer bound.* And: *a rigid-shape conclusion is the set of equality cases of a chain of inequalities.*

6. Decision rule: **when a strict inequality has the form $X < n+1$ with n an integer and the data are integral, find an integer N with $n \leq N \leq X$; then $N = n$, and every inequality used to get from N to X and from n to N becomes an equality.**

▷ **Cue:** lattice points and a strict inequality with a "$+1$" $\longrightarrow$ **Move:** list the integer-valued quantities; aim to trap an integer in $[n, n+1)$

▷ **Cue:** squared distance between lattice points; area of a lattice triangle $\longrightarrow$ **Move:** $|AB|^2 \in \mathbb{Z}$ (Pythagoras); $2[ABC] \in \mathbb{Z}$ (shoelace)

▷ **Cue:** a product of two sides next to an area $\longrightarrow$ **Move:** $ac \geq ac\sin\beta = 2K$, equality iff right angle

▷ **Cue:** conclusion is a rigid shape (square, equilateral, right isosceles) $\longrightarrow$ **Move:** choose inequalities whose equality cases are the shape's features, and squeeze the chain to equality

▷ **Cue:** $(a+c)^2$ with a^2, c^2 integers $\longrightarrow$ **Move:** split off the non-integer cross term $2ac$ and bound it by an integer

Putnam 1998 — Paper B

1998 B1

1998 B1. Find the minimum value of

$$\frac{(x+1/x)^6-(x^6+1/x^6)-2}{(x+1/x)^3+(x^3+1/x^3)}$$

for $x>0$.

Fields: algebra, difference of squares **Position:** B1 — the opener **Answer:** The minimum value is 6, achieved at $x=1$, because the whole fraction collapses to $3(x+1/x)$ and $x+1/x\geq 2$ for $x>0$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The expression is begging to be attacked with the substitution $u=x+1/x$, since $u\geq 2$ for $x>0$ (AM-GM) and the answer is presumably some clean function of u. The honest first move is to expand everything in terms of u directly.

$(x+1/x)^6$ expands by the binomial theorem into seven terms; grouping the symmetric pairs, $(x+1/x)^6=x^6+6x^4+15x^2+20+15/x^2+6/x^4+1/x^6$. Subtracting x^6+1/x^6 and 2 from this gives numerator $=6x^4+15x^2+18+15/x^2+6/x^4=6(x^4+1/x^4)+15(x^2+1/x^2)+18$.

Now $x^2+1/x^2=u^2-2$ and $x^4+1/x^4=(x^2+1/x^2)^2-2=(u^2-2)^2-2=u^4-4u^2+2$. So the numerator becomes $6(u^4-4u^2+2)+15(u^2-2)+18=6u^4-24u^2+12+15u^2-30+18=6u^4-9u^2$.

The denominator is $(x+1/x)^3+(x^3+1/x^3)$. Since $x^3+1/x^3=u^3-3u$, the denominator is $u^3+(u^3-3u)=2u^3-3u$.

So the ratio is $\dfrac{6u^4-9u^2}{2u^3-3u}=\dfrac{3u^2(2u^2-3)}{u(2u^2-3)}=3u$ (for $u\neq 0$ and $2u^2\neq 3$). This works, and gives the same answer as below — but it required expanding a sixth power by hand, tracking four separate symmetric-power identities (u^2, u^3, u^4 in terms of u), and only cancelling the common factor $2u^2-3$ after all of that was in front of you. A single sign slip in the binomial expansion, or in any of the three power-reduction identities, is invisible until the very end.

The model chain. Look at the shapes of numerator and denominator before touching a single expansion. Set $A=(x+1/x)^3$ and $B=x^3+1/x^3$. The denominator is exactly $A+B$. The numerator is $(x+1/x)^6-(x^6+1/x^6)-2$; since $(x+1/x)^6=A^2$ and $(x^3+1/x^3)^2=x^6+2+1/x^6$, i.e. $B^2=x^6+1/x^6+2$, the numerator is exactly A^2-B^2.

So the whole fraction is

$$\frac{A^2-B^2}{A+B}=A-B$$

with no expansion at all, valid whenever $A+B\neq 0$ (true here, since $A+B\geq 4>0$ once $x>0$). Now use the one identity actually needed: $x^3+1/x^3=(x+1/x)^3-3(x+1/x)$, i.e. $B=u^3-3u$ where $u=x+1/x$. Then

$$A-B=u^3-(u^3-3u)=3u=3\left(x+\frac{1}{x}\right).$$

By AM-GM, $x + 1/x \geq 2$ for $x > 0$, with equality at $x = 1$. So the minimum of the whole expression is $3 \cdot 2 = 6$, attained at $x = 1$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains reach the same intermediate fact — the answer is $3u$ for $u = x + 1/x$ — but they get there by entirely different routes. The natural attempt substitutes u into the numerator and denominator *separately*, expands each fully, and only discovers the common factor $2u^2 - 3$ by inspection after the algebra is finished. The model chain never expands the sixth power at all: it stops at the line "the denominator is $A + B$" and asks whether the numerator is $A^2 - B^2$ for that same A, B, before doing any arithmetic.

> *The denominator has the form (cube of a sum) + (sum of cubes), and the numerator has a sixth power sitting next to a sixth-power sum, with the sixth power being exactly the square of the cube already in the denominator. A number raised to the sixth power next to the same base's cube is a standing invitation to write it as a square: $x^6 = (x^3)^2$, $(x + 1/x)^6 = ((x + 1/x)^3)^2$. If the denominator is $A + B$ for those same two cubed quantities, I should check whether the numerator is $A^2 - B^2$ before expanding anything, because $(A^2 - B^2)/(A + B) = A - B$ is free.*

Layer 3 THE WHY

Why is checking for $A^2 - B^2$ over $A + B$ the move that has to occur to a solver here, rather than merely a shortcut noticed after the fact? Because the exponent structure of the problem is not incidental: the numerator is built from a sixth power and a pair of sixth-power reciprocals, and the denominator from the matching cube and cube-reciprocal sum. Sixth powers are squares of cubes by definition — $a^6 = (a^3)^2$ — so any expression that mixes a sixth power with a cube automatically presents a difference or sum of two squares, whether or not the numerator was designed that way. The defining property of a rational function $\frac{P}{Q}$ where $P = A^2 - B^2$ and $Q = A + B$ is that Q divides P identically, as polynomials, with no restriction on x beyond $Q \neq 0$ — this is not a numerical coincidence to be verified by expanding, it is algebraic identity, true for every value of A and B simultaneously. Once a solver has the vocabulary to see $A = (x + 1/x)^3$, $B = x^3 + 1/x^3$ as *objects* rather than as instructions to expand, the factorisation is not a trick: it is the only sensible reading of "sixth power next to cube." The natural attempt's failure is not an error, it is simply substituting a numerical value for u (via full expansion) instead of leaving A and B as symbols long enough to notice the shared factor $A + B$ cancels exactly.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that the numerator contains a sixth power and a sum of sixth-power reciprocals, while the denominator contains the matching cube and cube-reciprocal sum. On seeing it, write $A = (x + 1/x)^3$, $B = x^3 + 1/x^3$ and check whether numerator $= A^2 - B^2$ before expanding either.
2. That cue suggests the difference-of-squares factorisation $A^2 - B^2 = (A - B)(A + B)$, cancelling directly against the denominator $A + B$.
3. Hesitation occurs at the decision to substitute $u = x + 1/x$ immediately and expand, rather than first testing the A, B structure symbolically. It is reinforced away by treating any expression of the form (power)^6 next to (power)^3 of the same base as an automatic candidate for $a^6 = (a^3)^2$, checked before any binomial expansion is written down.
4. The trap is that the substitution $u = x + 1/x$ genuinely works and produces the right answer, so a solver who expands first is never told they took the expensive path — only that it cost four extra symmetric-power identities and a hidden common factor discovered by luck at the end.
5. Recognise this family faster by scanning the numerator and denominator for matching exponents before choosing any substitution: if the denominator looks like $A + B$ for two visible sub-expressions, check numerator $\stackrel{?}{=} A^2 - B^2$ as the very first move.

6. Reusable rule: whenever a fraction pairs a sixth power (or any even power $2k$) with the corresponding k-th power of the same base in both numerator and denominator, test for $A^2 - B^2$ over $A + B$ (or the analogous higher-degree factorisation) symbolically before expanding either side. Leaving compound quantities as named objects until a cancellation is confirmed is cheaper than reducing everything to one variable and hoping a common factor turns up.

▷ **Cue:** a numerator contains an even power $2k$ of an expression next to a k-th-power-based term, and the denominator visibly sums the corresponding k-th powers $\longrightarrow$ **Move:** name the two k-th-power quantities A, B and test numerator $= A^2 - B^2$, denominator $= A + B$, before expanding; the ratio is then simply $A - B$

▷ **Cue:** $x + 1/x = u$ appears with both a cube and a sixth power of the same base $\longrightarrow$ **Move:** convert only the minimal identity needed, $x^3 + 1/x^3 = u^3 - 3u$, rather than also deriving $x^2 + 1/x^2$ and $x^4 + 1/x^4$ for a full expansion

1998 B2

1998 B2. Given a point (a, b) with $0 < b < a$, determine the minimum perimeter of a triangle with one vertex at (a, b), one on the x-axis, and one on the line $y = x$. You may assume that a triangle of minimum perimeter exists.

Fields: Euclidean geometry, reflection **Position:** B2 **Answer:** The minimum perimeter is $\sqrt{2a^2 + 2b^2}$, because reflecting the fixed vertex across both mirror lines turns the broken three-segment path into a single straight segment between the two reflections.

Layer 1 THE WHAT — surface analysis

The natural attempt. Call the fixed point $A = (a, b)$, let $P = (p, 0)$ be the vertex on the x-axis and $Q = (q, q)$ the vertex on $y = x$. The perimeter is

$$f(p, q) = |AP| + |PQ| + |QA| = \sqrt{(a-p)^2 + b^2} + \sqrt{(p-q)^2 + q^2} + \sqrt{(a-q)^2 + (b-q)^2}.$$

This is a genuine two-variable minimisation with no constraint beyond $p, q \in \mathbb{R}$, so the honest first move is calculus: set $\partial f/\partial p = 0$ and $\partial f/\partial q = 0$ and solve the resulting simultaneous equations.

$$\partial f/\partial p = \frac{p-a}{\sqrt{(a-p)^2 + b^2}} + \frac{p-q}{\sqrt{(p-q)^2 + q^2}} = 0, \text{ and } \partial f/\partial q = \frac{q-p}{\sqrt{(p-q)^2 + q^2}} \cdot \frac{\partial}{\partial q}(p - q) + \ldots$$

— already, differentiating the middle term with respect to q requires care since q appears in *both* coordinates of $Q = (q, q)$, so the term $\sqrt{(p-q)^2 + q^2}$ differentiates to $\frac{-(p-q) + q}{\sqrt{(p-q)^2 + q^2}}$, and the third term $\sqrt{(a-q)^2 + (b-q)^2}$ differentiates to $\frac{-(a-q) - (b-q)}{\sqrt{(a-q)^2 + (b-q)^2}}$.

Each derivative, set to zero, is a statement that a unit vector along one segment plus a unit vector along an adjacent segment sums to a vector normal to the constraint line at that vertex — in other words, each equation is secretly the law of reflection (angle of incidence equals angle of reflection) in disguise, but nothing about the algebra reveals that geometric fact. The two equations are transcendental in p, q jointly, do not decouple, and solving them by substitution is realistically impossible by hand within competition time; a solver who starts here has correctly identified that a minimum exists and that calculus finds it, and has bought a system with no closed form in sight.

The model chain. Reflect A across the x-axis to get $D = (a, -b)$, and reflect A across the line $y = x$ to get $E = (b, a)$. For any point P on the x-axis, reflection across the x-axis preserves distances to points on that axis, so $|AP| = |DP|$. Likewise for any point Q on $y = x$, $|AQ| = |EQ|$.

So the perimeter of any admissible triangle is

$$|AP| + |PQ| + |QA| = |DP| + |PQ| + |QE|,$$

which is the length of a path from D to E that touches the x-axis at P and the line $y = x$ at Q, in that order. Any path from D to E has length at least $|DE|$, with equality exactly when the path is the straight segment DE itself — so the minimum perimeter is $|DE|$, achieved by choosing P, Q to be the two points where segment DE crosses the x-axis and the line $y = x$ respectively (these crossings exist and occur in the right order because $0 < b < a$ places D below the x-axis and E above the line $y = x$, on opposite sides of both mirrors from A).

Compute:

$$|DE|^2 = (a - b)^2 + (a - (-b))^2 = (a - b)^2 + (a + b)^2 = 2a^2 + 2b^2.$$

So the minimum perimeter is $\sqrt{2a^2 + 2b^2}$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains treat the perimeter as a sum of three segment lengths to be minimised over two free points P, Q. They split at the very first line: the natural attempt writes $f(p, q)$ as an explicit two-variable function of the coordinates and reaches for $\partial f/\partial p = \partial f/\partial q = 0$. The model chain never introduces coordinates for P and Q into a function to be differentiated at all — it instead asks what happens to the fixed vertex A under reflection in each mirror line, before writing down any expression to minimise.

> *Each of P and Q is free to slide along its own line, and the length of any segment from A to a point on a line is unchanged if I reflect A across that line instead — so I do not need A itself, I need a point that is at the same distance from every point of the mirror line as A is, and a reflection is exactly a tool for manufacturing that. If I replace A by its reflection at each end of the broken path, the whole three-segment perimeter becomes a single path between two fixed points, D and E, that happens to pass through the two mirror lines — and the shortest path between two fixed points, full stop, is the straight line between them.*

Layer 3 THE WHY

Why does reflecting A have to occur to a solver here, rather than being a trick pulled from a hat? Because the object being minimised is a sum of distances from a moving point to two *fixed* points (A, and the other end of the current segment), and distance is by definition the thing reflection preserves along a mirror: a reflection is precisely the isometry that fixes every point of the mirror line and swaps the two half-planes, so for any point M on the mirror, $|AM| = |A'M|$ where A' is A's reflection, is not a coincidence to verify — it is what "reflection across a line" means. Once that is available, the perimeter $|AP| + |PQ| + |QA|$ rewrites as $|DP| + |PQ| + |QE|$ with D, E fixed, and the defining property of a straight line — that it is the unique shortest connection between two points, with every other path strictly longer — settles the minimisation with no calculus at all. The natural attempt's calculus does not fail; it recovers the same critical points, but only after burying the reflection law inside derivative equations (each stationarity condition, unpacked, literally says "the incoming and outgoing unit vectors make equal angles with the mirror," which is the law of reflection restated as an algebraic identity nobody can read off the raw equations). The lesson is not that calculus is wrong, but that when a quantity to be minimised is manifestly a sum of distances to fixed points from points constrained to a line, that structure — not the coordinates — is the object to act on first.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that the perimeter is a sum of distances $|AP| + |PQ| + |QA|$ where P, Q each range over a fixed line and A is fixed. On seeing it, write down: "distance from a

moving point on a line to a fixed point equals distance to that fixed point's reflection across the line" before setting up any coordinates.

2. That cue suggests reflecting the fixed endpoint(s) across the constraint line(s) to turn the broken path into a path between two fixed points, whose minimum length is simply the straight-line distance between them.

3. Hesitation occurs at the decision of *which* endpoint to reflect and across which line, when there are two constraint lines rather than one (the classical single-mirror billiard problem generalises here to two mirrors). This is reinforced away by reflecting A separately across each mirror — once for the segment touching the x-axis, once for the segment touching $y = x$ — since each reflection only needs to fix the mirror it is meant to act on.

4. The trap is treating this as an unconstrained two-variable calculus problem the moment two free points appear, rather than checking first whether the quantity being minimised is a sum of distances to fixed points — the signature that reflection applies.

5. Recognise this family faster by checking, whenever a perimeter or path length is to be minimised with vertices confined to one or more fixed lines, whether every term is a distance to a point that does not move; if so, reflect the fixed points across the constraining lines before writing any derivative.

6. Reusable rule: a sum of distances from points constrained to lines, to fixed points, is minimised by reflecting each fixed point across its adjacent mirror line and connecting the resulting fixed points by a straight segment; the answer is that segment's length, and the optimal positions of the free points are its intersections with the mirror lines, in order. This generalises the single-mirror shortest-path (billiard) trick to any number of mirrors in sequence.

▷ **Cue:** a perimeter or path length is a sum of distances from constrained (line-bound) points to fixed points $\longrightarrow$ **Move:** reflect each fixed endpoint across the line constraining its adjacent free vertex, reducing the broken path to a straight segment between the reflections

▷ **Cue:** two free points, each constrained to a different line, appear in an expression to be minimised $\longrightarrow$ **Move:** before differentiating, check whether the expression is purely a sum of distances — if so, apply reflection separately at each mirror rather than solving a joint calculus system

1998 B3

1998 B3. let H be the unit hemisphere $\{(x, y, z) : x^2 + y^2 + z^2 = 1, z \geq 0\}$, C the unit circle $\{(x, y, 0) : x^2 + y^2 = 1\}$, and P the regular pentagon inscribed in C. Determine the surface area of that portion of H lying over the planar region inside P, and write your answer in the form $A \sin \alpha + B \cos \beta$, where A, B, α, β are real numbers.

Fields: solid geometry, spherical surface area **Position:** B3 **Answer:** The area is $5\pi \cos(\pi/5) - 3\pi$, i.e. $A \sin \alpha + B \cos \beta$ with $A = -3\pi$, $\alpha = \pi/2$, $B = 5\pi$, $\beta = \pi/5$, because the hemisphere minus the pentagon's shadow splits into five congruent pieces, each exactly half of a classical spherical cap.

Layer 1 THE WHAT — surface analysis

The natural attempt. The region "over P" is a subset of the hemisphere described by a planar boundary of five straight edges, so the honest first move is a direct surface integral. Parametrise the hemisphere by $(x, y) \mapsto (x, y, \sqrt{1 - x^2 - y^2})$ over the disc, with surface element $dS = \dfrac{dx\,dy}{\sqrt{1 - x^2 - y^2}}$, and integrate over the pentagon P:

$$\text{Area} = \iint_P \frac{dx\,dy}{\sqrt{1 - x^2 - y^2}}.$$

Switch to polar coordinates $x = r\cos\theta$, $y = r\sin\theta$: $dS = \dfrac{r\,dr\,d\theta}{\sqrt{1-r^2}}$. By the five-fold symmetry of the pentagon, the integral is five times the integral over one wedge, say $\theta \in [-\pi/5, \pi/5]$, where the outer boundary is no longer a fixed radius but the pentagon's edge — for a regular pentagon with apothem $\cos(\pi/5)$, the edge in that wedge satisfies $r\cos\theta = \cos(\pi/5)$, i.e. $r(\theta) = \cos(\pi/5)/\cos\theta$.

So one wedge contributes

$$\int_{-\pi/5}^{\pi/5}\int_{0}^{\cos(\pi/5)/\cos\theta} \frac{r\,dr}{\sqrt{1-r^2}}\,d\theta = \int_{-\pi/5}^{\pi/5}\left[1-\sqrt{1-r(\theta)^2}\right]d\theta,$$

using $\int_0^{r_0} r/\sqrt{1-r^2}\,dr = 1-\sqrt{1-r_0^2}$. This leaves

$$\int_{-\pi/5}^{\pi/5}\left(1-\sqrt{1-\frac{\cos^2(\pi/5)}{\cos^2\theta}}\right)d\theta,$$

a single-variable integral, but one with $\sqrt{\cos^2\theta-\cos^2(\pi/5)}/\cos\theta$ under it once simplified — not a standard elementary antiderivative on sight, and pursuing it by hand risks a substitution that does not close in elementary functions at all, let alone in the required form $A\sin\alpha + B\cos\beta$. A solver who starts here has correctly reduced the problem to one integral but has bought a genuinely harder computation than the one actually needed.

The model chain. Do not integrate the pentagon; integrate its complement in the circle, which is five congruent circular segments, and use the classical fact about spherical zones (Archimedes' hat-box theorem): for the unit sphere, the surface area of the zone $\{z \geq z_0\}$ is $2\pi(1-z_0)$, independent of which axis is called "z" — it depends only on the perpendicular distance z_0 from the centre to the cutting plane.

Each side of the pentagon is a chord of C at perpendicular distance from the centre equal to the pentagon's apothem, $z_0 = \cos(\pi/5)$ (the half-angle subtended at the centre by one side is $\pi/5$, since there are five sides). Erect the vertical plane through that chord (perpendicular to the xy-plane, containing the chord). This plane is at perpendicular distance $\cos(\pi/5)$ from the sphere's centre along the horizontal axis normal to it, so by the same hat-box theorem (applied with that horizontal axis playing the role of "z"), the region of the *full sphere* beyond this plane, away from the centre, is a genuine spherical cap of area $2\pi(1-\cos(\pi/5))$.

That cutting plane is vertical, hence symmetric under $z \mapsto -z$; so the full-sphere cap it cuts off splits evenly between the upper hemisphere ($z \geq 0$) and the lower one, each carrying exactly half: $\pi(1-\cos(\pi/5))$. This half-cap is precisely the piece of H lying beyond that one pentagon edge, outside P.

The five such pieces — one per edge — are pairwise disjoint (the five circular segments outside the pentagon but inside the circle meet only along the pentagon's vertices, a set of measure zero), so

$$\text{Area}(H \setminus P\text{-shadow}) = 5\cdot\pi\left(1-\cos\frac{\pi}{5}\right).$$

The hemisphere's total area is 2π (the case $z_0 = 0$ of the same theorem). So the desired area is

$$2\pi - 5\pi\left(1-\cos\frac{\pi}{5}\right) = 5\pi\cos\frac{\pi}{5} - 3\pi.$$

Writing $-3\pi = 3\pi\cdot(-1)$... more directly, $-3\pi = (-3\pi)\sin(\pi/2)$ and $5\pi\cos(\pi/5)$ is already of the required shape, this is $A\sin\alpha + B\cos\beta$ with $A = -3\pi$, $\alpha = \pi/2$, $B = 5\pi$, $\beta = \pi/5$.

Caveat. The hat-box theorem itself ($2\pi(1 - z_0)$ for a unit-sphere zone, independent of axis by rotational symmetry) is standard and I take it as known rather than re-deriving it by calculus here, as the official solution also does; everything downstream of it — the apothem value, the $z \mapsto -z$ symmetry argument for halving the cap, and the disjointness of the five segments — I have verified directly, and a numerical check (pentagon covers about 75.7% of the disc's planar area, but the computed spherical area of $5\pi\cos(\pi/5) - 3\pi \approx 3.28$ is only about 52% of the hemisphere's $2\pi \approx 6.28$, consistent with the rim regions carrying disproportionately more spherical area than planar area, since dS diverges near $r = 1$) is consistent with the result.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains recognise the five-fold symmetry of the pentagon and reduce the problem to handling one wedge or one edge. They split on *which region to compute*: the natural attempt computes the target region (inside P) directly, which forces a genuine bounded-by-an-edge polar integral with no clean closed form on sight; the model chain computes the complement (outside P, inside C) instead, because that complement is exactly the shape a named theorem already answers.

> *The boundary of the region I actually want is a pentagon edge, but the boundary of what's* left over *is also a pentagon edge — and a straight chord, extended to a vertical plane, is exactly the kind of cutting plane the spherical-cap theorem was built for. I don't have to integrate the piece the problem names; I can integrate the piece whose boundary matches a theorem I already have, and subtract.*

Layer 3 THE WHY

Why does switching to the complement have to occur to a solver, rather than being a lucky reframing? Because the classical result about spherical zones is stated for a plane cutting the sphere at a fixed perpendicular distance from the centre — that is its entire domain of applicability, and a pentagon edge (a chord of the equatorial circle) is precisely such a plane once extended vertically. The pentagon's boundary, taken edge by edge, is five instances of exactly the object the theorem consumes; the pentagon's interior, taken as a whole, is not a single instance of anything with a name. Choosing which of two complementary regions to compute is not cosmetic: a closed-form theorem exists for one description of the boundary (a chord at fixed distance from the centre) and not for the other (a polygon's interior as a function of angle), so the correct move is to ask, before integrating anything, which of the region and its complement is expressible as a union of objects a known theorem already closes. The $z \mapsto -z$ symmetry step is a second, smaller instance of the same discipline: rather than deriving the cap area newly for the hemisphere, recognise that the vertical cutting plane is invariant under reflection through $z = 0$, so the hemisphere's share of the cap is exactly half of the whole-sphere cap the theorem gives directly — no new integral, just a symmetry argument applied to an already-known quantity.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that the target region's boundary is a polygon inscribed in a circle, five straight chords. On seeing it, ask whether the *complement* of the region (inside the circle, outside the polygon) is a simpler union of named pieces than the region itself.
2. That cue suggests computing the complement via the classical spherical-zone (Archimedes hat-box) theorem, since each chord extended to a vertical plane is exactly the boundary that theorem is stated for, then subtracting from the known total hemisphere area 2π.
3. Hesitation occurs at the decision to trust a symmetry halving ($z \mapsto -z$) instead of directly computing the hemisphere's cap area by a fresh calculation. This is reinforced away by checking explicitly that the cutting plane, being vertical, is invariant under reflection

through $z = 0$ — so the upper and lower pieces of the full-sphere cap are congruent by that reflection, not merely "probably equal."

4. The trap is starting the integral over the named region (the pentagon) simply because it is the region the problem asks for, without first checking whether its complement admits a closed form the region itself does not.

5. Recognise this family faster by treating "area bounded by a polygon inscribed in a circle, on a sphere or elsewhere" as a cue to compute the polygon's complement in the circle whenever the complement decomposes into congruent pieces matching a named theorem (segments, caps, sectors), then subtract from the whole.

6. Reusable rule: when a region's own boundary does not match any closed-form theorem but its complement (within a larger, fully-understood shape) does, compute the complement and subtract. This is a special case of a broader habit: before setting up an integral, ask which of the region and its complement is literally the object a known formula was built to consume.

▷ **Cue:** a region on a sphere is bounded by chords of the equatorial circle (an inscribed polygon) $\longrightarrow$ **Move:** compute the complement (circle minus polygon) as a union of spherical caps cut by the chords' vertical planes, via the zone-area theorem $2\pi(1 - z_0)$ with z_0 = the chord's distance from the centre, then subtract from the total

▷ **Cue:** a cutting plane for a spherical cap is vertical (contains the polar axis' perpendicular direction), and only a hemisphere is wanted $\longrightarrow$ **Move:** use the $z \mapsto -z$ symmetry of the vertical plane to take exactly half the full-sphere cap area, instead of re-deriving the cap area for the hemisphere alone

1998 B4

1998 B4. Find necessary and sufficient conditions on positive integers m and n so that

$$\sum_{i=0}^{mn-1} (-1)^{\lfloor i/m \rfloor + \lfloor i/n \rfloor} = 0.$$

Fields: combinatorics; number theory (2-adic valuation); symmetry of sums **Position:** B4 — mid-paper; three short arguments glued by a descent, but only once you have computed enough small cases to guess the answer **Answer:** The sum is zero if and only if the highest powers of 2 dividing m and n are different.

Layer 1 THE WHAT — surface analysis

The natural attempt. Write $S(m, n)$ for the sum and $f(i) = \lfloor i/m \rfloor + \lfloor i/n \rfloor$. The strong student sees floors and a sign, and starts to compute: lay out $i = 0, 1, \ldots, mn - 1$ in blocks of length m (on which $\lfloor i/m \rfloor$ is constant) and try to count, within each block, how many i have $\lfloor i/n \rfloor$ even. For $m = 1$ this is fine: $S(1, n) = \sum_{i=0}^{n-1} (-1)^{i+\lfloor i/n \rfloor} = \sum_{i=0}^{n-1} (-1)^i$, which is 0 for n even and 1 for n odd. For general m, n the blocks of length m and the blocks of length n interleave according to how m and n sit relative to each other, and the count depends on $\gcd(m, n)$ in a way that does not produce a closed form without real work. The student writes a page of case analysis in m mod n, loses track, and has not yet even guessed the answer. That is the wall: an attempt to *evaluate* a sum whose question only asks *when it vanishes*.

The model chain. **Step 0: data.** Before proving anything, tabulate $S(m, n)$ for small m, n. One finds $S(1, 2) = S(2, 3) = S(1, 4) = S(4, 6) = 0$ but $S(1, 1) = S(1, 3) = S(3, 5) = 1$, $S(2, 2) = S(2, 6) = 4$, $S(4, 12) = 16$. The pattern that fits is: zero exactly when m and n have different 2-adic valuation v_2. Now prove it in three cases.

Case 1: m, n **both odd.** Then mn is odd, and a sum of an odd number of terms ± 1 is odd, hence nonzero.

Case 2: m, n **of opposite parity.** Pair i with $i' = mn - 1 - i$; this is a bijection of $\{0, \ldots, mn-1\}$ to itself. Write $i = qm + r$ with $0 \le r \le m-1$. Then $i' = (n-1-q)m + (m-1-r)$ with $0 \le m-1-r \le m-1$, so $\lfloor i'/m \rfloor = n-1-q$ and

$$\lfloor i/m \rfloor + \lfloor i'/m \rfloor = n - 1.$$

By the same computation with the roles swapped, $\lfloor i/n \rfloor + \lfloor i'/n \rfloor = m - 1$. Adding, $f(i) + f(i') = m + n - 2$, which is odd. So the terms for i and i' are negatives of each other. Since mn is even, $i \neq i'$ always ($i = i'$ would need $2i = mn - 1$), so the terms cancel in pairs and $S(m, n) = 0$.

Case 3: $m = 2k$, $n = 2l$ **both even.** For any integer j, $\lfloor (2j+1)/(2k) \rfloor = \lfloor 2j/(2k) \rfloor = \lfloor j/k \rfloor$: if $j = qk + r$ with $0 \le r \le k-1$ then $2j + 1 = 2qk + (2r+1)$ and $2r + 1 \le 2k - 1$. The same holds with l. So the terms at $i = 2j$ and $i = 2j + 1$ are equal, and

$$S(2k, 2l) = 2 \sum_{j=0}^{2kl-1} (-1)^{\lfloor j/k \rfloor + \lfloor j/l \rfloor}.$$

Split this range into $0 \le j \le kl - 1$ and $j + kl$. Since $\lfloor (j+kl)/k \rfloor = \lfloor j/k \rfloor + l$ and $\lfloor (j+kl)/l \rfloor = \lfloor j/l \rfloor + k$, the second half is $(-1)^{k+l}$ times the first, so

$$S(2k, 2l) = 2\,(1 + (-1)^{k+l})\, S(k, l).$$

If $k + l$ is even, $S(2k, 2l) = 4S(k, l)$, zero iff $S(k, l)$ is. If $k + l$ is odd, $S(2k, 2l) = 0$, and $S(k, l) = 0$ by Case 2. Either way: $S(2k, 2l) = 0 \iff S(k, l) = 0$.

Descent. Halving both arguments preserves both "$S = 0$" and "$v_2(m) \neq v_2(n)$". Repeat until one argument is odd. If the other is then even, v_2's differ and $S = 0$ (Case 2); if it is odd, v_2's agree and $S \neq 0$ (Case 1). Hence $S(m, n) = 0$ if and only if $v_2(m) \neq v_2(n)$. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains begin with the same sum. They split at the first decision: (N) asks "what is $S(m, n)$?" and starts counting; (M) asks "when is $S(m, n) = 0$?", computes a table to guess the answer, and then looks for *structure that forces cancellation* rather than a formula. Once the answer "different powers of 2" is on the page, the three cases write themselves: odd count, a pairing, and a halving.

> *The question is only whether this sum vanishes, not what it is. I will not evaluate it. I'll compute a small table, guess the condition, and then look for a reason it cancels.*

> *The answer depends on powers of* 2*. That means the proof is a descent: I need one case where I can see the sum is zero, one where I can see it isn't, and a move that halves m and n together.*

> *A sum over* $0, \ldots, N-1$ *with floors: try the reflection* $i \mapsto N - 1 - i$*. Floors reflect cleanly,* $\lfloor i/m \rfloor + \lfloor (mn - 1 - i)/m \rfloor = n - 1$*, so the exponent flips parity exactly when* $m + n$ *is odd.*

- ▷ **Cue:** a sum of ± 1 over a range of length N, asked when it is zero $\longrightarrow$ **Move:** check parity of N first: an odd number of ± 1's cannot vanish
- ▷ **Cue:** a sum over $i = 0, \ldots, N-1$ involving $\lfloor i/m \rfloor$ with $m \mid N$ $\longrightarrow$ **Move:** pair i with $N - 1 - i$; the floors sum to $N/m - 1$

▷ **Cue:** the answer to a problem about m, n seems to depend on powers of 2 $\longrightarrow$ **Move:** find a relation between $S(2k, 2l)$ and $S(k, l)$ and descend

▷ **Cue:** an "iff" condition on two integer parameters with no obvious formula $\longrightarrow$ **Move:** tabulate small cases before choosing a proof strategy

Layer 3 THE WHY

Why is the reflection $i \mapsto mn - 1 - i$ natural rather than lucky? Go to what a floor is. $\lfloor i/m \rfloor$ counts how many complete blocks of length m lie before i. The range $0, \ldots, mn - 1$ consists of exactly n blocks of length m and exactly m blocks of length n. Reading the range backwards turns "blocks before i" into "blocks after i", and before plus after equals total minus one (the block containing i). So the exponent $f(i)$ has a mirror partner with $f(i) + f(i')$ fixed at $m + n - 2$. The whole question is then the parity of $m + n - 2$. That is where the reflection comes from: the range is a union of whole blocks for *both* moduli simultaneously, which is exactly why the upper limit is $mn - 1$.

Why a descent by 2? Because the table says so, and because of what the reflection could not do. The reflection settles opposite parity; the odd count settles both odd. What remains, both even, cannot be settled by either tool, and the only natural thing to do with two even numbers is to halve them. Halving works because floors by an even number do not distinguish $2j$ from $2j + 1$ — the sign pattern is the pattern for (k, l) with every entry doubled. Look at the resulting identity: it reduces the both-even case to a smaller instance and nothing else. That is the signature of a v_2 answer.

Kedlaya's solution states $S(2k, 2l) = S(k, l)(1 + (-1)^{k+l})$. The correct identity has a factor 2 in front, as computed above (check $k = l = 1$: the exponents $f(0), \ldots, f(3)$ are $0, 0, 2, 2$, so $S(2, 2) = 4 = 2 \cdot 2 \cdot S(1, 1)$, whereas the official identity would give 2). The missing 2 changes nothing in the conclusion, since only vanishing matters. Their phrase "piecing our cases together, we easily deduce" hides the descent; state it, because a grader wants to see why the halving eventually stops at a case already settled.

The natural attempt is zombie maths in its purest form: evaluating when you were asked only about vanishing. The problem punishes it directly — the actual values of S for both-even m, n grow with no simple formula, and nobody needs them.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "necessary and sufficient conditions $\ldots = 0$". On seeing it, write a table of $S(m, n)$ for $m, n \leq 6$ and circle the zeros.

2. The table suggests $v_2(m) \neq v_2(n)$, which suggests three tools: a parity count, a pairing (reflection of the index range), and a descent $S(2k, 2l) \leftrightarrow S(k, l)$.

3. Hesitation comes when the reflection works for opposite parity but visibly fails for both even. It is reinforced away by recognising that "both even" is precisely the case a v_2-answer expects you to reduce by halving.

4. The trap is trying to evaluate $S(m, n)$ in closed form via block counting and $\gcd(m, n)$. A secondary trap is omitting the argument that the halving recursion preserves the condition in both directions, including the subcase $k + l$ odd.

5. Recognise faster by asking of any "when is this zero" problem: what involution of the index set would flip the sign of each term? Here the reversal of $\{0, \ldots, mn - 1\}$ is the obvious map that carries whole m-blocks to whole m-blocks and whole n-blocks to whole n-blocks at the same time.

6. Reusable decision rule: *when asked when a signed sum vanishes, do not evaluate it; find a sign-reversing involution for the zero cases, a parity or size argument for the nonzero cases, and a reduction map linking the rest to a smaller instance.*

1998 B5

1998 B5. Let N be the positive integer with 1998 decimal digits, all of them 1; that is,

$$N = 1111\cdots11.$$

Find the thousandth digit after the decimal point of $\sqrt{N}$.

Fields: number theory; asymptotic expansion; decimal expansions **Position:** B5 — nominally late in the paper, but short once N is written in closed form; the difficulty is entirely in controlling the error honestly **Answer:** The thousandth digit is 1.

Layer 1 THE WHAT — surface analysis

The natural attempt. The strong student starts from small cases: $\sqrt{11} = 3.3166\ldots$, $\sqrt{1111} = 33.33166\ldots$, $\sqrt{111111} = 333.333166\ldots$. The pattern is visible: k threes before the point, then k threes after, then a 1, then sixes. With $2k = 1998$, $k = 999$, so the digits after the point would be 999 threes, then 1 in position 1000. The answer "1" is guessed correctly in five minutes.

Then the student stalls, because a pattern in three examples is not a proof, and the obvious way to make it one — long division square-root algorithm, digit by digit — is hopeless at 1998 digits. Some students write the answer with "by the pattern" and lose most of the marks. Others try to square $333\ldots3.333\ldots31666\ldots$ and compare with N, and drown in carries.

The model chain. **Step 1: closed form.** $N = \dfrac{10^{1998}-1}{9}$, so

$$\sqrt{N} = \frac{10^{999}}{3}\sqrt{1-x}, \qquad x = 10^{-1998}.$$

Step 2: two-sided bound on $\sqrt{1-x}$. For $0 < x < 1$,

$$1 - \tfrac{x}{2} - \tfrac{x^2}{2} \;<\; \sqrt{1-x} \;<\; 1 - \tfrac{x}{2}.$$

Proof of both: all quantities are positive for our tiny x, so square. $(1-\frac{x}{2})^2 = 1 - x + \frac{x^2}{4} > 1 - x$ gives the right inequality. For the left, $(1-\frac{x}{2}-\frac{x^2}{2})^2 = 1 - x - \frac{3x^2}{4} + \frac{x^3}{2} + \frac{x^4}{4} < 1 - x$ when $x < 1$ (since $\frac{x^3}{2} + \frac{x^4}{4} < \frac{3x^2}{4}$). So

$$\sqrt{N} = \frac{10^{999}}{3} - \frac{10^{-999}}{6} - \varepsilon, \qquad 0 < \varepsilon < \frac{10^{999}}{3}\cdot\frac{10^{-3996}}{2} < 10^{-2997}.$$

Step 3: read off digits. $10^{999}/3 = 33\ldots3.333\ldots$ with 999 threes before the point and threes forever after. Next, $\frac{10^{-999}}{6} = 10^{-1000}\cdot\frac{10}{6} = 10^{-1000}\cdot 1.666\ldots$: its first 999 decimal digits are 0, digit 1000 is 1, and digits 1001, 1002, … are 6. Subtracting,

$$0.\underbrace{33\ldots3}_{999}3333\ldots \;-\; 0.\underbrace{00\ldots0}_{999}1666\ldots \;=\; 0.\underbrace{33\ldots3}_{999}1666\ldots,$$

since $0.333\ldots - 0.1666\ldots = \frac{1}{3} - \frac{1}{6} = \frac{1}{6}$ with no borrow into position 999. Finally subtract $\varepsilon < 10^{-2997}$: it removes less than one unit in position 2997 from a tail of 6s, so digits up to position 2996 are unchanged (a borrow would need a run of zeros, and there is none). Hence digit 1000 is 1. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains see the same pattern. They split at how to certify it. (N) attacks the *digits* of $\sqrt{N}$ directly — the digit-by-digit algorithm, or squaring a guessed decimal. (M) attacks the *number*: it writes N in closed form, factors out the huge part $10^{999}/3$, and treats the rest as 1 minus something microscopic. At that point the digits are read off a two-term expansion plus an error bar.

> *A repunit is* $(10^k-1)/9$*. Then* $\sqrt{N}$ *is* $10^{999}/3$ *times* $\sqrt{1-10^{-1998}}$*, and* $\sqrt{1-x}$ *for tiny* x *is* $1-x/2$ *plus something I can bound.*

> *I need exactly one digit, at position* 1000*. So I only need* $\sqrt{N}$ *to within about* 10^{-1001}*, and I must check no borrow from the error reaches position* 1000*.*

- ▷ **Cue:** a number with a long run of equal digits ⟶ **Move:** write it in closed form as a geometric sum, e.g. $(10^k-1)/9$
- ▷ **Cue:** $\sqrt{A^2-\delta}$ or $\sqrt{A^2(1-x)}$ with x tiny ⟶ **Move:** factor out A and use $1-\frac{x}{2}-\frac{x^2}{2} < \sqrt{1-x} < 1-\frac{x}{2}$, proved by squaring
- ▷ **Cue:** a specific digit of a real number is asked for ⟶ **Move:** find an approximation with an explicit two-sided error bar, then check the digits between the target and the error are not all 0 or all 9 (no carry or borrow reaches the target)

Layer 3 THE WHY

Why is the closed form the natural first move? Because the question is about $\sqrt{\cdot}$, and square root behaves well on products and badly on sums. "1111…1" is a sum of 1998 powers of 10; as a sum, you cannot take its square root. Rewritten as $(10^{1998}-1)/9$, it is $\frac{1}{9}\cdot 10^{1998}\cdot(1-10^{-1998})$ — a product — and the square root distributes: $\frac{1}{3}\cdot 10^{999}\cdot\sqrt{1-10^{-1998}}$. Every factor is now understood. That is not a trick; it is the only form of N on which $\sqrt{}$ acts.

Why does the digit come out as 1, and why at position 1000? The leading term $10^{999}/3$ has threes everywhere. The correction $-\frac{x}{2}\cdot\frac{10^{999}}{3} = -\frac{10^{-999}}{6}$ has size about 1.7×10^{-1000} — its first significant digit sits *exactly* at position 1000. The problem is set so that the first-order correction lands on the digit asked for; $1998 = 2\times 999$ and "thousandth" $= 999+1$ are chosen together. Seeing that tells you the problem expects a first-order expansion and no more, and that the second-order term (size 10^{-2997}) is there only to be bounded.

Why insist on the carry/borrow check? A digit is not a continuous function of the number. $0.1999\ldots9$ and $0.2000\ldots0$ are within 10^{-100} of each other and disagree in the first digit. So an approximation to within 10^{-2997} determines digit 1000 only if the digits of the approximation between positions 1001 and 2997 are not all 0 (for a subtracted error). They are all 6, so the error cannot propagate. Kedlaya's solution writes "$r < 10^{-2000}$" and reads off the digits without comment; the reason it is safe is this run of sixes, and the proof is incomplete without saying so. Note also that the official r is, strictly, negative ($\sqrt{1-x} < 1-\frac{x}{2}$); what matters is its size.

The natural attempt found the right answer and could not prove it because it worked with digits instead of with the number. Digit-level computation is zombie maths here: 1998 digits of long division, when a two-line expansion does it.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "all of them 1". On seeing it, write $N=(10^{1998}-1)/9$.

2. Closed form, factor out the dominant square, first-order expansion $\sqrt{1-x}\approx 1-\frac{x}{2}$ with a rigorous two-sided bound.

3. Hesitation comes after guessing the answer from $\sqrt{11}$, $\sqrt{1111}$: the pattern is convincing but unproved. It is reinforced away by recognising that the pattern *is* the first two terms of an expansion, so proving it means bounding the third.

4. The trap is proof by pattern. A subtler trap is forgetting that an approximation fixes a digit only when no carry or borrow can reach it.

5. When a digit of an irrational number is asked for, the setter has arranged a clean leading term plus a correction landing at the asked position. Look for the correction's size first: $10^{999}\cdot 10^{-1998} = 10^{-999}$ points straight at position 1000.

6. Reusable decision rule: *to find a specific digit of f(huge number), rewrite the number as a product (main term) $\times(1+\text{tiny})$, expand to first order, bound the second-order term, and verify that the digits between the target and the error bar forbid any carry.*

1998 B6

1998 B6. Prove that, for any integers a, b, c, there exists a positive integer n such that $\sqrt{n^3+an^2+bn+c}$ is not an integer.

Fields: number theory; quadratic residues modulo 4; finite differences **Position:** B6 — last problem, but the shortest proof is four lines; the difficulty is choosing the modulus, not executing the argument
Answer: Proof: if $n \not\equiv b \pmod 2$, then $p(n)$ and $p(n+2)$ cannot both be perfect squares.

Layer 1 THE WHAT — surface analysis

The natural attempt. Write $p(n) = n^3+an^2+bn+c$. Suppose, for contradiction, that $p(n)$ is a perfect square for every $n \geq 1$. The strong student reaches for growth: $p(n)$ is roughly n^3, so $\sqrt{p(n)} \approx n^{3/2}$, and consecutive squares near n^3 are about $2n^{3/2}$ apart, while $p(n+1)-p(n) \approx 3n^2$. That comparison is not a contradiction — p jumps past many squares, and nothing stops it landing on one each time. The next idea is to show p must be the square of a polynomial, which a cubic cannot be; but "takes square values at all positive integers $\Rightarrow$ square of a polynomial" is a real theorem, not a one-liner. (The official third solution does make an analytic argument work, by restricting to n a large even perfect square so that $n^{3/2}$ is an integer and sandwiching $p(n)$ between $(n^{3/2}+\frac{a}{2}n^{1/2}\pm 1)^2$; it needs care with parity of a and with "for sufficiently large n".) Another reach is the full generality of the second official solution — primes dividing values of P, discriminants, Hensel-type lifting — a correct and powerful argument that is far more than this problem asks for. The student spends forty minutes on growth estimates that never quite close.

The model chain. Let $p(n) = n^3+an^2+bn+c$. We use two facts.

(i) *Every perfect square is $\equiv 0$ or $1 \pmod 4$, and it is 0 if even, 1 if odd.* Indeed $(2k)^2 = 4k^2$ and $(2k+1)^2 = 4(k^2+k)+1$.

(ii) *The difference $p(n+2)-p(n)$.* Expanding,

$$p(n+2)-p(n) = (6n^2+12n+8)+a(4n+4)+2b \equiv 2n^2+2b \pmod 4.$$

In particular $p(n+2)-p(n)$ is even, so $p(n)$ and $p(n+2)$ have the same parity.

Argument. Choose a positive integer n of parity opposite to b: $n=1$ if b is even, $n=2$ if b is odd. Then $n^2+b \equiv n+b \equiv 1 \pmod 2$, so $2n^2+2b \equiv 2 \pmod 4$, and $p(n+2)-p(n) \equiv 2 \pmod 4$.

Suppose $\sqrt{p(n)}$ and $\sqrt{p(n+2)}$ were both integers. Then $p(n)$ and $p(n+2)$ are perfect squares of the same parity, so by (i) they are congruent mod 4, and $4 \mid p(n+2)-p(n)$. Contradiction. So at least one of the positive integers n, $n+2$ has $\sqrt{p(\cdot)}$ not an integer. (If $p(\cdot)<0$ there, the square root is not even real, which is certainly not an integer.) $\square$

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start from "suppose every value is a square". They split at what to exploit. (N) exploits *size*: squares are sparse among large numbers, so a cubic should miss them. (M) exploits *residues*: squares are sparse modulo 4 (only $0, 1$), so two values of p whose difference is $2 \pmod 4$ cannot both be squares. The model needs two values, not infinitely many, and no growth at all.

> *Squares are rare in two ways: they are far apart, and they occupy few residue classes. The first needs asymptotics. The second needs one congruence. Try the cheap one first, modulo* 4.

> *Modulo* 4 *a square is* 0 *or* 1 *according to its parity. So two squares of the same parity are congruent mod* 4. *I want two values of* p *with the same parity whose difference is* 2 mod 4. *Shifting by* 2 *keeps parity; compute* $p(n+2) - p(n)$ *mod* 4.

- ▷ **Cue:** show a polynomial with integer coefficients does not always take square values $\longrightarrow$ **Move:** look for a small modulus (4, 8, 3, 5) and two inputs whose outputs cannot both be squares modulo it
- ▷ **Cue:** two integers known to have the same parity, and both supposed to be squares $\longrightarrow$ **Move:** they must then be congruent mod 4 (even squares are 0, odd squares are 1)
- ▷ **Cue:** a polynomial p with integer coefficients and a shift h $\longrightarrow$ **Move:** $p(n+h) - p(n)$ is divisible by h; choosing $h = 2$ preserves parity, and the next power of 2 is where information lives

Layer 3 THE WHY

Why modulo 4, and why shift by 2? Go back to what the problem allows. There are three arbitrary integers a, b, c. Any congruence argument has to work for every choice of them. A single value $p(n)$ can be made to be anything modulo anything by choosing c, so no single-value argument can work; you need a *relation between two values* in which c cancels. Differences do that: $p(n+h) - p(n)$ does not involve c at all.

Now which modulus? The residues of squares modulo 4 are $\{0, 1\}$, and they are determined by parity. That turns "both squares" into a statement about parity plus one bit mod 4. So you want two values of the same parity — a shift by an even h — and then you need their difference not divisible by 4. With $h = 2$, the difference is $2\big(3n^2 + 6n + 4 + 2a(n+1) + b\big)$, and the bracket is $\equiv n^2 + b \pmod 2$; the a-term drops out because it carries an extra factor 2. So the whole thing depends on $n + b$ modulo 2, and you can choose n to make it odd. Every choice was forced: the difference kills c, the shift by 2 kills parity mismatch and a, and the free choice of n kills b.

Kedlaya's first solution is these four lines, without the motivation. Here is where they come from: c is uncontrollable, so compare two values; squares mod 4 depend only on parity, so compare values of equal parity; a cubic's second difference structure makes $p(n+2) - p(n) \equiv 2(n^2 + b)$, and one bit of freedom in n does the rest.

The problem is B6 and punishes the solver who believes a last problem needs deep tools. The growth argument and the prime-divisor argument both prove more (the latter for every non-square polynomial), and both cost far more time. The modular argument is specific to this problem and is the cheap route; checking a small modulus before anything else is exactly the rerouting habit.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "for any integers $a, b, c \ldots$ not an integer": a statement that a polynomial is not a square everywhere. On seeing it, write the residues of squares modulo 4 and 8.

2. A congruence obstruction: find two inputs whose outputs cannot both be squares modulo a small number, with the uncontrolled constant c cancelling in a difference.

3. Hesitation comes from the size of the paper position — "B6 cannot be a mod-4 argument". It is reinforced away by the rule of trying the cheapest obstruction first; a modular argument takes two minutes to test and either works or fails visibly.

4. The trap is growth estimates: $p(n)$ jumps over many squares, so size alone gives no contradiction for a single n, and making the analytic route rigorous needs a careful sandwich on a special subsequence. A secondary trap is a single-value congruence argument, which fails because c is arbitrary.

5. Ask: which parameters are arbitrary, and what operation eliminates them? A difference removes the constant term; a shift by 2 removes the parity issue; the remaining parameter is absorbed by the free choice of n.

6. Reusable decision rule: *to show a family of integer polynomials is not always a square, compare two values whose difference is independent of the uncontrolled coefficients, and pick a modulus (usually* 4 *or* 8*) where squares of known parity have a forced residue.*