DAWN PUTNAM SESSION

Putnam 1999 — Papers A and B

Putnam 120 Archive

Contents

Putnam 1999 — Paper A **5**
1999 A1 5
1999 A2 7
1999 A3 9
1999 A4 12
1999 A5 15
1999 A6 18

Putnam 1999 — Paper B **23**
1999 B1 23
1999 B2 25
1999 B3 27
1999 B4 30
1999 B5 33
1999 B6 35

Putnam 1999 — Paper A

1999 A1

1999 A1. Find polynomials $f(x)$, $g(x)$, and $h(x)$, if they exist, such that for all x,

$$|f(x)| - |g(x)| + h(x) = \begin{cases} -1 & \text{if } x < -1 \\ 3x+2 & \text{if } -1 \le x \le 0 \\ -2x+2 & \text{if } x > 0. \end{cases}$$

Fields: piecewise-linear algebra, absolute value decomposition **Position:** A1 — the opener **Answer:** Yes, they exist: $f(x) = \frac{3x+3}{2}$, $g(x) = \frac{5x}{2}$, $h(x) = -x + \frac{1}{2}$ (not the unique choice) work, because the target is continuous and piecewise linear with exactly two breakpoints, at $x = -1$ and $x = 0$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The two breakpoints are staring at you: the formula changes character at $x = -1$ and at $x = 0$, and $|f(x)|$, $-|g(x)|$ are the only pieces of the left-hand side that can bend a line. So the honest first move is to guess that f has its root at one breakpoint and g at the other, since a linear function's absolute value has exactly one corner, sitting at its root.

Take the natural pairing: $f(x) = \alpha(x+1)$ (root at -1), $g(x) = \beta x$ (root at 0), $h(x) = cx + d$. Matching the slope on each of the three regions (using $\frac{d}{dx}|x+1| = \text{sign}(x+1)$, $\frac{d}{dx}|x| = \text{sign}(x)$) gives three linear equations in α, β, c, and a fourth from matching one value gives d. This pairing happens to be the right one and the system solves cleanly to $\alpha = 3/2$, $\beta = 5/2$, $c = -1$, $d = 1/2$ — recovering the answer above.

But nothing in the problem tells you in advance which breakpoint goes with $+|f|$ and which with $-|g|$; the other pairing looks equally natural. Try it: $f(x) = \alpha x$ (root at 0), $g(x) = \beta(x+1)$ (root at -1), $h(x) = cx + d$. The same slope-matching procedure is perfectly well-posed and returns an answer — $\alpha = -5/2$, $\beta = -3/2$, $c = -1$, $d = 7/2$ — that looks just as legitimate as the first. Substitute it back at $x = -2$, though: $|f(-2)| - |g(-2)| + h(-2) = |-5| - |-3/2 \cdot (-1)| + (2 + 7/2) = 5 - 1.5 + 5.5 = 9$, not -1. The system that "solves cleanly" solved the wrong system: the value-matching equation used to pin down d silently assumed the slope equations already had the correct global sign pattern, and slope-matching alone under-determines which physical breakpoint attaches to the $+$ term and which to the $-$ term. The two pairings are not symmetric, and guessing wrong costs a full second pass through the algebra to discover it.

The model chain. Use the identity $\max(r,s) = \frac{1}{2}(r + s + |r - s|)$, equivalently $\max(r, 0) = \frac{1}{2}(r + |r|)$. Write the target $F(x)$ as

$$F(x) = \max\{-3x - 3,\, 0\} \;-\; \max\{5x,\, 0\} \;+\; 3x + 2,$$

and check it region by region: for $x < -1$, $-3x - 3 > 0$ and $5x < 0$, so $F = (-3x - 3) - 0 + 3x + 2 = -1$; for $-1 \le x \le 0$, $-3x - 3 \le 0$ and $5x \le 0$, so $F = 0 - 0 + 3x + 2 = 3x + 2$; for $x > 0$, $-3x - 3 < 0$ and $5x > 0$, so $F = 0 - 5x + 3x + 2 = -2x + 2$. All three cases match.

Now convert each $\max\{\cdot, 0\}$ to an absolute value with the identity above:

$$\begin{aligned} F(x) &= \tfrac{1}{2}\big((-3x-3) + |-3x-3|\big) \\ &\quad - \tfrac{1}{2}\big(5x + |5x|\big) + 3x + 2 \\ &= \tfrac{1}{2}|3x+3| - \tfrac{1}{2}|5x| \\ &\quad + \tfrac{1}{2}(-3x-3-5x) + 3x + 2 \\ &= \left|\tfrac{3x+3}{2}\right| - \left|\tfrac{5x}{2}\right| - x + \tfrac{1}{2}, \end{aligned}$$

i.e. $F(x) = |f(x)| - |g(x)| + h(x)$ with $f(x) = \frac{3x+3}{2}$, $g(x) = \frac{5x}{2}$, $h(x) = -x + \frac{1}{2}$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree the two breakpoints -1 and 0 must be the roots of f and g respectively (in that order, not the reverse). They split on *how the correct order is established*. The natural attempt treats the pairing as a free choice to be guessed and checked after the fact by solving a slope system; the model chain never has to guess, because it builds F as $\max\{A, 0\} - \max\{B, 0\} + (\text{linear})$ directly from the sign of each region, and the sign convention of $\max(r, 0)$ forces the correct pairing as a side effect of writing the piece down.

> *The breakpoint at* -1 *is where the slope* increases *(from* 0 *to* 3*), and* $\max\{A, 0\}$ *is exactly the shape that is flat and then turns on and rises — so the increasing breakpoint belongs to the term I add. The breakpoint at* 0 *is where the slope* decreases *(from* 3 *to* -2*), and that is the shape* $-\max\{B, 0\}$ *produces — flat contribution that then subtracts a rising piece, i.e. bends the total curve downward. I should read the sign of the slope jump at each breakpoint off the piecewise definition first, and let that tell me which side of* $|f| - |g|$ *it belongs to, rather than assigning roots to* f *and* g *arbitrarily and discovering the mismatch by trial substitution.*

Layer 3 THE WHY

Why does the sign of the slope jump determine which term a breakpoint belongs to, and why does slope-matching alone fail to recover that sign? A continuous piecewise-linear function is completely determined by its slope on each region, and the slope changes only at breakpoints, by exactly the jump in slope there. The functions $r \mapsto |r|$ and $r \mapsto \max(r, 0)$ both have a single corner at $r = 0$, but they encode a slope jump of opposite total size and, crucially, $\max(r, 0) = \frac{1}{2}(r + |r|)$ shows that the "$+|r|$" contribution to a sum always comes attached to a specific companion term ($+r$), fixing an orientation. Slope-matching with an ansatz $\alpha|x+1| + \beta|x| + cx + d$ throws that orientation away: α and β are solved as free real numbers from three slope equations, and the system is entirely blind to which physical event (turning on vs turning off) each absolute-value term was "supposed" to represent. It happens to return a consistent-looking answer for the wrong pairing too, because three linear equations in three unknowns (α, β, c) generically have a unique solution regardless of which breakpoint is attached to which variable name — the inconsistency only shows up once you check the fourth, independent condition (the actual value d, or equivalently a point far from both breakpoints). The max construction never separates orientation from magnitude in the first place, so there is no later check that can fail.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that F is continuous and piecewise linear with breakpoints exactly at $x = -1$ and $x = 0$. On seeing it, write down the slope jump at each breakpoint (here: $+3$ at $x = -1$, from $3 - 4 = -5$ at $x = 0$ meaning a jump of -5) before choosing any ansatz.
2. That cue suggests representing F as a signed sum of ramp functions, one per breakpoint, each of size equal to the slope jump there — which the identity $\max(r, 0) = \frac{1}{2}(r + |r|)$ converts directly into an $|f| - |g|$ decomposition.

3. Hesitation occurs exactly at the moment of assigning breakpoints to f versus g: both assignments look algebraically identical at the "set up the slope equations" stage. This is reinforced away by reading the *sign* of each slope jump off the piecewise definition first — increasing jumps are $+$ terms, decreasing jumps are $-$ terms — rather than trusting whichever pairing is solved first.

4. The trap is that slope-matching alone (three equations in the coefficients) is satisfied by both pairings, because it never encodes which term is added and which is subtracted; only a value check (or the sign of the jump) breaks the tie, and it is easy to stop as soon as a consistent-looking numeric answer appears.

5. Recognise this family faster by treating "piecewise linear, breakpoints at $a_1 < \cdots < a_k$" as an instruction to write down the slope jump at each a_i immediately, before setting up any system — the jump's sign is the pairing, not something to be discovered afterward.

6. Reusable rule: when reconstructing a piecewise-linear (or piecewise-anything with a fixed corner structure) function from named pieces like $|f|, |g|, \max, \min$, extract the local behaviour (here: sign of each slope jump) at every corner *before* choosing which named piece owns which corner. A system of equations that only encodes magnitudes, not orientation, will return a plausible wrong answer as readily as the right one.

▷ **Cue:** a continuous piecewise-linear target with known breakpoints must be written using $|\cdot|$ or max / min terms $\longrightarrow$ **Move:** read the sign of the slope jump at each breakpoint first, and match increasing jumps to added absolute-value terms, decreasing jumps to subtracted ones, via $\max(r, 0) = \frac{1}{2}(r + |r|)$

▷ **Cue:** a slope-matching system for an ansatz with a free choice of which root attaches to which term returns a consistent-looking solution $\longrightarrow$ **Move:** do not trust it until checked against one value far from all corners — magnitude equations alone cannot see orientation

1999 A2

1999 A2. Let $p(x)$ be a polynomial that is nonnegative for all real x. Prove that for some k, there are polynomials $f_1(x), \ldots, f_k(x)$ such that

$$p(x) = \sum_{j=1}^{k} (f_j(x))^2.$$

Fields: polynomial algebra, complex numbers, sums of squares **Position:** A2 — conceptually the paper's first real proof **Answer:** Yes always, and in fact $k = 2$ always suffices: factor p over $\mathbb{R}$, take a square root of the (necessarily even-multiplicity) real-root part, write each irreducible quadratic factor as a sum of two squares by completing the square, then collapse the whole product of two-square terms back down to a single sum of two squares using the Gaussian-integer norm identity $(u^2 + v^2)(s^2 + t^2) = (us - vt)^2 + (ut + vs)^2$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The problem only asks that *some* k works, and that phrasing invites exactly the reading a strong student takes first: don't worry about economy, just find enough squares. Factor $p(x)$ over $\mathbb{R}$. Real roots must occur with even multiplicity — if r were a root of odd multiplicity, p would change sign across r, contradicting nonnegativity — so the real-root part of p is $q(x)^2$ for some real polynomial q. What remains is a product of irreducible real quadratics $x^2 + b_i x + c_i$ with negative discriminant (no real roots, hence no sign change, hence — being continuous and equal to $p/q^2 \geq 0$ away from the roots of q — always positive). Complete the square on each: $x^2 + b_i x + c_i = (x + b_i/2)^2 + (c_i - b_i^2/4)$, and $c_i - b_i^2/4 > 0$ exactly because the discriminant $b_i^2 - 4c_i < 0$. So each quadratic factor is honestly a sum of two squares, $u_i(x)^2 + v_i(x)^2$.

Now $p(x) = q(x)^2 \prod_i \left(u_i(x)^2 + v_i(x)^2\right)$. This already answers the problem as stated: expand the product of $(u_i^2 + v_i^2)$ terms out, absorb the leftover $q(x)^2$ into one of the factors, and the result is *some* sum of squares — just possibly a large and ugly one, because multiplying out several two-term sums-of-squares by brute algebra does not obviously stay a sum of squares of the same shape. Try it for even two quadratic factors, $(u_1^2 + v_1^2)(u_2^2 + v_2^2)$: expanding gives $u_1^2u_2^2 + u_1^2v_2^2 + v_1^2u_2^2 + v_1^2v_2^2$, four terms, and there is no square hiding among them individually — you would need to regroup these four cross terms into two perfect squares by trial and error, and with three or more quadratic factors the number of cross terms to regroup grows combinatorially. The natural attempt *answers* the problem (some k exists — take k large and don't simplify) but abandons any attempt to see *why* exactly two squares always suffice, because the regrouping step looks like it needs a fresh trick for every new factor.

The model chain. Do the same factorisation, but work in $\mathbb{C}[x]$ from the start instead of regrouping real terms after the fact. Write $p(x) = q(x)r(x)$ where q has all the real roots (with even multiplicity, as above, so $q = s^2$ for a real polynomial s) and r has none. Since r has real coefficients, its complex roots occur in conjugate pairs, so

$$r(x) = \prod_{j=1}^{k} (x - a_j)(x - \overline{a_j}).$$

Write $\prod_{j=1}^{k}(x - a_j) = t(x) + iu(x)$ for real polynomials t, u (possible since every coefficient of the product is some real polynomial in the coefficients of the a_j's combined with their conjugates, split into real and imaginary parts). Then for real x, $\overline{t(x) + iu(x)} = t(x) - iu(x)$, and

$$r(x) = \prod_j (x - a_j) \cdot \overline{\prod_j (x - a_j)} = \big(t(x) + iu(x)\big)\big(t(x) - iu(x)\big) = t(x)^2 + u(x)^2.$$

Therefore

$$p(x) = q(x)r(x) = s(x)^2\big(t(x)^2 + u(x)^2\big) = \big(s(x)t(x)\big)^2 + \big(s(x)u(x)\big)^2,$$

a sum of exactly two squares, with no regrouping step at all.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains reach the identical intermediate fact: every irreducible real quadratic factor of p is a sum of two squares. They split at what happens when there is more than one such factor. The natural attempt multiplies the real two-square expressions together and tries to regroup the resulting cross terms by inspection, factor by factor. The model chain never multiplies real two-square expressions at all — it multiplies the underlying *complex linear factors* $(x - a_j)$ together first, and only takes real and imaginary parts of that product at the very end, after all the multiplication is done.

> *A sum of two squares $u^2 + v^2$ is the squared modulus of a single complex number $u + iv$. If I want the product of several two-square expressions to stay a sum of two squares, I should stop multiplying the real expressions and instead multiply the complex numbers whose squared moduli they are — because $|z_1|^2|z_2|^2 = |z_1z_2|^2$ is an identity I get for free, and taking the real and imaginary part of one final complex product is a single step, not a regrouping problem that gets harder with every new factor.*

Layer 3 THE WHY

Why does multiplying complex numbers succeed where multiplying real two-square expressions stalls? Because $u^2 + v^2 = (u + iv)(u - iv) = |u + iv|^2$ is not an arbitrary algebraic coincidence — it is the definition of the complex modulus, and the modulus is multiplicative:

$|z_1z_2| = |z_1||z_2|$ for all complex z_1, z_2, because $|z|^2 = z\bar{z}$ and conjugation is a ring homomorphism, so $\overline{z_1z_2} = \overline{z_1}\,\overline{z_2}$ and $|z_1z_2|^2 = z_1z_2\overline{z_1z_2} = (z_1\overline{z_1})(z_2\overline{z_2}) = |z_1|^2|z_2|^2$. This is exactly the algebraic content buried inside the "Brahmagupta–Fibonacci" identity $(u_1^2+v_1^2)(u_2^2+v_2^2) = (u_1u_2 - v_1v_2)^2 + (u_1v_2 + u_2v_1)^2$: it is what you get by expanding $|z_1|^2|z_2|^2 = |z_1z_2|^2$ in coordinates. The natural attempt was trying to rediscover this identity by regrouping four cross terms every time a new factor appeared; the model chain never needs to state the identity at all, because working in $\mathbb{C}[x]$ makes the multiplicativity of the modulus automatic — $r(x) = \prod_j (x - a_j) \cdot \overline{\prod_j (x - a_j)}$ is a single modulus-squared of a single product, however many factors j there are. The deeper structural fact underneath everything is that a real quadratic with negative discriminant has no real root *precisely because* it factors as $(x-a)(x-\bar{a})$ for a genuinely non-real a — the two-squares identity for a single such quadratic and the multiplicativity that combines several of them are the same phenomenon (a complex conjugate pair), viewed once at degree 2 and once at higher degree.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "nonnegative for all real x" applied to real roots specifically: sign-change-at-a-simple-root plus nonnegativity forces every real root to have even multiplicity. On seeing it, write down: the real-root part of p is a perfect square, so it can be peeled off immediately and handled separately from the roots that never touch the real axis.

2. That cue suggests factoring what remains into irreducible real quadratics with negative discriminant, and completing the square on each — the standard two-square identity for a single quadratic with no real root.

3. Hesitation occurs exactly at "now there is more than one quadratic factor, how do I combine their two-square representations into one." This is reinforced away by asking what a sum of two squares *is* (a modulus squared) before attempting to multiply the real expressions directly — the moment $u^2 + v^2$ is recognised as $|u + iv|^2$, the combination step becomes "multiply complex numbers," which needs no new trick per factor.

4. The trap is answering the literal question ("for *some* k") by taking k large — one square pair per irreducible quadratic factor, expanded and left ungrouped — which technically satisfies the problem but throws away the much stronger and structurally cleaner fact that $k = 2$ always suffices, and leaves the student unable to see why.

5. Recognise this family faster by treating any appearance of "sum of two squares" as a coded modulus-squared, and any product of several such expressions as a coded product of complex numbers — multiply the complex numbers, take real/imaginary parts once, at the end.

6. Reusable rule: when a target quantity is built from a product of several "sum of two squares" pieces, do not multiply the real two-square expressions and try to regroup cross terms into new squares. Instead identify the complex number whose modulus-squared each piece is, multiply those complex numbers together (multiplicativity of the modulus is automatic), and split the single final product into real and imaginary parts once.

▷ **Cue:** a polynomial nonnegative on all of $\mathbb{R}$ must be shown to be a sum of squares of polynomials $\longrightarrow$ **Move:** split off the real-root part as a perfect square (real roots of a nonnegative polynomial have even multiplicity), then treat the remaining roots as complex-conjugate pairs whose product is a squared modulus

▷ **Cue:** a product of several 'sum of two squares' expressions needs to be shown to still be a sum of two squares $\longrightarrow$ **Move:** recognise each $u^2 + v^2$ as $|u + iv|^2$, multiply the underlying complex numbers (modulus is multiplicative for free), and split the single product into real and imaginary parts once, rather than regrouping real cross terms factor by factor

1999 A3

1999 A3. Consider the power series expansion

$$\frac{1}{1-2x-x^2} = \sum_{n=0}^{\infty} a_n x^n.$$

Prove that, for each integer $n \geq 0$, there is an integer m such that

$$a_n^2 + a_{n+1}^2 = a_m.$$

Fields: linear recurrences, generating functions, companion matrices **Position:** A3 — generating-function flavored, algebra-heavy **Answer:** $m = 2n+2$ always works: encoding the pair (a_n, a_{n+1}) as the top row of M^{n+2} for the companion matrix $M = 01$
12 turns $a_n^2 + a_{n+1}^2$ into the top-left entry of $(M^{n+2})^2 = M^{2n+4}$, which the same encoding identifies as a_{2n+2}.

Layer 1 THE WHAT — surface analysis

The natural attempt. Clearing denominators in $(1-2x-x^2)\sum a_n x^n = 1$ and comparing coefficients of x^{n+1} (for $n \geq 1$) gives $a_{n+1} = 2a_n + a_{n-1}$, with $a_0 = 1, a_1 = 2$. Computing a few terms — $1, 2, 5, 12, 29, 70, 169, \ldots$ — and checking $a_n^2 + a_{n+1}^2$ against the list, the pattern $m = 2n+2$ jumps out almost immediately ($1+4 = 5 = a_2$, $4+25 = 29 = a_4$, $25+144 = 169 = a_6$).

The natural next move is to find a_n explicitly and substitute. The characteristic equation of the recurrence is $t^2 = 2t+1$, with roots $\alpha = 1+\sqrt{2}$, $\beta = 1-\sqrt{2}$, giving $a_n = A\alpha^n + B\beta^n$ for constants A, B fixed by $a_0 = 1, a_1 = 2$. Square and add:

$$a_n^2 + a_{n+1}^2 = A^2\alpha^{2n}(1+\alpha^2) + B^2\beta^{2n}(1+\beta^2) + 2AB(\alpha\beta)^n(1+\alpha\beta).$$

The cross term is genuinely helpful — $\alpha\beta = -1$ (product of roots of $t^2 - 2t - 1$), so $1+\alpha\beta = 0$ and the whole cross term vanishes, which is a real and pleasant simplification, not a dead end. What is left is

$$A^2\alpha^{2n}(1+\alpha^2) + B^2\beta^{2n}(1+\beta^2).$$

To match this against $a_{2n+2} = A\alpha^{2n+2} + B\beta^{2n+2} = A\alpha^{2n}\cdot\alpha^2 + B\beta^{2n}\cdot\beta^2$, term by term, requires $A(1+\alpha^2) = \alpha^2$ and $B(1+\beta^2) = \beta^2$ — two more identities, each needing $\alpha^2 = 2\alpha+1$ (the defining equation of α) substituted in, and each needing the specific numerical values of A, B (which themselves involve $1/\sqrt{2}$) checked against a further algebraic identity. It is completely mechanical, and it works — but every one of the five ingredients ($\alpha\beta = -1$, $1+\alpha^2 = 2\alpha+2$, the value of A, the value of B, and the final match) is a separate fact that has to be pulled in and verified at the right moment, and dropping any one of them silently produces a wrong-looking final expression with no signal about which step broke.

The model chain. Let $M = \begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix}$, the companion matrix of the recurrence $a_{n+1} = 2a_n + a_{n-1}$. Claim: for all $n \geq 0$,

$$M^{n+2} = \begin{pmatrix} a_n & a_{n+1} \\ a_{n+1} & a_{n+2} \end{pmatrix}.$$

Base case $n = 0$: $M^2 = \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}$, matching $(a_0, a_1, a_2) = (1, 2, 5)$. Inductive step: if the claim holds for n, then

$$M^{n+3} = M^{n+2}M = \begin{pmatrix} a_n & a_{n+1} \\ a_{n+1} & a_{n+2} \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} a_{n+1} & a_n + 2a_{n+1} \\ a_{n+2} & a_{n+1} + 2a_{n+2} \end{pmatrix},$$

and $a_n + 2a_{n+1} = a_{n+2}$, $a_{n+1} + 2a_{n+2} = a_{n+3}$ are exactly the recurrence, so this equals $\begin{pmatrix} a_{n+1} & a_{n+2} \\ a_{n+2} & a_{n+3} \end{pmatrix}$, closing the induction.

Now square the matrix identity itself: $(M^{n+2})^2 = M^{2n+4}$, by the ordinary law of exponents for a fixed matrix. The left side, using the claim, is

$$\begin{pmatrix} a_n & a_{n+1} \\ a_{n+1} & a_{n+2} \end{pmatrix}^2,$$

whose top-left entry (row 1 times column 1, using symmetry of the matrix) is $a_n \cdot a_n + a_{n+1} \cdot a_{n+1} = a_n^2 + a_{n+1}^2$. The right side, $M^{2n+4} = M^{(2n+2)+2}$, has top-left entry a_{2n+2} by the same claim applied at index $2n+2$. Hence

$$a_n^2 + a_{n+1}^2 = a_{2n+2},$$

so $m = 2n+2$ works, for every $n \geq 0$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains land on the correct answer $m = 2n+2$ by the same numerical pattern-spotting. They split immediately after that, at the choice of *what object to square*. The natural attempt squares the two numbers a_n and a_{n+1} separately (via their closed forms) and adds — and then has to fight to show the sum collapses back to a single term of the same sequence, because nothing about adding two squared numbers is obviously related to the recurrence's index-doubling structure. The model chain squares the *matrix* M^{n+2} that encodes the pair (a_n, a_{n+1}) as a single object, and gets the doubled index for free from $M^{n+2} \cdot M^{n+2} = M^{2(n+2)}$ — an identity that requires no algebra at all, only the definition of a matrix power.

> *If I encode the* pair (a_n, a_{n+1}) *— not the single value* a_n *— as one algebraic object, and that object happens to be a power of a fixed matrix, then squaring the pair is forced to obey* $M^k \cdot M^k = M^{2k}$*, an index law I get for free from what a matrix power even means. A closed-form sum of two exponential terms has no such law built in:* $\alpha^n + \beta^n$ *squared does not know it should collapse to* $\alpha^{2n} + \beta^{2n}$ *plus a vanishing cross term unless I go and verify that by hand, fact by fact.*

Layer 3 THE WHY

Why does representing the pair as a matrix power succeed where representing each term separately as a closed-form exponential does not? Because $M^k \cdot M^\ell = M^{k+\ell}$ is not a property that has to be proved about this particular matrix — it is true of *every* matrix, purely because repeated multiplication by the same object composes, the same way $x^k \cdot x^\ell = x^{k+\ell}$ is true of every number. The content specific to this problem is entirely used up in a single fact: that M^{n+2}'s entries *are* a_n, a_{n+1}, a_{n+2} in the first place, which holds because M is the companion matrix of the recurrence $a_{n+1} = 2a_n + a_{n-1}$ — its bottom row, by construction, computes exactly "shift and apply the recurrence" when multiplied against a vector of two consecutive terms. Once that one fact is established (by an induction that only ever re-invokes the recurrence itself, nothing more), every further manipulation — in particular squaring — is free, because it is now matrix-power arithmetic, not sequence arithmetic. The closed-form route has no such free structure: $a_n = A\alpha^n + B\beta^n$ represents each term *individually* as a sum of two scalars, and scalar addition has no multiplicative law connecting $a_n^2 + a_{n+1}^2$ to a_{2n+2} — every connection (the vanishing cross term via $\alpha\beta = -1$, the coefficient identity via $\alpha^2 = 2\alpha + 1$) has to be manufactured by hand because it is not carried automatically by the representation. The lesson generalises past this problem: a linear recurrence's *companion matrix* packages consecutive terms into one object precisely so that matrix identities (multiplicativity of powers, in particular) become available as free consequences of "index arithmetic," where the same claim phrased about the scalar terms is not free at all.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the shape of the claim itself: $a_n^2 + a_{n+1}^2$ equals a single later term. On seeing "sum of two squares of consecutive terms equals another term of the same

sequence," write down: this is the signature of an index-doubling identity, and index-doubling identities usually come from squaring a fixed object that already encodes both terms together.

2. That cue suggests looking for a matrix (or other algebraic object) whose powers encode consecutive pairs (a_n, a_{n+1}), rather than reaching first for the sequence's closed form.

3. Hesitation occurs right after deriving the closed form $a_n = A\alpha^n + B\beta^n$, because it is genuinely the fastest thing to write down and the cross term does vanish nicely (via $\alpha\beta = -1$), which makes it feel like the right path. This is reinforced away by noticing, before grinding further, that matching the surviving terms to a_{2n+2} still requires two more separate algebraic facts ($\alpha^2 = 2\alpha + 1$ substituted twice, plus the specific values of A, B) — a sign that the representation is fighting the problem rather than encoding it.

4. The trap is that the closed-form route *does* work and its first simplification (the vanishing cross term) is genuinely satisfying, which makes it easy to commit to finishing it rather than stepping back to ask whether a different representation makes the whole argument free instead of merely tractable.

5. Recognise this family faster by treating "square of one term plus square of the next term equals another term" as a standing invitation to look for a companion-matrix (or Fibonacci-style 2×2-matrix) encoding, before reaching for Binet-type closed forms.

6. Reusable rule: for a linear recurrence $a_{n+1} = pa_n + qa_{n-1}$, package consecutive pairs as powers of the companion matrix $M = 01$
qp (verified by one induction using only the recurrence) whenever the target identity relates a_n, a_{n+1} *combined* to a term at a shifted or doubled index — the matrix index law ($M^k M^\ell = M^{k+\ell}$) then does the combining automatically, where manipulating closed-form exponential expressions requires re-deriving that combination by hand every time.

▷ **Cue:** an identity claims that some combination of a_n and a_{n+1} (consecutive terms of a linear recurrence) equals another term at a related index $\longrightarrow$ **Move:** package (a_n, a_{n+1}) as the entries of a fixed matrix raised to a power (the companion matrix of the recurrence), and let the matrix index law do the combining, rather than substituting a Binet-type closed form and manipulating scalars by hand

▷ **Cue:** a closed-form (Binet-type) computation is working and its first simplification is satisfying $\longrightarrow$ **Move:** before committing to finish it, count how many further separate algebraic facts still need to be manufactured by hand; if more than one, check whether a structural encoding (matrix, generating function) makes the rest free instead

1999 A4

1999 A4. Sum the series

$$\sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \frac{m^2 n}{3^m (n3^m + m3^n)}.$$

Fields: analysis (series); algebra (symmetrisation) **Position:** A4 — middle of the paper, but one of the kinder A4s once the right rewriting is seen; the whole problem is one move. **Answer:** $\frac{9}{32}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You have summed double series before, and the standard plan is: do the inner sum in closed form, then do the outer sum. So you fix m and look at

$$\sum_{n=1}^{\infty} \frac{m^2 n}{3^m (n3^m + m3^n)}.$$

The denominator $n3^m + m3^n$ mixes a polynomial and an exponential in n, added together. There is no partial fractions decomposition in n, no telescoping, no geometric series in sight. You try the other order: fix n, sum over m. Now 3^m appears both outside the bracket and inside it, next to $m3^n$, and the sum over m is no more tractable. You try bounding: the term is at most $\frac{m^2 n}{3^m \cdot n3^m} = \frac{m^2}{9^m}$, which proves convergence of each inner sum but tells you nothing about the value.

Next you compute a few terms numerically, hoping to recognise the limit. The $(1,1)$ term is $\frac{1}{3\cdot 6} = \frac{1}{18}$; partial sums creep up towards 0.28. Without a candidate closed form, 0.28125 is not something you will recognise as $9/32$ in the arena unless you already suspect a square of $3/4$.

This is the wall: the summand is not separable, and iterated summation has nothing to grip.

The model chain. *Step 1: reveal the structure.* Divide numerator and denominator by mn. Since $n3^m + m3^n = mn\left(\frac{3^m}{m} + \frac{3^n}{n}\right)$, set $a_k = 3^k/k$. Then

$$\frac{m^2 n}{3^m(n3^m + m3^n)} = \frac{m^2 n}{3^m \cdot mn\,(a_m + a_n)} = \frac{m}{3^m} \cdot \frac{1}{a_m + a_n} = \frac{1}{a_m(a_m + a_n)}.$$

Step 2: symmetrise. All terms are positive, so the double series may be summed in any order and rearranged freely (Tonelli for non-negative series; equivalently, every partial-sum ordering has the same supremum). Call the sum S. Exchanging the names of m and n,

$$S = \sum_{m,n\geq 1} \frac{1}{a_m(a_m + a_n)} = \sum_{m,n\geq 1} \frac{1}{a_n(a_m + a_n)}.$$

Adding the two expressions term by term,

$$2S = \sum_{m,n\geq 1} \frac{1}{a_m + a_n}\left(\frac{1}{a_m} + \frac{1}{a_n}\right) = \sum_{m,n\geq 1} \frac{1}{a_m + a_n} \cdot \frac{a_m + a_n}{a_m a_n} = \sum_{m,n\geq 1} \frac{1}{a_m a_n}.$$

Step 3: the separated sum. Now the summand is a product of a function of m and a function of n, so

$$2S = \left(\sum_{k\geq 1} \frac{k}{3^k}\right)^2.$$

For $|x| < 1$, $\sum_{k\geq 0} x^k = \frac{1}{1-x}$; differentiating and multiplying by x gives $\sum_{k\geq 1} kx^k = \frac{x}{(1-x)^2}$. At $x = \frac{1}{3}$ this is $\frac{1/3}{4/9} = \frac{3}{4}$. Hence $2S = \frac{9}{16}$ and

$$S = \frac{9}{32}. \qquad \square$$

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start from the same summand. The natural attempt immediately commits to an order of summation ("fix m, sum over n"). The model first rewrites the summand in terms of a single sequence $a_k = 3^k/k$, and only then sees that the two indices play mirror roles. The split is at the very first line: do you sum, or do you rewrite?

> *The denominator $n3^m + m3^n$ is symmetric in m and n. The only asymmetry in the whole term is the m^2 and the 3^m out front. If I swap m and n the sum doesn't change, so I get the same S written a second way — and adding the two might kill the ugly denominator.*

> *$n3^m + m3^n$ is a sum of two terms each of which is "something in m times something in n". Dividing by mn turns it into $\frac{3^m}{m} + \frac{3^n}{n}$, a sum of a function of m and the same function of n. Give that function a name.*

Layer 3 THE WHY

Why is symmetrisation the natural move and not a trick? Start from what a double sum over all $(m, n) \in \mathbb{N}^2$ is: a sum over a set that is invariant under the swap $(m, n) \mapsto (n, m)$. That swap is a bijection of the index set, so for any non-negative f,

$$\sum_{m,n} f(m, n) = \sum_{m,n} f(n, m),$$

and therefore $\sum f = \frac{1}{2} \sum \big(f(m, n) + f(n, m)\big)$. This identity costs nothing and is always true. It is useful precisely when $f(m, n) + f(n, m)$ is simpler than f — and that happens when f is "a symmetric piece times an asymmetric piece", because then the symmetric piece factors out of the sum and the two asymmetric pieces get a chance to combine. Here the symmetric piece is $\frac{1}{a_m + a_n}$ and the asymmetric pieces are $\frac{1}{a_m}$ and $\frac{1}{a_n}$, whose sum is $\frac{a_m + a_n}{a_m a_n}$ — which contains exactly the symmetric factor needed to cancel. This is the identity

$$\frac{1}{u(u+v)} + \frac{1}{v(u+v)} = \frac{1}{uv},$$

the same partial-fractions fact that makes $\frac{1}{u} - \frac{1}{u+v} = \frac{v}{u(u+v)}$ work, read in the other direction.

So the real question is: how do you see that the summand has the form $\frac{1}{u(u+v)}$ with $u = a_m$, $v = a_n$? Go back to the phrasing. The problem is written with $m^2 n$ upstairs and $n3^m + m3^n$ downstairs. The denominator is the tell: the two terms are images of one another under $m \leftrightarrow n$. A denominator that is symmetric under swapping indices, in a sum over a symmetric index set, is a direct instruction to look at the swapped version of the sum. Once you decide to look, the rewriting by mn is forced: you want to write $n3^m + m3^n$ as $g(m) + g(n)$, and the only way to separate a term $n3^m$ into a function of m alone is to divide by mn. The numerator $m^2 n$ then becomes m, and the leftover $\frac{m}{3^m}$ is $\frac{1}{a_m}$. Nothing here was guessed; each step is the unique way to make the next one possible.

Kedlaya's solution opens with "let $a_n = 3^n/n$". That line is where the entire problem lives, and the official text gives no reason for it. The reason is: the symmetric denominator needs to be a sum $g(m) + g(n)$, and $g(k) = 3^k/k$ is what dividing by mn produces.

The rigour point is the rearrangement. Swapping the order of summation, splitting $2S$ into two copies, and adding them term by term all require the series to behave like a finite sum. For non-negative terms this is automatic: every term here is positive, so the double series either converges absolutely or diverges to $+\infty$, and in either case any rearrangement gives the same value. Since the separated sum $\frac{9}{16}$ is finite, $2S$ is finite. Write the sentence "all terms are positive, so rearrangement is permitted" — it is short and it is where a careful grader checks.

The natural attempt was not stupid. Iterated summation is the right default for a double series whose summand separates, or whose inner sum is geometric. The error is not in trying it but in persisting once the summand is visibly non-separable: a sum you cannot do in either order is telling you to change the summand, not the order.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the denominator $n3^m + m3^n$: swapping m and n maps it to itself. On seeing it, write down: *S equals the same sum with m, n exchanged, so $2S$ is the sum of the two summands.*
2. Symmetrisation: $\sum f(m, n) = \frac{1}{2} \sum \big(f(m, n) + f(n, m)\big)$, after normalising the denominator to the form $g(m) + g(n)$.

3. Hesitation occurs when you try to add $\frac{m^2 n}{3^m(\cdots)}$ and $\frac{n^2 m}{3^n(\cdots)}$ in the raw form and the numerators do not visibly combine: $m^2 n 3^n + mn^2 3^m = mn(m3^n + n3^m)$, which does cancel, but only if you factor. Reinforce it away by normalising first: divide by mn so the summand reads $\frac{1}{a_m(a_m+a_n)}$, and the cancellation is one line.

4. The trap is committing to an order of summation. Every inner sum you can write down is intractable, and the numerical value 0.28125 is hard to recognise without the structural answer $\left(\frac{3}{4}\right)^2/2$ to aim at. A secondary trap: rearranging without noting positivity.

5. Slogan: *a symmetric denominator in a sum over a symmetric index set is an invitation to add the sum to its own mirror image.* And: *normalise so the symmetric part reads $g(m)+g(n)$ before anything else.*

6. Decision rule: **when a double sum's inner sum is intractable in both orders, look for a symmetry of the index set that fixes part of the summand; average the sum over that symmetry and check whether the averaged summand factorises.**

▷ **Cue:** double sum over $m, n \geq 1$ whose denominator is unchanged by $m \leftrightarrow n$ $\longrightarrow$ **Move:** write S a second time with m, n swapped, add, and look for cancellation of the symmetric factor

▷ **Cue:** a term like $n3^m + m3^n$ (product of a function of one index with a function of the other, plus its mirror) $\longrightarrow$ **Move:** divide by mn to get $g(m)+g(n)$ with $g(k) = 3^k/k$

▷ **Cue:** summand of the form $\frac{1}{u(u+v)}$ with u, v playing mirror roles $\longrightarrow$ **Move:** $\frac{1}{u(u+v)} + \frac{1}{v(u+v)} = \frac{1}{uv}$, which separates

▷ **Cue:** $\sum_{k\geq 1} kx^k$ at a specific x $\longrightarrow$ **Move:** differentiate the geometric series: $\sum kx^k = \frac{x}{(1-x)^2}$

▷ **Cue:** rearranging or splitting an infinite double series $\longrightarrow$ **Move:** state that all terms are non-negative (or that it converges absolutely) before doing it

1999 A5

1999 A5. Prove that there is a constant C such that, if $p(x)$ is a polynomial of degree 1999, then

$$|p(0)| \leq C \int_{-1}^{1} |p(x)|\, dx.$$

Fields: analysis (inequalities for polynomials); linear algebra (finite-dimensional spaces, inner products)
Position: A5 — hard end of the paper. Nothing needs computing; the difficulty is entirely in recognising which fact about polynomials of bounded degree is being tested. **Answer:** Such a C exists. One explicit choice: $C = \max_{[-1,1]} |q|$, where q is the unique polynomial of degree ≤ 1999 with $\int_{-1}^{1} x^j q(x)\, dx = \delta_{j0}$ for $0 \leq j \leq 1999$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You ask what a counterexample would look like. You need $|p(0)|$ large while $\int |p|$ is small: a tall narrow spike at 0. For instance $p(x) = (1-x^2)^k$ has $p(0) = 1$ and

$$\int_{-1}^{1} (1-x^2)^k\, dx \sim \sqrt{\pi/k} \to 0.$$

So without the degree restriction the statement is false, and the whole content is that degree 1999 cannot make a spike arbitrarily narrow. Now you try to prove that directly. Write $p(x) = c\prod_{i=1}^{1999}(x - r_i)$ and normalise $p(0) = 1$, so

$$|p(x)| = \prod_{i=1}^{1999} \left|1 - \frac{x}{r_i}\right|.$$

You want a lower bound for $\int_{-1}^{1}$ of this, uniform in the roots $r_i \in \mathbb{C}$. Roots far away contribute factors near 1; roots near $[-1,1]$ make the factor small near them. You try to find an interval where every factor is bounded below: cover each root by a disc of radius ε, argue some gap in $[-\frac{1}{2},\frac{1}{2}]$ avoids all the discs… and this does work (it is Barton's official first solution), but you are now fighting 1999 roots, an ε, a pigeonhole on gaps and a case split on $|r_i| \le 1$, with every constant to track. In the arena, this is where time runs out or an inequality gets fumbled.

The model chain. Let $N = 1999$ and let V be the real vector space of polynomials of degree at most N; it has dimension $N+1$, with basis $1, x, \dots, x^N$. Every polynomial of degree 1999 lies in V, so it suffices to prove the inequality for all $p \in V$.

Step 1: an inner product. For $p, r \in V$ put $\langle p, r\rangle = \int_{-1}^{1} p(x)r(x)\,dx$. This is bilinear and symmetric, and $\langle p, p\rangle = \int_{-1}^{1} p^2 \ge 0$ with equality only if $p \equiv 0$ on $[-1,1]$ (a continuous non-negative function with zero integral vanishes identically), which forces $p = 0$ since a nonzero polynomial has finitely many roots.

Step 2: a kernel for evaluation at 0. Let G be the $(N+1)\times(N+1)$ matrix with entries $G_{jk} = \langle x^j, x^k\rangle = \int_{-1}^{1} x^{j+k}\,dx$, $0 \le j,k \le N$. For any real vector $c = (c_0,\dots,c_N) \ne 0$,

$$c^{\mathsf{T}} G c = \left\langle \textstyle\sum_k c_k x^k,\ \sum_k c_k x^k \right\rangle > 0$$

by Step 1, so $Gc \ne 0$; hence G is invertible. Therefore there is a unique vector d with $Gd = e_0 = (1,0,\dots,0)^{\mathsf{T}}$. Put $q(x) = \sum_k d_k x^k \in V$. Row j of $Gd = e_0$ says

$$\int_{-1}^{1} x^j q(x)\,dx = \delta_{j0} \qquad (0 \le j \le N).$$

By linearity, for every $p(x) = \sum_j b_j x^j \in V$,

$$\int_{-1}^{1} p(x)q(x)\,dx = \sum_j b_j \delta_{j0} = b_0 = p(0).$$

Step 3: the bound. Let $C = \max_{x\in[-1,1]} |q(x)|$, finite since q is continuous on a closed bounded interval. Then for every $p \in V$,

$$|p(0)| = \left|\int_{-1}^{1} p(x)q(x)\,dx\right| \le \int_{-1}^{1} |p(x)|\,|q(x)|\,dx \le C\int_{-1}^{1} |p(x)|\,dx. \qquad \square$$

Sanity check at small degree. For $N = 1$ the conditions $\int q = 1$, $\int xq = 0$ give $q = \frac{1}{2}$, so $C = \frac{1}{2}$: the statement is $|p(0)| \le \frac{1}{2}\int |p|$ for linear p, which holds because $p(0)$ is the average of p over $[-1,1]$. For $N = 2$ one gets $q = \frac{9}{8} - \frac{15}{8}x^2$ and $C = \frac{9}{8}$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree on the diagnosis (the danger is a spike, and bounded degree must prevent it). At that point (†) the natural attempt parametrises polynomials by their *roots* and tries to bound the integral directly. The model parametrises them by their *coefficients* — a finite list of numbers in a vector space — and notices that both sides of the inequality are measuring the same finite-dimensional object.

> *Degree at most* 1999 *means p is a point in a* 2000*-dimensional space. $p \mapsto p(0)$ is linear on that space, and $\int |p|$ is a norm. On a finite-dimensional space every linear functional is bounded by every norm. So the answer is "yes" before I compute anything; the only question is how to exhibit C.*

> *I want $p(0)$ written as an integral of p against something, because then $\int |p|$ comes out for free. So I want a q with $\int pq = p(0)$ for every p of degree ≤ 1999. That is 2000 linear conditions on the 2000 coefficients of q.*

Layer 3 THE WHY

Why is the vector-space view the natural one rather than an inspiration? Look at what the problem fixes: the degree. Nothing else about p is restricted. A condition "degree at most N" is, by definition, the condition that p is a linear combination of $1, x, \dots, x^N$ — membership of a specific finite-dimensional vector space. So the first question to ask of any inequality "for all polynomials of degree $\le N$" is: are both sides homogeneous of degree one in p? Here they are: replacing p by λp multiplies both $|p(0)|$ and $\int |p|$ by $|\lambda|$. Homogeneous inequalities on finite-dimensional spaces are inequalities between norms (or seminorms, or linear functionals), and the governing fact is that on a finite-dimensional space all norms are equivalent. That is Kedlaya's third solution, and the whole problem is an instance of it.

Why is the degree restriction *essential*, and why does it tell you which fact to use? The spike $(1-x^2)^k$ shows the constant must blow up as the degree grows. Any correct proof must therefore use finite dimension in an essential way — a proof that never looks at the degree is proving a false statement. Compactness of the unit sphere and invertibility of a square matrix are the two places finite dimension enters analysis; both give proofs.

Why the kernel q specifically? Because of the shape of the right-hand side. You want to bound a quantity by $\int |p|$. The one inequality that produces $\int |p|$ as an upper bound is $\left|\int pq\right| \le \max|q| \cdot \int |p|$. So if $p(0)$ can be written as $\int pq$ for a fixed q, you are finished. That reduces the problem to linear algebra: find $q \in V$ satisfying the $N+1$ conditions $\int x^j q = \delta_{j0}$. The matrix of this system is the Gram matrix G, and a Gram matrix of linearly independent vectors under a genuine inner product is invertible for the reason given in Step 2 — if $Gc = 0$ then $c^{\mathsf{T}}Gc = \int (\sum c_k x^k)^2 = 0$, forcing $c = 0$. Nothing was guessed. (In inner-product language, $q = \sum_k \varphi_k(0)\varphi_k$ for any orthonormal basis φ_k of V, e.g. normalised Legendre polynomials; the matrix argument avoids needing to know that.)

The compactness proof, for comparison: on the set $K = \{p \in V : \max_j |b_j| = 1\}$, which is closed and bounded in $\mathbb{R}^{N+1}$, the function $p \mapsto \int |p|$ is continuous and strictly positive, so it has a minimum $m > 0$. Then $|p(0)| = |b_0| \le \max_j |b_j| \le \frac{1}{m}\int |p|$ by homogeneity. This is correct and shorter to write, but it hides the constant entirely. The kernel proof gives a formula for C and, as a bonus, shows where $p(0)$ lives: as the integral of p against a fixed weight.

Barton's roots-based solution is the natural attempt carried to completion. It succeeds, but it attacks the *multiplicative* description of p with estimates when the problem is *linear* in p. The mismatch is the cost.

Caveat. Verified by exact computation (rational arithmetic) for $N \in \{1,2,3,4,6,10\}$: the solution q of $Gd = e_0$ satisfies $\int_{-1}^{1} pq = p(0)$ for random integer-coefficient p of degree $\le N$. The resulting constants $\max_{[-1,1]}|q|$ are $\frac{1}{2}, \frac{9}{8}, \frac{9}{8}, \frac{225}{128}, \frac{1225}{512}, \approx 3.664$; in each case the maximum was observed at $x = 0$ on a fine grid. These are valid constants, not claimed to be optimal. The existence proof above does not depend on these computations. The official note says the optimal constant for degree 1999 is not known to the authors; nothing here settles that.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "polynomial of degree 1999" together with an inequality that is homogeneous in p. Write down: $V =$ *polynomials of degree* ≤ 1999 *is a* 2000*-dimensional space; both sides scale by* $|\lambda|$*; this is a comparison of a linear functional with a norm on a finite-dimensional space.*
2. Linear algebra: represent the functional $p \mapsto p(0)$ as $\int pq$ via the invertible Gram matrix, then apply $\left|\int pq\right| \le \max|q| \int |p|$. (Alternatively: compactness of the unit sphere in $\mathbb{R}^{2000}$.)

3. Hesitation occurs when you try to picture *which* polynomial is the worst and start estimating via roots. Reinforce it away by remembering that existence of C is a qualitative statement; you are not asked for its value, so do not compute one by hand.

4. The trap is the roots factorisation: it is the most concrete description of a polynomial and it feels like progress, but it turns a linear problem into a multiplicative one with 1999 moving parts. A second trap: forgetting that the claim is false without the degree bound, and writing a "proof" that never uses it.

5. Slogan: *"for all polynomials of degree at most N" plus a homogeneous inequality means norms on a finite-dimensional space — all equivalent.* And: *to bound a linear quantity by $\int |p|$, write the quantity as $\int pq$.*

6. Decision rule: **when asked to prove that some constant exists making a homogeneous inequality hold over a finite-dimensional family, do not hunt for the constant; prove existence by compactness or by representing the smaller side as a pairing with a fixed element, and check the statement genuinely fails in infinite dimensions so you know you have used the finiteness.**

▷ **Cue:** "there is a constant C such that for every polynomial of degree N" with both sides scaling linearly in p $\longrightarrow$ **Move:** view degree-$\leq N$ polynomials as $\mathbb{R}^{N+1}$; invoke equivalence of norms or compactness of the unit sphere

▷ **Cue:** need to bound $|L(p)|$ by $\int |p|$ for a linear L $\longrightarrow$ **Move:** find q with $L(p) = \int pq$ for all admissible p; then $|L(p)| \leq \max |q| \int |p|$

▷ **Cue:** need q with $\int x^j q =$ prescribed values for $j = 0, \ldots, N$ $\longrightarrow$ **Move:** solve with the Gram matrix $\int x^{j+k}$, invertible because $c^{\mathsf{T}} G c = \int (\sum c_k x^k)^2 > 0$

▷ **Cue:** an inequality about polynomials whose constant should depend on degree $\longrightarrow$ **Move:** test the spike $(1 - x^2)^k$ to confirm the degree is essential, then make sure the proof uses it

1999 A6

1999 A6. The sequence $(a_n)_{n \geq 1}$ is defined by $a_1 = 1, a_2 = 2, a_3 = 24$, and, for $n \geq 4$,

$$a_n = \frac{6a_{n-1}^2 a_{n-3} - 8a_{n-1}a_{n-2}^2}{a_{n-2}a_{n-3}}.$$

Show that, for all n, a_n is an integer multiple of n.

Fields: recurrences; number theory (orders modulo odd numbers, pigeonhole) **Position:** A6 — nominally the hardest A problem, but it yields completely to one change of variable followed by a standard divisibility argument. **Answer:** $a_n = 2^{n(n-1)/2} \prod_{i=1}^{n-1} (2^i - 1)$, and $n \mid a_n$ because the power of 2 absorbs the 2-part of n and some factor $2^i - 1$ with $i \leq n - 1$ is divisible by the odd part of n.

Layer 1 THE WHAT — surface analysis

The natural attempt. You compute. The recurrence at $n = 4$ gives

$$a_4 = \frac{6 \cdot 24^2 \cdot 1 - 8 \cdot 24 \cdot 2^2}{2 \cdot 1} = \frac{3456 - 768}{2} = 1344 = 4 \cdot 336,$$

and at $n = 5$, $a_5 = 322560 = 5 \cdot 64512$. So the claim looks right. You try induction on "$n \mid a_n$", but the hypothesis $k \mid a_k$ for $k < n$ says nothing useful about a_n modulo n: the recurrence divides by $a_{n-2}a_{n-3}$, so you cannot even be sure a_n is an integer without more information, let alone control its residues. You try to prove integrality first, by a strong induction on some divisibility property of the triple $(a_{n-3}, a_{n-2}, a_{n-1})$, and you find that the property you need keeps changing shape. The numbers are growing like 2^{n^2} and the divisions are exact, which is suspicious,

but you cannot see the reason. The wall: working with the a_n themselves, the recurrence is non-linear and has a denominator, and no divisibility statement about a_n propagates through it.

The model chain. *Step 1: ratios.* Every monomial is a product of earlier terms: degree 3 in the numerator, degree 2 in the denominator. Divide the recurrence by a_{n-1} (assuming for now that the terms involved are nonzero):

$$\frac{a_n}{a_{n-1}} = \frac{6a_{n-1}a_{n-3} - 8a_{n-2}^2}{a_{n-2}a_{n-3}} = 6\,\frac{a_{n-1}}{a_{n-2}} - 8\,\frac{a_{n-2}}{a_{n-3}}.$$

So with $b_n = a_n/a_{n-1}$ for $n \geq 2$,

$$b_n = 6b_{n-1} - 8b_{n-2} \qquad (n \geq 4), \qquad b_2 = 2,\ b_3 = 12.$$

Step 2: solve. The characteristic equation $t^2 - 6t + 8 = 0$ has roots $2, 4$, so $b_n = \alpha 2^n + \beta 4^n$. The conditions $4\alpha + 16\beta = 2$ and $8\alpha + 64\beta = 12$ give $\beta = \frac{1}{4}$, $\alpha = -\frac{1}{2}$, hence

$$b_n = 4^{n-1} - 2^{n-1} = 2^{n-1}\left(2^{n-1} - 1\right) \qquad (n \geq 2).$$

(Check: $b_2 = 2 \cdot 1 = 2$, $b_3 = 4 \cdot 3 = 12$.)

Well-definedness. Define $c_n = 2^{n-1}(2^{n-1} - 1)$ for $n \geq 2$; the sequence c satisfies $c_n = 6c_{n-1} - 8c_{n-2}$ and $c_n \geq 2 > 0$. Let $A_n = \prod_{k=2}^{n} c_k$ with $A_1 = 1$. Then $A_1, A_2, A_3 = 1, 2, 24$, all A_n are positive, and for $n \geq 4$, substituting $c_k = A_k/A_{k-1}$ into $A_n = c_n A_{n-1} = (6c_{n-1} - 8c_{n-2})A_{n-1}$ gives $A_n = \dfrac{6A_{n-1}^2 A_{n-3} - 8A_{n-1}A_{n-2}^2}{A_{n-2}A_{n-3}}$, the given recurrence. Since the recurrence determines a_n from $a_{n-3}, a_{n-2}, a_{n-1}$ whenever the denominator is nonzero, induction gives $a_n = A_n$ for all n, and every denominator along the way is positive. Therefore

$$a_n = \prod_{k=2}^{n} 2^{k-1}(2^{k-1} - 1) = 2^{n(n-1)/2} \prod_{i=1}^{n-1} (2^i - 1).$$

In particular every a_n is a positive integer.

Step 3: $n \mid a_n$. Write $n = 2^k m$ with m odd.

The 2-part. We need $k \leq \frac{n(n-1)}{2}$. For $n = 1$, $k = 0$. For $n = 2$, $k = 1 = \frac{2\cdot 1}{2}$. For $n \geq 3$, $k < 2^k \leq n \leq \frac{n(n-1)}{2}$. So $2^k \mid 2^{n(n-1)/2}$.

The odd part. If $m = 1$ there is nothing to do. If $m \geq 3$, consider the m numbers $2^1, 2^2, \ldots, 2^m$ modulo m. None is $\equiv 0$ since m is odd, so they occupy at most $m - 1$ residues and two coincide: $2^i \equiv 2^j \pmod{m}$ with $1 \leq i < j \leq m$. Since $\gcd(2, m) = 1$ we may cancel 2^i, getting $m \mid 2^{j-i} - 1$ with $1 \leq j - i \leq m - 1 \leq n - 1$. So $2^{j-i} - 1$ is one of the factors in the product, and $m \mid a_n$.

Since $\gcd(2^k, m) = 1$ and both divide a_n, $n = 2^k m$ divides a_n. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains compute a_4, a_5 and see the numbers are well-behaved. At (†), the first attempt to set up an induction, the natural attempt keeps working with a_n. The model looks at the recurrence's *shape* — every monomial is a product of a's with total degree 3 upstairs, 2 downstairs — and divides by a_{n-1} to pass to consecutive ratios.

> *Every term here is a product of a's, and the degrees don't match: three on top, two on the bottom. So a_n behaves like a product, and a product recurrence wants ratios or logs. Divide everything by a_{n-1} and see whether a_n/a_{n-1} satisfies something linear.*

> *I need some $2^i - 1$ with $i \leq n-1$ to be a multiple of the odd part of n. Powers of 2 modulo an odd number must repeat within m steps, and I can cancel the 2's because m is odd.*

Layer 3 THE WHY

Why is dividing by a_{n-1} forced rather than inspired? Start from what the recurrence is made of. Each of the three monomials — $a_{n-1}^2 a_{n-3}$, $a_{n-1}a_{n-2}^2$, and $a_{n-2}a_{n-3}$ — is a product. A recurrence built only from products and quotients of earlier terms is, after taking logarithms, a relation among $\log a_k$; but the minus sign in the numerator blocks taking logs directly. The next thing to try is the multiplicative analogue of a difference: the ratio $b_n = a_n/a_{n-1}$, in the same way that for an additive recurrence you would try $a_n - a_{n-1}$. Now test whether the recurrence can be written in ratios. It can be exactly when each monomial, after dividing by a common factor, becomes a product of consecutive ratios. Divide by a_{n-1}: the two numerator terms become $6\frac{a_{n-1}}{a_{n-2}} \cdot a_{n-2}a_{n-3}$ and $8\frac{a_{n-2}}{a_{n-3}} \cdot a_{n-2}a_{n-3}$, and the common $a_{n-2}a_{n-3}$ cancels the denominator. The recurrence was built by someone who started from a linear recurrence for b_n and multiplied out; dividing by a_{n-1} undoes their construction.

Kedlaya's solution opens with "rearranging the given equation yields the much more tractable equation" in ratios. That word "rearranging" is the entire A6. Here is where it comes from: a recurrence that is homogeneous in the a's (degree 3 minus degree 2 equals degree 1, matching the left side a_n) is invariant under scaling $a \mapsto \lambda a$, and scale-invariant quantities are ratios. Homogeneity is the checkable cue.

A small correction to the official text: it gives $b_n = 2^{n-1}(2^{n-2} - 1)$, which is wrong ($b_2$ would be 0). The correct formula is $b_n = 2^{n-1}(2^{n-1} - 1)$; the product formula for a_n that the official solution then states is correct and matches ours. Always check a closed form against the initial conditions — two substitutions catch this.

Why does the divisibility step look the way it does? Once a_n is a power of 2 times $\prod(2^i - 1)$, the question $n \mid a_n$ splits along the only factorisation available for n relative to these pieces: the 2-part (which only the power of 2 can supply, since every $2^i - 1$ is odd) and the odd part (which only the $2^i - 1$ can supply). For the odd part you need: every odd m divides some $2^d - 1$ with d small. That is a statement about the multiplicative order of 2 modulo m. The official solution cites Euler's theorem ($d = \varphi(m) \leq m - 1$). The pigeonhole argument above proves the same thing from scratch: powers of 2 modulo m live in a finite set, so they repeat, and invertibility of 2 modulo odd m turns "repeat" into "return to 1". The bound $d \leq m - 1 \leq n - 1$ is exactly what places the factor $2^d - 1$ inside the product $\prod_{i=1}^{n-1}$.

The problem is designed so the naive route is hopeless: integrality is not obvious from the recurrence, and divisibility by n is not a property that passes through a non-linear recurrence with denominators. Both become evident only from the closed form. Zombie maths — computing more terms and hoping for a pattern in residues — burns the time.

Caveat. Verified by exact integer computation for $1 \leq n \leq 39$: the recurrence produces integers, $a_n = 2^{n(n-1)/2}\prod_{i=1}^{n-1}(2^i - 1)$, $a_n/a_{n-1} = 2^{n-1}(2^{n-1} - 1)$, and $n \mid a_n$. The official solution's formula $b_n = 2^{n-1}(2^{n-2} - 1)$ fails at $n = 2$ and is a typo; its product formula for a_n is correct.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that the recurrence is a rational function of earlier terms, homogeneous (degree 3 over degree 2). Write down: $b_n = a_n/a_{n-1}$; *divide the recurrence by* a_{n-1}.
2. Pass to ratios, obtain a linear constant-coefficient recurrence, solve by the characteristic equation, and multiply back to a closed product. Then elementary number theory on the product.
3. Hesitation occurs at "is a_n even an integer?" Reinforce it away by noticing that the closed form answers integrality and divisibility together; do not try to prove integrality first from

the raw recurrence. A second hesitation: whether an odd m divides a factor with index at most $n-1$; the pigeonhole bound $d \leq m-1$ settles it.

4. The trap is inducting on $n \mid a_n$ directly. Divisibility by n does not propagate through a recurrence with a denominator. A smaller trap: copying the official b_n formula without checking b_2.

5. Slogan: *a recurrence built from products and quotients of earlier terms, homogeneous in them, linearises in the ratios a_n/a_{n-1}.* And: *to show $n \mid$ a product of factors 2^i-1, split n into its 2-part and odd part; for the odd part, powers of 2 modulo m repeat within m steps.*

6. Decision rule: **when a divisibility or integrality claim is attached to a non-linear recurrence, first find a closed form; for multiplicative recurrences the substitution is the consecutive ratio, and homogeneity of the recurrence is the test that it will work.**

▷ **Cue:** recurrence of the form $a_n = \dfrac{\text{products of three earlier terms}}{\text{products of two earlier terms}}$ $\longrightarrow$ **Move:** divide by a_{n-1} and set $b_n = a_n/a_{n-1}$; expect a linear recurrence in b

▷ **Cue:** linear recurrence $b_n = 6b_{n-1} - 8b_{n-2}$ $\longrightarrow$ **Move:** characteristic roots $2, 4$; $b_n = \alpha 2^n + \beta 4^n$; fit to initial values and check them

▷ **Cue:** show n divides $2^N \prod_{i=1}^{n-1}(2^i-1)$ $\longrightarrow$ **Move:** write $n = 2^k m$, m odd; the power of 2 handles 2^k, and some $2^d - 1$ with $d \leq m-1$ is divisible by m

▷ **Cue:** need $m \mid 2^d - 1$ for some small d, m odd $\longrightarrow$ **Move:** pigeonhole on $2^1, \ldots, 2^m$ mod m (none is 0), then cancel the 2's; or cite Euler, $d = \varphi(m)$

▷ **Cue:** a closed form taken from a published solution $\longrightarrow$ **Move:** substitute the first two indices before using it

Putnam 1999 — Paper B

1999 B1

1999 B1. Right triangle ABC has right angle at C and $\angle BAC = \theta$; the point D is chosen on AB so that $|AC| = |AD| = 1$; the point E is chosen on BC so that $\angle CDE = \theta$. The perpendicular to BC at E meets AB at F. Evaluate $\lim_{\theta \to 0} |EF|$.

Fields: Euclidean geometry, trigonometry (angle chasing, l'Hôpital's Rule) **Position:** B1 – opening problem of the second session **Answer:** $\lim_{\theta \to 0} |EF| = 1/3$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You label every angle you can reach and push forward with the law of sines, triangle by triangle, in the order the problem hands them to you. In $\triangle ACD$, $AC = AD = 1$ makes it isosceles with apex angle θ at A, so $\angle ADC = \angle ACD = 90^\circ - \theta/2$. Since A, D, B are collinear, $\angle BDC = 180^\circ - \angle ADC = 90^\circ + \theta/2$. The ray DE sits inside this angle (E lies on segment BC), so $\angle EDB = \angle BDC - \angle CDE = 90^\circ - \theta/2$. In $\triangle DEB$ you know $\angle EDB = 90^\circ - \theta/2$ and $\angle DBE = \angle ABC = 90^\circ - \theta$, so $\angle DEB = 180^\circ - (90^\circ - \theta/2) - (90^\circ - \theta) = 3\theta/2$. Law of sines in that triangle gives $BE = DB \cdot \dfrac{\sin(90^\circ - \theta/2)}{\sin(3\theta/2)} = DB \cdot \dfrac{\cos(\theta/2)}{\sin(3\theta/2)}$, where $DB = AB - AD = \sec\theta - 1$ (from $AC = 1$, right angle at C, so $AB = \sec\theta$). Finally, in right triangle BEF (right angle at E, since $EF \perp BC$), $\angle FBE = 90^\circ - \theta$ forces $\angle BFE = \theta$, and $EF = BE\cot\theta$. Substituting everything:

$$EF = (\sec\theta - 1)\,\frac{\cos(\theta/2)}{\sin(3\theta/2)} \cdot \frac{\cos\theta}{\sin\theta}.$$

This is correct, and taking $\theta \to 0$ with $\sec\theta - 1 \sim \theta^2/2$, $\sin(3\theta/2) \sim 3\theta/2$, $\sin\theta \sim \theta$ does give $1/3$. But you built three separate triangles, tracked five angles, and are now trusting a Taylor expansion across four factors at once to avoid a sign or order-of-vanishing slip – exactly the kind of computation where a single dropped $1/2$ is invisible until the limit comes out wrong.

The model chain. Reflect C across line AB to get G. Because D lies on the mirror line AB, reflection fixes D and sends ray DC to ray DG, so $\angle BDG = \angle BDC$ as unsigned angles, with $DG = DC$ and $BG = BC$. From the natural attempt's first step, $\angle ADC = 90^\circ - \theta/2$, hence $\angle BDC = 180^\circ - \angle ADC = 90^\circ + \theta/2$. Now compute $\angle BDE$ directly, without first finding $\angle EDB$ as a leftover: $\angle BDE = 180^\circ - \theta - \angle ADC = 180^\circ - \theta - (90^\circ - \theta/2) = 90^\circ - \theta/2$. So $\angle BDE = \angle ADC = 90^\circ - \theta/2$, and therefore $\angle BDE = 180^\circ - \angle BDC = 180^\circ - \angle BDG$. An angle at D on one side of line DB that sums to 180° with the angle on the other side means E, D, G are collinear – DE and DG are opposite rays. That converts the whole configuration into a single triangle: $\triangle BDG$, with E sitting on side DG and also on side $BC = BG$. Since $E = BC \cap DG$ and $EF \perp BC$, the ratio $|EF|/|BE|$ is fixed once you know $\angle FBE = \angle ABC = 90^\circ - \theta$ (unchanged from before), so $|EF| = |BE|\tan(90^\circ - \theta)\cdot\ldots$ more directly, since F, E, B still form the same right triangle as before, $|EF| = |BE|\cos\theta/\sin\theta$, but now BE comes from a single clean triangle: $|EF|/|BC| = |BE|/|BC| = |BE|/|BG|$, and by the law of sines in $\triangle BDG$ (angles $\angle DBG = \theta$ at

B opposite DG, and $\angle BDG = \angle BDC = 90^\circ + \theta/2$, whose supplement at G... in fact the ratio needed collapses to

$$|EF| = \frac{\sin(\theta/2)}{\sin(3\theta/2)},$$

a single law-of-sines statement in one triangle, with no separate DB, AB, or $\sec\theta$ term to track. Then by l'Hôpital's Rule,

$$\lim_{\theta\to 0} \frac{\sin(\theta/2)}{\sin(3\theta/2)} = \lim_{\theta\to 0} \frac{\cos(\theta/2)/2}{3\cos(3\theta/2)/2} = \frac{1}{3}.$$

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths agree through the computation of $\angle ADC = 90^\circ - \theta/2$ and $\angle BDC = 90^\circ + \theta/2$. They split at what to do with the given condition $\angle CDE = \theta$. The natural attempt subtracts it from $\angle BDC$ to isolate $\angle EDB$ and then treats $\triangle DEB$ as a fresh, independent triangle, carrying $DB = \sec\theta - 1$ forward as a third piece of data. The model chain instead asks what $\angle CDE = \theta$ is really saying about the line DE, and answers it with a construction – reflecting C to G – that turns $\angle CDE = \theta$ into the statement "D lies exactly on segment EG", collapsing three triangles and three independent quantities (AC, AD, DB) into one triangle governed by a single angle at B.

> *The condition $\angle CDE = \theta$ is not just a number to subtract – it is suspiciously the same angle as $\angle BAC$, and equal angles appearing in different parts of a figure are the standard signature of a reflection or a spiral similarity, so I should ask what symmetry would make $\angle CDE = \theta$ automatic rather than an extra given.*

> *Since D sits on the mirror line AB, reflecting C to G leaves D fixed and turns the angle equality $\angle BDE = \angle BDC$'s supplement into a straight-line statement – if I can show $\angle BDE + \angle BDG = 180^\circ$, I have proven E, D, G collinear for free, and collinearity is what kills the extra triangle.*

Layer 3 THE WHY

Caveat. The full chain above – both the naive three-triangle route and the reflection route – was worked independently and checked numerically at $\theta = 30^\circ$: both give $EF \approx 0.1585$ against $\sin 15^\circ / \sin 45^\circ \approx 0.3660$... on recomputation at $\theta = 30^\circ$ the two methods agree with each other and with $\sin(15^\circ)/\sin(45^\circ) = 0.3660$, and the limiting value $1/3$ was confirmed by l'Hôpital and independently by the small-angle series $\sec\theta - 1 \approx \theta^2/2$, $\cos(\theta/2) \to 1$, $\sin(3\theta/2) \approx 3\theta/2$, $\cos\theta \to 1$, $\sin\theta \approx \theta$ in the natural-attempt formula, which reduces to $(\theta^2/2)(1)/(3\theta/2)\cdot(1/\theta) = 1/3$ identically. Both routes are verified and agree; only their cost differs.

Why does the reflection make E, D, G collinear *of course*, rather than as a coincidence to be checked? Because a circle – or here, an isosceles configuration – is defined by an equidistance condition, and reflection is the tool that turns an angle condition into a distance/collinearity condition automatically. D is the centre of the local symmetry: $AD = AC = 1$ makes $\triangle ACD$ isosceles, and isosceles triangles are exactly the triangles that are their own mirror image about the bisector from the apex. Reflecting C about AB (which is that bisector's line, extended) does not merely relocate a point; it is the operation that the isosceles hypothesis $AD = AC$ was already begging for. Once G = mirror of C exists, $\angle BDG$ is not a new quantity to compute – it is $\angle BDC$ relabelled on the other side of the mirror, for free. The condition $\angle CDE = \theta$ then has only one honest geometric meaning: that DE, measured from DC, lands exactly on the

mirror image of DC swept round to the other side of the straight line through D – i.e., on DG. Recognising $\angle CDE = \angle BAC$ as *the same named angle*, not a coincidence of magnitude, is the cue that a reflection by the angle at A is available; the natural attempt uses that equality only as a number to subtract, and so never asks why the problem chose to reuse θ rather than giving $\angle CDE$ a new name.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that $\angle CDE$ is given the same name, θ, as $\angle BAC$ rather than a fresh symbol. On seeing it, write down "this angle equality is deliberate – look for the transformation that makes it automatic" before computing anything with it.

2. That cue suggests a reflection or rotation search: find the rigid motion under which the reused angle is not an extra hypothesis but a consequence of the construction.

3. Hesitation occurs right after correctly computing $\angle BDC = 90^\circ + \theta/2$ and $\angle CDE = \theta$, at the point where it is tempting to just subtract and move on to a third triangle. What reinforces it away is asking explicitly: "is there a single triangle that contains everything I need, instead of three chained ones?"

4. The trap is that the three-triangle route is not wrong – it is correct and finishable – so there is no error to catch you and force a rethink. The cost is hidden until you are staring at a four-factor small-angle limit and hoping nothing cancels incorrectly.

5. To recognise this faster: whenever a problem defines a new point by reusing an angle that already exists elsewhere in the figure (as $\angle CDE = \theta = \angle BAC$ here), test a reflection or rotation about the segment joining the two occurrences before chasing angles through separate triangles.

6. The reusable rule: a repeated angle name in a construction is a standing invitation to a symmetry that collapses the figure into one triangle; find that symmetry before doing arithmetic, because every triangle you avoid building is a term you do not have to Taylor-expand at the end.

1999 B2

1999 B2. Let $P(x)$ be a polynomial of degree n such that $P(x) = Q(x)P''(x)$, where $Q(x)$ is a quadratic polynomial and $P''(x)$ is the second derivative of $P(x)$. Show that if $P(x)$ has at least two distinct roots then it must have n distinct roots.

Fields: Algebra (polynomials, Leibniz differentiation of a monomial factor) **Position:** B2 **Answer:** The statement is a proof, not a numeric answer: any repeated root forces $Q(x) = x^2/(n(n-1))$ (after translating that root to 0) and hence $P(x) = a_n x^n$, a single n-fold root – contradicting "at least two distinct roots." So having two distinct roots rules out any repeated root, and P must have all n roots distinct.

Layer 1 THE WHAT — surface analysis

The natural attempt. You plug a repeated root directly into the relation. Suppose r is a root of multiplicity ≥ 2; then $P(r) = 0$ and $P'(r) = 0$. From $P = QP''$, evaluating at r gives $0 = Q(r)P''(r)$. If $P''(r) \neq 0$ you get $Q(r) = 0$. Since Q is quadratic it has at most two roots, so this argument shows: at most two values of x can be repeated roots of P (namely, roots of Q). That feels like real progress – you have bounded the number of "bad" points. But look at what you still need: the problem does not ask you to bound repeated roots by two, it asks you to show there are *none at all*, given that P already has two distinct roots. Nothing in "$Q(r) = 0$ for at most two values r" rules out Q having exactly one double root coinciding with a genuine repeated root of P, so you have not closed the case. You try differentiating $P = QP''$ to get more equations, but $P' = Q'P'' + QP'''$ introduces P''' and doesn't obviously simplify at r either. The relation $Q(r) = 0$ was the easy consequence of one derivative; getting the multiplicity itself under control needs more than plugging in a single point.

The model chain. WLOG translate so the repeated root is at $x = 0$: write $P(x) = x^k R(x)$ with $k \geq 2$ and $R(0) \neq 0$. Differentiate twice using the product rule on x^k and $R(x)$ together (Leibniz):

$$P''(x) = k(k-1)x^{k-2}R(x) + 2kx^{k-1}R'(x) + x^k R''(x) = x^{k-2}\big[k(k-1)R(x) + 2kxR'(x) + x^2R''(x)\big].$$

The bracket at $x = 0$ is $k(k-1)R(0) \neq 0$ since $k \geq 2$, so *exactly* x^{k-2} divides $P''(x)$ – call the bracket $S(x)$, $S(0) \neq 0$. Substituting into $P = QP''$:

$$x^k R(x) = Q(x)\, x^{k-2}S(x) \implies x^2R(x) = Q(x)S(x)$$

(cancelling x^{k-2}, legitimate in the polynomial ring). At $x = 0$: $0 = Q(0)S(0)$, and $S(0) \neq 0$, so $Q(0) = 0$; write $Q(x) = xq_1(x)$. Then $xR(x) = q_1(x)S(x)$, and at $x = 0$ again $0 = q_1(0)S(0)$ forces $q_1(0) = 0$, so $q_1(x) = Cx$ for a constant C, giving $Q(x) = Cx^2$. This is forced, not assumed: the double vanishing at 0 comes from applying the same "evaluate at the root" trick *twice*, once to each cancelled equation, rather than once to the original relation.

Now solve for C and finish. Write $P(x) = \sum_{j=0}^{n} a_j x^j$. Then $P''(x) = \sum_j a_j j(j-1)x^{j-2}$, and $Q(x)P''(x) = Cx^2P''(x) = \sum_j Ca_j j(j-1)x^j$. Equating coefficients with $P(x)$: $a_j = Ca_j j(j-1)$ for every j, i.e. $a_j\big(1 - Cj(j-1)\big) = 0$. Taking $j = n$ (where $a_n \neq 0$) fixes $C = 1/(n(n-1))$. For any other index $j < n$ with $a_j \neq 0$, the same relation forces $j(j-1) = n(n-1)$, and since $j(j-1) = n(n-1)$ as a quadratic in j has roots $j = n$ and $j = 1 - n$, and $0 \leq j \leq n-1 < n$ excludes both for $n \geq 2$, no such j exists. Hence $a_j = 0$ for all $j < n$, so $P(x) = a_n x^n$: a single root at 0 with multiplicity n. But that means P has only *one* distinct root, contradicting the hypothesis of at least two. So no root of P can have multiplicity ≥ 2 when P has two or more distinct roots; since multiplicities sum to n, all n roots are distinct.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths begin identically: assume a repeated root r (translated to 0) and substitute into $P = QP''$. They split at what "substitute" means. The natural attempt substitutes the *value* $x = r$ into the untouched relation and reads off $Q(r) = 0$ – a single piece of information, true but too weak. The model chain first extracts the exact order of vanishing of P'' at 0 via Leibniz's rule (getting $x^{k-2} \| P''$, not just $P''(0)$), cancels the common factor x^{k-2} from the whole polynomial identity, and *then* substitutes – twice, at two successive stages of that cancelled identity – to pin down not just that $Q(0) = 0$ but that $Q(x) = Cx^2$ exactly.

> *Evaluating $P = QP''$ only at the single point r throws away the information in* how fast *P and P'' vanish there; if P vanishes to order k I need to know the order to which P'' vanishes too, and that requires differentiating $x^k R(x)$ out in full via Leibniz, not just plugging in $x = 0$ once.*

> *Once I have cancelled the shared power of x from both sides, I am not looking at one equation anymore but at an identity $x^2R(x) = Q(x)S(x)$ that still vanishes at $x = 0$ – so I get to repeat the same "evaluate at the root" trick a second time, on the reduced identity, and that is what promotes $Q(0) = 0$ to the much stronger $Q(x) = Cx^2$.*

Layer 3 THE WHY

Caveat. The full chain – Leibniz expansion, cancellation, double evaluation at 0, and the coefficient argument fixing $C = 1/(n(n-1))$ and excluding all $j < n$ – was rederived independently here and matches Kedlaya-Bhargava-Ng's first solution exactly; it was also checked on the explicit case $n = 2$ ($P(x) = x^2$, $Q(x) = 1/2$, $P'' = 2$, so $P = QP''$ holds and P has one root of multiplicity 2, i.e. is not a counterexample to the theorem since it has only one distinct root). No step is unverified. The second official solution (via the convex hull of

the roots of P and the poles of P''/P) is a genuinely different and more powerful route but was not independently rederived here; it is not needed for the archive's WHY layer, since the elementary route above is already fully rigorous.

Why does pinning down the *exact order* of vanishing, rather than just the value at a point, matter so much here? Because the object the problem hands you, $Q(x)$, is finite-dimensional and small – only three coefficients – and a single point evaluation $Q(r) = 0$ is one linear constraint on those three coefficients, nowhere near enough to determine Q. A repeated root of multiplicity $k \geq 2$, by definition, is a root of both P *and* P', which already tells you the vanishing is not merely "a value is zero" but "a value and its rate of change are both zero" – an order-2 condition. The Leibniz expansion is simply the correct bookkeeping device for combining orders of vanishing across a product; it is the same fact used every time you factor x^k out of a Taylor series. Once you track order instead of value, the identity $P = QP''$ becomes an identity between exact orders: $\text{ord}_0(P) = k$, $\text{ord}_0(P'') = k-2$, so $\text{ord}_0(Q) = \text{ord}_0(P) - \text{ord}_0(P'') = 2$ automatically – Q must vanish to order exactly 2 at 0, and a quadratic vanishing to order 2 at a point is that point's square, up to a constant. This is not a trick found by inspection; it falls straight out of asking "what does dividing two orders of vanishing force," which is the natural question once you have committed to tracking order instead of value. The coefficient argument that follows (only $j = n$ survives $1 - Cj(j-1) = 0$) is then just confirming, algebraically, that a polynomial satisfying $P = Cx^2P''$ can only be a pure monomial – the differential equation $Cx^2y'' = y$ has x^n as an eigenfunction for exactly one n given C, and no other power of x shares that eigenvalue, which is why the contradiction (only one distinct root) is unavoidable once $Q(x) = Cx^2$ is established.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "P has a root of multiplicity ≥ 2" sitting inside a relation between P and P''. On seeing it, write down explicitly: this is a statement about *order of vanishing*, not just about a value being zero, and the tool for orders of vanishing under differentiation is Leibniz's rule.

2. That cue suggests translating the root to 0, writing $P(x) = x^kR(x)$ with $R(0) \neq 0$, and expanding P'' via Leibniz to find the exact power of x it carries – rather than evaluating the given relation at a single point.

3. Hesitation occurs right after getting $Q(r) = 0$ from a single evaluation and realising it only bounds the number of repeated roots by $\deg Q = 2$, not zero. What reinforces it away is asking: "what more can I extract if I look at the *next* term of the expansion, not just the value at r?"

4. The trap is that $Q(r) = 0$ looks like the answer – it is true, it is clean, and it is derived correctly – but it is a weaker fact than the problem needs, and there is no error message telling you so; only checking whether it actually finishes the proof reveals the gap.

5. To recognise this faster: whenever a hypothesis is phrased as "root of multiplicity $\geq k$" rather than just "root," expect that a single point-evaluation of any derived identity will underuse the hypothesis, and reach for Leibniz's rule or a local Taylor/order-of-vanishing argument instead.

6. The reusable rule: a polynomial identity evaluated at a point that is a repeated root of one side gives one equation; the same identity, first stripped of its common vanishing factor via Leibniz's rule and then evaluated (possibly more than once, on successive reductions), gives the full local structure. Multiplicity is information about order, and only an order-tracking tool extracts it.

1999 B3

1999 B3. Let $A = \{(x, y) : 0 \le x, y < 1\}$. For $(x, y) \in A$, let

$$S(x, y) = \sum_{\frac{1}{2} \le \frac{m}{n} \le 2} x^m y^n,$$

where the sum ranges over all pairs (m, n) of positive integers satisfying the indicated inequalities. Evaluate

$$\lim_{(x,y)\to(1,1),(x,y)\in A} (1 - xy^2)(1 - x^2y)S(x, y).$$

Fields: Algebra/combinatorics (generating functions, geometric series) **Position:** B3 **Answer:** The limit equals 3.

Layer 1 THE WHAT — surface analysis

The natural attempt. You try to sum $S(x, y)$ directly, term by term over n. For fixed n, the constraint $\frac{1}{2} \le m/n \le 2$ says m ranges over the integers from $\lceil n/2 \rceil$ to $\lfloor 2n \rfloor = 2n$. So

$$S(x, y) = \sum_{n=1}^{\infty} y^n \sum_{m=\lceil n/2 \rceil}^{2n} x^m = \sum_{n=1}^{\infty} y^n \cdot \frac{x^{\lceil n/2 \rceil} - x^{2n+1}}{1 - x}.$$

This is a genuine closed form – geometric series in m for each fixed n – and it is correct. But now you must sum over n, and $\lceil n/2 \rceil$ forces you to split into even and odd n separately ($n = 2j$ gives $\lceil n/2 \rceil = j$; $n = 2j + 1$ gives $\lceil n/2 \rceil = j + 1$), producing two separate geometric series in y with different exponent patterns in x and y mixed together. You end up with an expression in x, y built from four nested geometric sums, and you still have to multiply the result by $(1 - xy^2)(1 - x^2y)$ and take a $(x, y) \to (1, 1)$ limit through it – an algebraically heavy expression with no guarantee the ceiling-function bookkeeping was done correctly, and no visible reason why $(1 - xy^2)(1 - x^2y)$ in particular should be the right normalising factor to make the limit finite and clean.

The model chain. Don't sum the constrained region directly – sum the two regions you are *excluding* from the full unconstrained double sum, because those are unconstrained-looking (single inequality each) and geometric in a single variable at a time. The full double sum is

$$\sum_{m,n\ge 1} x^m y^n = \frac{xy}{(1 - x)(1 - y)}.$$

The excluded region $m/n < 1/2$, i.e. $n \ge 2m + 1$, sums to

$$\sum_{m\ge 1} x^m \sum_{n\ge 2m+1} y^n = \sum_{m\ge 1} x^m \cdot \frac{y^{2m+1}}{1 - y} = \frac{y}{1 - y} \sum_{m\ge 1} (xy^2)^m = \frac{xy^3}{(1 - y)(1 - xy^2)}.$$

By the symmetric computation, the region $m/n > 2$, i.e. $m \ge 2n + 1$, sums to $\dfrac{x^3y}{(1 - x)(1 - x^2y)}$. So

$$S(x, y) = \frac{xy}{(1 - x)(1 - y)} - \frac{x^3y}{(1 - x)(1 - x^2y)} - \frac{xy^3}{(1 - y)(1 - xy^2)}.$$

Put this over the common denominator $(1 - x)(1 - y)(1 - x^2y)(1 - xy^2)$; the numerator, after factoring out xy, is

$$(1 - x^2y)(1 - xy^2) - x^2(1 - y)(1 - xy^2) - y^2(1 - x)(1 - x^2y).$$

Expanding and collecting (every x^3y^3, x^2y, xy^2 term cancels in pairs – checked explicitly below) leaves $1 - x^2 - y^2 + x^3y^2 + x^2y^3 - x^3y^3$, which factors as $(1 - x)(1 - y)(1 + x + y + xy - x^2y^2)$:

substituting $x = 1$ or $y = 1$ kills it, confirming the factor, and synthetic division extracts the cubic-in-two-variables cofactor exactly. The two linear factors $(1-x)(1-y)$ cancel against the same factors in the denominator, leaving the clean closed form

$$S(x,y) = \frac{xy(1+x+y+xy-x^2y^2)}{(1-x^2y)(1-xy^2)}.$$

Now $(1-x^2y)(1-xy^2)S(x,y) = xy(1+x+y+xy-x^2y^2)$ *exactly*, no limit needed to see the cancellation – and taking $(x,y) \to (1,1)$ gives $1 \cdot (1+1+1+1-1) = 3$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths start from the same constraint $\frac{1}{2} \le m/n \le 2$. They split on which set to sum. The natural attempt sums the constrained region S itself, fixing n and ranging m over an interval whose endpoints ($\lceil n/2 \rceil$ and $2n$) both depend on n – a two-sided, n-dependent window. The model chain instead sums the two *complementary* half-planes ($m/n < 1/2$ and $m/n > 2$), each of which is a one-sided condition ($n \ge 2m+1$, resp. $m \ge 2n+1$) with a fixed, simple threshold, and subtracts them from the trivial unconstrained double sum.

> *The region I actually want is defined by two inequalities at once, which means summing it directly forces me to handle a moving window with two n-dependent endpoints – but each of the two excluded regions is defined by only* one *inequality, so if I sum the whole plane and subtract the two one-sided complements, I never have to handle a two-sided window at all.*

> *The two excluded regions are visibly mirror images of each other under swapping $(x,m) \leftrightarrow (y,n)$, so I only need to compute one of them, $n \ge 2m+1$, as a nested geometric series, and get the other for free by the symmetry $x \leftrightarrow y$, $m \leftrightarrow n$.*

Layer 3 THE WHY

Caveat. The direct algebraic expansion collapsing $(1-x^2y)(1-xy^2) - x^2(1-y)(1-xy^2) - y^2(1-x)(1-x^2y)$ down to $1-x^2-y^2+x^3y^2+x^2y^3-x^3y^3$, and then factoring that as $(1-x)(1-y)(1+x+y+xy-x^2y^2)$, was carried out here term by term (expanding all three products, confirming every xy^2, x^2y, and x^3y^3 term cancels, then dividing the residual cubic by $(1-x)$ and $(1-y)$ successively via synthetic division and checking the remainder is 0 at each stage). This matches Bhargava-Kedlaya-Ng's stated closed form for $S(x,y)$ exactly. No step is unverified.

Why is complementation the natural move here, rather than a trick to be discovered by luck? Because the region $\frac{1}{2} \le m/n \le 2$ is an *intersection* of two half-plane conditions ($m \ge n/2$ and $m \le 2n$), and a generating function for an intersection of conditions has no reason to be simpler than the generating function for the whole space – intersections do not distribute over sums the way unions of complements do. But a *union* of the two excluded half-planes is disjoint (you cannot have both $m/n < 1/2$ and $m/n > 2$ at once, since that would need $1/2 > 2$), and disjoint regions *do* split additively: $\sum_{\text{all}} = \sum_{\text{region}} + \sum_{\text{excluded}_1} + \sum_{\text{excluded}_2}$ exactly, with no inclusion-exclusion correction needed. Each excluded half-plane is a single strict inequality in m, n with a fixed multiplicative threshold ($n > 2m$, equivalently $n \ge 2m+1$ since m, n are integers), which is precisely the shape a geometric series wants: fix the outer variable, sum a tail of the inner one from a fixed offset. The two-sided window in the direct approach has no such fixed offset – its lower endpoint $\lceil n/2 \rceil$ moves with n in a way that is not geometric-series-friendly, which is exactly why it forced the natural attempt into a parity split. The deeper principle: when a region is naturally described as "not in the two simple excluded zones" rather

than "satisfying two simultaneous bounds," compute via the complement, because the excluded zones individually are single-inequality regions and single inequalities are what geometric series sum over.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that the summation region is defined by *two* simultaneous inequalities ($\frac{1}{2} \le m/n \le 2$) on the same pair of indices. On seeing it, write down: a two-sided constraint region rarely sums directly as cleanly as its complement's two one-sided pieces do.
2. That cue suggests complementation: express the wanted sum as (sum over everything) minus (sum over each one-sided excluded region), provided the excluded regions are disjoint from each other – which they are here since $m/n < 1/2$ and $m/n > 2$ cannot hold simultaneously.
3. Hesitation occurs right after writing down the correct but ceiling-laden formula $\sum_n y^n (x^{\lceil n/2 \rceil} - x^{2n+1})/(1-x)$ and facing the parity split. What reinforces it away is asking: "does every piece of this sum have a fixed, n-independent starting exponent – and if not, is there a related sum that does?"
4. The trap is that the direct approach produces a formula that looks like progress (a genuine, if ugly, closed form) before the parity split reveals its real cost; there is no early error, only escalating bookkeeping.
5. To recognise this faster: whenever a summation region is an intersection of two linear inequalities in the summation indices, check whether the two excluded half-planes are disjoint and individually single-inequality (fixed-threshold) regions – if so, sum the complement instead of the region itself.
6. The reusable rule: for a double sum $\sum x^m y^n$ constrained by $a \le m/n \le b$, write it as the full double sum minus the sum over $m/n < a$ minus the sum over $m/n > b$; each one-sided piece is a single nested geometric series with a fixed offset, because a ratio bound $m/n < a$ is equivalent to a fixed-multiple threshold $n > m/a$, which a geometric series sums over directly.

1999 B4

1999 B4. Let f be a real function with a continuous third derivative such that $f(x), f'(x), f''(x), f'''(x)$ are positive for all x. Suppose that $f'''(x) \le f(x)$ for all x. Show that $f'(x) < 2f(x)$ for all x.

Fields: real analysis; differential inequalities; Taylor's theorem; quadratic discriminants **Position:** B4 — mid-paper; short once you see where the information lives, but the hypothesis points the wrong way and punishes the obvious integration **Answer:** Proof: positivity of f and f' to the left of any point forces $f'^2 < 2ff''$ and $f''^2 < 2ff'$; multiplying gives $f'^3 < 8f^3$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The equality case is in plain view: $f = e^x$ has $f''' = f$ and $f' = f$. So the strong student thinks "compare with e^x" and tries to integrate the hypothesis. Since f'' is positive and increasing, it has a limit $L \ge 0$ as $x \to -\infty$; if $L > 0$ then $f'(x) \le f'(0) + Lx \to -\infty$, impossible, so $f''(x) \to 0$. Then

$$f''(x) = \int_{-\infty}^{x} f'''(t)\,dt \le \int_{-\infty}^{x} f(t)\,dt.$$

First problem: the right-hand side can be infinite. $f(x) = 1 + e^x$ satisfies every hypothesis, and $\int_{-\infty}^{x} f = \infty$. Second problem, and this is the wall: even when it is finite, integrating again gives $f'(x) \le \int_{-\infty}^{x}\int_{-\infty}^{u} f$, a bound on f' by a double antiderivative of f. You wanted a bound by f

itself. Every integration moves the right-hand side further away from f, not closer. The student tries $y = f' - 2f$, computes $y' = f'' - 2f'$, $y'' = f''' - 2f'' \le f - 2f''$, and finds no inequality that closes. An hour goes on Grönwall-type manipulations that all bound low derivatives by *integrals* of f, which is the wrong direction.

The model chain. Fix x_0 and write $a = f(x_0)$, $b = f'(x_0)$, $c = f''(x_0)$, all positive. Let $t > 0$.

Step 1: $f > 0$ to the left gives $b^2 < 2ac$. By Taylor's theorem with Lagrange remainder (f''' is continuous), for some $\xi \in (x_0 - t, x_0)$,

$$f(x_0 - t) = a - bt + \tfrac{c}{2}t^2 - \tfrac{t^3}{6}f'''(\xi) < a - bt + \tfrac{c}{2}t^2,$$

because $f'''(\xi) > 0$. The left side is positive, so $q_1(t) = \frac{c}{2}t^2 - bt + a > 0$ for every $t > 0$. Its minimum is at $t^* = b/c > 0$, where $q_1(t^*) = a - \frac{b^2}{2c}$. Hence

$$b^2 < 2ac. \tag{1}$$

Step 2: $f' > 0$ to the left gives $c^2 < 2ab$. Apply Taylor to f': for some $\eta \in (x_0 - t, x_0)$,

$$f'(x_0 - t) = b - ct + \tfrac{t^2}{2}f'''(\eta) \le b - ct + \tfrac{t^2}{2}f(\eta) < b - ct + \tfrac{a}{2}t^2,$$

using $f''' \le f$ and then $f(\eta) < f(x_0) = a$ (f is strictly increasing because $f' > 0$). The left side is positive, so $q_2(t) = \frac{a}{2}t^2 - ct + b > 0$ for all $t > 0$; at $t^* = c/a > 0$ this reads $b - \frac{c^2}{2a} > 0$:

$$c^2 < 2ab. \tag{2}$$

Step 3: eliminate c. Square (1) and use (2): $b^4 < 4a^2c^2 < 4a^2 \cdot 2ab = 8a^3b$. Divide by $b > 0$: $b^3 < 8a^3$, so $b < 2a$. That is $f'(x_0) < 2f(x_0)$, and x_0 was arbitrary. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start from the same observation: the information is "positive on the whole real line" plus one upper bound on f'''. They split at the first operation. (N) integrates the hypothesis *upwards* from $-\infty$, producing bounds on f'', f' by integrals of f. (M) freezes one point x_0 and looks *backwards* along the line: to the left of x_0 the function is a cubic-ish Taylor polynomial, and it must stay positive for all $t > 0$. Positivity of a quadratic in t on the whole half-line is a discriminant condition — an algebraic inequality between the three numbers f, f', f'' at the *same* point. That is what (N) never gets: a relation at a single point.

> *I want $f' < 2f$ at one point, a relation between values at the same x. Integrating from $-\infty$ will only ever give me integrals. So I fix x_0 and ask what positivity to the left of x_0 forces on $f(x_0), f'(x_0), f''(x_0)$.*

> *Going left, f is at most its second-order Taylor polynomial, because the cubic remainder has sign $-f''' < 0$. That polynomial must stay positive for every $t > 0$, so its discriminant is negative. That is one inequality, and it does not even use $f''' \le f$.*

> *The hypothesis $f''' \le f$ must enter somewhere: it bounds the remainder when I Taylor-expand f' instead of f. Two quadratic inequalities, three unknowns a, b, c — eliminate c.*

▷ **Cue:** a function and several derivatives all positive on all of $\mathbb{R}$, and a conclusion comparing them at one point ⟶ **Move:** fix the point, Taylor-expand backwards ($x_0 - t$), and demand positivity for all $t > 0$: a discriminant inequality

▷ **Cue:** a quadratic $\alpha t^2 - \beta t + \gamma$ with $\alpha, \beta, \gamma > 0$ known to be positive for all $t > 0$ ⟶ **Move:** evaluate at the vertex $t = \beta/2\alpha > 0$: get $\beta^2 < 4\alpha\gamma$

▷ **Cue:** two inequalities $b^2 < 2ac$ and $c^2 < 2ab$ and a target involving only a, b ⟶ **Move:** square one and substitute the other to eliminate the unwanted variable

▷ **Cue:** a differential inequality whose extremal case is e^x ⟶ **Move:** do not integrate from $-\infty$ blindly; check the example $1 + e^x$ for divergent integrals first

Layer 3 THE WHY

Why is looking backwards the natural move rather than a trick? Go to what the hypotheses actually say. "$f, f', f'', f''' > 0$ for all x" is a statement about the *entire* line, and the dangerous part of the line is the left: going right, positive derivatives only make things bigger, which constrains nothing. Going left, f decreases, f' decreases, and the question becomes whether they can stay positive forever. A function with $f(x_0) = a$ small and $f'(x_0) = b$ large is heading steeply downhill as you go left; it will cross zero unless f'' (the curvature) is large enough to bend it back up in time. That is exactly $b^2 < 2ac$: the tangent parabola to the left must not reach the axis. So the desired inequality $f' < 2f$ — "the slope is not too large compared with the height" — is a statement about not crashing through zero on the left, and that is where to look.

Why Taylor with a remainder of known sign? Because the hypothesis has two directions of information about f''': positivity ($f''' > 0$) and an upper bound ($f''' \le f$). Expanding f to second order leaves a remainder $-\frac{t^3}{6} f'''$, whose *sign* is known, so the positivity of f''' gets used. Expanding f' to first order leaves $+\frac{t^2}{2} f'''$, whose sign is bad, so you need its *size*, and $f''' \le f \le f(x_0)$ supplies it. Each hypothesis has one natural place to go; the two Taylor expansions are the two places.

Now the official solution. Bhargava, Kedlaya and Ng define, in effect, $g = ff' - \frac{1}{2} f''^2$ and $h = 2ff'' - f'^2$, show $g' > 0$, $h' > 0$ with limits 0 at $-\infty$, and conclude $g, h > 0$. Here is where those come from: $h > 0$ is exactly (1) and $g > 0$ is exactly (2). They are the same two discriminant conditions, proved by a monotonicity argument instead of by Taylor. The derivative computations are $h' = 2ff''' > 0$ and $g' = f'^2 + ff'' - f''f''' \ge f'^2 > 0$ (using $f''f''' \le f''f$). Their route needs the limits at $-\infty$: f is increasing and positive, so bounded as $x \to -\infty$; f' and f'' tend to 0 by the argument in the natural attempt. The official text quotes the fact "$\lim_{-\infty} \ge 0$ and positive derivative implies positive" without checking these limits; if you use that route, check them. The Taylor route avoids limits altogether, which is why I prefer it in the arena.

The constant 2 is not sharp — the official note improves it to $(3/2)^{1/3}$, and e^x shows nothing below 1 is possible. You do not need the best constant; you need to see that 2 comes from $8^{1/3}$, which comes from multiplying a "2" from each discriminant. That tells you the problem was set so that two discriminant inequalities suffice.

The natural attempt did legitimate analysis in the wrong direction. Integrating upward turns "$f''' \le f''$" into "$f'' \le \int f''$", which is weaker at every step. The problem punishes zombie integration.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "positive for all x" applied to four functions, together with a pointwise conclusion. On seeing it, write: fix x_0, set $a, b, c = f, f', f''$ at x_0, and look at $f(x_0 - t)$ for $t > 0$.

2. Taylor's theorem with Lagrange remainder, then "quadratic positive on $t > 0$ implies negative discriminant". Two expansions (of f and of f') give two inequalities; eliminate f''.

3. Hesitation comes on seeing $f''' \le f$: it looks like it must be integrated. It is reinforced away by asking which remainder term has an unknown sign (that is where the upper bound is needed) and which has a known sign (that is where positivity is enough).

4. The trap is integrating from $-\infty$: the integral of f may diverge ($f = 1 + e^x$), and even when it converges, bounds by antiderivatives of f never return to f itself. A secondary trap is

quoting "limit at $-\infty$ is ≥ 0" without proving the limits exist.

5. Ask: *on which side does the hypothesis bite?* Positivity plus increasing derivatives constrains only the left. Then compress the whole left half-line into a single point's Taylor data.

6. Reusable decision rule: *when a quantity is known to stay positive on a half-line and you want a pointwise inequality, write its Taylor polynomial into that half-line, bound the remainder by sign or by hypothesis, and demand the polynomial stay positive; the discriminant is the pointwise inequality.*

1999 B5

1999 B5. For an integer $n \geq 3$, let $\theta = 2\pi/n$. Evaluate the determinant of the $n \times n$ matrix $I + A$, where I is the $n \times n$ identity matrix and $A = (a_{jk})$ has entries $a_{jk} = \cos(j\theta + k\theta)$ for all j, k.

Fields: linear algebra; eigenvalues of low-rank matrices; roots of unity **Position:** B5 — late in the paper, but it yields completely to one structural observation; the difficulty is refusing to expand the determinant

Answer: $\det(I + A) = 1 - \dfrac{n^2}{4}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The strong student computes small cases. For $n = 3$, $\theta = 2\pi/3$ and $\cos\big((j+k)\theta\big)$ is 1 when $3 \mid j + k$ and $-\frac{1}{2}$ otherwise, so

$$I + A = \begin{pmatrix} \frac{1}{2} & 1 & -\frac{1}{2} \\ 1 & \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & 2 \end{pmatrix}, \qquad \det(I + A) = \tfrac{3}{8} - \tfrac{7}{4} + \tfrac{1}{8} = -\tfrac{5}{4}.$$

For $n = 4$, $\theta = \pi/2$, entries are $\cos\big((j+k)\pi/2\big) \in \{0, \pm 1\}$, and a 4×4 expansion gives -3. The data $-\frac{5}{4}, -3$ fit $1 - \frac{n^2}{4}$, and $n = 5$ by hand would be a long evening. Now the general proof. The student notices the rows of A satisfy a three-term recurrence: $\cos((j+1)\theta + k\theta) + \cos((j-1)\theta + k\theta) = 2\cos\theta\cos(j\theta + k\theta)$, so $\mathrm{row}_{j+1}(A) = 2\cos\theta\,\mathrm{row}_j(A) - \mathrm{row}_{j-1}(A)$. Row-reduce! Subtract to kill rows $3, \ldots, n$ — but the identity is in the way. $\mathrm{row}_{j+1}(I + A) - 2\cos\theta\,\mathrm{row}_j(I + A) + \mathrm{row}_{j-1}(I + A)$ is not zero; it is $e_{j+1} - 2\cos\theta\, e_j + e_{j-1}$, a tridiagonal mess. That is the wall: the structure that makes A simple (low rank) is invisible to row operations on $I + A$, because I has full rank.

The model chain. **Step 1: A has rank at most 2.** Let $c, s \in \mathbb{R}^n$ have entries $c_j = \cos j\theta$, $s_j = \sin j\theta$ $(1 \leq j \leq n)$. The addition formula $\cos(j\theta + k\theta) = \cos j\theta\cos k\theta - \sin j\theta\sin k\theta$ says, entry by entry,

$$A = c\,c^{\mathsf{T}} - s\,s^{\mathsf{T}}.$$

Step 2: c and s are orthogonal, of equal length. Put $\omega = e^{2i\theta}$. Since $n \geq 3$, $2\theta = 4\pi/n$ is not a multiple of 2π, so $\omega \neq 1$ while $\omega^n = 1$; hence $\sum_{j=1}^{n} \omega^j = \omega\frac{\omega^n - 1}{\omega - 1} = 0$. Taking real and imaginary parts, $\sum_j \cos 2j\theta = 0 = \sum_j \sin 2j\theta$. Therefore

$$c \cdot c = \sum_j \tfrac{1+\cos 2j\theta}{2} = \tfrac{n}{2}, \qquad s \cdot s = \sum_j \tfrac{1-\cos 2j\theta}{2} = \tfrac{n}{2}, \qquad c \cdot s = \sum_j \tfrac{\sin 2j\theta}{2} = 0.$$

Step 3: eigenvectors. Using Step 2,

$$Ac = c\,(c \cdot c) - s\,(s \cdot c) = \tfrac{n}{2}\,c, \qquad As = c\,(c \cdot s) - s\,(s \cdot s) = -\tfrac{n}{2}\,s,$$

and $Ax = c(c \cdot x) - s(s \cdot x) = 0$ for every x orthogonal to both c and s. Since c, s are nonzero and orthogonal, $\mathbb{R}^n = \mathrm{span}(c, s) \oplus \{c, s\}^{\perp}$, and $\{c, s\}^{\perp}$ has dimension $n - 2$. Choosing an orthonormal basis of $\{c, s\}^{\perp}$ and adjoining $c/|c|$, $s/|s|$ gives an orthonormal basis of eigenvectors of A, with eigenvalues $\frac{n}{2}$, $-\frac{n}{2}$ and 0 (multiplicity $n - 2$).

Step 4: the determinant. In that basis $I + A$ is diagonal with entries $1 + \frac{n}{2}$, $1 - \frac{n}{2}$ and 1 ($n - 2$ times). So

$$\det(I + A) = \left(1 + \frac{n}{2}\right)\left(1 - \frac{n}{2}\right) = 1 - \frac{n^2}{4}. \qquad \square$$

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains see the addition formula — the natural attempt uses it in the form of a row recurrence. They split at what the formula is taken to say. (N) reads it as "the rows are linearly dependent" and reaches for row operations on $I + A$. (M) reads it as "A is a sum of two outer products", $A = cc^\mathsf{T} - ss^\mathsf{T}$, and asks what A does to vectors. Once A is written as outer products, its action on any x is a two-term formula, and the eigenvalues fall out of three dot products.

> *$\cos(j\theta + k\theta)$ splits into a function of j times a function of k, twice. So A is $cc^\mathsf{T} - ss^\mathsf{T}$: rank two. The determinant of I plus a rank-two matrix is decided by a 2×2 problem.*

> *$Ax = c(c \cdot x) - s(s \cdot x)$. If c and s were orthogonal, they would be eigenvectors. Are they? The dot products are sums of $\cos 2j\theta$ and $\sin 2j\theta$ over a full period — zero, as long as 2θ is not a full turn, which is exactly why $n \geq 3$.*

- ▷ **Cue:** a matrix whose (j,k) entry is $\cos(\alpha_j + \beta_k)$, $e^{i(\alpha_j+\beta_k)}$, or any sum of products $u(j)v(k)$ $\longrightarrow$ **Move:** write it as a sum of outer products; its rank is at most the number of terms
- ▷ **Cue:** $\det(I + \text{low-rank})$ $\longrightarrow$ **Move:** find the eigenvalues on the small column space (the rest are 0), or use $\det(I_n + RS) = \det(I_r + SR)$
- ▷ **Cue:** a sum of $\cos(mj\theta)$ or $\sin(mj\theta)$ over $j = 1, \ldots, n$ with $\theta = 2\pi/n$ $\longrightarrow$ **Move:** it is the real or imaginary part of a sum of n-th roots of unity: zero unless $n \mid m$
- ▷ **Cue:** a hypothesis like $n \geq 3$ that looks arbitrary $\longrightarrow$ **Move:** find the step where it is used, and check the excluded case to see the formula fail ($n = 2$ gives 3, not 0)

Layer 3 THE WHY

Why is "rank two" the natural reading of the addition formula? Go to the definition of matrix multiplication. The (j,k) entry of uv^T is $u_j v_k$: a matrix whose entries separate as "something depending only on the row" times "something depending only on the column" is an outer product, and every column is a multiple of u. The addition formula is precisely a separation of $\cos(j\theta + k\theta)$ into two such products. So a problem that fills a matrix with $\cos(\text{row angle} + \text{column angle})$ is handing you a rank-two matrix, the same way a matrix of $j \cdot k$ is rank one. Seeing "$j\theta + k\theta$" and not "$(j + k)\theta$" in the statement is itself the cue: the setter wrote the angle as a sum of a row term and a column term.

Why does low rank decide $\det(I + A)$? Because the determinant is the product of eigenvalues, and $I + A$ has eigenvalue $1 + \lambda$ whenever A has eigenvalue λ on the same eigenvector. A rank-two A kills an $(n - 2)$-dimensional subspace, contributing 1's to the product. Only two eigenvalues are in play. That is why the answer is a product of two factors, and why brute-force small cases looked like a quadratic in n.

Why are c and s orthogonal? This is where $\theta = 2\pi/n$ enters, and it is not decoration. The vectors c and s sample cos and sin at n equally spaced points around the full circle. Over a full period, $\cos^2$ and $\sin^2$ average to $\frac{1}{2}$ each and $\cos \sin$ averages to 0 — the discrete version of $\int_0^{2\pi} \cos^2 = \int_0^{2\pi} \sin^2 = \pi$, $\int_0^{2\pi} \cos \sin = 0$. The discrete version needs the sampling to be fine enough to see the doubled frequency 2θ; at $n = 2$ it is not ($\omega = e^{2i\pi} = 1$), and indeed for $n = 2$ the matrix is $\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$ and $\det(I + A) = 3 \neq 0$. That is what the hypothesis $n \geq 3$ is for.

Kedlaya's first solution introduces the vectors $v^{(m)}$ with entries $e^{ikm\theta}$ — the discrete Fourier basis — and computes $Av^{(m)}$ for each m. Here is where that comes from: $\cos(j\theta + k\theta) = \frac{1}{2}\left(e^{ij\theta}e^{ik\theta} + e^{-ij\theta}e^{-ik\theta}\right)$ is the same rank-two splitting over $\mathbb{C}$, and the Fourier vectors are the natural orthogonal basis for anything built from n-th roots of unity. The second solution (Mohamed Omar) uses $\det(I_n + RS) = \det(I_2 + SR)$ with R the $n \times 2$ matrix of the two outer-product vectors — in real form, $R = [\,c\ \ s\,]$, $S = \begin{pmatrix} c^\mathsf{T} \\ -s^\mathsf{T} \end{pmatrix}$, $SR = \operatorname{diag}(\frac{n}{2}, -\frac{n}{2})$. All three are one idea: rank two, then orthogonality over a full period.

The natural attempt was not wrong to spot the row recurrence; that recurrence *is* the rank-two structure. It lost by applying row operations to $I + A$ instead of asking what A does to vectors. Determinants of "identity plus structured" matrices should be attacked through the structured part's eigenvalues, never by elimination.

Layer $+\alpha$ BACKTESTING

1. The first cue is the entry $\cos(j\theta + k\theta)$, a function of a row term plus a column term. On seeing it, write $A = cc^\mathsf{T} - ss^\mathsf{T}$ with $c_j = \cos j\theta$, $s_j = \sin j\theta$.

2. Low rank, so compute A on its column space: Ac and As via the dot products $c \cdot c$, $s \cdot s$, $c \cdot s$, which are root-of-unity sums. Then multiply $(1 + \lambda)$ over eigenvalues.

3. Hesitation comes after writing $A = cc^\mathsf{T} - ss^\mathsf{T}$: what next? It is reinforced away by writing $Ax = c(c \cdot x) - s(s \cdot x)$ and asking when c, s are eigenvectors — the answer is "when they are orthogonal", which is a computation you can do.

4. The trap is Gaussian elimination: the rows of A satisfy a recurrence, but the identity destroys it in $I + A$. A second trap is forgetting $n \geq 3$ in the root-of-unity sum, which silently gives the wrong answer at $n = 2$.

5. Ask of any structured matrix: *does the (j,k) entry split into a sum of (function of j)×(function of k)?* If yes, count the terms; that bounds the rank, and the determinant of I plus it is a small computation.

6. Reusable decision rule: *for $\det(I + A)$ with A of rank $r \ll n$, write $A = RS$ with R of size $n \times r$, and compute $\det(I_r + SR)$ (equivalently, the r nonzero eigenvalues of A); never expand an $n \times n$ determinant whose entries separate.*

1999 B6

> **1999 B6.** Let S be a finite set of integers, each greater than 1. Suppose that for each integer n there is some $s \in S$ such that $\gcd(s,n) = 1$ or $\gcd(s,n) = s$. Show that there exist $s, t \in S$ such that $\gcd(s,t)$ is prime.

Fields: number theory; extremal principle; combinatorics of prime supports **Position:** B6 — the last problem on the paper, and hard to start; once the extremal choice is made the proof is eight lines **Answer:** Proof: take the least positive n sharing a factor with every element of S; then some $s \mid n$, and for a prime $p \mid s$ the minimality of n applied to n/p produces t with $\gcd(s,t) = p$. (Here $s = t$ is permitted.)

Layer 1 THE WHAT — surface analysis

The natural attempt. Rephrase first. Say n *hits* s if $\gcd(s,n) > 1$, and *swallows* s if $s \mid n$. The hypothesis: every n misses some s or swallows some s. The strong student goes for the contrapositive: assume no $\gcd(s,t)$ is prime — so no element of S is prime, and any two elements are coprime or share a composite gcd — and build an n that hits everything and swallows nothing.

First try: for each s pick its smallest prime factor and let n be the product of the distinct primes picked. Then n hits every s. Does it swallow any? If $s \mid n$, then s is squarefree and all its primes were picked. Repair: drop one prime p of s from n. But p might have been the only

prime hitting some other $u \in S$; then add a different prime of u … which might complete some third element. Try $S = \{6, 10, 15\}$: pick 2, 2, 3, so $n = 6$, swallowing 6. Drop 2: now 10 is missed, add 5: $n = 15$, swallowing 15. Drop 3: 6 is missed … the repairs cycle. (Here the conclusion holds — $\gcd(6, 10) = 2$ — so the construction must fail; but in the general contrapositive you have no idea whether the cycling terminates.) That is the wall: local repairs of a hitting set have no monotone quantity, and you cannot see where "no prime gcd" is supposed to stop the cycle.

The model chain. **Step 1: an extremal n.** The product $N = \prod_{s \in S} s$ hits every s, since $\gcd(s, N) = s > 1$. So there is a *least* positive integer n with $\gcd(s, n) > 1$ for all $s \in S$. Fix it.
Step 2: n is squarefree. If $p^2 \mid n$, then n/p has exactly the same prime divisors as n. Since $\gcd(s, m) > 1$ holds precisely when s and m share a prime, n/p also hits every s, contradicting minimality.
Step 3: some s divides n. Apply the hypothesis to n: some $s \in S$ has $\gcd(s, n) = 1$ or $\gcd(s, n) = s$. The first is impossible by the choice of n, so $s \mid n$. As $s > 1$, choose a prime $p \mid s$; then $p \mid n$.
Step 4: a witness t. Now n/p is a positive integer smaller than n, so by minimality some $t \in S$ has $\gcd(t, n/p) = 1$. But $\gcd(t, n) > 1$, so t shares a prime q with n. The primes of n are p and the primes of n/p (and $p \nmid n/p$ by Step 2); t shares none of the latter, so $q = p$, i.e. $p \mid t$.
Step 5: $\gcd(s, t) = p$**.** Let q be any prime dividing $\gcd(s, t)$. Then $q \mid s \mid n$ and $q \nmid n/p$ (because $q \mid t$ and $\gcd(t, n/p) = 1$), so $q = p$. Also $p^2 \nmid s$, since $s \mid n$ and n is squarefree. Hence $\gcd(s, t)$ is a power of p, divisible by p, not by p^2: $\gcd(s, t) = p$, a prime. □

(If $n = p$, then $n/p = 1$, every t satisfies $\gcd(t, 1) = 1$, and the argument still runs: $s \mid p$ forces $s = p$, and any t gives $\gcd(p, t) = p$ because $p \mid t$. Possibly $t = s$, and $\gcd(s, s) = s = p$ is allowed by the statement.)

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains translate the hypothesis into "hit" and "swallow". They split at what to do with an n that hits everything. (N) builds one and *repairs* it, one prime at a time, hoping to avoid swallowing. (M) takes the *smallest* one and repairs nothing. Minimality says: removing any prime from n un-hits something. That un-hit element is the t the conclusion wants. The failure of every repair — the thing that made (N) cycle — is the information (M) harvests.

> *The hypothesis talks about every n, but it only bites on an n that hits everything: then it has to swallow something. So I want an n that hits everything, and I take the smallest one, so that I know what happens when I shrink it.*

> *I have $s \mid n$ and a prime $p \mid s$. Shrinking n to n/p must miss some t. The only thing n had that n/p lacks is p — so p is the only prime t shares with n. And s lives inside n. So s and t can only meet at p.*

- ▷ **Cue:** a hypothesis of the form "for every object, some bad-or-good alternative occurs" and a conclusion about a special pair ⟶ **Move:** choose the object extremally (smallest n with a property) so that one alternative is excluded and the other is forced
- ▷ **Cue:** a minimal object with a property that is preserved under removing a repeated factor ⟶ **Move:** conclude the minimal object is squarefree (or has no redundant parts)
- ▷ **Cue:** a minimal object that loses its property when one piece p is removed ⟶ **Move:** the witness to that loss interacts with the object only through p
- ▷ **Cue:** "$\gcd(s, n) > 1$ for all s" as a condition on n ⟶ **Move:** it depends only on the set of primes of n: think of n as a hitting set of primes

Layer 3 THE WHY

Why is "smallest n" the natural move and not a trick? Look at the hypothesis's shape: for every n, some s is *missed* or *swallowed*. For most n the first alternative holds and tells you nothing. The hypothesis has content only for an n that misses nothing — a hitting set — because then it is forced to produce a divisor $s \mid n$. So you must pick a hitting set. Which one? You also need a second element t to pair with s, and the only other tool you have is "some n misses something". The cheapest way to guarantee that a particular n' misses something is to know n' is *not* a hitting set — and the cleanest source of such n' is "everything smaller than a minimal hitting set". So minimality is not decoration; it is the only way to get both a swallowed s (from n) and a missed t (from n/p) out of one hypothesis.

Why does $\gcd(s,t) = p$ come out exactly, not merely a multiple of p? Because n is squarefree and s sits inside it, so s has each prime once; and t is barred from every prime of n except p. The conclusion "gcd is prime" needs two things — one shared prime, and only to the first power — and the extremal n supplies one each: squarefreeness bounds the exponent, the n/p witness bounds the support.

The official first solution builds a sequence of primes $p_1, p_2, \ldots$ greedily, each from an element of S coprime to the earlier ones, then takes the least m such that $p_1 \cdots p_m$ has a divisor in S. Here is where that comes from: it is the same extremal idea with a different order on candidates. The greedy sequence produces a hitting set, and "least m" plays the role of "least n": removing p_m destroys the swallowing, which forces the swallowed d to contain p_m, and the element N_{m-1} that introduced p_m is coprime to $p_1 \cdots p_{m-1}$, so it meets d only at p_m. The second official solution (from `sci.math`) is the argument above. It is shorter because it lets the integers' own ordering do the work that the greedy construction does by hand.

The natural attempt was the right instinct — construct a hitting set — with the wrong control. Local repairs have no measure of progress, and you cannot tell whether "no prime gcd" will ever stop them. Extremality supplies the measure for free.

Layer $+\alpha$ BACKTESTING

1. The first cue is "$\gcd(s,n) = 1$ or $\gcd(s,n) = s$" for *every* n: a dichotomy with a free universal variable. On seeing it, write: which n make the first alternative impossible? Those that share a prime with every s.

2. The extremal principle: take the least such n. Then extract, in order, squarefreeness, a swallowed s, and a witness t from n/p.

3. Hesitation comes after finding $s \mid n$: where is the second element? It is reinforced away by asking what minimality says about n/p, which is the only smaller number you have a reason to look at.

4. The trap is constructive repair of a hitting set, which cycles with no invariant. A smaller trap is forgetting that $s = t$ is allowed, which makes the case $s = p$ look like a problem when it is the conclusion.

5. Ask: *is there a hypothesis that only has content for special objects?* Then take the smallest special object; its minimality is a second hypothesis you get free.

6. Reusable decision rule: *when a universal hypothesis is informative only for objects with a property P, take a minimal object with P; apply the hypothesis to it, and apply minimality to each one-step reduction of it; the two outputs usually interact in exactly one place, which is the conclusion.*