# DAWN PUTNAM SESSION

Putnam 2000 — Papers A and B

Putnam 120 Archive

# Contents

**Putnam 2000 — Paper A** **5**
2000 A1 . . . . . 5
2000 A2 . . . . . 7
2000 A3 . . . . . 9
2000 A4 . . . . . 12
2000 A5 . . . . . 15
2000 A6 . . . . . 18

**Putnam 2000 — Paper B** **21**
2000 B1 . . . . . 21
2000 B2 . . . . . 23
2000 B3 . . . . . 25
2000 B4 . . . . . 28
2000 B5 . . . . . 31
2000 B6 . . . . . 33

# Putnam 2000 — Paper A

## 2000 A1

**2000 A1.** Let $A$ be a positive real number. What are the possible values of $\sum_{j=0}^{\infty} x_j^2$, given that $x_0, x_1, \ldots$ are positive numbers for which $\sum_{j=0}^{\infty} x_j = A$?

**Fields:** sequences and series, elementary inequalities **Position:** A1 — the opener **Answer:** Every value in the open interval $(0, A^2)$, and no others.

### Layer 1 THE WHAT — surface analysis

*The natural attempt.* The instinct on seeing $\sum x_j = A$ and $\sum x_j^2$ side by side is Cauchy–Schwarz, or its cousin, the identity $\left(\sum x_j\right)^2 = \sum x_j^2 + 2\sum_{j<k} x_j x_k$. Squaring the given sum and expanding is the honest first move, and it does produce something real: since every $x_j > 0$, every cross term $x_j x_k$ is positive, so $A^2 = \sum x_j^2 + (\text{a positive quantity})$, hence $\sum x_j^2 < A^2$. That is a genuine upper bound, obtained cheaply.

The trouble starts at the lower bound and at the question of which values in between are actually achieved. A natural next guess is that the answer is simply "$(0, A^2)$" by a hand-wave — $\sum x_j^2$ can plainly be made small by spreading $A$ thin across many terms, and large by concentrating it in one term, so surely everything between is hit "by continuity". But there is no function of a real parameter on the table yet, no continuity to invoke, and no single family of examples has been exhibited that sweeps the whole interval. A student who stops here has correctly bounded the set of possible values but has not shown the set actually equals $(0, A^2)$; they have proved a necessary condition, not exhibited it as sufficient. Worse, some students try to patch this by writing down two or three specific examples (all mass in $x_0$; mass split evenly among $n$ terms) and asserting "clearly everything in between is possible", which is exactly the kind of gap the Putnam grader is trained to dock.

*The model chain.* Fix a single one-parameter family that can be pushed continuously between the two extremes, and read off its range. Take a geometric progression: $x_j = x_0 d^j$ for $j = 0, 1, 2, \ldots$, with $0 < d < 1$ (needed so all $x_j > 0$ and the series converges) and $x_0 > 0$. Then

$$\sum_{j=0}^{\infty} x_j = \frac{x_0}{1-d} = A \implies x_0 = A(1-d),$$

and

$$\sum_{j=0}^{\infty} x_j^2 = \frac{x_0^2}{1-d^2} = \frac{A^2(1-d)^2}{(1-d)(1+d)} = A^2 \cdot \frac{1-d}{1+d}.$$

The map $d \mapsto \frac{1-d}{1+d}$ is continuous and strictly decreasing on $(0,1)$ (its derivative is $-2/(1+d)^2 < 0$), with

$$\lim_{d \to 0^+} \frac{1-d}{1+d} = 1, \qquad \lim_{d \to 1^-} \frac{1-d}{1+d} = 0.$$

So as $d$ ranges over the open interval $(0,1)$ — and every such $d$ gives a genuine sequence of positive numbers summing to $A$ — the value $A^2 \cdot \frac{1-d}{1+d}$ ranges over the full open interval $(0, A^2)$, hit exactly once each by the Intermediate Value Theorem applied to a strictly monotonic continuous function. Combined with the upper bound $\sum x_j^2 < A^2$ (from the cross-term argument above, which is still needed, since the geometric family alone does not rule out some sequence achieving $A^2$ or more by other means) and the trivial lower bound $\sum x_j^2 > 0$ (each $x_j$ is positive, so $x_0^2 > 0$ already), the possible values are exactly $(0, A^2)$.

**Layer 2 THE HOW — the bifurcation**

**Bifurcating point.** Both chains reach the same upper bound $\sum x_j^2 < A^2$ by the same cross-term argument — that part is not where they split. They split at the point where the natural attempt, having the bound in hand, is tempted to declare the whole interval achieved by inspection: "the sum can be made small or large, so everything between must occur." The model chain instead asks for a single continuous one-parameter family whose value at every parameter is a legitimate example, and lets the Intermediate Value Theorem — applied honestly, to an explicit continuous function of one real variable — do the work that "clearly everything in between" was trying to do by assertion.

> *I have an upper bound and I have intuition that small and large values are both possible, but intuition about the endpoints is not a proof that the middle is reachable. I need one formula, in one real parameter, that produces a genuine example for every value of that parameter, and whose output value is a continuous, monotonic function of the parameter — only then can I invoke the Intermediate Value Theorem instead of asserting it.*

**Layer 3 THE WHY**

Why does a geometric progression, specifically, supply that one-parameter family? Because the problem gives exactly two pieces of data to control — the sum $A$ and the sum of squares — and a two-term-recursive family like $x_j = x_0 d^j$ has exactly two free parameters, $x_0$ and $d$, which is precisely enough freedom to pin down $A$ (fixing $x_0$ in terms of $d$) and leave one genuinely free real parameter, $d$, to sweep out a range of values for the second quantity. Any family with only one free parameter total (e.g. $x_j$ all equal) collapses to a single point once $A$ is fixed, and cannot sweep an interval at all. A family with the sum $\sum x_j^2 / (\sum x_j)^2$ expressed as a ratio of two homogeneous quadratic-and-linear functionals of $(x_0, x_1, \ldots)$ is scale-invariant in exactly the way that makes "normalise $\sum x_j = A$, then vary the shape of the sequence" the right axis to move along — and the geometric family is the simplest infinite family whose "shape" is controlled by a single number $d$.

The deeper reason the interval is open, not closed, is worth isolating from the mechanics: $\sum x_j^2 = A^2$ would require all the cross terms $2x_j x_k$ to vanish, i.e. at most one $x_j$ nonzero, which is disallowed by the hypothesis that every $x_j$ is positive (an infinite sequence of positive numbers, not merely nonnegative). Similarly $\sum x_j^2 = 0$ would require every $x_j = 0$. So the openness of $(0, A^2)$ is not a technicality of the geometric-progression computation; it is forced by the problem's own hypothesis that every term is strictly positive, and the geometric family merely realises how close to each endpoint one can get without ever reaching it, since $d$ itself only ranges over the open interval $(0,1)$ for the same reason ($d = 0$ would force $x_1 = x_2 = \cdots = 0$, and $d = 1$ would fail to converge).

**Layer $+\alpha$ BACKTESTING**

**1.** The first useful cue is a sum and a sum-of-squares of the same sequence, presented together. On seeing it, write down the expansion $\left(\sum x_j\right)^2 = \sum x_j^2 + 2\sum_{j<k} x_j x_k$ immediately — it is the only identity that relates the two quantities directly, and it produces the upper bound for free.

**2.** That cue suggests bounding $\sum x_j^2$ from the square-of-the-sum identity, then searching for an explicit one-parameter family of admissible sequences to determine exactly which bounded values are attained — not stopping at the bound and asserting the rest.

**3.** Hesitation occurs right after the bound $\sum x_j^2 < A^2$ is secured, where it is tempting to treat "I can make it small" and "I can make it close to $A^2$" as sufficient evidence for the whole interval. What reinforces the correct move is asking: is there one formula, in one real parameter, whose value sweeps continuously between these extremes? If the answer is "I haven't written one down", the argument is not finished.

**4.** The trap is answering with a closed interval $[0, A^2]$ or with the half-open $(0, A^2]$, either by forgetting that $x_j > 0$ strictly rules out concentrating all the mass in one term, or by forgetting that spreading the mass over finitely many terms (rather than the full infinite sequence) is not the same construction as letting $d \to 1^-$ over positive infinite sequences.

**5.** Recognise it faster by treating any "sum is fixed, describe the range of the sum of squares (or similar homogeneous quantity)" problem as an instruction to reach for a geometric progression or other single-parameter scaling family as the default sweeping tool, before reaching for case-by-case examples.

**6.** Reusable rule: to prove a claimed range of values $(m, M)$ for some functional of a constrained sequence is fully achieved — not merely bounded — exhibit one continuous, monotonic, one-real-parameter family of valid instances and invoke the Intermediate Value Theorem on that explicit function; separately, always check that the endpoints $m, M$ themselves are excluded or included by re-examining whether the problem's strict hypotheses (here, $x_j > 0$, not $x_j \geq 0$) can actually be met at the boundary.

▷ **Cue:** a fixed sum $\sum x_j = A$ and a bound on the sum of squares $\sum x_j^2$ are wanted together $\longrightarrow$ **Move:** expand $(\sum x_j)^2 = \sum x_j^2 + 2\sum_{j<k} x_j x_k$ to get $\sum x_j^2 < A^2$ for free from positivity of the cross terms

▷ **Cue:** a claimed interval of achievable values needs to be shown fully achieved, not just bounded $\longrightarrow$ **Move:** construct one continuous one-parameter family (e.g. a geometric progression) realising every value, and invoke the Intermediate Value Theorem on that explicit function rather than asserting the middle values by inspection

## 2000 A2

**2000 A2.** Prove that there exist infinitely many integers $n$ such that $n, n+1, n+2$ are each the sum of the squares of two integers. [Example: $0 = 0^2 + 0^2$, $1 = 0^2 + 1^2$, $2 = 1^2 + 1^2$.]

**Fields:** elementary number theory, sums of two squares **Position:** A2 **Answer:** Yes: take $n = 4k^4 + 4k^2$ for any nonnegative integer $k$, giving $n = (2k^2)^2 + (2k)^2$, $n+1 = (2k^2+1)^2$, $n+2 = (2k^2+1)^2 + 1^2$.

### Layer 1 THE WHAT — surface analysis

*The natural attempt.* The example in the problem statement ($n = 0$: $0, 1, 2$) invites case-checking: compute which small $n$ have $n, n+1, n+2$ all expressible as sums of two squares, and look for a pattern. This turns up $n = 0, 2, 8, 16, 18, \ldots$ fairly quickly, and a student with the sum-of-two-squares theorem in reach (a positive integer is a sum of two squares iff every prime $\equiv 3 \pmod 4$ in its factorisation appears to an even power) can in principle check each candidate this way. But this is a search, not a construction: nothing about scanning outputs of a theorem case by case tells you the search never terminates, and "the pattern seems to continue" is not a proof of infinitude.

A more structural instinct is to reach for the sum-of-two-squares theorem directly and try to force it: pick an even $a$ with $a^2 + 1$ composite and not a sum of two squares' worth of obstruction, factor $a^2 + 1 = rs$, and build $n = x^2 - 1$ from $x = (r+s)/2$. This works (it is one

of the official routes) but it requires the sum-of-two-squares theorem as a prerequisite, plus a divisibility argument to guarantee infinitely many valid $a$ exist, plus checking $r, s$ have the right parity to make $(r+s)/2$ an integer. It is correct machinery, but heavy for what turns out to be, underneath, a much smaller algebraic fact.

*The model chain.* Notice first that the problem's three conditions are not equally hard. If $n+1$ happens to be a perfect square, say $n+1 = x^2$, then $n+1 = x^2 + 0^2$ is automatically a sum of two squares — free. And then $n+2 = x^2+1 = x^2+1^2$ is *also* automatically a sum of two squares — also free, for any $x$ at all. So two of the three conditions are satisfied the moment $n+1$ is chosen to be a perfect square $x^2$; the entire problem collapses to a single remaining question: find infinitely many integers $x$ for which $x^2-1$ is a sum of two squares.

Factor $x^2-1 = (x-1)(x+1)$ and look for a value of $x$ that makes this factorisation transparently a product of sums of two squares, since a product of two sums of two squares is again a sum of two squares (Brahmagupta–Fibonacci identity: $(a^2+b^2)(c^2+d^2) = (ac-bd)^2 + (ad+bc)^2$). Take $x = 2k^2+1$ for an integer $k \geq 0$. Then

$$x-1 = 2k^2 = (2k)(k), \qquad x+1 = 2k^2+2 = 2(k^2+1),$$

so

$$x^2-1 = (x-1)(x+1) = 4k^2(k^2+1) = \big((2k)^2+0^2\big)\big(k^2+1^2\big).$$

Applying the identity with $(a,b) = (2k,0)$ and $(c,d) = (k,1)$:

$$x^2-1 = (2k\cdot k - 0\cdot 1)^2 + (2k\cdot 1 + 0\cdot k)^2 = (2k^2)^2 + (2k)^2.$$

So, setting $n = x^2-1 = 4k^4+4k^2$, we have $n = (2k^2)^2+(2k)^2$, $n+1 = x^2 = (2k^2+1)^2+0^2$, and $n+2 = x^2+1 = (2k^2+1)^2+1^2$ — all three verified directly, with no appeal to the sum-of-two-squares theorem at all. As $k$ ranges over $0,1,2,\ldots$, $n = 4k^4+4k^2$ takes infinitely many distinct values (it is strictly increasing in $k$), giving infinitely many valid $n$.

### Layer 2 THE HOW — the bifurcation

**Bifurcating point.** Both chains eventually use the algebraic fact that a product of sums of two squares is a sum of two squares — the search-based attempt uses it implicitly through the sum-of-two-squares theorem (which is proved using exactly this multiplicativity), and the model chain uses it explicitly and by hand. The chains genuinely split earlier than that, at the very first move: the search-based attempt treats all three conditions ($n$, $n+1$, $n+2$ each a sum of two squares) as three separate obstacles to satisfy simultaneously. The model chain asks whether the three conditions are really independent, and notices that fixing $n+1 = x^2$ discharges the $n+1$ and $n+2$ conditions for free, for *every* $x$, reducing three conditions to one.

> *Three simultaneous conditions on consecutive integers sounds like it should require juggling three separate constructions, but $n+1$ and $n+2$ are consecutive to each other too — and a perfect square plus $0^2$, and a perfect square plus $1^2$, are both automatically sums of two squares no matter what the perfect square is. If I just force $n+1$ to be a perfect square, two of my three conditions vanish for free, and I only have one condition left to actually engineer.*

### Layer 3 THE WHY

Why does forcing $n+1 = x^2$ collapse the problem this cleanly? Because "is a sum of two squares" is a condition satisfied trivially whenever one of the two squares is allowed to be $0^2$ or $1^2$ — and a perfect square $x^2$ is, tautologically, $x^2+0^2$. The problem's own phrasing gives this away: the worked example $0,1,2$ has $n+1 = 1 = 1^2$, a perfect square, which is exactly the free move being generalised here. Recognising that the example already contains the reduction is the difference between treating the example as decoration and treating it as the seed of the construction.

Having reduced to "find infinitely many $x$ with $x^2-1$ a sum of two squares", the factorisation $x^2-1=(x-1)(x+1)$ is the natural move because a difference of squares is the one factorisation identity that applies to $x^2-1$ without further hypotheses, and it converts an additive question (is this number a sum of two squares) into a multiplicative one (is this product of two numbers a sum of two squares), which is exactly where the Brahmagupta–Fibonacci identity lives: sums of two squares are closed under multiplication precisely because $\mathbb{Z}[i]$ is closed under multiplication and $a^2+b^2=|a+bi|^2$, so a product of norms is the norm of a product. Choosing $x=2k^2+1$ specifically is what makes $x-1=2k^2$ manifestly of the form $(2k)^2+0^2$ and $x+1=2(k^2+1)$ manifestly 2 times a sum of two squares $k^2+1^2$ — the choice is reverse-engineered from wanting both factors to already be recognisable as (twice) a sum of two squares, so that the identity can be applied by inspection rather than searched for.

**Layer $+\alpha$ BACKTESTING**

**1.** The first useful cue is three consecutive integers each required to satisfy the same kind of condition. On seeing it, write down: are these three conditions really independent, or does satisfying one of them for a general reason automatically satisfy a neighbour?

**2.** That cue suggests looking specifically at $n+1$, the middle term, and asking what special form (perfect square, in this case) would make both its neighbours' conditions free consequences rather than separate constraints.

**3.** Hesitation occurs right after the reduction to "find infinitely many $x$ with $x^2-1$ a sum of two squares", where the instinct is to reach for the sum-of-two-squares theorem (prime factorisation, primes $\equiv 3 \pmod 4$) as the only tool available. What reinforces the cheaper move is remembering that $x^2-1$ factors, and that a factored form is an invitation to use multiplicativity of sums of two squares directly, without classifying primes at all.

**4.** The trap is treating the problem as needing infinitely many independent verifications (one full sum-of-two-squares check per candidate $n$), when a single algebraic identity, verified once and parametrised by $k$, supplies all of them simultaneously and requires no case-by-case checking.

**5.** Recognise it faster by treating "several consecutive integers must each satisfy the same divisibility/representability condition" as an instruction to look for a term (often the middle one) that can be forced into a canonical form making its neighbours' conditions automatic.

**6.** Reusable rule: when a target integer property (sum of two squares, perfect $k$-th power plus a fixed correction, etc.) is trivially true for numbers of the form $x^2$, $x^2\pm1$, or similar, and the problem demands the property on a short run of consecutive integers, search for a substitution that forces the *easiest* member of the run into that canonical form first, then verify only the one remaining member — usually via a factorisation identity rather than a general classification theorem.

▷ **Cue:** several consecutive integers must each satisfy the same number-theoretic representability condition $\longrightarrow$ **Move:** check whether forcing one member (often the middle one) into a canonical form – perfect square, etc. – makes its neighbours' conditions automatic, reducing several constraints to one

▷ **Cue:** a number of the form $x^2-1$ (or similar difference of squares) is asked to be a sum of two squares $\longrightarrow$ **Move:** factor as $(x-1)(x+1)$ and choose $x$ so each factor is visibly (a multiple of) a sum of two squares, then apply the Brahmagupta–Fibonacci identity $(a^2+b^2)(c^2+d^2)=(ac-bd)^2+(ad+bc)^2$ by inspection

## 2000 A3

**2000 A3.** The octagon $P_1P_2P_3P_4P_5P_6P_7P_8$ is inscribed in a circle, with the vertices around the circumference in the given order. Given that the polygon $P_1P_3P_5P_7$ is a square of area 5,

and the polygon $P_2P_4P_6P_8$ is a rectangle of area 4, find the maximum possible area of the octagon.

**Fields:** plane geometry, constrained optimisation **Position:** A3 **Answer:** $3\sqrt{5}$.

**Layer 1 THE WHAT — surface analysis**

*The natural attempt.* The natural first move is to pin down the circle. A square of area 5 inscribed in a circle of radius $r$ has diagonal $2r$, and a square's area is half its diagonal squared, so $2r^2 = 5$, giving $r^2 = 5/2$. The rectangle, inscribed in the same circle, has its diagonal also equal to $2r$, so if its sides are $p, q$: $p^2 + q^2 = 4r^2 = 10$ and $pq = 4$ (the given area). Solving, $(p+q)^2 = p^2 + q^2 + 2pq = 18$ and $(p-q)^2 = 10 - 8 = 2$, giving $p = 2\sqrt{2}$, $q = \sqrt{2}$. This much is clean and forced — there is no choice involved yet.

The trouble starts at "find the maximum area of the octagon". A strong student's honest instinct is to set up eight points on the circle with angular coordinates, write the octagon's area as a sum of 8 triangles from the centre ($\frac{1}{2}r^2 \sin\theta_i$ for each central angle $\theta_i$ between consecutive vertices), impose the constraints that four alternating angles sum in a way that forces a square and the other four force the given rectangle, and then run Lagrange multipliers or substitution over the remaining degrees of freedom. This is not wrong, but it is heavy machinery for a competition, it obscures which two quantities are actually free, and it is easy to lose track of which angle differences are genuinely independent versus rigidly forced by "these four points form an exact square." Many attempts along this route stall in algebra before reaching a clean answer, or arrive at a correct-looking but under-justified optimum by symmetry intuition without checking that both proposed optimal points can be achieved *simultaneously* by a single configuration.

*The model chain.* Split the octagon, not into eight thin triangles from the centre, but into the rectangle $P_2P_4P_6P_8$ (area 4, fixed) plus the four "corner" triangles $P_2P_3P_4$, $P_4P_5P_6$, $P_6P_7P_8$, $P_8P_1P_2$ cut off outside the rectangle's sides. Because both the square and the rectangle are inscribed in the same circle, both are centred at the circle's centre $O$, so the whole configuration has point symmetry about $O$: rotating by $180^\circ$ sends $P_1 \leftrightarrow P_5$, $P_2 \leftrightarrow P_6$, $P_3 \leftrightarrow P_7$, $P_4 \leftrightarrow P_8$. This sends triangle $P_2P_3P_4$ onto $P_6P_7P_8$ and $P_4P_5P_6$ onto $P_8P_1P_2$, so those pairs are congruent, and

$$[\text{octagon}] = [P_2P_4P_6P_8] + 2[P_2P_3P_4] + 2[P_4P_5P_6],$$

where $[\cdot]$ denotes area. Only two independent quantities remain to maximise: $[P_2P_3P_4]$, a triangle on the fixed base $P_2P_4$ (one side of the rectangle, length $2\sqrt{2}$) with apex $P_3$ free to move on the arc between $P_2$ and $P_4$; and $[P_4P_5P_6]$, a triangle on base $P_4P_6$ (length $\sqrt{2}$) with apex $P_5$ free on the arc between $P_4$ and $P_6$. For a fixed chord, the triangle's area is maximised when the third point sits at the arc's midpoint, since that is where its distance from the chord (the height) is greatest. With chord length $s$ subtending a chord-to-centre distance $d = \sqrt{r^2 - (s/2)^2}$, the maximum height above the chord (on the far arc) is $r - d$, so the maximum triangle area is $\frac{1}{2}s(r-d)$.

For $P_2P_4$: $s = 2\sqrt{2}$, $(s/2)^2 = 2$, $d = \sqrt{r^2 - 2} = \sqrt{5/2 - 2} = \sqrt{1/2}$, giving

$$[P_2P_3P_4]_{\max} = \tfrac{1}{2} \cdot 2\sqrt{2}\left(\sqrt{5/2} - \sqrt{1/2}\right) = \sqrt{2}\left(\sqrt{5/2} - \sqrt{1/2}\right) = \sqrt{5} - 1.$$

For $P_4P_6$: $s = \sqrt{2}$, $(s/2)^2 = 1/2$, $d = \sqrt{5/2 - 1/2} = \sqrt{2}$, giving

$$[P_4P_5P_6]_{\max} = \tfrac{1}{2}\sqrt{2}\left(\sqrt{5/2} - \sqrt{2}\right) = \frac{\sqrt{5}}{2} - 1.$$

So the octagon's area is at most $4 + 2(\sqrt{5} - 1) + 2\left(\frac{\sqrt{5}}{2} - 1\right) = 4 - 2 - 2 + 2\sqrt{5} + \sqrt{5} = 3\sqrt{5}$, *provided* both maximising positions for $P_3$ and $P_5$ can be attained at once by a single valid square $P_1P_3P_5P_7$ — which needs an explicit check, not an assumption, since $P_3$ and $P_5$ are two vertices of the *same rigid square* and cannot be moved independently.

**Caveat.** The official solution states this simultaneous attainability without proof. I verified it directly by placing coordinates: circle centred at $O$, rectangle $P_2 = (\sqrt{2}/2, -\sqrt{2})$, $P_4 = (\sqrt{2}/2, \sqrt{2})$, $P_6 = (-\sqrt{2}/2, \sqrt{2})$, $P_8 = (-\sqrt{2}/2, -\sqrt{2})$ (so $P_2P_4 = 2\sqrt{2}$, $P_4P_6 = \sqrt{2}$, as required, all on a circle of radius $r = \sqrt{5/2}$), giving angular positions $P_2 \approx -63.43°$, $P_4 \approx 63.43°$, $P_6 \approx 116.57°$, $P_8 \approx 243.43°$. The midpoint of arc $P_2P_4$ (the short way, through angle 0°) is exactly 0°; the midpoint of arc $P_4P_6$ is exactly 90° — and these two required positions for $P_3$ and $P_5$ are exactly 90° apart, matching the square's own vertex spacing. Placing the square at $P_1 = 270°, P_3 = 0°, P_5 = 90°, P_7 = 180°$ satisfies the ordering constraint $P_1, \ldots, P_8$ around the circle and realises both maxima simultaneously. So the bound $3\sqrt{5}$ is genuinely attained, not merely an upper bound.

### Layer 2 THE HOW — the bifurcation

**Bifurcating point.** Both chains agree, and must agree, on the forced first step: extracting $r^2 = 5/2$ and the rectangle's sides $\sqrt{2}, 2\sqrt{2}$ from the two given areas. The chains split at the decomposition of the octagon's area. The natural attempt decomposes into eight triangles from the centre, treating all eight vertices' angular positions as the variables to be jointly optimised. The model chain decomposes into the fixed rectangle plus two independently-varying corner triangles, using the point symmetry of the configuration to halve the work and to isolate exactly two free real parameters (the arc-positions of $P_3$ and $P_5$) out of what looked like eight.

> *Eight points on a circle sounds like eight degrees of freedom, but four of them are rigidly the square and four are rigidly the rectangle — and because both shapes share the same centre as the circle, rotating the whole picture by* 180° *must send the configuration to itself in a specific way. If I decompose the octagon's area by that symmetry instead of by uniform central angles, I should end up with far fewer independent quantities to maximise than eight.*

### Layer 3 THE WHY

Why does "both inscribed in the same circle" force the shared centre, and why does that matter? A square's diagonals bisect each other at its own centre, and since the square is inscribed in the circle with its diagonal equal to the circle's diameter, that centre coincides with the circle's centre $O$ — the same is true of the rectangle, whose diagonal is also a diameter. Two shapes centred at the same point, one with 4-fold symmetry and one with 2-fold symmetry, combine to give the whole eight-point configuration exactly 2-fold (point) symmetry about $O$: this is not an assumption to be checked case by case, it follows immediately from both diagonals being diameters of the same circle. Once that symmetry is recognised, splitting the octagon's boundary into a rectangle plus four corner triangles that pair up under the 180° rotation is the natural decomposition, because it is the decomposition that respects the symmetry the problem already has, rather than one (uniform central angles) that ignores it.

The deeper reason the "maximise a triangle on a fixed chord by moving the apex to the arc midpoint" fact is the right tool here is that area is linear in height for a fixed base, and the height of a point above a chord, as that point moves along the far arc, is a strictly concave function of arc-position with its unique maximum exactly at the midpoint (by the reflective symmetry of a circle about the perpendicular bisector of the chord) — this is the same "distance from a fixed line, on a circle" extremal principle that underlies many inscribed-polygon maximisation problems, and recognising the corner triangles as instances of it is what converts the problem from "optimise eight coupled angles" into "apply a one-variable fact twice." Finally, the archive caveat above is not a formality: the two triangles $P_2P_3P_4$ and $P_4P_5P_6$ share the vertex $P_4$ and are cut from vertices of the *same rigid square*, so their apexes $P_3, P_5$ cannot be chosen independently in general — the fact that the two required optimal positions happen to be exactly 90° apart, matching the square's own spacing, is what makes the bound achievable, and this is a genuine

feature of the specific numbers 5 and 4 in the problem, not a general phenomenon that would survive changing the areas.

**Layer $+\alpha$ BACKTESTING**

**1.** The first useful cue is two inscribed polygons on the same circle, given by area, with the actual shape of an outer polygon depending on their relative orientation. On seeing it, write down: both diagonals are diameters of the same circle, so both polygons share the circle's centre — this is a symmetry fact, not just a coincidence of the picture.

**2.** That cue suggests decomposing the compound polygon's area along the symmetry (here, 180° point symmetry) rather than by uniform central angles, in order to reduce the number of genuinely independent optimisation variables before any calculus or Lagrange-multiplier machinery is reached for.

**3.** Hesitation occurs right after the two independent triangle maxima are found on paper, where it is tempting to add them up and declare the answer without checking that both maximising configurations for $P_3$ and $P_5$ can hold *simultaneously*, since they are vertices of the same rigid square and cannot be moved one at a time. What reinforces the correct move is remembering that a bound obtained by maximising two quantities separately is only an upper bound until a single configuration achieving both is exhibited.

**4.** The trap is treating $P_3$ and $P_5$ as free independent variables because they appear in different triangles, when in fact they are two vertices of one rigid square and their positions are linked by a fixed 90° rotation; failing to verify simultaneous attainability turns a correct number into an unjustified one.

**5.** Recognise it faster by treating any "polygon built from two inscribed shapes sharing a circle" problem as an instruction to look first for the symmetry forced by both diagonals being diameters, and to decompose the area along that symmetry rather than vertex by vertex.

**6.** Reusable rule: when a bound is obtained by maximising several sub-quantities separately, but those sub-quantities are controlled by a shared rigid object (here, one square supplying two of the free points), the bound is only a candidate maximum until an explicit single configuration is exhibited achieving all the separate maxima at once — check this with coordinates or angles before declaring the bound the answer.

▷ **Cue:** two polygons are separately inscribed in the same circle $\longrightarrow$ **Move:** note both diagonals are diameters of that circle, so both polygons share the circle's centre, giving the combined figure a forced rotational symmetry to decompose along

▷ **Cue:** a triangle has one fixed chord as base and its third vertex free to move on the circle's far arc $\longrightarrow$ **Move:** maximum area occurs with the third vertex at the arc's midpoint; compute via height $= r - \sqrt{r^2 - (s/2)^2}$ for chord length $s$ and radius $r$

▷ **Cue:** a bound is obtained by maximising several sub-quantities that are secretly linked through one shared rigid object $\longrightarrow$ **Move:** exhibit one explicit configuration (coordinates or angles) achieving all the separate maxima simultaneously before declaring the summed bound to be the actual answer

## 2000 A4

**2000 A4.** Show that the improper integral

$$\lim_{B\to\infty} \int_0^B \sin(x)\sin(x^2)\,dx$$

converges.

**Fields:** real analysis (improper integrals, oscillatory integrals, integration by parts) **Position:** A4 — middle of the paper. Routine for anyone who has met oscillatory integrals; a long grind for anyone who

has not, because the obvious substitution leaves a factor that refuses to decay. **Answer:** The limit exists. (Numerically it is about 0.4917; the problem asks only for existence.)

### Layer 1 THE WHAT — surface analysis

*The natural attempt.* You know the model problem: $\int_0^\infty \sin(x^2)\,dx$ converges because of cancellation, not because the integrand is small, and the standard proof substitutes $u = x^2$ to get $\int \frac{\sin u}{2\sqrt{u}}\,du$, a decaying amplitude times a pure oscillation. So you do the same here. With $u = x^2$, $dx = du/(2\sqrt{u})$,

$$\int_0^B \sin x\,\sin x^2\,dx = \int_0^{B^2} \frac{\sin\sqrt{u}}{2\sqrt{u}}\,\sin u\,du.$$

Now you want the Dirichlet test: $\sin u$ has bounded antiderivative, so you need the amplitude $g(u) = \frac{\sin\sqrt{u}}{2\sqrt{u}}$ to be monotone and to tend to 0. It tends to 0. It is not monotone — $\sin\sqrt{u}$ keeps changing sign. So you integrate by parts by hand, with $v = -\cos u$:

$$\int g(u)\sin u\,du = \big[-g(u)\cos u\big] + \int g'(u)\cos u\,du, \qquad g'(u) = \frac{\cos\sqrt{u}}{4u} - \frac{\sin\sqrt{u}}{4u^{3/2}}.$$

The boundary term dies. The second piece of $g'$ is $O(u^{-3/2})$, absolutely integrable. The first is $\frac{\cos\sqrt{u}}{4u}$, and $\int^\infty \frac{du}{u}$ diverges. You are left with $\int \frac{\cos\sqrt{u}\cos u}{4u}\,du$ — the same kind of object you started with, one power of $u$ better. A second integration by parts does finish it (the next derivative has a term $\frac{\sin\sqrt{u}}{8u^{3/2}}$, which is fine), and this is exactly the official first solution. It works. It costs two rounds of product rule, sign bookkeeping, and the nagging worry that a third round will be needed.

*The model chain.* *Step 1: separate the frequencies.* By the product-to-sum identity $\sin\alpha\sin\beta = \frac{1}{2}[\cos(\alpha-\beta) - \cos(\alpha+\beta)]$,

$$\sin x\,\sin x^2 = \tfrac{1}{2}\cos(x^2 - x) - \tfrac{1}{2}\cos(x^2 + x).$$

The integrand is continuous on $[0,1]$, so it suffices to show that $\lim_{B\to\infty}\int_1^B \cos\varphi(x)\,dx$ exists for $\varphi(x) = x^2 - x$ and for $\varphi(x) = x^2 + x$.

*Step 2: one integration by parts.* For either choice, on $[1,\infty)$ we have $\varphi'(x) = 2x \mp 1 \geq 1 > 0$ and $\varphi''(x) = 2$. Write $\cos\varphi = \frac{1}{\varphi'}\cdot\varphi'\cos\varphi$ and integrate by parts with $u = 1/\varphi'$, $dv = \varphi'\cos\varphi\,dx$, $v = \sin\varphi$:

$$\int_1^B \cos\varphi(x)\,dx = \left[\frac{\sin\varphi(x)}{\varphi'(x)}\right]_1^B + \int_1^B \frac{\varphi''(x)}{\varphi'(x)^2}\,\sin\varphi(x)\,dx.$$

(The sign is $+$ because $\frac{d}{dx}\frac{1}{\varphi'} = -\frac{\varphi''}{\varphi'^2}$.)

*Step 3: each piece converges.* The boundary term at $B$ is at most $\frac{1}{2B-1}$ in absolute value, so it tends to 0; the term at 1 is a constant. The integrand of the remaining integral is bounded by $\frac{2}{(2x\mp1)^2} \leq \frac{2}{(2x-1)^2}$, and $\int_1^\infty \frac{2\,dx}{(2x-1)^2} = \left[-\frac{1}{2x-1}\right]_1^\infty = 1 < \infty$. An absolutely convergent improper integral converges, so $\lim_{B\to\infty}\int_1^B \frac{\varphi''}{\varphi'^2}\sin\varphi\,dx$ exists.

Hence both limits exist, and so does $\lim_{B\to\infty}\int_0^B \sin x\sin x^2\,dx$. □

### Layer 2 THE HOW — the bifurcation

**Bifurcating point.** Both chains start from the same diagnosis: this converges by cancellation, and cancellation is harvested by integration by parts against the oscillating factor. They split at the very first line, (†): *what do I treat as the oscillation?* The natural attempt treats $\sin x^2$ as the oscillation and $\sin x$ as part of the amplitude. The model refuses to split the product at all and rewrites it as two pure oscillations with phases $x^2 \pm x$.

> *Integration by parts only gains if the amplitude I differentiate gets smaller. Here the amplitude contains* $\sin x$*, and differentiating* $\sin x$ *gives* $\cos x$ *— no smaller. So I will pay the same price every round. I want an integrand with no slowly-oscillating amplitude at all: combine the two sines into single cosines, and then the only thing left to differentiate is* $1/\varphi'$*, which decays.*

> *After product-to-sum I have* $\cos(x^2 - x)$*, and its phase derivative* $2x - 1$ *vanishes at* $x = \frac{1}{2}$*. Dividing by it there is illegal, so I cut the integral at* $x = 1$ *and let continuity handle* $[0, 1]$*.*

**Layer 3 THE WHY**

Go to the mechanism of integration by parts on an oscillatory integral. If $\varphi'$ is non-zero, $\int a(x)\cos\varphi(x)\,dx = \int \frac{a}{\varphi'}\,d(\sin\varphi)$, and one integration by parts trades it for $\int \left(\frac{a}{\varphi'}\right)'\sin\varphi\,dx$. The whole gain is in the word *derivative*: you win exactly when $\left(\frac{a}{\varphi'}\right)'$ is smaller than $a$ by an integrable margin. For $a = 1$ and $\varphi = x^2 \pm x$, the new amplitude is $\frac{2}{(2x\pm1)^2}$ — decisively integrable, done in one step.

Now look at what the natural attempt fed into that machine. With the phase $x^2$ and amplitude $a(x) = \sin x$, the new amplitude is $\left(\frac{\sin x}{2x}\right)' = \frac{\cos x}{2x} - \frac{\sin x}{2x^2}$. The second term gained a power; the first did not gain enough, because the derivative of $\sin x$ is as large as $\sin x$. *An amplitude that itself oscillates is not an amplitude.* It is a second frequency hiding in the wrong slot. That is the structural fact: $\sin x \sin x^2$ is not "a slow envelope times a fast wave", it is the superposition of two fast waves whose frequencies ($2x - 1$ and $2x + 1$, the phase derivatives) are almost equal. Product-to-sum is not a trick you remember; it is the move that puts each frequency in its own term, so that every term has the shape the integration-by-parts machine is built for: constant amplitude, phase with derivative growing to infinity.

That also tells you exactly what the problem is testing, and why it is an A4 and not an A1. The convergence criterion hiding here is:

$$\int^{\infty} \cos\varphi(x)\,dx \text{ converges whenever } \varphi' \text{ is monotone and } \varphi' \to \infty,$$

because then $1/\varphi'$ is monotone and tends to 0, so the boundary term dies and $\int \left|(1/\varphi')'\right| = 1/\varphi'(1) - \lim 1/\varphi'$ is finite. (Monotonicity of $1/\varphi'$ is what makes the total variation telescope; here $\varphi'' = 2 > 0$ gives it for free.) The phase $x^2 \pm x$ qualifies; the phase $x$ alone does not ($\int \cos x$ diverges), which is why one cannot treat $\sin x$ as "oscillation" either.

Kedlaya's solution gives the double integration by parts first and the product-to-sum route as "an alternate approach". The order is historical, not a ranking. The first route is correct and fully rigorous — every step checks — but it is the zombie route: it starts computing before asking what shape the integrand is. One small note on the printed alternate: the display there reads $-\left[\frac{\sin(x^2+x)}{2x+1}\right] - \int \frac{2\sin(x^2+x)}{(2x+1)^2}$; by the sign computation in Step 2 above both signs should be $+$. Convergence is unaffected, since only absolute values are used, but copy the signs from Step 2, not from the printed line.

Finally, why the left endpoint needs any attention at all. The official solution replaces 0 by $\epsilon$ "to avoid improper integrals at 0" — that is an artefact of dividing by $2x$ in its first route. In the product-to-sum route the only singular point is where $\varphi' = 0$, namely $x = \frac{1}{2}$ for $x^2 - x$; the integrand itself is continuous and bounded on $[0, 1]$. Cutting at 1 is the cheapest cure.

**Caveat.** The value $\approx 0.4917$ was checked numerically: the closed form $\frac{1}{2}\left(\int_0^{\infty}\cos(x^2 - x)\,dx - \int_0^{\infty}\cos(x^2 + x)\,dx\right)$, evaluated by completing the square and using Fresnel integrals, gives $0.491700\ldots$, and direct quadrature of $\int_0^B$ for $B = 10, 20, 40, 80$ gives $0.516, 0.504, 0.497, 0.486$, oscillating about it with amplitude of order $1/(2B)$, as the boundary term predicts. The

value is not required by the problem.

**Layer $+\alpha$ BACKTESTING**

**1.** The first useful cue is a product of two oscillating factors inside an integral to infinity, with the problem saying only "converges" (not "absolutely"). Write down at once: $\sin x \sin x^2 = \frac{1}{2}[\cos(x^2 - x) - \cos(x^2 + x)]$.

**2.** Integration by parts against the phase: $\cos\varphi = \frac{1}{\varphi'}(\sin\varphi)'$, leaving an absolutely convergent remainder $\int \frac{\varphi''}{\varphi'^2}\sin\varphi$.

**3.** Hesitation comes after the first integration by parts in the substitution route, when $\int \frac{\cos\sqrt{u}\cos u}{u}\,du$ appears and it is unclear whether the process terminates. Reinforce it away by checking, before integrating, whether the amplitude you will differentiate is monotone; if it oscillates, recombine first.

**4.** Two traps. First, comparing $|\sin x \sin x^2|$ with something integrable: the integral does not converge absolutely, so no comparison test can work, and time spent there is lost. Second, dividing by $\varphi' = 2x - 1$ across $x = \frac{1}{2}$.

**5.** Slogan: *an amplitude that oscillates is a hidden frequency — move it into the phase.* And: *convergence without absolute convergence means integration by parts; aim to leave exactly one oscillating factor with nothing to differentiate but* $1/\varphi'$.

**6.** Decision rule: **before integrating by parts on an oscillatory integral, rewrite the integrand as a sum of terms $a(x)\cos\varphi(x)$ with $a$ monotone (ideally constant) and $\varphi'$ monotone and unbounded; one integration by parts then finishes.**

▷ **Cue:** an integral to infinity of a product of two trigonometric factors, e.g. $\sin x \sin x^2$, asked only to converge $\longrightarrow$ **Move:** product-to-sum into cos(phase) terms, then integrate each by parts against $\varphi'$

▷ **Cue:** $\int^{\infty}\cos\varphi(x)\,dx$ with $\varphi'$ monotone and $\to\infty$ $\longrightarrow$ **Move:** write $\cos\varphi = \frac{1}{\varphi'}(\sin\varphi)'$; boundary term $\to 0$, remainder absolutely convergent

▷ **Cue:** after one integration by parts the remainder is "as bad as before" (an $O(1/x)$ amplitude times an oscillation) $\longrightarrow$ **Move:** the amplitude you differentiated was itself oscillating; recombine frequencies instead of integrating by parts again

▷ **Cue:** dividing by $\varphi'$ in an integration by parts $\longrightarrow$ **Move:** locate the zeros of $\varphi'$ and cut the integral away from them; the finite piece is continuous

## 2000 A5

**2000 A5.** Three distinct points with integer coordinates lie in the plane on a circle of radius $r > 0$. Show that two of these points are separated by a distance of at least $r^{1/3}$.

**Fields:** geometry (triangle area, circumradius); lattice points **Position:** A5 — the hard end of the paper by position, though the proof is four lines once the right formula is on the page. The difficulty is entirely in the choice of what to compute. **Answer:** With side lengths $a, b, c$ of the triangle of the three points, $abc = 4r \cdot \text{Area} \geq 2r$, so the longest side is at least $(2r)^{1/3} > r^{1/3}$.

**Layer 1 THE WHAT — surface analysis**

*The natural attempt.* Integer coordinates and a circle: the reflex is number theory. You write the circle as $(x-p)^2 + (y-q)^2 = r^2$ and hope the centre $(p, q)$ is a lattice point, so that $r^2$ is a sum of two squares and the three points are solutions of $u^2 + v^2 = r^2$. Then you look for a reason why two solutions of $u^2 + v^2 = N$ cannot be close together. But the problem never said the centre is a lattice point, and it need not be: $(0,0), (1,0), (0,1)$ lie on the circle with centre $(\frac{1}{2}, \frac{1}{2})$ and $r^2 = \frac{1}{2}$. You can prove the centre is rational (it solves two linear equations with integer

coefficients, the perpendicular bisectors), and so $r^2$ is rational, and at this point you are clearing denominators and tracking gcds with no idea what inequality you are aiming at. That is the wall: the arithmetic is true and inert.

A stronger student abandons the arithmetic and thinks about shape. Suppose all three distances are less than some $d$ that is small compared with $r$. Then the three points sit on a short arc, and a short arc of a large circle is nearly straight: the arc bulges away from its chord by only about $d^2/(8r)$ (the sagitta). So the triangle is a sliver, with base at most $d$ and height about $d^2/r$, and area of order $d^3/r$. That is the right scale — $d^3 \gtrsim r$ is exactly $d \gtrsim r^{1/3}$ — but you now need a lower bound on the area of a sliver with integer vertices, and to turn every "about" into an inequality with constants, including the case where the arc is not short at all.

*The model chain.* Let the three points be $P, Q, R$, and let $a = |QR|$, $b = |RP|$, $c = |PQ|$. Let $K$ be the area of triangle $PQR$.

*Step 1: the triangle is non-degenerate.* A line meets a circle in at most two points, so $P, Q, R$ are not collinear and $K > 0$.

*Step 2: $K \geq \frac{1}{2}$.* With $P = (x_1, y_1)$, $Q = (x_2, y_2)$, $R = (x_3, y_3)$,

$$K = \tfrac{1}{2}\big|(x_2 - x_1)(y_3 - y_1) - (x_3 - x_1)(y_2 - y_1)\big|.$$

The quantity inside the absolute value is an integer, and it is non-zero by Step 1, so it is at least 1 in absolute value. Hence $K \geq \frac{1}{2}$.

*Step 3: $K = \dfrac{abc}{4r}$.* The circle through $P, Q, R$ is the circumcircle of the triangle, so its radius $r$ is the circumradius. Let $C$ be the angle of the triangle at $R$. By the extended sine rule, $c = 2r \sin C$. (Proof: the diameter through $P$ meets the circle again at $P'$; the inscribed angle $\angle PP'Q$ equals $C$ or $\pi - C$ according to which arc $P'$ lies on, and triangle $PP'Q$ has a right angle at $Q$, so $c = |PQ| = 2r \sin C$.) Also $K = \frac{1}{2}ab \sin C$. Eliminating $\sin C$,

$$K = \tfrac{1}{2}ab \cdot \frac{c}{2r} = \frac{abc}{4r}.$$

*Step 4: conclude.* From Steps 2 and 3, $abc \geq 2r$. If $M = \max\{a, b, c\}$ then $M^3 \geq abc \geq 2r$, so

$$M \geq (2r)^{1/3} > r^{1/3}. \qquad \square$$

### Layer 2 THE HOW — the bifurcation

**Bifurcating point.** Both chains know the three points are on a circle and have integer coordinates. They split at the first line written, (†): the natural attempt writes down *the equation of the circle*; the model writes down *a quantity that involves both r and the side lengths and is controlled by the lattice*. The second, geometric version of the natural attempt actually finds that quantity (the area) and the right scaling; it stalls only because it estimates the area from the shape of an arc instead of using the exact formula that already relates area, sides and circumradius.

> *I need an inequality with the three distances on one side and r on the other. What single formula contains the sides of a triangle and the radius of its circumcircle? $abc = 4rK$. And what does integrality of the coordinates buy me about a triangle? Its area is a positive multiple of $\frac{1}{2}$. So $abc \geq 2r$ — and a cube root appears, which is exactly the exponent in the question.*

> *The exponent $\frac{1}{3}$ is a cube root. Three distances, one cube root: the product abc is the thing to bound, and the maximum of three numbers is at least the cube root of their product.*

### Layer 3 THE WHY

Why is the area formula the object of interest, not a trick? Read the problem's two hypotheses and ask what each one controls.

"Integer coordinates" controls very little by itself. Distances between lattice points are square roots of integers, which gives $|PQ| \geq 1$ and nothing better. The one robust lower bound that integrality gives in the plane is on *area*: the shoelace determinant is an integer, so any non-degenerate lattice triangle has area at least $\frac{1}{2}$. That is the lattice's only lever here, and a student should know it the way they know $|n| \geq 1$ for a non-zero integer $n$ — it *is* $|n| \geq 1$, applied to a determinant.

"On a circle of radius $r$" is a statement about the circumradius of the triangle $PQR$. What connects the circumradius to anything else about a triangle? The extended sine rule, $a/\sin A = 2r$. And the moment you have $\sin C$ from the sine rule, you have the area, because $K = \frac{1}{2}ab\sin C$. So the formula $K = abc/(4r)$ is not a memorised curiosity; it is what you get by writing down the only two formulas that contain $\sin C$ and eliminating it.

Put the two levers side by side: one hypothesis bounds $K$ below, the other expresses $K$ exactly in terms of the sides and $r$. That is a complete proof skeleton before any computation.

And the target tells you the same thing from the other end. The answer is $r^{1/3}$. Where does a cube root come from in plane geometry? From a product of three lengths set against one length: $abc \sim r$. If you see a cube root and three objects, bound their product. The geometric version of the natural attempt found this scaling honestly ($d \cdot d^2/r$) — it is the same fact, since a thin triangle inscribed in a large circle has area roughly base times sagitta. The exact identity $abc = 4rK$ is the rigorous version of that estimate, valid for every triangle, short arc or not, so no case split is needed.

The number-theoretic reflex fails for a structural reason worth naming. The equation of the circle uses the centre, and the centre is not a lattice point; the hypothesis "integer coordinates" applies to the three points only. So any argument should be written in terms of the three points and quantities intrinsic to them — their pairwise distances and the area they span — and $r$ should enter only through a formula that is also intrinsic to the triangle. The circumradius formula is exactly that.

Kedlaya's solution is three lines and cites Pick's theorem for $K \geq \frac{1}{2}$. Pick is overkill: the shoelace formula does it, and it is the version to write, because it makes visible that the bound is "a non-zero integer is at least 1". Note also that the proof gives $(2r)^{1/3}$, strictly more than asked. Slack in the stated constant is common on the Putnam; do not read it as a sign that your stronger bound must be wrong.

**Caveat.** Checked numerically: for 200,000 random non-collinear triples of lattice points in $[-6,6]^2$, the ratio $\max\{a,b,c\}/(2r)^{1/3}$ was always at least $1.2599\ldots = 2^{1/3}$, with the minimum attained by, for example, $(0,0),(1,0),(0,1)$, where $\max\{a,b,c\}^3 = (\sqrt{2})^3 = 2\sqrt{2} = 4r$. So on that sample the longest side is in fact at least $(4r)^{1/3}$; the proof above gives only $(2r)^{1/3}$, because $\max \geq (abc)^{1/3}$ is tight only for equilateral triangles, which do not occur on the lattice. The sharper constant is an observation on a finite sample, not a verified theorem. Whether the exponent $\frac{1}{3}$ is optimal for circles of large radius is a separate question (it concerns lattice points on short arcs) and is not verified here.

### Layer $+\alpha$ BACKTESTING

**1.** The first useful cue is the combination "integer coordinates" + "circle of radius $r$" + an exponent $\frac{1}{3}$ in the answer. On seeing it, write $K = \dfrac{abc}{4r}$ and $K \geq \frac{1}{2}$.

**2.** Squeeze a single quantity (the area) between a lattice lower bound and an exact geometric formula; then use max $\geq$ geometric mean.

**3.** Hesitation occurs when the circle's centre turns out not to be a lattice point, and the student wonders whether integrality is usable at all. Reinforce it away with the fact that integrality always gives area $\geq \frac{1}{2}$ for a non-degenerate lattice triangle (or volume $\geq \frac{1}{n!}$ for a lattice simplex in $n$ dimensions).

**4.** The trap is the equation $(x-p)^2+(y-q)^2=r^2$: it invites number theory about sums of two squares, which assumes a lattice centre that the problem never grants.

**5.** Slogan: *a fractional exponent $1/k$ in the answer means "bound a product of $k$ things".* And: *the lattice's cheapest lower bound in the plane is on area, not on distance.*

**6.** Decision rule: **when a problem mixes a discrete hypothesis with a continuous quantity, find one quantity that the discrete hypothesis bounds below and that has an exact formula in terms of the continuous data; set the two against each other.**

▷ **Cue:** three lattice points / a lattice triangle and something to be shown "large" $\longrightarrow$ **Move:** shoelace: the area is a positive multiple of $\frac{1}{2}$

▷ **Cue:** a triangle's side lengths and the radius of the circle through its vertices appear together $\longrightarrow$ **Move:** $abc = 4rK$ (from $c = 2r\sin C$ and $K = \frac{1}{2}ab\sin C$)

▷ **Cue:** the answer has an exponent $1/3$ (or $1/k$) $\longrightarrow$ **Move:** bound a product of three ($k$) quantities and use max $\geq$ geometric mean

▷ **Cue:** "points with integer coordinates on a circle" with no condition on the centre $\longrightarrow$ **Move:** do not write the circle's equation; work with quantities intrinsic to the points (distances, area)

## 2000 A6

**2000 A6.** Let $f(x)$ be a polynomial with integer coefficients. Define a sequence $a_0, a_1, \ldots$ of integers such that $a_0 = 0$ and $a_{n+1} = f(a_n)$ for all $n \geq 0$. Prove that if there exists a positive integer $m$ for which $a_m = 0$ then either $a_1 = 0$ or $a_2 = 0$.

**Fields:** number theory (divisibility, integer polynomials); dynamics (periodic orbits) **Position:** A6 — the last problem of the morning. Hard because nothing in the statement tells you which divisibility to use; once the right one is chosen the proof is a short chain of forced steps. **Answer:** Consecutive differences $b_n = a_{n+1} - a_n$ satisfy $b_n \mid b_{n+1}$; on a cycle this forces $|b_n|$ constant, the walk must reverse somewhere, and a reversal forces period 1 or 2. (More generally: every periodic orbit of an integer polynomial acting on $\mathbb{Z}$ has period 1 or 2.)

### Layer 1 THE WHAT — surface analysis

*The natural attempt.* The one fact everyone knows about integer polynomials is

$$x - y \text{ divides } f(x) - f(y) \quad \text{for all integers } x, y,$$

because $x^k - y^k = (x-y)(x^{k-1} + \cdots + y^{k-1})$. So you apply it, and the natural pair to use is the one involving the special value $a_0 = 0$: taking $y = 0$,

$$a_n \mid f(a_n) - f(0) = a_{n+1} - a_1 \quad \text{for every } n.$$

Now feed in the hypothesis. With $n = m-1$: $a_{m-1} \mid a_m - a_1 = -a_1$. With $n = 1$: $a_1 \mid a_2 - a_1$, so $a_1 \mid a_2$. Iterating, $a_1 \mid a_n$ for all $n$. These are all true, and you can generate many more, but they relate terms at *different offsets* ($a_n$ against $a_{n+1} - a_1$), and nothing closes up. You try small cases: a 3-cycle $0 \to u \to v \to 0$ gives $u \mid v - u$, $v \mid -u$, $u \mid v$ from the $y = 0$ pairs, and with the extra pair $(u, v)$, $u - v \mid v - 0$ — you can kill $m = 3$ by hand. But $m = 4$ has more pairs, and you have no organising principle that works for every $m$. That is the wall.

*The model chain.* *Setup.* Put $b_n = a_{n+1} - a_n$ for $n \geq 0$.

*Step 1: $b_n \mid b_{n+1}$.* Since $x - y \mid f(x) - f(y)$ for integers $x, y$, taking $x = a_{n+1}$, $y = a_n$ gives $a_{n+1} - a_n \mid f(a_{n+1}) - f(a_n) = a_{n+2} - a_{n+1}$. (Here $0 \mid t$ means $t = 0$.)

*Step 2: the sequence is periodic.* Since $a_m = a_0$ and each term is determined by the previous one, induction gives $a_{n+m} = a_n$ for all $n \geq 0$. In particular $b_m = b_0$.

*Step 3: either $a_1 = 0$, or all $|b_n|$ are equal and non-zero.* If $b_0 = 0$ then $a_1 = a_0 = 0$ and we are done. Suppose $b_0 \neq 0$. If some $b_i = 0$ with $0 \leq i < m$, then Step 1 gives $b_{i+1} = 0$, then $b_{i+2} = 0$, and so on up to $b_m = 0$, contradicting $b_m = b_0 \neq 0$. So $b_0, \ldots, b_m$ are all non-zero, and divisibility between non-zero integers gives $|b_0| \leq |b_1| \leq \cdots \leq |b_m| = |b_0|$. Hence $|b_n| = d := |b_0| > 0$ for $0 \leq n \leq m$, so each $b_n \in \{d, -d\}$.

*Step 4: a reversal.* Now $b_0 + b_1 + \cdots + b_{m-1} = a_m - a_0 = 0$. A sum of terms each equal to $\pm d$ with $d > 0$ can vanish only if both signs occur. Hence there is an index $k$ with $1 \leq k \leq m - 1$ and $b_{k-1} = -b_k$ (between a $+d$ and a $-d$ in the list, some adjacent pair differs). That says $a_k - a_{k-1} = -(a_{k+1} - a_k)$, that is, $a_{k+1} = a_{k-1}$.

*Step 5: period two from the reversal.* If $a_{j+2} = a_j$ then applying $f$ gives $a_{j+3} = a_{j+1}$; so by induction $a_{n+2} = a_n$ for all $n \geq k - 1$. Since $k - 1 \leq m - 2 < m$, this applies with $n = m$: $a_{m+2} = a_m$. By Step 2, $a_{m+2} = a_2$ and $a_m = a_0 = 0$. Therefore $a_2 = 0$. □

**Layer 2 THE HOW — the bifurcation**

**Bifurcating point.** Both chains reach for $x - y \mid f(x) - f(y)$ at once. They split at the choice of pair, (†). The natural attempt plugs in $y = a_0 = 0$, because 0 is the special value in the statement. The model plugs in *consecutive terms* $x = a_{n+1}$, $y = a_n$, because that pair is carried by $f$ to the next consecutive pair — so the divisibility becomes a chain.

> *The divisibility relates $x - y$ to $f(x) - f(y)$. Which pair $(x, y)$ does $f$ send to a pair of the same kind? Consecutive terms: $f$ sends $(a_n, a_{n+1})$ to $(a_{n+1}, a_{n+2})$. So the differences divide each other in a chain, $b_0 \mid b_1 \mid b_2 \mid \cdots$.*

> *The sequence goes round a cycle, so the chain of divisibilities comes back to where it started. A quantity that can only grow, on a cycle, cannot grow at all: every $|b_n|$ is the same.*

> *Now I have a walk on $\mathbb{Z}$ with steps of fixed length $d$ that returns to 0. It has to turn round somewhere. When it turns round, it revisits the point it was at two steps earlier — and the map is deterministic, so from then on it bounces between two values forever.*

**Layer 3 THE WHY**

Why is "consecutive differences" the natural choice rather than an inspired one? Look at what the hypothesis $a_{n+1} = f(a_n)$ actually is: a rule that moves you one step along the sequence. Any property of $f$ is useful exactly to the extent that it can be *transported along* the sequence, i.e. applied repeatedly. The divisibility $x - y \mid f(x) - f(y)$ is a property of a *pair*. For it to be applied repeatedly you need a family of pairs that $f$ maps into itself. The pairs $(a_n, 0)$ are not such a family: $f$ sends $(a_n, 0)$ to $(a_{n+1}, a_1)$, not to $(a_{n+1}, 0)$ — which is exactly why the natural attempt kept producing relations with a stray $a_1$ in them. The pairs $(a_{n+1}, a_n)$ are such a family. So the choice is forced by the definition of the sequence, the way the centre of a circle is forced by the definition of a circle.

Once you have $b_0 \mid b_1 \mid \cdots$, the second step also comes from the structure, not from cleverness. Divisibility among non-zero integers is a partial order that forces $|b_n|$ to be non-decreasing. Periodicity says the chain returns to $b_0$. A non-decreasing quantity that returns to its start is constant. This is the general principle of monovariants on cycles, and it is worth stating in that generality: *on a periodic orbit, every monotone quantity is constant*. (The same principle proves, for instance, that a strictly increasing function has no cycles other than fixed points.)

The third step — the reversal — is the only place that uses integers as numbers on a line rather than as divisibility objects. Steps of constant size $d$ summing to 0 must include both $+d$ and $-d$, and the first sign change is a point where the walk steps back to where it was. Determinism then does the rest: $a_{k+1} = a_{k-1}$ means $f(f(a_{k-1})) = a_{k-1}$, so the orbit from $a_{k-1}$ on has period dividing 2, and since the whole sequence is periodic this period governs $a_0$ as well.

Kedlaya's solution follows this route and compresses Step 5 into "for $m = n$" (a typo: the intended substitution is $n = m$). The step that deserves care is the transfer from "period 2 from index $k-1$ onward" to "$a_2 = a_0$". It uses that $k - 1 \leq m$, so that index $m$ lies in the range where period two holds, and then Step 2 to pull $a_{m+2}$ and $a_m$ back to $a_2$ and $a_0$. The official text also asserts "$|b_0| = |b_1| = \cdots$" directly from $b_0 = b_m$; the missing case is that some $b_i$ might be 0, since 0 is divisible by everything and the size comparison fails. Step 3 closes it: a zero propagates forward and would force $b_m = 0$.

Finally, notice the theorem you have proved is stronger and cleaner than the problem: *for $f \in \mathbb{Z}[x]$, every periodic point of $f$ in $\mathbb{Z}$ has period 1 or 2.* Nothing used $a_0 = 0$ except as a name for the starting point. The problem's phrasing ("$a_1 = 0$ or $a_2 = 0$") is that statement in disguise, and recognising the disguise is itself a cue: a conclusion of the form "$a_1 = a_0$ or $a_2 = a_0$" is a statement about periods.

**Caveat.** Checked by exhaustive search: for every polynomial of degree at most 3 with integer coefficients in $[-4, 4]$, iterating from 0 for up to 11 steps, the orbit of 0 never returned to 0 at a step other than 1 or 2. This is a sanity check, not a proof; the proof above is complete.

### Layer $+\alpha$ BACKTESTING

**1.** The first useful cue is "polynomial with integer coefficients" together with an iteration $a_{n+1} = f(a_n)$. On seeing it write $x - y \mid f(x) - f(y)$ and immediately $b_n = a_{n+1} - a_n$, $b_n \mid b_{n+1}$.

**2.** A divisibility chain along the orbit, closed up by periodicity into equality of absolute values; then a sign/reversal argument; then determinism.

**3.** Hesitation occurs at the choice of pair in $x - y \mid f(x) - f(y)$. Reinforce it away by asking which pairs $f$ maps to pairs of the same form; for an iteration, that is always consecutive terms (and more generally $(a_{n+j}, a_n)$ for fixed $j$).

**4.** The trap is the special value 0: it looks like the natural anchor, but the pairs $(a_n, 0)$ are not preserved by $f$ and generate relations that never close. A second trap is dividing by $b_i$ without checking it is non-zero.

**5.** Slogan: *for an iteration, apply every two-variable property of the map to consecutive terms, so that it chains.* And: *on a cycle, a monotone quantity is constant.*

**6.** Decision rule: **for $a_{n+1} = f(a_n)$ with a periodicity hypothesis, find a quantity $Q(a_n, a_{n+1})$ that $f$ makes monotone; periodicity forces it constant; then extract the rigid structure that constancy imposes.**

▷ **Cue:** integer polynomial iterated, $a_{n+1} = f(a_n)$ $\longrightarrow$ **Move:** set $b_n = a_{n+1} - a_n$; then $b_n \mid b_{n+1}$

▷ **Cue:** a sequence that returns to its start ($a_m = a_0$) plus a monotone quantity along it $\longrightarrow$ **Move:** the quantity is constant on the whole cycle

▷ **Cue:** steps of equal size $\pm d$ that sum to zero $\longrightarrow$ **Move:** there is an adjacent pair of opposite signs: a reversal, so the walk revisits a point two steps apart

▷ **Cue:** $a_{k+1} = a_{k-1}$ in a deterministic iteration $\longrightarrow$ **Move:** period 2 from index $k-1$ on; transfer back to the start by periodicity

▷ **Cue:** conclusion "$a_1 = a_0$ or $a_2 = a_0$" $\longrightarrow$ **Move:** it is a statement that every period is 1 or 2; prove it for an arbitrary periodic point

# Putnam 2000 — Paper B

## 2000 B1

**2000 B1.** Let $a_j, b_j, c_j$ be integers for $1 \le j \le N$. Assume for each $j$, at least one of $a_j, b_j, c_j$ is odd. Show that there exist integers $r$, $s$, $t$ such that $ra_j + sb_j + tc_j$ is odd for at least $4N/7$ values of $j$, $1 \le j \le N$.

**Fields:** Combinatorics, linear algebra over $\mathbb{F}_2$ (double counting, pigeonhole) **Position:** B1 – opening problem of the second session **Answer:** Such $r, s, t$ exist. Take $(r, s, t)$ ranging over the seven nonzero triples in $\{0,1\}^3$; a double-counting argument shows the total number of odd hits, summed over all seven, is exactly $4N$, so by pigeonhole some one triple achieves at least $4N/7$.

### Layer 1 THE WHAT — surface analysis

*The natural attempt.* The phrase "there exist integers $r, s, t$" reads, on first contact, as a search over all of $\mathbb{Z}^3$. The natural move is to try the cheapest candidates first. Take $r = 1, s = t = 0$: does $a_j$ alone land odd for at least $4N/7$ of the indices? You test this against an adversarial input where oddness is spread evenly – a third of the $j$ have only $a_j$ odd, a third have only $b_j$ odd, a third have only $c_j$ odd. No single coordinate then covers more than $N/3$, and $N/3 < 4N/7$, so the pure-coordinate guess fails outright. You try $r = s = t = 1$ next, looking at $a_j + b_j + c_j$, which is odd exactly when one or three of $a_j, b_j, c_j$ are odd. Another adversarial input – every $j$ has exactly two of the three odd – makes this sum even for *every* $j$: zero hits out of $N$, a complete failure. Two of the most obvious guesses are now dead, and what's left is trial and error over the handful of remaining small triples, with no principle for which to try next and no argument that any single fixed triple is safe against every possible input. For any one candidate $(r, s, t)$ you can, in principle, construct an input on which it performs badly – so a proof by exhibiting one clever triple cannot work in general. The problem needs a guarantee that holds for every input simultaneously, which means the argument has to operate on all the candidate triples at once, not hunt for the best one.

*The model chain.* Since only the parity of $ra_j + sb_j + tc_j$ matters, only the parities of $r, s, t$ matter: replace the search over $\mathbb{Z}^3$ by a search over $\mathbb{F}_2^3$, which has exactly 8 elements. For each $j$, let $v_j = (a_j, b_j, c_j) \bmod 2 \in \mathbb{F}_2^3$; the hypothesis that at least one of $a_j, b_j, c_j$ is odd says exactly $v_j \neq 0$. For a candidate $(r, s, t) \in \mathbb{F}_2^3$, write $\mathrm{hit}(r, s, t) = \#\{j : (r, s, t) \cdot v_j = 1 \text{ in } \mathbb{F}_2\}$, the count of indices this triple odd-hits. Sum $\mathrm{hit}(r, s, t)$ over the seven *nonzero* triples $(r, s, t)$. Fix $j$ and count how many of these seven contribute to the sum: as $(r, s, t)$ ranges over all eight elements of $\mathbb{F}_2^3$, the map $(r, s, t) \mapsto (r, s, t) \cdot v_j$ is a nonzero linear functional (because $v_j \neq 0$), so it takes the value 1 on exactly half its domain – 4 of the 8 triples. The triple $(0, 0, 0)$ always gives 0, so exactly 4 of the 7 nonzero triples pair to 1 with this particular $v_j$, regardless of what $v_j$ actually is. Summing over all $j = 1, \ldots, N$:

$$\sum_{(r,s,t) \neq (0,0,0)} \mathrm{hit}(r, s, t) \;=\; \sum_{j=1}^{N} 4 \;=\; 4N.$$

Seven nonnegative integers summing to $4N$ must include one that is at least the average, $4N/7$. Any integer lift of that winning $(r, s, t) \in \{0,1\}^3$ is the required triple.

**Layer 2 THE HOW — the bifurcation**

**Bifurcating point.** Both readings start from the same phrase, "there exist integers $r, s, t$." The natural attempt reads it as licence to search $\mathbb{Z}^3$ one candidate at a time. The model chain reads it as a search over the eight-element group $\mathbb{F}_2^3$ and, crucially, stops testing individual candidates: it sums the hit count of *every* nonzero candidate at once via double counting.

> $ra_j + sb_j + tc_j \bmod 2$ *depends only on* $r, s, t \bmod 2$, *so I am not really choosing three integers – I am choosing one of eight parity patterns, seven of them nonzero, and I can afford to reason about all seven simultaneously instead of guessing which one is good.*

> *Instead of asking whether one specific triple works, I should ask how many of the seven nonzero triples odd-hit a given index* $j$ *– and because* $v_j$ *is guaranteed nonzero, that count is exactly* 4 *for every single* $j$*, regardless of what* $v_j$ *happens to be. So I can sum over* $j$ *first to get* $4N$ *total hits, then divide by the seven candidates second, and pigeonhole hands me a winner without my ever testing a triple by hand.*

**Layer 3 THE WHY**

Why does reducing to $\mathbb{F}_2^3$ turn an unbounded search into a two-line argument? Because "odd" is entirely a statement about residues mod 2: the moment the problem's hypothesis is phrased as a parity condition ("at least one of $a_j, b_j, c_j$ is odd"), the only information that can ever be extracted about $r, s, t$ is their parity too, so the true search space was never $\mathbb{Z}^3$ – it was the 8-element group $\mathbb{F}_2^3$ all along, and the infinitude of $\mathbb{Z}^3$ was cosmetic.

Why does every nonzero $v_j$ pair to 1 with exactly half of $\mathbb{F}_2^3$? This is the defining property of a nonzero linear functional on a vector space over $\mathbb{F}_2$: its kernel is a subspace of index 2 (codimension 1), so it splits the ambient space exactly in half. This is not a coincidence to verify case by case for each $v_j$ – it holds for *every* nonzero $v_j$ uniformly, which is exactly why the count "4 nonzero triples odd-hit $j$" does not depend on the actual values of $a_j, b_j, c_j$, only on the hypothesis $v_j \neq 0$. That uniformity is what makes summing over $j$ before summing over triples legitimate: you are allowed to swap the order of the double sum because each individual term is controlled the same way regardless of the input.

And why does averaging over seven candidates suffice to beat $4N/7$, when no case-by-case reasoning could pin down a single winning triple in advance? Because pigeonhole needs no structural information about which candidate wins – only that the total, $4N$, is fixed and known, and a fixed total spread over seven boxes forces some box to hold at least the average. The zero triple is excluded from the seven precisely because it can never contribute a hit (its dot product with anything is 0), so its zero contribution is not allowed to drag the average down; that exclusion is what makes $4N/7$, rather than $4N/8$, the correct bound.

**Layer $+\alpha$ BACKTESTING**

**1.** The first useful cue is that "at least one of $a_j, b_j, c_j$ is odd" is a statement entirely about residues mod 2. On seeing it, write down immediately: replace $a_j, b_j, c_j$ by their parities and treat $(a_j, b_j, c_j) \bmod 2$ as a nonzero vector in $\mathbb{F}_2^3$.

**2.** That cue suggests working inside the finite group $\mathbb{F}_2^3$ rather than $\mathbb{Z}^3$, and because the goal is a guaranteed bound achieved by *some* element of a small finite set, it suggests double counting / averaging (pigeonhole) rather than a direct construction of the winning triple.

**3.** Hesitation occurs right after realising $r, s, t$ only matter mod 2: it is tempting to keep guessing among the now-finite eight candidates rather than summing over all of them. Push past this by asking, for one fixed index $j$, exactly how many of the seven nonzero candidates hit it – that count turns out to be the same number (4) for every $j$, which is the fact that licenses summing globally instead of case-checking.

**4.** The trap is assuming one of the three coordinate vectors ($r = 1, s = t = 0$ etc.) must already achieve $4N/7$ because it looks like the "biggest slice" available – an adversarial input defeats every single fixed triple, and only the average over all seven is guaranteed.

**5.** To recognise this faster: whenever a hypothesis is a parity (or more generally a mod-$p$) condition on several quantities, and the question asks for a bound achieved by *some* choice among finitely many discrete coefficient patterns, expect a double count over the whole finite pattern space, not a search among the patterns individually.

**6.** The reusable rule: if the only information a hypothesis supplies is mod-$p$ information, shrink the search space to the corresponding finite vector space over $\mathbb{F}_p$ immediately; then prove existence of a good element by summing a per-element count over the whole space and dividing by the space's size, rather than by exhibiting or checking individual elements.

## 2000 B2

**2000 B2.** Prove that the expression

$$\frac{\gcd(m,n)}{n}\binom{n}{m}$$

is an integer for all pairs of integers $n \geq m \geq 1$.

**Fields:** Number theory, combinatorics (Bézout's identity, binomial absorption identity) **Position:** B2
**Answer:** The expression is always an integer. Writing $\gcd(m,n) = um + vn$ by Bézout's identity and using $\frac{m}{n}\binom{n}{m} = \binom{n-1}{m-1}$, the expression equals $u\binom{n-1}{m-1} + v\binom{n}{m}$, an integer linear combination of integers.

### Layer 1 THE WHAT — surface analysis

*The natural attempt.* The expression has $n$ sitting in a denominator underneath a gcd, which reads as a divisibility statement: you need $n \mid \gcd(m,n)\binom{n}{m}$. The direct route is to attack this prime by prime. Expand $\binom{n}{m} = n!/(m!(n-m)!)$ and, for each prime $p$ dividing $n$, use Legendre's formula to compute the exact power of $p$ in $n!$, $m!$, and $(n-m)!$, then compare against the power of $p$ in $\gcd(m,n)$ (which is $\min(v_p(m), v_p(n))$). Kummer's theorem tells you $v_p\binom{n}{m}$ equals the number of carries when adding $m$ and $n-m$ in base $p$, so in principle you can track, for each prime $p \mid n$, whether the carry count plus $\min(v_p(m), v_p(n))$ is enough to cover $v_p(n)$. This is a real strategy and it can be pushed through, but it forces you to run a separate carry-counting argument for every prime dividing $n$ separately, and $n$ can have arbitrarily many distinct prime factors. You have traded one global divisibility claim for an unbounded number of local ones, each requiring its own base-$p$ bookkeeping, with no single argument that closes all of them at once. Worse, nothing about the carry-counting approach explains *why* the statement is true – it verifies it prime by prime without ever exposing the identity that makes the whole thing work in one line.

*The model chain.* By Bézout's identity, since $\gcd(m,n)$ is the smallest positive integer expressible as an integer linear combination of $m$ and $n$, there exist integers $u, v$ with $\gcd(m,n) = um + vn$. Substitute:

$$\frac{\gcd(m,n)}{n}\binom{n}{m} = \frac{um+vn}{n}\binom{n}{m} = u \cdot \frac{m}{n}\binom{n}{m} + v \cdot \binom{n}{m}.$$

The second term is already an integer. For the first, use the absorption identity

$$\frac{m}{n}\binom{n}{m} = \binom{n-1}{m-1},$$

which follows directly from the factorial definition: $m\binom{n}{m} = m \cdot \frac{n!}{m!(n-m)!} = \frac{n!}{(m-1)!(n-m)!} = n \cdot \frac{(n-1)!}{(m-1)!(n-m)!} = n\binom{n-1}{m-1}$, so dividing both sides by $n$ gives the claim. Substituting back,

$$\frac{\gcd(m,n)}{n}\binom{n}{m} = u\binom{n-1}{m-1} + v\binom{n}{m},$$

an integer linear combination of the two integers $\binom{n-1}{m-1}$ and $\binom{n}{m}$, hence an integer.

**Layer 2 THE HOW — the bifurcation**

**Bifurcating point.** Both readings agree that $\binom{n}{m} = \frac{n!}{m!(n-m)!}$ is the object to unpack, and both agree the claim is a divisibility-by-$n$ statement. The natural attempt keeps $\gcd(m,n)$ as a single opaque number and tries to prove $n$ divides $\gcd(m,n)\binom{n}{m}$ multiplicatively, prime by prime. The model chain instead *replaces* $\gcd(m,n)$ with its Bézout expansion $um + vn$ before doing anything else, turning one multiplicative divisibility claim into two additive pieces.

> *$\gcd(m,n)$ is not just a number I need to divide something by – it is, by definition, an integer combination $um + vn$ of $m$ and $n$ themselves. So instead of asking why $n$ divides $\gcd(m,n)\binom{n}{m}$ as one opaque product, I should substitute the combination in and see what the two resulting pieces, $\frac{um}{n}\binom{n}{m}$ and $\frac{vn}{n}\binom{n}{m}$, become separately.*

> *The awkward piece $\frac{m}{n}\binom{n}{m}$ is not a fraction I need to bound or estimate – it is literally equal to $\binom{n-1}{m-1}$, an integer I already know exists, via the absorption identity $m\binom{n}{m} = n\binom{n-1}{m-1}$. Once I see that, the whole expression collapses to an integer combination of two binomial coefficients, and I never touch a prime factorisation.*

**Layer 3 THE WHY**

Why is it legitimate, and not a trick, to replace $\gcd(m,n)$ by $um + vn$? Because that is what $\gcd(m,n)$ *is*: in the ring $\mathbb{Z}$, the set $\{um + vn : u, v \in \mathbb{Z}\}$ is an ideal, and since $\mathbb{Z}$ is a principal ideal domain that ideal is generated by a single positive integer, which is exactly $\gcd(m,n)$ by definition. Writing $\gcd(m,n) = um + vn$ is not manipulating the problem – it is unpacking what the symbol $\gcd(m,n)$ already means. Any proof that treats $\gcd(m,n)$ as an unanalysed number and tries to fight the divisibility multiplicatively is declining to use the one fact about gcd that is actually algebraic, which is why it is forced into prime-by-prime bookkeeping instead.

Why does dividing $m\binom{n}{m}$ by $n$ ever produce an integer, when division does not generally preserve integrality? Because $m\binom{n}{m} = n\binom{n-1}{m-1}$ is not a divisibility claim to be verified – it is a combinatorial identity, provable by double counting: choosing an $m$-subset of an $n$-set and then marking one of its $m$ elements can be counted two ways – $\binom{n}{m}$ ways to choose the subset times $m$ ways to mark an element within it, or $n$ ways to choose which element is marked first, times $\binom{n-1}{m-1}$ ways to fill out the remaining $m-1$ slots from the other $n-1$ elements. These count the same set of (subset, marked element) pairs, so they are equal as integers before any division happens; dividing by $n$ afterwards is legitimate precisely because the identity already exhibits the quotient, $\binom{n-1}{m-1}$, as a genuine integer, not as a formal simplification.

So the whole proof rests on stacking two facts that are each definitional or combinatorial, never analytic: $\gcd(m,n)$ *is* an integer combination of $m, n$ by the structure of $\mathbb{Z}$ as a PID, and $\frac{m}{n}\binom{n}{m}$ *is* an integer, $\binom{n-1}{m-1}$, by a counting argument that never divides at all. Integrality of the target expression is then inherited for free from closure of $\mathbb{Z}$ under addition and multiplication – nothing about primes, carries, or exponents ever needs to be examined.

**Layer $+\alpha$ BACKTESTING**

**1.** The first useful cue is a $\gcd(m,n)$ sitting as a coefficient multiplying a binomial coefficient in a divisibility claim. On seeing it, write Bézout's identity immediately: $\gcd(m,n) = um + vn$ for some integers $u, v$.

**2.** That substitution suggests splitting a single fraction into two integer-combination pieces (linearity over $\mathbb{Z}$), rather than attacking the divisibility multiplicatively or prime by prime.
**3.** Hesitation occurs right after substituting $um + vn$, when the piece $\frac{um}{n}\binom{n}{m}$ still looks like a fraction and it is tempting to fall back on expanding factorials from scratch. Push past this by recalling (or re-deriving by double counting) the absorption identity $m\binom{n}{m} = n\binom{n-1}{m-1}$ before touching any prime factorisation.
**4.** The trap is reaching for Legendre's formula and Kummer's theorem (prime-by-prime carry counting) as the default tool for binomial-coefficient divisibility – it is correct but needlessly heavy here, and it is easy to mishandle when $n$ has several distinct prime factors, since each one needs its own separate verification.
**5.** To recognise this faster: whenever a gcd (or anything else definable as an integer linear combination) appears as a coefficient in front of an expression that can independently be written as a combination of two related, already-integer quantities, look for the substitution before reaching for any factorisation.
**6.** The reusable rule: $\gcd(m, n)$ is never just a number – it is definitionally $um + vn$ for some integers $u, v$. Substituting this turns a divisibility question "does $n$ divide $\gcd(m, n) \cdot X$?" into two separate questions, "is $\frac{m}{n}X$ an integer?" and "is $X$ an integer?", and combinatorial absorption identities (or double counting) often answer the first directly without any prime factorisation.

## 2000 B3

**2000 B3.** Let $f(t) = \sum_{j=1}^{N} a_j \sin(2\pi jt)$, where each $a_j$ is real and $a_N$ is not equal to 0. Let $N_k$ denote the number of zeroes (including multiplicities) of $\frac{d^k f}{dt^k}$. Prove that

$$N_0 \leq N_1 \leq N_2 \leq \cdots \text{ and } \lim_{k\to\infty} N_k = 2N.$$

[Editorial clarification: only zeroes in $[0, 1)$ should be counted.]

**Fields:** Analysis (Rolle's theorem on a circle, trigonometric polynomials via $z = e^{2\pi it}$) **Position:** B3
**Answer:** $N_k$ is non-decreasing, because a cyclic, multiplicity-aware application of Rolle's theorem to the period-1 function $f_k$ forces a genuine zero of $f_{k+1}$ between (and inherited at) every consecutive pair of zeroes of $f_k$; and $\lim_k N_k = 2N$, because $f_k$, rewritten via $z = e^{2\pi it}$, is $z^{-N}$ times a polynomial of degree $2N$ (giving $N_k \leq 2N$ always), while for $k$ large enough the $j = N$ term dominates at its own peak points, forcing $2N$ genuine sign changes and hence $N_k = 2N$.

### Layer 1 THE WHAT — surface analysis

*The natural attempt.* The instinct is to compute. For $k = 1$, $f'(t) = \sum_j 2\pi j a_j \cos(2\pi jt)$, and you might check a toy case, $N = 1$: $f = a_1 \sin 2\pi t$ has $N_0 = 2$ zeroes in $[0, 1)$ (at $t = 0$ and $t = 1/2$), and $f' = 2\pi a_1 \cos 2\pi t$ also has 2 zeroes; the chain $N_0 \leq N_1$ holds trivially and the limit $2N = 2$ is already reached. This is reassuring but it does not obviously generalise – for larger $N$, computing $f_k$ explicitly gets unmanageable fast. For the monotonicity claim in general, the natural route is Rolle's theorem applied to the *listed* zeroes of $f_k$ in $[0, 1)$: between any two consecutive zeroes, $f_{k+1}$ has a zero. But $[0, 1)$ is a half-open interval standing in for a period-1 function, and it is easy to only check the open intervals between the zeroes you can see and forget the arc wrapping from the last zero back through $t = 1 \equiv t = 0$ to the first – an off-by-one that silently loses one guaranteed zero every time. It is equally easy to assume the zeroes are all simple and apply Rolle once per gap, forgetting that a repeated zero of $f_k$ needs its own argument to guarantee it survives (in reduced order) into $f_{k+1}$. For the upper bound $N_k \leq 2N$, the natural instinct is to try to bound the zero count of a sum of sines and cosines

directly – by counting sign changes of individual terms, or some calculus estimate – and this does not obviously terminate in a clean bound, because a sum of $N$ oscillating terms does not visibly "look like" a degree-$2N$ object until you change variables. And for the limit, the natural argument – "as $k \to \infty$ the $j = N$ term, growing like $N^{4k}$, eventually dominates the rest" – is true in spirit but is not yet a proof: without naming specific points and an explicit threshold on $k$, there is no way to convert "dominates" into an actual count of sign changes.

*The model chain.* **Monotonicity.** Treat $[0, 1)$ as a circle, since $f$ (and every derivative of $f$) is periodic with period 1. Suppose $f_k$ has zeroes $0 \le a_0 \le a_1 \le \cdots \le a_r < 1$ in $[0, 1)$, counted with multiplicity but listed as distinct points $a_0 < \cdots < a_r$; a zero of multiplicity $m$ at $a_i$ forces (by Rolle applied $m - 1$ times to $f_k$ itself near $a_i$, or directly: $f_k$ vanishes to order $m$ at $a_i$ means $f_k(t) = (t - a_i)^m g(t)$ locally with $g(a_i) \neq 0$, so $f_{k+1}(t)$ vanishes to order at least $m - 1$ there) a zero of $f_{k+1}$ of multiplicity at least $m - 1$ at the same point $a_i$. Between each consecutive pair $(a_i, a_{i+1})$ – including the wraparound pair $(a_r, a_0)$, read cyclically through $t = 1 \equiv 0$ – Rolle's theorem supplies a genuine zero of $f_{k+1}$ strictly inside, adopting the convention that a degenerate "interval" $(a_i, a_i)$ (when consecutive listed zeroes coincide, i.e. a multiplicity is being unpacked) simply contributes the point $a_i$ itself. Summing the zeroes inherited at each $a_i$ (with multiplicity dropped by 1) and the zeroes newly created strictly between consecutive points gives at least as many zeroes of $f_{k+1}$, counted with multiplicity, as $f_k$ had. Hence $N_{k+1} \ge N_k$.

**Upper bound.** Set $z = e^{2\pi i t}$, so $\sin(2\pi j t) = \frac{z^j - z^{-j}}{2i}$. Differentiating $\sin(2\pi j t)$ four times returns $(2\pi j)^4 \sin(2\pi j t)$, so $f_{4k}(t) = (2\pi)^{4k} \sum_{j=1}^{N} j^{4k} a_j \sin(2\pi j t)$; up to the nonzero constant $(2\pi)^{4k}$ and (for $k$ not a multiple of 4) a similar constant times cos, every derivative of $f$ is, after multiplying through by $z^N$, a genuine polynomial in $z$ of degree at most $2N$ – because the highest power of $z$ appearing is $z^N$ and the lowest is $z^{-N}$. Since $a_N \neq 0$, this polynomial is not identically zero, so by the fundamental theorem of algebra it has at most $2N$ roots (with multiplicity) on the unit circle, hence $f_k$ has at most $2N$ zeroes in $[0, 1)$: $N_k \le 2N$ for every $k$.

**Limit.** Write $f_k(t) = \sum_{j=1}^{N} j^{4\lfloor k/4 \rfloor} (\cdots) a_j$(sine or cosine of $2\pi j t$); restricting to $k = 4\ell$ for concreteness, $f_{4\ell}(t) = (2\pi)^{4\ell} \sum_j j^{4\ell} a_j \sin(2\pi j t)$. Evaluate at the $2N$ points $t_i = \frac{2i+1}{2N}$, $i = 0, 1, \ldots, 2N - 1$, chosen so that $\sin(2\pi N t_i) = \sin(\pi(2i + 1)) = 0$ – correction: choose instead $t_i = \frac{2i+1}{4N}$ so that $2\pi N t_i = \pi(2i + 1)/2$ and $\sin(2\pi N t_i) = \pm 1$ exactly, for $i = 0, \ldots, 2N - 1$; these $2N$ points are distinct in $[0, 1)$. At $t_i$, the $j = N$ term contributes $\pm N^{4\ell} a_N$. Once $\ell$ is large enough that

$$|a_N| N^{4\ell} > \sum_{j=1}^{N-1} |a_j| j^{4\ell},$$

which holds for all sufficiently large $\ell$ because $(j/N)^{4\ell} \to 0$ for every $j < N$, the sign of $f_{4\ell}(t_i)$ is forced to match the sign of the $j = N$ term, which alternates as $i$ runs from 0 to $2N - 1$. So $f_{4\ell}$ changes sign $2N$ times around the circle, giving at least $2N$ zeroes. Combined with the upper bound $N_{4\ell} \le 2N$, this forces $N_{4\ell} = 2N$ for all sufficiently large $\ell$; since $(N_k)$ is non-decreasing and bounded above by $2N$, and it equals $2N$ infinitely often, $N_k = 2N$ for all sufficiently large $k$, so $\lim_{k\to\infty} N_k = 2N$.

### Layer 2 THE HOW — the bifurcation

**Bifurcating point.** The three sub-claims split at three different lines. For monotonicity, both readings agree Rolle's theorem is the tool; they split at "between any two *listed* zeroes of $f_k$" – the natural attempt applies this to the open intervals it can see, the model chain applies it cyclically to the circle (adding the wraparound interval) and separately, at each zero, to the multiplicity itself. For the upper bound, both readings agree the count must be finite; they split at whether the trigonometric sum is left as-is (natural attempt) or rewritten via $z = e^{2\pi i t}$ into an honest polynomial (model chain), after which "how many zeroes" becomes "what is the polynomial's degree." For the limit, both readings agree the top-frequency term should eventually dominate; they split at whether that domination is asserted asymptotically (natural

attempt) or witnessed at $2N$ explicit points with an explicit threshold inequality on $\ell$ (model chain).

> *$f$ is periodic with period 1 and I am counting its zeroes in $[0,1)$ – that is really counting zeroes on a circle, so Rolle's theorem has to be applied cyclically, including the arc from the last listed zero back through $t = 1 \equiv 0$ to the first, or I will undercount by exactly one interval every single time I apply this.*

> *A zero of $f_k$ with multiplicity $m$ does not just guarantee a zero of $f_{k+1}$ somewhere strictly between it and the next zero – it directly forces $f_{k+1}$ to vanish to order at least $m-1$ at that same point, because locally $f_k(t) = (t-a_i)^m g(t)$ with $g(a_i) \neq 0$, and differentiating peels off exactly one power of $(t-a_i)$; multiplicities are inherited lower bounds at a fixed point, not just extra zeroes I need to find somewhere nearby.*

> *Writing $z = e^{2\pi i t}$ turns each $\sin(2\pi j t)$ into $\frac{z^j - z^{-j}}{2i}$, a genuinely algebraic object – so $f_k(t)$, multiplied through by $z^N$, becomes an honest polynomial in $z$ of degree at most $2N$, and "how many zeroes does this trigonometric sum have" becomes "how many roots does this polynomial have," which the fundamental theorem of algebra answers directly instead of by calculus estimation.*

> *To prove $f_k$ has at least $2N$ zeroes for large $k$, an asymptotic claim that "the $j = N$ term dominates as $k \to \infty$" is not yet a proof – I need $2N$ explicit sample points where the $j = N$ term takes its extreme values $\pm 1$, and one finite inequality, valid once $k$ passes some explicit threshold, that makes every other term too small to override the sign there.*

**Layer 3 THE WHY**

Why is it legitimate to apply Rolle's theorem around a wraparound interval at all? Rolle's theorem is a statement about a differentiable function on a closed interval with equal endpoint values; the leap from "an interval" to "the whole circle" is licensed precisely because $f_k$ is periodic with period 1 (every derivative of a period-1 trigonometric polynomial is again period-1), so $f_k(0) = f_k(1)$ genuinely, with no boundary discontinuity to worry about – going around the circle costs nothing extra, it is simply one more application of the same theorem to the interval $[a_r, 1] \cup [0 \equiv 1, a_0]$ glued at the periodic point. The convention that a degenerate interval $(a_i, a_i)$ contains the point $a_i$ itself is not a bookkeeping trick either: it is precisely what makes the multiplicity-tracking argument and the between-zeroes argument compatible, so that a repeated zero is counted the same way whether you think of it as "one point with high multiplicity" or as "several coincident listed zeroes."

Why does the substitution $z = e^{2\pi i t}$ turn a calculus problem into an algebra problem? Because a real trigonometric polynomial of frequency-degree $N$ – a finite sum of sin and cos at frequencies up to $N$ – is, term by term, built from $z^{\pm j}$ for $j \leq N$; multiplying by $z^N$ clears every negative power and leaves an honest polynomial of degree at most $2N$ in $z$. The bound $N_k \leq 2N$ then has nothing to do with sines, cosines, or calculus at all – it is the fundamental theorem of algebra applied to a specific nonzero polynomial (nonzero because $a_N \neq 0$ guarantees the top coefficient, $z^{2N}$'s coefficient, survives), restricted to the special locus $|z| = 1$. The role of $a_N \neq 0$ in the hypothesis is exactly this: it is what stops the "degree $2N$" polynomial from secretly being lower degree, which would break both the upper bound's tightness and the limit's target value.

Why must the top-frequency term eventually dominate, and why does that force the limit rather than just a suspicion? Differentiating four times multiplies each term's effective coefficient by $(2\pi j)^4$; iterating $\ell$ times multiplies by $(2\pi j)^{4\ell}$, so the ratio of the $j$-th term's peak size to the $N$-th term's peak size scales like $(j/N)^{4\ell}$, which tends to 0 for every $j < N$ because $j/N < 1$ – a

one-line comparison of geometric growth rates, not an appeal to intuition. This converts "the top term eventually dominates" into an explicit threshold: past that $\ell$, the sign of $f_{4\ell}$ at each of the $2N$ chosen peak points is forced to alternate, producing $2N$ sign-change zeroes there. Since the upper bound already caps $N_{4\ell}$ at $2N$, the lower bound of $2N$ sign changes pins $N_{4\ell}$ at exactly $2N$ – not approximately, not eventually in a vague sense, but exactly, for every sufficiently large multiple of 4. Monotonicity of $N_k$ (already established) then forces the same value for every sufficiently large $k$, not just multiples of 4, because a non-decreasing integer sequence bounded above by $2N$ that hits $2N$ infinitely often must equal $2N$ from that point on.

**Layer $+\alpha$ BACKTESTING**

1. The first useful cue is that $f$ is a finite sum of $\sin(2\pi jt)$ terms, period 1, and the question is about zero counts of successive derivatives. On seeing it, write down immediately: treat $[0,1)$ as a circle (every derivative of $f$ is again period-1), and note the problem is really three separate sub-claims – non-decreasing, bounded, and a matching limit – each wanting its own tool.
2. The cue suggests Rolle's theorem, applied cyclically and with explicit multiplicity tracking, for the monotonicity claim; and it suggests the substitution $z = e^{2\pi it}$, turning the trigonometric sum into an honest polynomial, for both the upper bound (via the fundamental theorem of algebra) and the limit (via evaluation at the top frequency's own peak points).
3. Hesitation occurs at exactly two places: the wraparound interval and points of higher multiplicity in the Rolle argument (a solver who checks only the open intervals between distinct listed zeroes will miscount by one), and the transition from "the top term dominates asymptotically" to an actual proof (a solver who stops at the intuition never produces the $2N$ specific points or the explicit threshold inequality that makes it rigorous).
4. The trap is treating "the $j = N$ term eventually dominates" as self-evidently a complete argument for the limit, without ever naming the $2N$ peak points $t_i = (2i+1)/4N$ or writing down the finite inequality $|a_N|N^{4\ell} > \sum_{j<N} |a_j| j^{4\ell}$ that makes the domination witnessed rather than asserted.
5. To recognise this faster: whenever a problem gives a finite trigonometric sum and asks about zero counts of derivatives, expect a fixed three-part toolkit every time – Rolle's theorem (cyclically, with multiplicity) for monotonicity, the substitution $z = e^{2\pi it}$ for an upper bound via polynomial degree, and evaluation at the top frequency's own peak points for the matching lower bound that pins the limit.
6. The reusable rule: a real trigonometric polynomial of frequency-degree $N$ is secretly a Laurent polynomial of degree $2N$ under $z = e^{2\pi it}$ – reach for this substitution the moment "how many roots" is asked of such an object, and reserve pure calculus (Rolle, sign-change arguments) for questions about how those root counts interlace or change under differentiation, never for bounding the total count itself.

## 2000 B4

**2000 B4.** Let $f(x)$ be a continuous function such that $f(2x^2 - 1) = 2xf(x)$ for all $x$. Show that $f(x) = 0$ for $-1 \le x \le 1$.

**Fields:** real analysis (functional equations, continuity); trigonometric substitution; density **Position:** B4 — the start of the hard half of paper B, but short once the equation is run in the right direction **Answer:** $f$ vanishes on the dense set $\{\cos(m\pi/2^k)\}$, hence, by continuity, on all of $[-1,1]$.

**Layer 1 THE WHAT — surface analysis**

*The natural attempt.* The strong student starts by plugging in the values that make the equation collapse. At $x = 1$: $f(1) = 2f(1)$, so $f(1) = 0$. At $x = 0$: $f(-1) = 0$. At $x = -\frac{1}{2}$, where

$2x^2-1=x$: $f(-\frac{1}{2})=-f(-\frac{1}{2})$, so $f(-\frac{1}{2})=0$. Three zeros, and a feeling that the equation is a machine for producing more.

The next move is to recognise $2x^2-1$ as the double-angle formula and put $x=\cos\theta$:

$$f(\cos 2\theta)=2\cos\theta\, f(\cos\theta). \tag{$*$}$$

Now the student iterates *forwards*, from $\theta$ to $2\theta$ to $4\theta$:

$$f(\cos 2^k\theta)=2^k\cos\theta\cos 2\theta\cdots\cos 2^{k-1}\theta\; f(\cos\theta)=\frac{\sin 2^k\theta}{\sin\theta}\,f(\cos\theta),$$

using the telescoping product $2\sin\theta\cos\theta=\sin 2\theta$ repeatedly. So

$$f(\cos\theta)=\frac{\sin\theta}{\sin 2^k\theta}\,f(\cos 2^k\theta).$$

The plan was to make the right side small. It is not small. $f$ is bounded on $[-1,1]$ by continuity, say by $M$, but $\sin 2^k\theta$ wanders over $[-1,1]$ and is sometimes tiny, so the bound on $|f(\cos\theta)|$ is $M|\sin\theta|/|\sin 2^k\theta|$, which does not go to zero along any sequence of $k$ that the student controls. The student has found a genuine invariant, $f(\cos\theta)/\sin\theta$ is unchanged under $\theta\mapsto 2\theta$, and is stuck on how to use it. That is the wall: forwards iteration from a generic point produces an identity, not an estimate.

*The model chain.* **Step 1: $f$ is odd.** Replace $x$ by $-x$: $f(2x^2-1)=-2xf(-x)$. Comparing with the given equation, $2xf(x)=-2xf(-x)$, so $f(-x)=-f(x)$ for every $x\neq 0$. Letting $x\to 0$ and using continuity, $f(0)=-f(0)$, so $f(0)=0$.

**Step 2: the endpoints.** At $x=1$ the equation reads $f(1)=2f(1)$, so $f(1)=0$; at $x=0$ it reads $f(-1)=0$.

**Step 3: zeros propagate to half-angles.** Let $Z=\{\theta\in\mathbb{R}: f(\cos\theta)=0\}$. Claim: if $\theta\in Z$ then $\theta/2\in Z$. Put $x=\cos(\theta/2)$; then $2x^2-1=\cos\theta$, and the equation gives

$$0=f(\cos\theta)=2\cos(\theta/2)\,f(\cos(\theta/2)).$$

If $\cos(\theta/2)\neq 0$, divide: $f(\cos(\theta/2))=0$. If $\cos(\theta/2)=0$, then $f(\cos(\theta/2))=f(0)=0$ by Step 1. Either way $\theta/2\in Z$.

**Step 4: induction.** Claim: $m\pi/2^k\in Z$ for all integers $m$ and all $k\geq 0$. For $k=0$, $\cos(m\pi)=\pm 1$ and Step 2 applies. If the claim holds for $k$, then $m\pi/2^{k+1}=(m\pi/2^k)/2\in Z$ by Step 3.

**Step 5: density.** The numbers $m\pi/2^k$ with $0\leq m\leq 2^k$ are dense in $[0,\pi]$. Given $y\in[-1,1]$, write $y=\cos\alpha$ with $\alpha\in[0,\pi]$ and pick $\theta_j=m_j\pi/2^{k_j}\to\alpha$. Then $\cos\theta_j\to y$ and $f(\cos\theta_j)=0$, so $f(y)=0$ by continuity of $f$. □

**Layer 2 THE HOW — the bifurcation**

**Bifurcating point.** Both chains agree through $(*)$: substitute $x=\cos\theta$ so the equation relates the angle $\theta$ to the angle $2\theta$. They split at the choice of direction. The natural attempt runs $(*)$ as "know $f(\cos\theta)$, deduce $f(\cos 2\theta)$" — doubling. The model runs it as "know $f(\cos 2\theta)=0$, deduce $f(\cos\theta)=0$" — halving. Doubling spreads a generic point around the circle chaotically; halving starting from a known zero produces the dyadic angles $m\pi/2^k$, a dense set, with no estimates at all.

> *I can't bound $f$ at a generic point. So where do I actually* know *$f$? At $\pm 1$, from plugging in. The equation says $2x\,f(x)=0$ whenever $f(2x^2-1)=0$, so a zero at $y$ hands me a zero at each square root of $(1+y)/2$. I run the equation backwards from the zeros I have.*

> *In angles, going backwards is halving. Halve* 0 *and* $\pi$ *forever and I get every* $m\pi/2^k$*. That set is dense, and* $f$ *is continuous. Done — as long as I have* $f(0) = 0$ *for the one case where I can't divide by* $2x$*.*

- ▷ **Cue:** a functional equation of the form $f(T(x)) = c(x)f(x)$ with $f$ continuous and the conclusion "$f = 0$" $\longrightarrow$ **Move:** find the points where $f$ is forced to be 0, then run the equation backwards through $T^{-1}$ to spread zeros to a dense set
- ▷ **Cue:** the expression $2x^2 - 1$ (or $4x^3 - 3x$, $x^2 - 2$) inside a functional equation $\longrightarrow$ **Move:** substitute $x = \cos\theta$ (or $x = 2\cos\theta$): the map becomes $\theta \mapsto 2\theta$ (or $3\theta$)
- ▷ **Cue:** an equation that is unchanged under $x \mapsto -x$ on one side $\longrightarrow$ **Move:** compare the two instances to extract a parity of $f$ for free
- ▷ **Cue:** a product $\cos\theta\cos 2\theta\cdots\cos 2^{k-1}\theta$ $\longrightarrow$ **Move:** multiply by $\sin\theta$ and telescope with $2\sin u\cos u = \sin 2u$

**Layer 3 THE WHY**

Why backwards and not forwards? Go to what the equation is. It is a statement about the map $T(x) = 2x^2 - 1$, which on $[-1, 1]$ is angle-doubling in disguise: $T(\cos\theta) = \cos 2\theta$. Angle-doubling on the circle is the standard example of a chaotic map: forward orbits of generic points are dense and unpredictable, which is exactly why the forward iteration produced $\sin 2^k\theta$ with no control. Its *inverse*, halving, is the opposite: every step contracts angles by a factor 2, and the backward orbits of a single point fill the circle in an orderly dyadic grid. A functional equation $f(T(x)) = c(x)f(x)$ transports information along orbits of $T$; the question is only which direction transports it somewhere useful. You want to reach *every* point of $[-1, 1]$, and the orbit structure that reaches every point in a controlled way is the backward tree of preimages, not the forward orbit.

Why start at $\pm 1$? Because the conclusion is $f = 0$ and zeros are the only thing the equation propagates unconditionally backwards: if $f(T(x)) = 0$ then $c(x)f(x) = 0$, and $c(x) = 2x$ vanishes at one point only. Nonzero values would need division and would carry the factor $2x$ along; zeros carry nothing. And the places where the equation hands you zeros for free are the fixed points and special points of $T$: $T(1) = 1$ gives $f(1) = 2f(1)$; $T(0) = -1$ gives $f(-1) = 0$. Plugging in these values was the right instinct. The mistake was treating them as curiosities rather than as seeds.

Why is $f(0)$ needed, and why is oddness the way to get it? The backward step divides by $2x$ with $x = \cos(\theta/2)$, which fails exactly when $\theta/2$ is an odd multiple of $\pi/2$, i.e. at the point $x = 0$. The equation at $x = 0$ says only $f(-1) = 0$ and is silent about $f(0)$. The only other source of information is the symmetry $x \mapsto -x$, which fixes $2x^2 - 1$ and flips $2x$. That symmetry says $f$ is odd, and a continuous odd function vanishes at 0.

Kedlaya's solution begins by setting $g(t) = f(\cos t)/\sin t$ and showing $g(2t) = g(t)$ and $g(t + \pi) = g(t)$. Here is where that comes from: it is the invariant the natural attempt found, $f(\cos\theta)/\sin\theta = f(\cos 2^k\theta)/\sin 2^k\theta$. The official solution then combines doubling with the period $\pi$: $g(1 + n\pi/2^k) = g(2^k + n\pi) = g(2^k) = g(1)$, so $g$ equals $g(1)$ on the dense set $\{1 + n\pi/2^k\}$. (Each intermediate value $2^j + n\pi/2^{k-j}$ is not a multiple of $\pi$, since $\pi$ is irrational, so $g$ is defined along the way.) Continuity makes $g$ constant on its domain, and oddness $g(-t) = -g(t)$ forces the constant to be 0. Notice what that argument does: it too moves backwards, from the fixed value $g(2^k)$ to the dyadic points $1 + n\pi/2^k$ that double onto it. The official route and the model route are the same idea; the model applies it to $f$ directly and needs no irrationality of $\pi$ and no division by $\sin t$. The natural attempt had the right invariant and was one change of direction away from the official proof.

**Layer $+\alpha$ BACKTESTING**

**1.** The first useful cue is $2x^2 - 1$. On seeing it, write $x = \cos\theta$ and the equation $f(\cos 2\theta) = 2\cos\theta\, f(\cos\theta)$, restricted to $[-1, 1]$, which is exactly the interval named in the conclusion.

**2.** A functional equation with a doubling map and a continuity hypothesis suggests: find forced values, propagate them along orbits of the map, finish by density plus continuity.
**3.** Hesitation occurs after forward iteration yields $\sin\theta/\sin 2^k\theta$ and no decay. It is reinforced away by asking "where do I already know $f$?" — the answer ($\pm 1$) is a zero, and zeros propagate backwards through $2x\,f(x) = 0$.
**4.** The trap is estimating: trying to show $|f(\cos\theta)| \to 0$ by bounding the forward iterate. A second, smaller trap is the division by $2x$ at $x = 0$; without Step 1 the induction silently breaks at $\cos(\pi/2)$.
**5.** Ask, for any equation $f(T(x)) = c(x)f(x)$: *is $T$ expanding or contracting, and in which direction do orbits fill the space in an orderly way?* Iterate in that direction, from the points the equation pins down.
**6.** Reusable decision rule: *to prove a continuous function vanishes identically from a functional equation, seed zeros at fixed or special points, propagate them through the inverse of the map to a dense set, and close with continuity; if a division step fails at an isolated point, get that point from a symmetry.*

## 2000 B5

**2000 B5.** Let $S_0$ be a finite set of positive integers. We define finite sets $S_1, S_2, \ldots$ of positive integers as follows: the integer $a$ is in $S_{n+1}$ if and only if exactly one of $a-1$ or $a$ is in $S_n$. Show that there exist infinitely many integers $N$ for which $S_N = S_0 \cup \{N + a : a \in S_0\}$.

**Fields:** combinatorics; algebra over $\mathbb{F}_2$ (polynomials mod 2); Pascal's triangle mod 2 **Position:** B5 — hard half of the paper, but a single translation turns it into a two-line computation **Answer:** Every power of two $N = 2^k \geq \max S_0$ works, because $(1+x)^{2^k} \equiv 1 + x^{2^k} \pmod 2$.

### Layer 1 THE WHAT — surface analysis

*The natural attempt.* The strong student experiments. Take $S_0 = \{1\}$. The rule "$a \in S_{n+1}$ iff exactly one of $a-1, a$ is in $S_n$" gives

$$S_1 = \{1,2\}, \quad S_2 = \{1,3\}, \quad S_3 = \{1,2,3,4\}, \quad S_4 = \{1,5\}, \quad S_5 = \{1,2,5,6\}, \ \ldots$$

Written as rows of 0s and 1s these are the rows of Pascal's triangle mod 2, and rows $1, 2, 4, 8$ have only their two end entries, so $S_N = \{1, N+1\}$ for $N$ a power of 2. That is the conjecture for a singleton. The student may even prove it: $a \in S_n$ iff $\binom{n}{a-1}$ is odd (induction, from Pascal's rule), and $\binom{2^k}{j}$ is odd only for $j = 0, 2^k$.

Then comes $S_0 = \{1,2\}$, and the singleton argument does not transfer directly. The student computes $S_1 = \{1,3\}$, $S_2 = \{1,2,3,4\}$, $S_3 = \{1,5\}$, $S_4 = \{1,2,5,6\}$, sees that $S_4 = S_0 \cup (4 + S_0)$ again, and suspects some superposition. But the rule mentions "exactly one", and the student in the arena cannot yet see why the evolution of $\{1,2\}$ should be the evolution of $\{1\}$ overlaid on the evolution of $\{2\}$: where two evolving patterns overlap, what happens? A page of case analysis on overlapping patterns follows, or an attempt at a general formula via Lucas's theorem with $S_0$ arbitrary. The wall is not the singleton; it is combining singletons, because the student has not named the operation that "exactly one" performs.

*The model chain.* **Step 1: encode sets as polynomials mod 2.** For a finite set $S$ of positive integers put $P_S(x) = \sum_{j \in S} x^j$, a polynomial with coefficients in $\mathbb{F}_2 = \{0,1\}$ (arithmetic mod 2). Distinct sets give distinct polynomials, and every polynomial over $\mathbb{F}_2$ with zero constant term is $P_S$ for exactly one finite set $S$ of positive integers.

**Step 2: the rule is multiplication by $1+x$.** The coefficient of $x^a$ in $(1+x)P_{S_n}(x) = P_{S_n}(x) + xP_{S_n}(x)$ is $[a \in S_n] + [a-1 \in S_n] \pmod 2$, where $[\cdot]$ is 1 if true and 0 if false. That sum is 1

exactly when one, and only one, of $a-1, a$ lies in $S_n$. So

$$P_{S_{n+1}}(x) = (1+x)\,P_{S_n}(x) \quad \text{in } \mathbb{F}_2[x].$$

(The constant term stays 0: the coefficient of $x^0$ is $[0 \in S_n] + [-1 \in S_n] = 0$, consistent with $S_{n+1}$ consisting of positive integers.) By induction, $P_{S_n} = (1+x)^n P_{S_0}$.

**Step 3: Frobenius.** In $\mathbb{F}_2[x]$, $(u+v)^2 = u^2 + 2uv + v^2 = u^2 + v^2$. Applying this $k$ times, $(1+x)^{2^k} = 1 + x^{2^k}$.

**Step 4: conclude.** Let $M = \max S_0$ and $N = 2^k$ with $N \geq M$. Then

$$P_{S_N}(x) = (1+x^N)P_{S_0}(x) = \sum_{a \in S_0} x^a + \sum_{a \in S_0} x^{N+a}.$$

The exponents in the first sum lie in $[1, M]$, those in the second in $[N+1, N+M]$, and $N+1 > M$, so no two terms cancel. Hence $S_N = S_0 \cup \{N + a : a \in S_0\}$. There are infinitely many such $k$. □

### Layer 2 THE HOW — the bifurcation

**Bifurcating point.** Both chains agree on the data: small cases look like Pascal's triangle mod 2, and powers of 2 are special. They split at the phrase "exactly one of $a-1$ or $a$". The natural attempt reads it as a rule about sets and tries to track shapes. The model reads it as arithmetic: "exactly one of two bits is 1" is the sum of the bits mod 2. Once the rule is addition mod 2 of a set and its shift by one, the evolution is multiplication by $1+x$, superposition is automatic (multiplication distributes over addition), and the power-of-two phenomenon is the identity $(1+x)^2 = 1 + x^2$.

> *"Exactly one of" is XOR, and XOR is addition mod* 2. *So $S_{n+1}$ is $S_n$ plus $S_n$ shifted by one, added mod* 2. *A shift is multiplication by $x$. I write each set as a polynomial over $\mathbb{F}_2$ and the whole process is $(1+x)^n$.*

> *I want $(1+x)^N$ to have two terms. Mod* 2, *squaring kills cross terms: $(1+x)^2 = 1 + x^2$. So powers of* 2, *and then the target set drops out as long as the shifted copy does not overlap the original — take $N \geq \max S_0$.*

- ▷ **Cue:** a rule of the form "$a$ is in the new set iff exactly one (or an odd number) of $a - c_1, \ldots, a - c_r$ is in the old set" ⟶ **Move:** encode sets as polynomials over $\mathbb{F}_2$; the step is multiplication by $x^{c_1} + \cdots + x^{c_r}$
- ▷ **Cue:** "exactly one of two" conditions ⟶ **Move:** it is addition mod 2 (XOR); the process is linear over $\mathbb{F}_2$, so superposition holds
- ▷ **Cue:** a power $(1+x)^N$ over $\mathbb{F}_p$ that should collapse to few terms ⟶ **Move:** take $N = p^k$: $(1+x)^{p^k} = 1 + x^{p^k}$ by Frobenius
- ▷ **Cue:** a union of a set and a translated copy defined through cancellation ⟶ **Move:** choose the shift larger than the diameter of the set so that no terms cancel

### Layer 3 THE WHY

Why is addition mod 2 the natural reading and not a trick? Go to the definition of "exactly one of $P, Q$". It is true when $(P, Q)$ is (true, false) or (false, true), and false when both or neither hold. Record true as 1 and false as 0: the outputs are $1+0=1, 0+1=1, 1+1=0, 0+0=0$ — the addition table of $\mathbb{F}_2$. There is nothing to notice; "exactly one" *is* mod-2 addition, written in words. And once the rule is "add two things", where the two things are the set and its shift, you are looking at a linear operation, and the linear-algebra reflex is to find the coordinates in which it is multiplication.

Why polynomials? Because the only other operation in sight is the shift $a \mapsto a+1$, and the object in which shifting is multiplication is the generating polynomial: $x \cdot x^a = x^{a+1}$. The characteristic function of a set, with the shift acting on it, is exactly a polynomial with the multiplication-by-$x$ operator. So the encoding $S \mapsto \sum_{j \in S} x^j$ is not imported from outside; it is the unique way to make both "add" and "shift" into ring operations at once.

Why powers of 2? The experiments already said so; the question is why it is inevitable. In $(1+x)^N$ over $\mathbb{F}_2$ the coefficient of $x^j$ is $\binom{N}{j}$ mod 2, and you want all middle coefficients even. Squaring a sum over $\mathbb{F}_2$ loses its cross term because the cross term has coefficient 2. Repeated squaring therefore preserves the two-term shape: $1+x \to 1+x^2 \to 1+x^4 \to \cdots$. This is the Frobenius map $u \mapsto u^2$, which is additive in characteristic 2 because the prime 2 divides every $\binom{2}{j}$ with $0 < j < 2$. The same fact for the prime $p$ says $\binom{p^k}{j} \equiv 0 \pmod{p}$ for $0 < j < p^k$, which is why rows $p^k$ of Pascal's triangle mod $p$ are "$1, 0, \ldots, 0, 1$".

Kedlaya's solution begins by writing $\sum_{j \in S_n} x^j \equiv (1+x) \sum_{j \in S_{n-1}} x^j \pmod 2$. Here is where that comes from: it is the definition of $S_n$, read with "exactly one" as addition mod 2 and "$a-1$" as a shift, with the shift made into multiplication. The official solution states it as if it were visible; it is visible only once you have translated the English into $\mathbb{F}_2$.

The natural attempt's singleton proof via $\binom{n}{a-1}$ mod 2 was correct. What it lacked was the one word *linear*: the evolution of $S_0$ is the sum mod 2 of the evolutions of its elements, and at $N = 2^k$ each singleton $\{a\}$ becomes $\{a, a+N\}$. Overlaps are not a case analysis — they are addition, and at $N \geq \max S_0$ there are none. The student who had said "exactly one means XOR" at line one would have had superposition before computing a single example.

**Layer $+\alpha$ BACKTESTING**

**1.** The first cue is the phrase "exactly one of $a-1$ or $a$". On seeing it, write $[a \in S_{n+1}] \equiv [a \in S_n] + [a-1 \in S_n] \pmod 2$.

**2.** A mod-2 recurrence involving shifts suggests generating polynomials over $\mathbb{F}_2$: the step is multiplication by $1+x$, so $P_{S_n} = (1+x)^n P_{S_0}$.

**3.** Hesitation occurs when passing from a singleton $S_0$ to a general one: will overlapping patterns interfere? It is reinforced away by the word *linear*: the rule is addition mod 2, so evolutions superpose.

**4.** The trap is simulation and pattern-matching on shapes (Sierpiński triangle), followed by case analysis of overlaps. A small second trap is forgetting that $N$ must be at least $\max S_0$, or else the two copies overlap and cancel.

**5.** Translate logical connectives into arithmetic before doing anything else: "exactly one" and "odd number of" are mod-2 sums; a shift is multiplication by $x$. Ask: *is this process linear over some field?*

**6.** Reusable decision rule: *when a process is defined by parity conditions on shifted copies, encode states as polynomials over $\mathbb{F}_2$; the process becomes multiplication by a fixed polynomial $Q$, the $n$-th state is $Q^n$ times the initial one, and $Q^{2^k}(x) = Q(x^{2^k})$ explains every power-of-two self-similarity.*

## 2000 B6

**2000 B6.** Let $B$ be a set of more than $2^{n+1}/n$ distinct points with coordinates of the form $(\pm 1, \pm 1, \ldots, \pm 1)$ in $n$-dimensional space with $n \geq 3$. Show that there are three distinct points in $B$ which are the vertices of an equilateral triangle.

**Fields:** combinatorics (double counting, pigeonhole); geometry of the hypercube $\{\pm 1\}^n$ **Position:** B6 — the last problem, hard to start and three lines to finish; all the difficulty is in reading the bound $2^{n+1}/n$ **Answer:** Some vertex $X$ of the cube (not necessarily in $B$) has three neighbours $P, Q, R$ in $B$; they are pairwise at distance $2\sqrt{2}$.

**Layer 1 THE WHAT — surface analysis**

*The natural attempt.* The strong student first translates distance. For $P, Q \in \{\pm 1\}^n$, each coordinate contributes $(p_i - q_i)^2 \in \{0, 4\}$, so

$$|PQ|^2 = 4\, d(P, Q),$$

where $d$ is the Hamming distance, the number of coordinates in which $P$ and $Q$ differ. An equilateral triangle is three points of $B$ pairwise at the same Hamming distance. So far, so good.

Then the student reaches for induction on $n$, because the cube splits into two half-cubes by the last coordinate. Let $B^+$ and $B^-$ be the points of $B$ with last coordinate $+1$ and $-1$. One of them has at least $|B|/2 > 2^n/n$ points, and deleting the last coordinate embeds it in $\{\pm 1\}^{n-1}$ without changing Hamming distances. The inductive hypothesis for $n-1$ needs more than $2^{(n-1)+1}/(n-1) = 2^n/(n-1)$ points. But the student has only $2^n/n$, and $2^n/n < 2^n/(n-1)$. The induction loses a factor $n/(n-1)$ at every step and cannot close. (It fails before the base case is even relevant.)

The student next tries to find a triangle inside $B$ directly: fix $P \in B$ and look for $Q, R$ at the same distance $d$ from $P$ and from each other, for some $d$. With $d$ free there are too many cases, and no counting argument suggests itself because the student is counting inside $B$ only. That is the wall: the student is searching $B$ for a triangle, while the bound is telling them to count something outside $B$.

*The model chain.* **Step 1: which triangle to look for.** As above, $|PQ|^2 = 4d(P, Q)$. Target triangles with $d = 2$: three points pairwise differing in exactly two coordinates, i.e. side $2\sqrt{2}$.

**Step 2: neighbourhoods.** For a vertex $P$ of the cube, let $S_P$ be the set of the $n$ vertices at Hamming distance 1 from $P$ (flip one coordinate).

**Step 3: double count.** Count pairs $(P, X)$ with $P \in B$ and $X \in S_P$. Each $P$ contributes $n$ pairs, so there are $n|B| > n \cdot 2^{n+1}/n = 2 \cdot 2^n$ pairs. The second entry $X$ ranges over the $2^n$ vertices of the cube. If every vertex $X$ lay in at most two of the sets $S_P$, there would be at most $2 \cdot 2^n$ pairs. So some vertex $X$ lies in $S_P, S_Q, S_R$ for three distinct $P, Q, R \in B$.

**Step 4: the triangle.** Each of $P, Q, R$ is $X$ with one coordinate flipped. Since $P, Q, R$ are distinct, the flipped coordinates $i, j, l$ are distinct. Then $P$ and $Q$ differ exactly in coordinates $i$ and $j$, so $d(P, Q) = 2$; likewise $d(Q, R) = d(P, R) = 2$. All three sides have length $2\sqrt{2}$, and $P, Q, R \in B$ are the vertices of an equilateral triangle. □

(The hypothesis $n \geq 3$ is where three distinct coordinates to flip exist; for $n \leq 2$ the bound $2^{n+1}/n \geq 2^n$ exceeds the number of vertices, so the statement would be vacuous.)

**Layer 2 THE HOW — the bifurcation**

**Bifurcating point.** Both chains agree on $|PQ|^2 = 4d(P, Q)$. They split at the question "where do I look?". The natural attempt looks for structure inside $B$ (split $B$, or fix a point of $B$). The model reads the bound: $2^{n+1}/n$ is $2 \times$ (number of vertices) $\div\, n$, and it looks for a vertex $X$ — possibly not in $B$ — that is hit three times when each point of $B$ fires $n$ arrows.

> *The bound is $2^{n+1}/n$. Rewrite it: $n|B| > 2 \cdot 2^n$. That is "n things per point of B, more than 2 per vertex of the cube". So each point of B sends out n of something into the $2^n$ vertices, and some vertex gets hit three times. What does each point have n of? Neighbours.*

> *Three points that are all neighbours of one vertex X: each is X with a different coordinate flipped, so any two differ in two coordinates. That's an equilateral triangle, and X doesn't have to be in B.*

> *Why side $2\sqrt{2}$? Hamming distance 1 can't give a triangle: around a triangle each coordinate contributes 0 or 2 to $d(P,Q)+d(Q,R)+d(R,P)$, so that sum is even and three equal odd sides are impossible. Distance 2 is the smallest that can work, so I try it first.*

- ▷ **Cue:** a threshold of the form $c \cdot (\text{size of the ambient space})/k$ ⟶ **Move:** double count: each element of the set generates $k$ objects in the ambient space; pigeonhole with capacity $c$ gives a point hit $c+1$ times
- ▷ **Cue:** an induction on $n$ that loses a factor like $n/(n-1)$ at each step ⟶ **Move:** abandon induction; the bound was designed for a single global count
- ▷ **Cue:** distances between points of $\{\pm 1\}^n$ or $\{0,1\}^n$ ⟶ **Move:** convert to Hamming distance: $|PQ|^2 = 4d(P,Q)$ (or $d(P,Q)$)
- ▷ **Cue:** need three points pairwise at Hamming distance 2 ⟶ **Move:** look for a common neighbour: three one-coordinate flips of a single vertex

**Layer 3 THE WHY**

Why read the bound as a double count? Go to what a bound of that shape can mean. A condition "$|B| > 2^{n+1}/n$" is equivalent to $n \cdot |B| > 2 \cdot 2^n$. The left side is a number of incidences (each point of $B$ times $n$); the right side is a capacity (each of $2^n$ boxes times 2). A strict inequality between incidences and capacity is the pigeonhole principle in its counting form: some box receives more than its capacity, i.e. at least 3. The problem setter did not choose $2^{n+1}/n$ at random; it is exactly the threshold at which "each vertex has at most two neighbours in $B$" becomes impossible. Once you rewrite the bound this way, you are asked two questions — what are the $n$ objects per point, and what are the boxes? — and on the cube there is one natural family of $n$ objects per vertex: its neighbours, which are themselves vertices, i.e. boxes.

Why are three co-neighbours an equilateral triangle, and why is that the only kind we need? Hamming distance is the cube's own metric, and a triangle of side $d$ in Hamming distance is a triangle of side $2\sqrt{d}$ in space. The parity argument above rules out odd $d$. For $d = 2$ there is more: every Hamming-2 triangle is a star around a common neighbour. Indeed, if $P, Q$ differ in $\{i,j\}$ and $Q, R$ differ in a two-element set $T$, then $P, R$ differ in the symmetric difference $\{i,j\} \triangle T$, which has size 2 only if $|\{i,j\} \cap T| = 1$, say $T = \{j,l\}$; then $X = P$ with coordinate $i$ flipped differs from $Q$ only in $j$ and from $R$ only in $l$. So "$B$ contains a side-$2\sqrt{2}$ triangle" is *equivalent* to "some vertex has three neighbours in $B$". The model is not a lucky construction; it is the exact reformulation of the smallest possible triangle.

Why did induction fail, and why was that informative? Halving the cube halves the number of points but divides the threshold by only $2(n-1)/n$, not by 2. That loss is a signal that the bound is not recursive in nature: it comes from a single global count in which $n$ appears as the degree of the cube graph. When the dimension appears in a denominator, ask where each point is counted $n$ times; do not peel off dimensions one at a time.

Kedlaya's solution begins by setting $S_P$, the set of points differing from $P$ in exactly one coordinate. Here is where that comes from: the bound, multiplied through by $n$, says $\sum_{P \in B} |S_P| > 2 \cdot 2^n$, and the sets $S_P$ are the only size-$n$ sets attached to a vertex that live in the same $2^n$-element space. The one conceptual step the official solution leaves implicit is that the common element $X$ need not belong to $B$; the counting runs over the whole cube, and the triangle's vertices $P, Q, R$ are the ones in $B$.

**Layer $+\alpha$ BACKTESTING**

**1.** The first useful cue is the bound $2^{n+1}/n$. On seeing it, write $n|B| > 2 \cdot 2^n$ and ask what $n$ objects each point of $B$ owns in a space of size $2^n$.

**2.** Double counting with pigeonhole: count incidences (point of $B$, neighbouring vertex); average more than 2 per vertex forces a vertex with three neighbours in $B$.

**3.** Hesitation comes from wanting the three points' common structure to be inside $B$. It is reinforced away by allowing the pigeonhole boxes to be all $2^n$ vertices, whether or not they

lie in $B$.

**4.** The trap is induction on $n$ by splitting the cube, which loses a factor $n/(n-1)$ and never closes; a second trap is searching over all Hamming distances $d$ instead of fixing the smallest feasible one, $d = 2$.

**5.** When a threshold has the dimension (or a degree) in the denominator and the size of the ambient space in the numerator, multiply it out and read it as "incidences exceed capacity".

**6.** Reusable decision rule: *to force a configuration from a density bound $|B| > c|V|/k$, find a structure attaching $k$ elements of $V$ to each point of $B$ such that $c+1$ points sharing an attached element form the configuration; then double count incidences and apply pigeonhole.*