DAWN PUTNAM SESSION

Putnam 1990 — Papers A and B

Putnam 120 Archive

Contents

Putnam 1990 — Paper A **5**
1990 A1 5
1990 A2 8
1990 A3 10
1990 A4 13
1990 A5 16
1990 A6 18

Putnam 1990 — Paper B **23**
1990 B1 23
1990 B2 25
1990 B3 27
1990 B4 30
1990 B5 33
1990 B6 36

Putnam 1990 — Paper A

1990 A1

1990 A1. Let

$$T_0 = 2, T_1 = 3, T_2 = 6,$$

and for $n \geq 3$,

$$T_n = (n+4)T_{n-1} - 4nT_{n-2} + (4n-8)T_{n-3}.$$

The first few terms are

$$2, 3, 6, 14, 40, 152, 784, 5168, 40576.$$

Find, with proof, a formula for T_n of the form $T_n = A_n + B_n$, where $\{A_n\}$ and $\{B_n\}$ are well-known sequences.

Fields: sequences and recurrences; linear recurrences with polynomial coefficients; induction **Position:** A1 — the opener, and a guess-then-verify problem: the whole difficulty is the guess, and the proof is a single algebraic check done twice. **Answer:** $T_n = n! + 2^n$ for every $n \geq 0$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You see a linear recurrence and reach for the tool you were trained on: the characteristic equation. Try $T_n = \lambda^n$. Substituting gives $\lambda^3 = (n+4)\lambda^2 - 4n\lambda + (4n-8)$, and the equation still contains n. There is no single λ that works for all n, because the coefficients are not constant. The method was built for constant coefficients and it has nothing to say here.

So you escalate to a generating function. Set $F(x) = \sum T_n x^n$. The factors of n in the coefficients turn into derivatives, $\sum nT_{n-1}x^n = x\,(xF)'$ and so on, and you get a first-order differential equation for F with polynomial coefficients. You grind through it for twenty minutes. Then the wall: if the answer has any factorial-sized term in it (and the ratios $784/152 \approx 5.2$, $5168/784 \approx 6.6$, $40576/5168 \approx 7.9$ are climbing, which says it does), then $\sum T_n x^n$ has radius of convergence 0. The generating function you are solving for is not a function. You can still work formally, but the ODE you have produced will not hand you "a well-known sequence" at the end without another half hour.

Or you guess, but you guess by checking terms: you find $n! + 2^n$ matches $2, 3, 6, 14, 40, 152, 784, 5168, 40576$ and write "by pattern, $T_n = n! + 2^n$". Nine matching terms are evidence, not proof, and the problem says *with proof*.

The model chain. *Guess.* The consecutive ratios T_n/T_{n-1} for $n = 6, 7, 8$ are about $5.2, 6.6, 7.9$: growing roughly like n. A sequence whose ratio grows like n is factorial-sized. Subtract $n!$:

n	0	1	2	3	4	5	6	7	8
T_n	2	3	6	14	40	152	784	5168	40576
$n!$	1	1	2	6	24	120	720	5040	40320
$T_n - n!$	1	2	4	8	16	32	64	128	256

The remainder is 2^n. Conjecture: $T_n = n! + 2^n$.

Proof. The recurrence is linear: if two sequences each satisfy

$$X_n = (n+4)X_{n-1} - 4nX_{n-2} + (4n-8)X_{n-3} \qquad (n \geq 3) \tag{R}$$

then so does their sum, because the right-hand side of (R) is a sum of terms each linear in one X_k. So it suffices to check (R) for $n!$ and for 2^n separately, for every $n \geq 3$.

For $X_n = 2^n$, factor out 2^{n-3}:

$$(n+4)2^{n-1} - 4n\,2^{n-2} + (4n-8)2^{n-3} = 2^{n-3}\big[4(n+4) - 8n + (4n-8)\big] = 2^{n-3} \cdot 8 = 2^n.$$

For $X_n = n!$, factor out $(n-3)!$, which is defined since $n \geq 3$:

$$(n+4)(n-1)! - 4n(n-2)! + (4n-8)(n-3)! = (n-3)!\,(n-2)\big[(n+4)(n-1) - 4n + 4\big].$$

The bracket is $n^2 + 3n - 4 - 4n + 4 = n^2 - n = n(n-1)$, so the whole expression is $(n-3)!(n-2)(n-1)n = n!$.

Hence $t_n = n! + 2^n$ satisfies (R) for all $n \geq 3$, and $t_0 = 1 + 1 = 2$, $t_1 = 1 + 2 = 3$, $t_2 = 2 + 4 = 6$ agree with T_0, T_1, T_2.

Finally, (R) with three initial values determines the sequence uniquely. By strong induction on n: if $t_k = T_k$ for all $k < n$ with $n \geq 3$, then t_n and T_n are given by the same expression in $t_{n-1}, t_{n-2}, t_{n-3}$ and $T_{n-1}, T_{n-2}, T_{n-3}$, so $t_n = T_n$. Therefore $T_n = n! + 2^n$ for all $n \geq 0$. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths read the recurrence and the listed terms. At the recurrence, the natural attempt tries to *solve* it — characteristic equation, then generating function. The model never solves it. It reads the phrase "$A_n + B_n$ where $\{A_n\}$ and $\{B_n\}$ are well-known" as an instruction to guess from the data, and treats the recurrence only as the thing the guess must be checked against. The split is at the first line: solve versus guess-and-verify.

> *The question has told me the answer is a sum of two sequences I already know. There are only a handful of those: powers, factorials, Fibonacci, binomials, polynomials. I don't need to solve this recurrence — I need to recognise the terms, then check.*

> *The ratios of consecutive terms are going up by about one each time. That is what $n!$ does. Subtract $n!$ and see what is left.*

> *Linear recurrence: I can check the two pieces separately and add. And three initial values plus a third-order recurrence pin the sequence down, so matching them is the whole proof.*

Layer 3 THE WHY

Why is guessing the right move, and not a gamble? Two facts about the problem make it natural.

First, the phrasing. "Find a formula of the form $A_n + B_n$ where both are well-known" is not a request to solve a recurrence; it is a promise that the answer lies in a very small search space. When the search space is small, search it. Solving a variable-coefficient recurrence from scratch is a research-level activity in general; recognising a sum of two familiar sequences is a five-minute activity. The characteristic equation fails for a precise reason — it relies on λ^n being an eigenvector of the shift with the *same* eigenvalue at every step, which needs constant coefficients — and once you see why it fails, you see that no off-the-shelf solving method was intended.

Second, growth rate is the fastest diagnostic you own. For a sequence X_n, the ratio X_n / X_{n-1} tells you its type: tending to a constant λ means geometric-like λ^n; tending to 1 means polynomial; growing like n means factorial. Here the ratios climb by about 1.3 per step and are near n, so the dominant term is $n!$. Once the dominant part is identified, subtracting it exposes the next part — the same move as peeling off the leading term of an asymptotic expansion. The residue $1, 2, 4, 8, \ldots$ is recognisable at sight. (Kedlaya–Poonen–Vakil note the same diagnostic from the other side: the terms become divisible by high powers of 2 and no other prime, because $n!$ and 2^n are both highly 2-divisible for large n.)

Why does the proof reduce to two separate checks? Because (R) is linear and homogeneous in the X_k: the set of sequences satisfying it is closed under addition and scaling. That is the definition of linearity doing the work, not a trick. And why is agreement on T_0, T_1, T_2 enough? Because (R) expresses X_n in terms of the three previous values, so the first three values determine everything; this is exactly strong induction.

There is also a deeper reason the recurrence has these two solutions, and it is where the official solution's second argument comes from. The official solution sets $t_n = n! + 2^n$ and computes $t_n - nt_{n-1} = 2^n - n2^{n-1}$, then observes that 2^n and $n2^{n-1}$ both satisfy $f_n - 4f_{n-1} + 4f_{n-2} = 0$. Here is where that comes from. Put $u_n = T_n - nT_{n-1}$. Then

$$u_n - 4u_{n-1} + 4u_{n-2} = T_n - (n+4)T_{n-1} + 4nT_{n-2} - (4n-8)T_{n-3},$$

which you can confirm by expanding: the T_{n-1} terms give $-n-4$, the T_{n-2} terms give $4(n-1)+4 = 4n$, the T_{n-3} term gives $-4(n-2)$. So (R) says precisely that u_n obeys the constant-coefficient recurrence $u_n = 4u_{n-1} - 4u_{n-2}$, whose characteristic polynomial is $(\lambda - 2)^2$. The operator *factorises*: first $X_n \mapsto X_n - nX_{n-1}$, which sends $n!$ to 0, then the constant-coefficient operator with characteristic polynomial $(\lambda - 2)^2$, which kills 2^n and $n2^n$. Feed in 2^n: the first step gives $2^n - n2^{n-1} = 2^{n-1}(2-n)$, a combination of 2^n and $n2^n$, which the second step kills. Recurrence (R) is exactly this composition, which is why $n!$ and 2^n fit. That is also why the characteristic-equation instinct was not entirely wrong — half the operator *is* constant-coefficient; it was hidden behind the variable-coefficient half.

Caveat. No official solution for 1990 exists in the archive's `.tex` sources; the proof above was constructed independently, then compared against the Kedlaya–Poonen–Vakil book (project PDF), which gives the same answer and the operator-factorisation argument discussed in this layer. Both verifications of (R) were also checked by machine for $0 \le n \le 29$ and symbolically.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "of the form $A_n + B_n$, where $\{A_n\}$ and $\{B_n\}$ are well-known sequences". On seeing it, write down the ratios T_n / T_{n-1} of the listed terms, and a short list of candidates: $n!$, 2^n, 3^n, Fibonacci, $\binom{2n}{n}$, polynomials.

2. Guess-and-verify: identify the dominant growth, subtract it, recognise the residue; then prove by checking each piece against the recurrence and invoking linearity plus uniqueness from initial values.

3. Hesitation: "guessing is not proof, so I must solve the recurrence". Reinforce it away with the fact that a guessed closed form becomes a proof the moment it is checked against the recurrence and the initial conditions — the guess is the method, the check is the rigour.

4. The trap is the characteristic equation and its big brother, the ordinary generating function: both assume structure the recurrence lacks, and the OGF of anything factorial-sized has radius of convergence 0. The second trap is stopping at "the pattern holds for nine terms".

5. Growth ratio near n means factorial; near a constant λ means λ^n; near 1 means polynomial. Subtract the dominant term and look again. A variable-coefficient recurrence handed to

you in a competition is usually a composition of simple operators, and its solutions are simple.

6. Decision rule: when the problem promises the answer lies in a small familiar family, search the family using the data before building any general machinery; then turn the guess into a proof by verification against the defining relation, splitting by linearity where you can.

▷ **Cue:** a recurrence whose coefficients depend on n, with the answer promised to be "well-known" $\longrightarrow$ **Move:** do not solve it; guess from the listed terms and verify against the recurrence

▷ **Cue:** consecutive ratios X_n/X_{n-1} growing roughly like n $\longrightarrow$ **Move:** the dominant term is $n!$; subtract it and inspect the remainder

▷ **Cue:** proving a guessed closed form for a linear recurrence of order k $\longrightarrow$ **Move:** check the recurrence for each summand separately, check the first k values, conclude by strong induction

▷ **Cue:** a variable-coefficient linear recurrence with a suspiciously clean solution $\longrightarrow$ **Move:** look for a factorisation of the operator, e.g. $u_n = T_n - nT_{n-1}$ satisfying a constant-coefficient recurrence

1990 A2

1990 A2. Is $\sqrt{2}$ the limit of a sequence of numbers of the form $\sqrt[3]{n} - \sqrt[3]{m}$ $(n, m = 0, 1, 2, \dots)$?

Fields: real analysis (sequences, limits, density); elementary algebra ($a^3 - b^3$ factorisation) **Position:** A2 — early-paper, and short once you stop treating $\sqrt{2}$ as special. **Answer:** Yes. In fact every real number is such a limit: the gaps $\sqrt[3]{k+1} - \sqrt[3]{k}$ tend to 0 while $\sqrt[3]{k} \to \infty$, so the differences $\sqrt[3]{n} - \sqrt[3]{m}$ are dense in $\mathbb{R}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You see a square root on one side and cube roots on the other, and the first instinct is algebraic: can $\sqrt[3]{n} - \sqrt[3]{m}$ ever *equal* $\sqrt{2}$? Cube $\sqrt[3]{n} = \sqrt[3]{m} + \sqrt{2}$ to get

$$n = m + 3\sqrt{2}\, m^{2/3} + 6m^{1/3} + 2\sqrt{2},$$

and you start wondering whether a degree-2 irrational can come out of degree-3 radicals. Some students stop here and answer "no" on the strength of that unease. That is the first wall, and it is the wrong wall: the question asks about a *limit*, not an equality, and equality was never needed.

The stronger student recovers and tries to aim at $\sqrt{2}$ directly. Take $m = j^3$, so $\sqrt[3]{m} = j$, and look for $n = j^3 + d$ with $\sqrt[3]{j^3 + d} - j \approx \sqrt{2}$. The binomial expansion gives

$$\sqrt[3]{j^3 + d} - j = j\left(1 + \tfrac{d}{j^3}\right)^{1/3} - j \approx \frac{d}{3j^2} - \frac{d^2}{9j^5} + \cdots,$$

so you want $d \approx 3\sqrt{2}\, j^2$; set $d = \lfloor 3\sqrt{2}\, j^2 \rfloor$. This works. But to make it a proof you now owe a remainder estimate for a binomial series with non-integer exponent, uniform enough to show the tail vanishes as $j \to \infty$, and you owe an argument that the floor costs nothing. It is correct and it is expensive, and every line of it is spent on the number $\sqrt{2}$, which the problem never needed.

The model chain. Step 1: the gaps shrink. For $k \geq 1$ put $a = \sqrt[3]{k+1}$, $b = \sqrt[3]{k}$. From $a^3 - b^3 = (a-b)(a^2 + ab + b^2)$ and $a^3 - b^3 = 1$,

$$0 < \sqrt[3]{k+1} - \sqrt[3]{k} = \frac{1}{a^2 + ab + b^2} < \frac{1}{3b^2} = \frac{1}{3k^{2/3}},$$

since $a > b > 0$. So the gaps tend to 0, while $\sqrt[3]{k} \to \infty$.

Step 2: walk up in small steps. Fix a positive integer j. Choose m_j with $3m_j^{2/3} > j$; then every gap $\sqrt[3]{k+1} - \sqrt[3]{k}$ with $k \geq m_j$ is less than $1/j$. Consider

$$\sqrt[3]{k} - \sqrt[3]{m_j}, \qquad k = m_j,\, m_j + 1,\, m_j + 2, \ldots$$

It starts at $0 < \sqrt{2}$, increases, and is unbounded. Let n_j be the least $k \geq m_j$ with $\sqrt[3]{k} - \sqrt[3]{m_j} \geq \sqrt{2}$. Then $n_j > m_j$, and $n_j - 1 \geq m_j$ fails the condition, so

$$\sqrt{2} \leq \sqrt[3]{n_j} - \sqrt[3]{m_j} = \left(\sqrt[3]{n_j - 1} - \sqrt[3]{m_j}\right) + \left(\sqrt[3]{n_j} - \sqrt[3]{n_j - 1}\right) < \sqrt{2} + \tfrac{1}{j}.$$

Step 3: conclude. The sequence $x_j = \sqrt[3]{n_j} - \sqrt[3]{m_j}$ satisfies $|x_j - \sqrt{2}| < 1/j$, so $x_j \to \sqrt{2}$. The answer is yes. Nothing used $\sqrt{2}$ except $\sqrt{2} > 0$; replacing it by any $r \geq 0$ gives the same, and swapping n and m handles $r < 0$. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree on the question up to the moment of reading it. At that moment the natural attempt fixes its eyes on the target $\sqrt{2}$ and tries to manufacture it — first exactly, then approximately by expansion. The model looks away from the target and at the set $\{\sqrt[3]{n} - \sqrt[3]{m}\}$ itself, asking how finely it covers the line. The split is the choice between "build a sequence converging to this number" and "show this set has no gaps of any fixed size".

> *Nothing about $\sqrt[3]{n} - \sqrt[3]{m}$ knows about $\sqrt{2}$. If the answer is yes for $\sqrt{2}$, it is almost certainly yes for every positive number, so I should be proving that these differences are dense, not hunting for $\sqrt{2}$.*

> *Consecutive cube roots get closer and closer together but the cube roots still go to infinity. So if I start at some large $\sqrt[3]{m}$ and walk up one integer at a time, my steps are tiny and I cannot jump over $\sqrt{2}$. I stop at the first step past it.*

> *Being asked about a limit, not an equality, means I only need to land within $1/j$ for each j. The algebraic question of whether $\sqrt{2}$ is ever hit exactly is irrelevant.*

Layer 3 THE WHY

Why is density the natural question and the shrinking gap the natural tool? Go back to what the question literally asks. "$\sqrt{2}$ is the limit of a sequence from the set S" means, by the definition of a limit, that for every $\varepsilon > 0$ some element of S lies within ε of $\sqrt{2}$. That is a statement about S near one point. If the property of S that makes it true is structural, it will hold near every point; so the honest first question is "what does S look like on the number line?", and a specific irrational target is a hint to ask exactly that. The setters chose $\sqrt{2}$ — an irrational of the wrong algebraic degree — to lure you into algebra. The degree mismatch obstructs equality and says nothing about approximation.

Now, what does S look like? Its elements are differences of terms of one sequence $c_k = \sqrt[3]{k}$. Fix the subtracted term c_m; then $c_k - c_m$ for $k \geq m$ is the same sequence shifted to start at 0. A sequence that climbs to infinity with steps smaller than ε must pass within ε of every level it crosses: this is the intermediate value theorem for sequences, and it is nothing but the observation that you cannot get from below $\sqrt{2}$ to above $\sqrt{2} + \varepsilon$ with a step shorter than ε. So the two facts to check are exactly $c_k \to \infty$ and $c_{k+1} - c_k \to 0$. Both are properties of $x \mapsto x^{1/3}$: it is unbounded and its derivative $\frac{1}{3}x^{-2/3}$ tends to 0. The $a^3 - b^3$ factorisation is the A-level way

of saying "the derivative tends to 0" without calculus: it rationalises the difference, the same move as multiplying $\sqrt{k+1}-\sqrt{k}$ by its conjugate.

Kedlaya, Poonen and Vakil's first solution partitions $[0, \sqrt[3]{m}]$ by the points $\sqrt[3]{m+i}-\sqrt[3]{m}$, $0 \le i \le 7m$, and observes the mesh is $O(m^{-2/3})$; their remark isolates the general principle (any $a_n \to +\infty$ with $a_{n+1}-a_n \to 0$ gives dense differences), which is what the model chain proves directly. Their second solution is the aimed construction $\sqrt[3]{\lfloor (n+r)^3 \rfloor}-\sqrt[3]{n^3} \to r$, a cleaner version of the natural attempt: it avoids the binomial series by noting $(n+r)^3-(n+r-\varepsilon)^3 = 3n^2\varepsilon + O(n) > 1$ for large n, so the floor cannot drop below $(n+r-\varepsilon)^3$. Their third solution notes $k(\sqrt[3]{n}-\sqrt[3]{m}) = \sqrt[3]{k^3 n}-\sqrt[3]{k^3 m}$, so S is closed under positive integer multiples; a set with arbitrarily small positive elements and that closure is dense. All three rest on the same fact: S contains arbitrarily small positive numbers. The solution above was constructed independently before comparison and agrees with their first solution and remark.

Rigour points. The inequality $a^2+ab+b^2 > 3b^2$ needs $a > b > 0$, which holds for $k \ge 1$. The "least k" must be shown to exist (unboundedness of $\sqrt[3]{k}$) and to exceed m_j (the difference at $k = m_j$ is $0 < \sqrt{2}$). The last step used must have index $\ge m_j$ so that it is below $1/j$.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "the limit of a sequence of numbers of the form". Write down: *I need, for each ε, an element of the set within ε of $\sqrt{2}$ — a density question.* Then write the gap $\sqrt[3]{k+1}-\sqrt[3]{k}$ and rationalise it.

2. Discrete intermediate value argument: an unbounded increasing sequence with steps below ε, started below the target, has a first term at or above the target, and that term overshoots by less than ε.

3. Hesitation occurs at the mismatch between $\sqrt{2}$ and cube roots, which invites the question "can this ever equal $\sqrt{2}$?". It is reinforced away by rereading the word *limit*: equality is not asked for, so algebraic degree is irrelevant.

4. The trap is the specific target. It tempts an algebraic "no" (degree mismatch) or an expensive aimed construction by binomial expansion whose remainder must then be controlled. Both spend effort on $\sqrt{2}$, which plays no role.

5. Slogan: *when a problem names one specific number as a limit or approximant, ask whether every number would do; if so, prove density.* And: *differences of a sequence tending to infinity with vanishing increments are dense.*

6. Decision rule: before constructing an approximation to a particular value, strip the particular value out and ask what structural property of the approximating set would give every value. Prove that property — typically "arbitrarily small steps and unbounded reach", or "arbitrarily small elements and closure under integer multiples".

▷ **Cue:** "is x the limit of numbers of the form …" with a specific irrational x $\longrightarrow$ **Move:** ask whether every real works; aim to prove the set is dense

▷ **Cue:** a set of differences $a_n - a_m$ of one sequence $\longrightarrow$ **Move:** check $a_n \to \infty$ and $a_{n+1}-a_n \to 0$; then walk up from a fixed a_m and stop at the first term past the target

▷ **Cue:** a difference of kth roots such as $\sqrt[3]{k+1}-\sqrt[3]{k}$ $\longrightarrow$ **Move:** rationalise with $a^3-b^3 = (a-b)(a^2+ab+b^2)$ to bound it by $1/(3k^{2/3})$

▷ **Cue:** a set containing arbitrarily small positive elements and closed under positive integer multiples $\longrightarrow$ **Move:** it is dense in $[0,\infty)$

1990 A3

1990 A3. Prove that any convex pentagon whose vertices (no three of which are collinear) have integer coordinates must have area greater than or equal to $5/2$.

Fields: combinatorial geometry (lattice polygons); number theory (parity, pigeonhole); extremal principle **Position:** A3 — middle of the morning paper; short once the right object is chosen, but the one-line parity idea has a hole that a careless write-up falls straight into. **Answer:** The bound holds and is sharp: a convex lattice pentagon of least area has a lattice point strictly inside (the midpoint of two vertices of equal parity type), and the fan of five lattice triangles from that point gives area at least $5 \cdot \frac{1}{2}$. Equality at $(0,0), (1,0), (2,1), (1,2), (0,1)$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You know two facts about lattice polygons. The shoelace formula: a triangle with integer vertices $(x_1, y_1), (x_2, y_2), (x_3, y_3)$ has area $\frac{1}{2}|(x_2 - x_1)(y_3 - y_1) - (x_3 - x_1)(y_2 - y_1)|$, a positive half-integer when the triangle is non-degenerate. And, if you have met it, Pick's theorem: a lattice polygon with i interior lattice points and b boundary lattice points has area $i + \frac{b}{2} - 1$.

So you write the first thing Pick gives you. Five vertices are boundary points, so $b \geq 5$ and

$$\text{area} \;\geq\; i + \tfrac{5}{2} - 1 \;=\; i + \tfrac{3}{2}.$$

Target $\frac{5}{2}$. You need $i \geq 1$: one lattice point strictly inside. (Without Pick the same count appears another way: cut the pentagon into three triangles by the diagonals from one vertex, each of area $\geq \frac{1}{2}$, total $\geq \frac{3}{2}$. Same deficit of exactly 1.)

How do you force an interior lattice point? You look for one. Perhaps the parity idea surfaces: a point (x, y) has one of four parity types $(x \bmod 2,\ y \bmod 2)$; five vertices, four types, so two vertices P, Q share a type, and then $\frac{1}{2}(P + Q)$ has integer coordinates. If PQ is a diagonal, its midpoint is inside the pentagon and you are done.

Then the wall. Nothing stops P and Q from being *adjacent*. Then the midpoint sits on a side, it is a boundary point, not an interior one, and Pick gives $b \geq 6$ and area $\geq i + 2$ — still short by $\frac{1}{2}$ if $i = 0$. You try a different pair; there may not be one. You try to handle the "bad" pentagon by coordinates, setting one vertex at the origin, and the case analysis grows without end.

The model chain. *Areas are half-integers.* Triangulating a convex lattice polygon by the diagonals from one vertex expresses its area as a sum of lattice-triangle areas, each a positive multiple of $\frac{1}{2}$ by the shoelace formula. So the area of every convex lattice pentagon lies in $\{\frac{1}{2}, 1, \frac{3}{2}, \dots\}$, a set with a least element among any non-empty subset.

Take a smallest one. Let $ABCDE$ (vertices in order around the boundary) be a convex lattice pentagon, no three vertices collinear, of least possible area. It suffices to show this area is at least $\frac{5}{2}$.

No side of the minimal pentagon contains a lattice point in its interior. Suppose F is a lattice point strictly between A and B. I claim $AFCDE$ is again a convex lattice pentagon with no three vertices collinear. Its hull is contained in that of $ABCDE$, so A, C, D, E, being extreme points of the larger set, are extreme points of the smaller. For F: the diagonal AC of the convex pentagon has B strictly on one side and D, E strictly on the other; F is on the open segment from A (on the line) to B, so F is strictly on B's side, hence outside the hull of A, C, D, E, hence extreme. Five points each of which is an extreme point of their hull are in convex position with no three collinear (a middle point of three collinear points is not extreme). And its area is the old area minus the area of triangle FBC, which is positive since F is not on line BC. This contradicts minimality. Applying the same argument to every side: the only lattice points on the boundary of $ABCDE$ are its five vertices.

Pigeonhole on parity. Each vertex has a parity type in $\{0,1\}^2$, four types, five vertices; two vertices $P \neq Q$ share a type. Then $M = \frac{1}{2}(P + Q)$ has integer coordinates. M is a lattice point on the open segment PQ, so PQ is not a side (by the previous paragraph); it is a diagonal. In a convex polygon with no three vertices collinear, the open segment of a diagonal lies in the interior. So M is a lattice point strictly inside $ABCDE$.

Fan from M. Join M to A, B, C, D, E. This cuts the pentagon into triangles MAB, MBC, MCD, MDE, MEA, which overlap only on edges and whose areas sum to the area of the pentagon. None is degenerate, because M is strictly inside and so lies on no side line. Each has lattice vertices, so each has area $\geq \frac{1}{2}$. Total area $\geq \frac{5}{2}$.

Sharpness. $(0,0), (1,0), (2,1), (1,2), (0,1)$ is convex with no three collinear (consecutive turns all left), and shoelace gives $\frac{1}{2}(0+1+3+1+0) = \frac{5}{2}$. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains reach the same sentence: "two vertices have the same parity type, so their midpoint is a lattice point." At that sentence the natural attempt is holding *an arbitrary* pentagon and must now fight the case where the midpoint lands on a side. The model chain had already, one step earlier, replaced the arbitrary pentagon by *the smallest* one — and for the smallest one the bad case is impossible, because a lattice point on a side lets you cut off a triangle and get a smaller pentagon. The split happens before the parity step, at the choice of which pentagon to talk about.

> *The parity trick works except when the midpoint lands on a side. But a lattice point on a side is a gift, not an obstacle: I can move a vertex to it and get a smaller lattice pentagon. So I should not prove this for every pentagon; I should prove it for the smallest one, where that gift has already been spent.*

> *Pick says I am exactly one interior lattice point short. So the whole problem is: find one lattice point strictly inside. Where do lattice points come from that I did not put there? Midpoints of lattice points with matching parities.*

Layer 3 THE WHY

Why is "take the minimal pentagon" natural rather than a trick? Look at what the statement is. It is a lower bound: "area $\geq \frac{5}{2}$" for all members of a family. A lower bound over a family is the same thing as a statement about the family's minimum. So the question "what does the minimiser look like?" is not an extra idea; it is the problem restated. The minimum exists here for a reason you can point to: areas of lattice polygons are multiples of $\frac{1}{2}$ (the shoelace determinant is an integer), and a set of positive multiples of $\frac{1}{2}$ has a least element. Where there is a well-ordered quantity, there is a minimal counterexample, and a minimal object inherits every property that "could have been used to make it smaller".

That last phrase is the mechanism. Why does the bad case of the parity trick fall to minimality? Because the bad case is *itself* a smaller-making move. A lattice point F in the middle of side AB gives a new lattice pentagon $AFCDE$ sitting inside the old one; the old one minus triangle FBC. The obstacle to the direct argument is exactly an instruction for descent. Whenever the failure of your argument hands you a strictly smaller object of the same kind, the argument should be run on the smallest object. This is the extremal principle, and it is the same logic as infinite descent in number theory (Fermat's proof that $x^4 + y^4 = z^2$ has no solutions: a solution yields a smaller one).

Why parity in the first place? Drive to the definition. We want a lattice point we were not given. The lattice $\mathbb{Z}^2$ is closed under addition but not under halving; $\frac{1}{2}(P+Q)$ is a lattice point exactly when $P + Q \in 2\mathbb{Z}^2$, i.e. when $P \equiv Q \pmod 2$ coordinatewise. There are $2^2 = 4$ residue classes mod 2, and a pentagon has 5 vertices. The number five in the problem is not decoration: it is one more than the number of parity classes in the plane. That is also why a lattice *quadrilateral* can have area 1 (the unit square: four vertices, four different parity types, no midpoint forced) and why the bound jumps at five. The problem's own numbers are pointing at pigeonhole on $\mathbb{Z}^2/2\mathbb{Z}^2$.

Why the fan and not Pick? Both work. Pick with $i \geq 1$, $b \geq 5$ gives $1 + \frac{5}{2} - 1 = \frac{5}{2}$ in one line; the fan from M gives the same number from the shoelace formula alone, which is why the Kedlaya–Poonen–Vakil solution uses it. The fan has one advantage worth naming: it uses only that M is strictly inside, so it survives if you forget Pick in the exam hall.

The official solution (Kedlaya, Poonen, Vakil) takes the same route: minimal pentagon, clear the sides, pigeonhole on parity, fan. It states in one clause that $AFCDE$ is "a convex lattice pentagon". That clause is where marks are lost: you must check that F is genuinely a vertex (not swallowed by the hull of the others) and that no three of the new vertices are collinear. The extreme-point argument above supplies it, and it uses the diagonal AC — the same fact about convex polygons (a diagonal separates the vertices on its two arcs) that later puts M in the interior. I solved and checked the argument independently before comparing; a brute-force search over all five-point subsets of the 4×4 grid confirms that the least area is $\frac{5}{2}$, attained by the pentagon above.

Layer $+\alpha$ BACKTESTING

1. The first cue is the conclusion "area $\geq \frac{5}{2}$" for lattice polygons. Write down immediately: areas are in $\frac{1}{2}\mathbb{Z}$; with Pick, area $= i + \frac{b}{2} - 1 \geq i + \frac{3}{2}$; so I need exactly one interior lattice point.

2. Manufacture a lattice point by averaging: pigeonhole five vertices into four parity classes, take the midpoint of a matching pair. Then fan-triangulate from it.

3. Hesitation comes when you see that the matching pair may be adjacent, putting the midpoint on a side. It is reinforced away by noticing that such a point lets you shrink the pentagon, so a minimal pentagon never has it.

4. The trap is writing "the midpoint is inside the pentagon" without checking that the pair forms a diagonal. A second trap is the replacement step: asserting $AFCDE$ is a convex pentagon without showing F is a genuine vertex and no three are collinear.

5. Slogan: *a lower bound over a family is a statement about the minimiser — go and look at it.* And: *if the failure case of your argument produces a smaller object of the same kind, run the argument on the smallest one.*

6. Decision rule: when a direct proof has a bad case, ask whether the bad case yields a strictly smaller instance (in a well-ordered measure). If yes, assume minimality from the start and the bad case disappears. Separately: when you need a point of a lattice that you were not given, count residue classes mod 2 against the number of points you have.

▷ **Cue:** lower bound on area of a lattice polygon $\longrightarrow$ **Move:** areas lie in $\frac{1}{2}\mathbb{Z}$; use Pick $A = i + \frac{b}{2} - 1$ or fan triangulation, and find how many interior points are missing

▷ **Cue:** need a lattice point you were not given; more than 2^d lattice points in $\mathbb{Z}^d$ $\longrightarrow$ **Move:** pigeonhole on parity classes; the midpoint of a matching pair is a lattice point

▷ **Cue:** a direct argument fails in a case that produces a smaller object of the same kind $\longrightarrow$ **Move:** choose a minimal object (extremal principle) so that the bad case contradicts minimality

▷ **Cue:** a lattice point strictly inside a convex lattice polygon $\longrightarrow$ **Move:** join it to every vertex; each fan triangle has area at least $\frac{1}{2}$

1990 A4

1990 A4. Consider a paper punch that can be centered at any point of the plane and that, when operated, removes from the plane precisely those points whose distance from the center is irrational. How many punches are needed to remove every point?

Fields: geometry (distances in the plane); number theory (rational versus irrational); algebra of squared distances **Position:** A4 — mid-paper, and one of the shortest A4s on record once you ask what a point that *survives* looks like. **Answer:** Three. One or two punches always leave a point at rational distance

from every centre; three punches centred at $(-\alpha, 0)$, $(0,0)$, $(\alpha, 0)$ with α^2 irrational (for instance $\alpha = \sqrt[4]{2}$) remove every point.

Layer 1 THE WHAT — surface analysis

The natural attempt. You read "removes points at irrational distance" and start thinking about how many points one punch removes: nearly all of them. Only the circles of rational radius about the centre survive, a countable family of circles. So you picture the survivors of punch one as a target of concentric rational circles, and the survivors of punch two as a second target, and you try to show the two targets have a common point — or do not. You draw circles of radius $1, 2, 3$ about P and Q and look for intersections. They plainly intersect. So two punches fail, and you move to three.

With three centres you try the same picture and get lost. Three families of concentric circles; does some circle from each family pass through a common point? You try an equilateral triangle of side 1, compute distances from a sample point, get expressions like $\sqrt{x^2+y^2}$, $\sqrt{(x-1)^2+y^2}$, and cannot see how to rule out all three being rational at once. You wonder whether the answer is four, or infinitely many, and the counting intuition ("each punch leaves a measure-zero set, so surely two or three suffice") tells you nothing about which.

Where this stalls: the attempt is trying to control the square roots. It never removes them.

The model chain. *A point survives every punch iff its distance to every centre is rational.* So the question is: what is the least number of centres such that no point of the plane is at rational distance from all of them?

One punch fails: the centre itself is at distance 0 from itself, and 0 is rational.

Two punches fail. Let the centres be P, Q. If $P = Q$, the point P survives. Otherwise put $d = |PQ| > 0$ and choose a rational $r > d/2$. The circles of radius r about P and Q meet, because $|r - r| = 0 < d < 2r = r + r$ (the triangle-inequality condition for two circles to intersect). A common point X has $|XP| = |XQ| = r \in \mathbb{Q}$, so X survives both punches.

Three punches suffice. Let $\alpha > 0$ with α^2 irrational; take $\alpha = \sqrt[4]{2}$, so $\alpha^2 = \sqrt{2}$. Centre the punches at

$$A = (-\alpha, 0), \qquad O = (0,0), \qquad B = (\alpha, 0).$$

For any point $X = (x, y)$ write $a = |XA|$, $c = |XO|$, $b = |XB|$. Then

$$a^2 = (x+\alpha)^2 + y^2, \qquad c^2 = x^2 + y^2, \qquad b^2 = (x-\alpha)^2 + y^2,$$

so the cross terms $\pm 2\alpha x$ cancel on adding:

$$a^2 + b^2 - 2c^2 = 2\alpha^2.$$

If a, b, c were all rational, the left side would be rational, so $\alpha^2 = \sqrt{2}$ would be rational — false. Hence every X has irrational distance from at least one of A, O, B and is removed by that punch. So three punches suffice, and the answer is 3. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree that a surviving point is one whose distances to all centres are rational. The natural attempt then works with the distances themselves, $\sqrt{(x-p)^2+(y-q)^2}$, and tries to control irrational square roots of unknown quantities. The model squares the distances at once and asks which *linear combination* of the squared distances kills the unknowns x, y. That combination is a constant built only from the centres, and making that constant irrational finishes the problem.

> *Rational distances means rational squared distances, and squared distances are polynomials in x and y with only $x^2 + y^2$ in the quadratic part. If I pick centres so that some combination of the*

> *squared distances loses x and y entirely, what is left is a number depending only on where I put the punches. I get to choose that number. I will choose it irrational.*

> *Two punches: I should look for a survivor, not argue about measure. Two equal rational radii large enough will do it.*

Layer 3 THE WHY

Why is squaring the right move, and why is the symmetric collinear triple the right configuration? Go back to what a distance is.

The distance from $X = (x, y)$ to a centre (p, q) is the square root of $(x-p)^2 + (y-q)^2 = (x^2+y^2) - 2px - 2qy + (p^2+q^2)$. Look at the shape of this. Every squared distance has the *same* quadratic part $x^2 + y^2$, then a part linear in (x, y), then a constant fixed by the centre. So the squared distances to k centres are k affine functions of the three unknowns $x^2 + y^2$, x, y. With three unknowns and three centres, generically you can solve for everything and there is no obstruction. But if the centres are collinear, the linear parts only involve one direction, and there are only two unknowns in play: $x^2 + y^2$ and x. Three affine expressions in two unknowns must satisfy a linear relation, and that relation has a constant term which depends only on the centres. That relation is the constraint a surviving point must obey. Here, with centres at $-\alpha, 0, \alpha$ on the x-axis, it is

$$a^2 + b^2 - 2c^2 = 2\alpha^2,$$

which you may recognise as Apollonius's theorem: the median XO of triangle XAB satisfies $a^2 + b^2 = 2c^2 + 2\alpha^2$. That is not a coincidence you have to remember; it is exactly what "three collinear points give a linear relation among squared distances" produces.

Now the rationality enters. Squaring preserves rationality in the direction that matters: rational distance implies rational squared distance. A linear relation with rational coefficients among rational numbers forces its constant to be rational. So a survivor can exist only if $2\alpha^2$ is rational. Choosing α^2 irrational destroys every possible survivor in one line. The irrationality has been moved off the unknown point (where you cannot control it) and onto the configuration (where you can).

This is also why the answer is not two. With two centres there are two squared distances and still two unknowns $x^2 + y^2$ and x (put both centres on the x-axis); two equations in two unknowns impose no relation at all, and indeed any two rational radii with the circles intersecting give a survivor. The count 3 is exactly one more than the number of free quantities.

There is no official solution in the archive's sources for 1990. The argument above was constructed and checked here: the identity $a^2 + b^2 - 2c^2 = 2\alpha^2$ is a direct expansion, and the two-punch construction uses only the intersection criterion for circles.

Rigour points. The one-punch and two-punch cases both need a stated survivor, not a heuristic. The case $P = Q$ must be handled (the point P survives). In the three-punch case you must say why rational a, b, c give rational $a^2 + b^2 - 2c^2$ (closure of $\mathbb{Q}$ under multiplication and addition), and you must exhibit a specific α with α^2 irrational, with a reason ($\sqrt{2}$ irrational).

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "distance from the center is irrational". Write down the complement: a point survives iff all its distances to the centres are rational; then write each squared distance as $x^2 + y^2 - 2px - 2qy + p^2 + q^2$.

2. Eliminate the unknown point from the squared distances by a linear combination, leaving a constant that depends only on the centres; then choose centres making that constant irrational. For the lower bound, construct a survivor.

3. Hesitation comes at three punches: three families of circles look uncontrollable. It is reinforced away by noticing every squared distance has the same quadratic part, so differences and combinations are linear, and collinear centres cut the linear part to one variable.

4. The trap is measure or cardinality thinking ("each punch leaves a null set") which says nothing about whether the survivor sets intersect. A second trap is working with unsquared distances and fighting nested square roots.

5. Slogan: *squared distances to several points differ by linear functions; with enough points they obey a linear relation whose constant is yours to choose.* And: *to prove "$k-1$ is not enough", exhibit the survivor.*

6. Decision rule: when a condition must hold for a set of quantities depending on an unknown, find a combination of them in which the unknown cancels; the cancelled combination is a certificate. Put the forbidden property (irrationality, non-integrality, odd parity) into that certificate.

▷ **Cue:** a point at rational distance from several given points $\longrightarrow$ **Move:** square the distances; squared distances share the quadratic part x^2+y^2 and differ by linear terms

▷ **Cue:** three collinear points and squared distances to a moving point $\longrightarrow$ **Move:** use the linear relation (Apollonius / Stewart): $a^2+b^2-2c^2=2\alpha^2$ for a midpoint

▷ **Cue:** need to show no point satisfies a rationality condition $\longrightarrow$ **Move:** find a combination eliminating the point; make the remaining constant irrational

▷ **Cue:** a question "how many are needed" $\longrightarrow$ **Move:** prove the lower bound by an explicit survivor, the upper bound by a certificate

1990 A5

1990 A5. If **A** and **B** are square matrices of the same size such that $\mathbf{ABAB}=\mathbf{0}$, does it follow that $\mathbf{BABA}=\mathbf{0}$?

Fields: linear algebra (nilpotent matrices, products AB versus BA); construction of counterexamples **Position:** A5 — late in the paper, but the difficulty is entirely in deciding which way the answer goes; once you decide "no" and know where to look, the counterexample is a few lines. **Answer:** No. For 3×3 matrices, $A=E_{23}+E_{32}$ and $B=E_{13}+E_{22}$ give $(AB)^2=0$ but $(BA)^2=E_{13}\neq 0$. The implication does hold for 1×1 and 2×2 matrices, and whenever A or B is invertible.

Notation: E_{ij} is the matrix with a 1 in row i, column j and 0 elsewhere, so $E_{ij}E_{kl}=E_{il}$ if $j=k$ and 0 otherwise; e_1,e_2,e_3 are the standard basis column vectors, with $E_{ij}e_j=e_i$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You treat it as an identity to prove, because $ABAB$ and $BABA$ look like the same word read from a different starting letter. You try to manipulate: $BABA=B(ABAB)\ldots$ — no, the letters do not line up. You try $BABA\cdot BABA=B(ABAB)(ABA)=0$, which gives $(BA)^4=0$, and then sharpen it to $(BA)^3=B(ABAB)A=0$. So BA is nilpotent. You feel close, and try to push from $(BA)^3=0$ to $(BA)^2=0$. Nothing in the algebra gives that last step. You try 2×2 examples to get a feel, and every one satisfies the implication, which makes you more confident it is true. You spend the rest of the time looking for the missing lemma.

Where this stalls: the 2×2 experiments are exactly the size in which the statement is true, so they reinforce the wrong belief. And the manipulation stops at $(BA)^3=0$ because that is genuinely all that follows.

The model chain. Step 1: what does follow. Always

$$(BA)^3=B\,(AB)(AB)\,A=B\,(ABAB)\,A=0,$$

so BA is nilpotent of index at most 3. The question is whether index 3 can occur.

Step 2: where the statement is true, and why — this locates the counterexample.

- If A is invertible, $BA=A^{-1}(AB)A$, so $(BA)^2=A^{-1}(AB)^2A=0$. If B is invertible, $BA=B(AB)B^{-1}$ and likewise $(BA)^2=0$. So a counterexample needs both A and B singular.

- If the size is 1, $(BA)^3 = 0$ forces $BA = 0$. If the size is 2, write $M = BA$. Then $\det M = 0$ (since $(\det M)^3 = \det M^3 = 0$), so the 2×2 Cayley–Hamilton identity $M^2 - (\operatorname{tr} M)M + (\det M)I = 0$ (checkable by expanding) gives $M^2 = (\operatorname{tr} M)M$, hence $0 = M^3 = (\operatorname{tr} M)^2 M$. Either $M = 0$ or $\operatorname{tr} M = 0$; in both cases $M^2 = (\operatorname{tr} M)M = 0$. So a counterexample needs size at least 3.

Step 3: build it in size 3. Aim for BA to be the 3×3 shift $N = E_{12} + E_{23}$, which satisfies $N^2 = E_{13} \neq 0$. Take A singular: let $A = E_{23} + E_{32}$, which sends $e_1 \mapsto 0$, $e_2 \mapsto e_3$, $e_3 \mapsto e_2$. We need $BAe_2 = Ne_2 = e_1$ and $BAe_3 = Ne_3 = e_2$, i.e. $Be_3 = e_1$ and $Be_2 = e_2$; set $Be_1 = 0$. So $B = E_{13} + E_{22}$.

Step 4: verify. Using $E_{ij}E_{kl} = \delta_{jk}E_{il}$:

$$AB = (E_{23} + E_{32})(E_{13} + E_{22}) = E_{32}, \qquad (AB)^2 = E_{32}E_{32} = 0;$$

$$BA = (E_{13} + E_{22})(E_{23} + E_{32}) = E_{12} + E_{23}, \qquad (BA)^2 = E_{13} \neq 0.$$

In full,

$$A = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad BABA = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

So $ABAB = 0$ does not imply $BABA = 0$. For any size $n \geq 3$, pad A and B with zeros to get an $n \times n$ counterexample. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains reach $(BA)^3 = 0$. At that line the natural attempt tries to improve the exponent from 3 to 2; the model asks whether the improvement is even true, and answers by finding the hypotheses under which it *is* true (invertible factor; size ≤ 2). Those proofs say exactly what a counterexample must avoid: both matrices singular, size at least 3. The search space collapses to 3×3 singular matrices with BA a nilpotent of index 3.

> *I have derived $(BA)^3 = 0$ and I cannot get to $(BA)^2 = 0$. Before I hunt for a cleverer manipulation I should ask: in what situations is it true, and does my proof in those situations use something the problem did not give me? If it uses invertibility, or uses size two, the counterexample lives exactly where those fail.*

> *My 2×2 experiments all worked. That is not evidence — in size two every nilpotent squares to zero, so the test was rigged.*

> *I want BA to be the shift. I will pick a singular A and solve for B column by column.*

Layer 3 THE WHY

Why is the answer "no", and why is "locate where the proof works" the natural move rather than a lucky one?

Start with the definitions. AB and BA are different matrices, but they are close relatives: $A(BA) = (AB)A$ and $(BA)B = B(AB)$. So A carries the action of BA to that of AB, and B carries it back. If either of these carrying maps is invertible, the two products are similar — same matrix in different coordinates — and every polynomial identity satisfied by one is satisfied by the other. That is the whole content of the invertible case.

When both A and B are singular, each carrying map loses a direction. The only thing that survives in general is what you get by sandwiching: $(BA)^{k+1} = B(AB)^k A$. This shows the nilpotency index of BA can exceed that of AB by at most one — and the problem is asking

whether it can exceed it by exactly one. The sandwich identity is tight precisely because A on the right and B on the left can each kill a vector; there is no reason the loss must be zero.

Now the size. A nilpotent $n \times n$ matrix has $M^n = 0$ (the kernels of $M, M^2, \ldots$ strictly grow until they stabilise at the whole space, and there are only n dimensions). For $(BA)^2 \neq 0$ with BA nilpotent you need $n \geq 3$. So the problem, by asking about "square matrices of the same size" without fixing the size, is inviting you to check whether small size is special. It is.

So the construction is forced rather than inspired. Pick the simplest nilpotent of index 3, the shift N with $e_3 \mapsto e_2 \mapsto e_1 \mapsto 0$. Factor it as $N = BA$ with A singular. The chosen A kills e_1 and swaps e_2, e_3; then B is read off from $BA = N$. Check the other product: AB sends $e_2 \mapsto e_3$ and kills e_1, e_3, so it is a single "arrow" $e_2 \to e_3$, which squares to zero. In chain language: BA is one chain $e_3 \to e_2 \to e_1 \to 0$ of length three, and AB is a chain of length two plus a fixed zero direction. The lost direction (e_1, killed by A) is exactly what shortens the chain.

There is no official solution in the archive's sources for 1990. The counterexample above was constructed and verified here by the matrix-unit calculation shown, and independently by direct numerical multiplication.

Rigour points. A "no" answer requires one explicit counterexample with both products computed; the prose about similarity and chain lengths is motivation, not proof. The Step 2 claims are not needed for the answer, but if you state them, the 2×2 case needs a real argument (the Cayley–Hamilton route above), not "nilpotent 2×2 matrices square to zero" by assertion.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the pair $ABAB$ versus $BABA$ — the same word with the letters cycled. Write down $(BA)^3 = B(ABAB)A = 0$ and ask whether $(BA)^2 = 0$ follows.

2. Decide the truth value by finding the conditions under which it holds (either factor invertible; size at most 2), then build a counterexample outside those conditions.

3. Hesitation occurs after $(BA)^3 = 0$, when the exponent will not drop. Reinforce it away with the question "is this true?", and by distrusting small cases that are structurally special.

4. The trap is testing only 2×2 matrices, which always satisfy the implication, and concluding it is a theorem. A second trap is quietly assuming invertibility when manipulating AB and BA.

5. Slogan: *AB and BA are similar when a factor is invertible; otherwise they can differ by one step in a nilpotent chain.* And: *when a proof stalls, prove the special cases and read off what a counterexample must avoid.*

6. Decision rule: for any "does P imply Q" with no size or dimension fixed, check the smallest case where the structure has room to fail (here the first size where a nilpotent can have index 3). If every natural proof uses an extra hypothesis, build the counterexample where that hypothesis fails.

▷ **Cue:** $ABAB = 0$ and a question about $BABA$ $\longrightarrow$ **Move:** sandwich: $(BA)^{k+1} = B(AB)^k A$, so BA is nilpotent with index at most one more than AB

▷ **Cue:** a statement about AB versus BA $\longrightarrow$ **Move:** if A or B is invertible they are similar; a counterexample needs both singular

▷ **Cue:** a nilpotent 2×2 matrix $\longrightarrow$ **Move:** it squares to zero (Cayley–Hamilton with det $= 0$); counterexamples to nilpotency statements need size ≥ 3

▷ **Cue:** need a matrix with prescribed product $BA = N$ $\longrightarrow$ **Move:** choose a simple singular A and solve for B column by column using $E_{ij}e_j = e_i$

▷ **Cue:** a manipulation that stalls one step short $\longrightarrow$ **Move:** prove the special cases where it succeeds and read off the shape of a counterexample

1990 A6

1990 A6. If X is a finite set, let X denote the number of elements in X. Call an ordered pair (S, T) of subsets of $\{1, 2, \ldots, n\}$ *admissible* if $s > |T|$ for each $s \in S$, and $t > |S|$ for each $t \in T$. How many admissible ordered pairs of subsets of $\{1, 2, \ldots, 10\}$ are there? Prove your answer.

Fields: combinatorics (counting, binomial identities, Fibonacci numbers) **Position:** A6 — the last problem of the morning. Guessing the answer takes ten minutes of small cases; proving it is the problem. **Answer:** $17711 = F_{22}$. In general the number of admissible pairs of subsets of $\{1, \ldots, n\}$ is the Fibonacci number F_{2n+2} (with $F_1 = F_2 = 1$).

The source statement reads "let X denote the number of elements in X"; the intended text is "let $|X|$ denote", and $|X|$ is used in that sense throughout. Write $[n] = \{1, \ldots, n\}$, $s = |S|$, $t = |T|$, and let $f(n)$ be the number of admissible pairs of subsets of $[n]$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You compute small cases. For $n = 0$ only $(\varnothing, \varnothing)$: $f(0) = 1$. For $n = 1$: $(\varnothing, \varnothing)$, $(\{1\}, \varnothing)$, $(\varnothing, \{1\})$ are admissible, $(\{1\}, \{1\})$ is not since $1 \not> 1$: $f(1) = 3$. With more patience, $f(2) = 8$ and $f(3) = 21$. The sequence $1, 3, 8, 21$ is every other Fibonacci number, and you guess $f(n) = F_{2n+2}$, so $f(10) = F_{22} = 17711$.

Now you try to prove it by a recursion. Every other Fibonacci number satisfies $F_{2n+2} = 3F_{2n} - F_{2n-2}$, so you look for $f(n) = 3f(n-1) - f(n-2)$ by asking what happens to the element n: in S, in T, or in neither. If n is in neither, (S, T) is admissible for $[n-1]$. If $n \in S$, removing it lowers $|S|$, which *relaxes* the condition on T, so removal lands in an admissible pair for $[n-1]$ but the map is not onto — adding n back to S can break $t > |S|$. The three cases do not decouple, and the minus sign in $3f(n-1) - f(n-2)$ has no obvious combinatorial meaning. You have a correct guess and no proof.

Where this stalls: the conditions couple S and T through their *sizes*, and the recursion on the largest element does not respect sizes.

The model chain. *Step 1: fix the sizes and the problem decouples.* Fix s and t. The condition "$x > t$ for every $x \in S$" says exactly $S \subseteq \{t+1, \ldots, n\}$, a set of $n - t$ elements; likewise $T \subseteq \{s+1, \ldots, n\}$, a set of $n - s$ elements. With the sizes fixed, the two choices are independent, so the number of admissible pairs with $|S| = s$, $|T| = t$ is

$$\binom{n-t}{s}\binom{n-s}{t},$$

which is nonzero only if $s + t \le n$. Hence

$$f(n) = \sum_{\substack{s,t \ge 0 \\ s+t \le n}} \binom{n-t}{s}\binom{n-s}{t}.$$

Step 2: group by $k = s + t$. For $0 \le k \le n$ put

$$g(n,k) = \sum_{s=0}^{k} \binom{n-k+s}{s}\binom{n-s}{k-s} = \sum_{s=0}^{k} \binom{n-k+s}{n-k}\binom{n-s}{n-k}.$$

Lemma. For integers $N, p, q \ge 0$,

$$\sum_{i=0}^{N} \binom{i}{p}\binom{N-i}{q} = \binom{N+1}{p+q+1}.$$

Proof. Count the $(p+q+1)$-element subsets of $\{0, 1, \ldots, N\}$ by their $(p+1)$-th smallest element i: the p smaller elements are chosen from the i numbers $0, \ldots, i-1$, and the q larger ones from the $N - i$ numbers $i+1, \ldots, N$. □

Apply it with $N = 2n-k$, $p = q = n-k$, $i = n-k+s$. Then $\binom{i}{p} = \binom{n-k+s}{n-k}$ and $\binom{N-i}{q} = \binom{n-s}{n-k}$. The terms of the lemma's sum with $i < n-k$ (i.e. $s < 0$) vanish because $\binom{i}{p} = 0$ for $i < p$, and those with $s > k$ vanish because then $0 \le n-s < n-k$. So

$$g(n,k) = \binom{2n-k+1}{2n-2k+1} = \binom{2n+1-k}{k}.$$

Step 3: sum the diagonal. $\binom{2n+1-k}{k}$ counts tilings of a $1 \times (2n+1)$ strip by 1×1 squares and 1×2 dominoes using exactly k dominoes (such a tiling has $2n+1-k$ tiles, and we choose which k are dominoes). Since $2k \le 2n+1$ forces $k \le n$, summing over $k = 0, \dots, n$ counts all tilings of the strip. The number τ_m of tilings of a $1\times m$ strip satisfies $\tau_0 = \tau_1 = 1$ and $\tau_m = \tau_{m-1} + \tau_{m-2}$ (the last tile is a square or a domino), so $\tau_m = F_{m+1}$. Therefore

$$f(n) = \sum_{k=0}^{n} \binom{2n+1-k}{k} = \tau_{2n+1} = F_{2n+2}.$$

For $n = 10$: $f(10) = F_{22} = 17711$. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree on the small cases and on the guess. The natural attempt recurses on the *largest element*; the model conditions on the *sizes* (s,t). The second split happens after the double sum is on the page: the natural move is to sum over s for fixed t (rows), which produces nothing recognisable; the model sums along $s+t=k$ (diagonals).

> *The condition on S only mentions $|T|$, and the condition on T only mentions $|S|$. If I fix the two sizes, each condition says "this set lives above a threshold" and the two sets stop talking to each other. Count with the sizes fixed first.*

> *The answer is Fibonacci, and Fibonacci numbers are diagonal sums of Pascal's triangle, $F_{m+1} = \sum_k \binom{m-k}{k}$. So I should group my double sum by $k = s+t$ and hope each diagonal collapses to a single binomial.*

> *A sum of products of two binomials whose tops add to a constant is a Vandermonde-type convolution. Choose a subset, look at its middle element.*

Layer 3 THE WHY

Why is conditioning on sizes natural, and why does the Fibonacci target dictate the grouping?

Look at the definition of admissible. Each element of S is compared not with an element of T but with the *number* $|T|$. So the interaction between S and T passes through two integers and nothing else. Whenever the coupling between two objects is mediated by a few summary numbers, fix those numbers first: the objects then become independent and the count factorises. That is the whole of Step 1, and it is not a trick; it is reading the definition for what it depends on. The largest-element recursion fails for the mirror-image reason: moving one element changes $|S|$ by one and silently changes the condition on every element of T.

Now the target. Once you have guessed F_{2n+2}, ask where Fibonacci numbers come from in a binomial world. The domino count gives the answer: a tiling of length m with k dominoes is a choice of k positions among $m-k$ tiles, so $F_{m+1} = \sum_k \binom{m-k}{k}$ — the shallow diagonals of Pascal's triangle. The target has a single sum over one index, and the index is the number of dominoes. Your double sum has two indices s,t. The only way to match is for some combination

of s and t to play the role of k and for the other direction to collapse. The symmetry $s \leftrightarrow t$ of the summand $\binom{n-t}{s}\binom{n-s}{t}$ says the collapsing direction should be symmetric too, so the natural candidate is $k = s + t$. The small case $n = 2$, $k = 2$ gives $1 + 1 + 1 = 3 = \binom{3}{2}$, which confirms $g(n,k) = \binom{2n+1-k}{k}$ before you prove it.

And the collapse itself is forced by shape. On the diagonal $s + t = k$, the two top indices are $n - t = n - k + s$ and $n - s$; they add to $2n - k$, a constant. Sums of $\binom{i}{p}\binom{N-i}{q}$ with i running are exactly what you get by splitting a subset of a line at one of its elements. Kedlaya's archive has no solution for this year; the lemma is the standard "upper Vandermonde" identity, and the proof by the $(p+1)$-th smallest element is the reason it holds.

There is no official solution in the archive's sources for 1990. The proof above was constructed here; each step was checked by hand, and the values $f(n) = F_{2n+2}$ for $n = 0, \ldots, 10$ and the diagonal identity $g(n,k) = \binom{2n+1-k}{k}$ for all $0 \le k \le n \le 10$ were additionally confirmed by exhaustive computer enumeration of admissible pairs.

Rigour points. The pattern $1, 3, 8, 21$ is a guess, not a proof; the problem says "Prove your answer" and scores nothing for the guess alone. In Step 1, state both directions: every admissible pair with sizes (s,t) arises from such a choice, and every such choice is admissible. In Step 2, handle the vanishing terms at the ends of the lemma's range explicitly. In Step 3, justify $k \le n$ so that the diagonal sum really is all tilings.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "$s > |T|$ for each $s \in S$" — a condition on elements of one set in terms of the *size* of the other. Write down: fix $|S| = s$, $|T| = t$; then $S \subseteq \{t+1, \ldots, n\}$ and $T \subseteq \{s+1, \ldots, n\}$.

2. Count with sizes fixed, $\binom{n-t}{s}\binom{n-s}{t}$, then evaluate the double sum. Small cases give the target F_{2n+2}, which tells you to sum along diagonals.

3. Hesitation comes with the double sum on the page and no closed form in sight. It is reinforced away by the question "what binomial sum gives Fibonacci numbers?" — diagonals of Pascal's triangle — and the check at $n = 2$.

4. The trap is stopping at the pattern-matched guess 17711 without a proof, or spending the time on a largest-element recursion that does not decouple.

5. Slogan: *when two objects interact only through their sizes, fix the sizes and the count factorises.* And: *Fibonacci in the answer means domino tilings, i.e.* $\sum_k \binom{m-k}{k}$.

6. Decision rule: when you have guessed a closed form, let its known combinatorial expansion dictate how to reorganise your sum; then prove the reorganised pieces equal term by term, using a standard convolution where the tops add to a constant.

▷ **Cue:** a condition comparing elements of one set with the size of another $\longrightarrow$ **Move:** fix the sizes; each set is then an arbitrary subset of a fixed range and the count factorises

▷ **Cue:** small cases $1, 3, 8, 21, 55$ $\longrightarrow$ **Move:** every other Fibonacci number, F_{2n+2}; recall $F_{m+1} = \sum_k \binom{m-k}{k}$ (domino tilings)

▷ **Cue:** a double sum symmetric in two indices whose target is a single sum $\longrightarrow$ **Move:** group by the sum of the indices

▷ **Cue:** $\sum_i \binom{i}{p}\binom{N-i}{q}$ with tops adding to a constant $\longrightarrow$ **Move:** upper Vandermonde: $\binom{N+1}{p+q+1}$, proved by conditioning on the $(p+1)$-th smallest chosen element

▷ **Cue:** a sequence satisfying $a_m = a_{m-1} + a_{m-2}$ combinatorially $\longrightarrow$ **Move:** tilings of a strip by squares and dominoes

Putnam 1990 — Paper B

1990 B1

1990 B1. Find all real-valued continuously differentiable functions f on the real line such that for all x,

$$(f(x))^2 = \int_0^x [(f(t))^2 + (f'(t))^2]\,dt + 1990.$$

Fields: real analysis (integral equations), differential equations **Position:** B1 – opening problem of the second session **Answer:** Exactly two functions: $f(x) = \sqrt{1990}\,e^x$ and $f(x) = -\sqrt{1990}\,e^x$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You see an integral equation with the variable in the upper limit, and the reflex from A-level is right: differentiate both sides. Since f is C^1, the integrand $f^2 + (f')^2$ is continuous, so the Fundamental Theorem of Calculus applies and

$$2f(x)f'(x) = f(x)^2 + f'(x)^2.$$

Now most students treat this as a differential equation to be solved, and that is where the time goes. It is nonlinear in f'. You try to solve for f' as a quadratic: $f'^2 - 2ff' + f^2 = 0$, so

$$f' = \frac{2f \pm \sqrt{4f^2 - 4f^2}}{2} = f.$$

That works, but many students do not take the quadratic route. They divide by f^2 and set $u = f'/f$, getting $2u = 1 + u^2$ – which needs $f \neq 0$ everywhere, a condition you have not established. Others try to guess a power series, or try $f = Ce^{kx}$ and find $2k = 1 + k^2$, so $k = 1$, and then are left wondering whether the exponentials are *all* the solutions. Then, having found $f = Ce^x$, the common slip is to forget that differentiation threw away information: the derived equation is satisfied by Ce^x for every C, and you still have to feed the constant back in. Some students stop at "$f = Ce^x$" and lose the mark for the answer.

The model chain. Put $x = 0$ in the equation: the integral vanishes, so $f(0)^2 = 1990$ and $f(0) = \pm\sqrt{1990}$.

The right-hand side is differentiable in x with derivative $f(x)^2 + f'(x)^2$, because the integrand is continuous (this is where "continuously differentiable" is used). The left-hand side is differentiable with derivative $2f(x)f'(x)$. Equating,

$$f^2 + f'^2 - 2ff' = 0, \qquad \text{that is} \qquad (f - f')^2 = 0.$$

A real square is zero only if its base is zero, so $f'(x) = f(x)$ for every x.

Let $g(x) = e^{-x}f(x)$. Then $g'(x) = e^{-x}(f'(x) - f(x)) = 0$ for all x, so g is constant by the Mean Value Theorem, and $f(x) = f(0)e^x = \pm\sqrt{1990}\,e^x$.

Conversely, if $f(x) = Ce^x$ with $C^2 = 1990$, then $f^2 + f'^2 = 2C^2e^{2t}$, and

$$\int_0^x 2C^2e^{2t}\,dt + 1990 = C^2(e^{2x} - 1) + 1990 = 1990\,e^{2x} = f(x)^2.$$

So both functions work, and there are no others.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths agree up to the derived equation $2ff' = f^2 + f'^2$. At that line the natural attempt treats it as a differential equation to be solved – dividing, substituting, guessing exponentials. The model reads it as an algebraic identity between two real numbers at each point, moves everything to one side, and sees a perfect square.

> *Before I solve anything, $f^2 + f'^2 - 2ff'$ is $(f - f')^2$. A sum that equals a cross term is a square being zero. So $f' = f$ exactly, with no case split on whether f vanishes.*

> *Differentiating lost the constant 1990. Whatever I get from the derived equation, I must go back to the original at $x = 0$ to fix it, and then check the candidates in the original equation, not in the derivative.*

Layer 3 THE WHY

Why is spotting the square the natural move and not a lucky one? Look at the shape of the original equation. The integrand is $f^2 + (f')^2$, a sum of two squares; the left side is f^2, whose derivative is $2ff'$. So the moment you differentiate, you are comparing $a^2 + b^2$ with $2ab$ for the real numbers $a = f(x)$, $b = f'(x)$. The inequality $a^2 + b^2 \geq 2ab$ is AM–GM in its most primitive form, and it is an equality exactly when $a = b$. The problem-setter built the equation around the equality case of that inequality. Once you see "sum of squares on one side, twice the product on the other", the only question is which square is zero.

There is a deeper reason too. Real numbers are ordered, and squares are non-negative; that is the one piece of structure the reals have that the equation can exploit. Any time an equation says "a non-negative quantity equals zero", each non-negative piece is forced to vanish separately. That is how one scalar equation can pin down a function completely. Treating $2ff' = f^2 + f'^2$ as a quadratic in f' reaches the same place (the discriminant is zero), but it is the same fact in disguise: a zero discriminant is a perfect square.

The second structural point is the loss of the constant. An equation of the form $F(x) = \int_0^x G + c$ is equivalent to the pair "$F' = G$ and $F(0) = c$". Differentiation keeps the first and discards the second. So the full solution always has two halves: the derivative equation gives the family, the value at the base point of the integral picks the member. The base point here is 0, so you evaluate at 0. That is why the model starts with $x = 0$ before differentiating – it records the information that differentiation is about to destroy.

Finally, the step from $f' = f$ to $f = Ce^x$ deserves its justification, since it is where "are these *all* the solutions?" is answered. Multiplying by the integrating factor e^{-x} turns $f' - f$ into $e^x(e^{-x}f)'$, and a function with zero derivative on an interval is constant. No existence–uniqueness theorem for ODEs is needed; the Mean Value Theorem suffices.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the variable upper limit $\int_0^x$ together with a constant 1990. On seeing it, write down two things: $f(0)^2 = 1990$, and the differentiated equation $2ff' = f^2 + f'^2$.

2. Differentiate to remove the integral, then read the resulting pointwise equation as algebra: rearrange to one side and factor.

3. Hesitation comes at $2ff' = f^2 + f'^2$, when it looks like a nonlinear ODE. It is reinforced away by the habit of always moving everything to one side and asking whether it factors before doing any calculus.

4. Two traps: dividing by f (or f^2) without knowing $f \neq 0$, and stopping at $f = Ce^x$ without using $f(0)^2 = 1990$ – the derived equation is blind to the constant, and the answer needs both signs.

5. Recognise faster by seeing "sum of two squares" on one side of an identity and a product of the same two quantities on the other. That is the equality case of $a^2 + b^2 \geq 2ab$ and it always forces $a = b$.

6. Decision rule: for an equation $F(x) = \int_a^x G + c$, split it into $F' = G$ and $F(a) = c$ immediately; solve the first, use the second to select, and verify in the original. And whenever an equation reduces to "non-negative expression $= 0$", write the expression as a sum of squares and set each to zero.

▷ **Cue:** an equation with $\int_0^x (\dots)\, dt$ plus a constant $\longrightarrow$ **Move:** set $x = 0$ to extract the initial value, then differentiate by the Fundamental Theorem of Calculus

▷ **Cue:** after rearranging, $a^2 + b^2 - 2ab = 0$ (or $a^2 + b^2 = 2ab$) for real a, b $\longrightarrow$ **Move:** write it as $(a - b)^2 = 0$ and conclude $a = b$

▷ **Cue:** a first-order linear ODE $f' = kf$ and a need to show all solutions are exponentials $\longrightarrow$ **Move:** differentiate $e^{-kx} f(x)$; zero derivative means constant

▷ **Cue:** a solution method passes through differentiation $\longrightarrow$ **Move:** check the candidates in the original undifferentiated equation; the constant of integration is where extra or missing solutions hide

1990 B2

1990 B2. Prove that for $|x| < 1$, $|z| > 1$,

$$1 + \sum_{j=1}^{\infty} (1 + x^j) P_j = 0,$$

where P_j is

$$\frac{(1-z)(1-zx)(1-zx^2)\cdots(1-zx^{j-1})}{(z-x)(z-x^2)(z-x^3)\cdots(z-x^j)}.$$

Fields: series (telescoping, q-series), algebra **Position:** B2 – early, but it punishes the wrong first move hard **Answer:** The identity holds: the N-th partial sum equals $\dfrac{(1 - zx^N)\, P_N}{1 - z}$, and $P_N \to 0$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The natural attempt is to expand. You write out the first terms,

$$1 + (1 + x)\frac{1 - z}{z - x} + (1 + x^2)\frac{(1 - z)(1 - zx)}{(z - x)(z - x^2)} + \cdots,$$

and try to expand each P_j in powers of x (or of $1/z$) and collect coefficients, hoping each coefficient of x^k vanishes. This is a real method – it is how identities of this kind are often discovered – but it is a swamp here. Every P_j contributes to every power of x, the coefficients are rational functions of z, and to prove that all of them vanish you need a closed form for the coefficient of x^k, which is harder than the original problem. A second natural attempt is to recognise a named q-series identity (the q-binomial theorem, Heine's transformation) and try to specialise it. That can work if you know the right identity cold, but it is a lookup, not a proof you can build in the room, and matching the parameters costs twenty minutes even when it succeeds.

The model chain. Throughout, x and z may be real or complex with $|x| < 1 < |z|$. Every denominator factor $z - x^k$ ($k \geq 1$) is nonzero because $|x^k| < 1 < |z|$, and $1 - z \neq 0$. Set $P_0 = 1$ (the empty product), and note the defining recurrence

$$(z - x^j)\, P_j = (1 - zx^{j-1})\, P_{j-1} \qquad (j \geq 1). \tag{$*$}$$

Let $S_N = 1 + \sum_{j=1}^{N} (1 + x^j) P_j$ be the partial sums.

Claim. For every $N \geq 0$,

$$S_N = \frac{(1 - zx^N)\, P_N}{1 - z}.$$

For $N = 0$ both sides equal 1. Suppose the claim holds for $N - 1$. Then

$$(1-z)S_N = (1 - zx^{N-1})P_{N-1} + (1-z)(1+x^N)P_N = (z - x^N)P_N + (1-z)(1+x^N)P_N$$

using $(*)$. The bracket is $z - x^N + 1 + x^N - z - zx^N = 1 - zx^N$, so $(1-z)S_N = (1-zx^N)P_N$, proving the claim by induction.

Limit. Since $1 - zx^N \to 1$, it remains to show $P_N \to 0$. The ratio of consecutive terms is controlled by

$$\left| \frac{1 - zx^{k-1}}{z - x^k} \right| \leq \frac{1 + |z|\,|x|^{k-1}}{|z| - |x|^k} \longrightarrow \frac{1}{|z|} < 1 \qquad (k \to \infty).$$

Fix r with $1/|z| < r < 1$. There is K such that every factor with $k \geq K$ has modulus at most r, so for $N \geq K$, $|P_N| \leq |P_{K-1}|\, r^{N-K+1} \to 0$. Hence $S_N \to 0$; that is, the series converges and its sum $1 + \sum_{j \geq 1} (1 + x^j) P_j$ equals 0.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths start by writing down the first two or three terms. The natural attempt then expands each term. The model instead computes the partial sums $S_0 = 1$ and S_1 as single fractions and compares them with the last term added: $S_1 = 1 + (1+x)\frac{1-z}{z-x} = \frac{1-zx}{z-x} = \frac{1-zx}{1-z}\, P_1$.

> *An infinite sum that is claimed to be exactly 0 for all x and z in a region is almost never proved coefficient by coefficient. It is a telescope. So I will compute S_1 as one fraction and see how it compares with P_1.*

> *$S_0 = 1$ and $S_1 = \frac{1-zx}{1-z} P_1$. With $P_0 = 1$, that is $S_0 = \frac{1-z}{1-z} P_0$ too. Guess $S_N = \frac{1-zx^N}{1-z} P_N$ and check it against the recurrence for P_N. Then the sum is zero if and only if $P_N \to 0$, and that is a ratio test.*

Layer 3 THE WHY

Why is "the partial sum is a multiple of the last term" the natural guess, rather than a rabbit from a hat? Look at the structure of P_j. Each P_j is P_{j-1} times one rational factor, $\frac{1-zx^{j-1}}{z-x^j}$. A sequence where consecutive terms have a simple rational ratio is the discrete analogue of an exponential: it is the kind of term ("hypergeometric", or here "q-hypergeometric") whose sums behave like integrals of exponentials. And the integral of $e^{\lambda t}$ is again $e^{\lambda t}$ times a constant. The discrete version of that fact is: if a sum of such terms has a closed form at all, the closed form is very often the last term times a simple rational factor. That is the principle behind Gosper's algorithm, and it is available to you as a heuristic without the algorithm: compute S_N / P_N for $N = 0, 1$ and see whether it looks like something.

Here it does. $S_0/P_0 = 1$ and $S_1/P_1 = \frac{1-zx}{1-z}$. The pattern to test is $\frac{1-zx^N}{1-z}$, which at $N = 0$ gives 1. The test is one line because the recurrence $(*)$ contains exactly the factor $1 - zx^{N-1}$ that the guess puts in front of P_{N-1} – the guess was chosen so that it cancels against the recurrence. That cancellation is the whole content of a telescope.

The second structural point is why the answer is 0 and not something else. Once $S_N = \frac{(1-zx^N)P_N}{1-z}$, the sum is $\lim S_N$, and the problem has been reduced to the size of P_N. Each new factor $\frac{1-zx^{k-1}}{z-x^k}$ tends to $\frac{1}{z}$ as $k \to \infty$ (the powers of x die because $|x| < 1$), and $|1/z| < 1$ because $|z| > 1$. So P_N decays geometrically. This is where both hypotheses are used, and you should

notice that they are used for exactly this: $|x| < 1$ to make the factor approach $1/z$, and $|z| > 1$ to make $1/z$ small. A hypothesis that appears only in the limit step is a strong hint that the finite part is an exact algebraic identity – and it is, the claim holds for any x, z with nonzero denominators.

The official route is not reproduced in the archive's sources for 1990, so the solution above is constructed here and verified independently, including numerical checks of the identity at several real and complex (x, z).

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "$= 0$" for an infinite sum whose terms are built as running products $P_j = P_{j-1} \cdot (\text{factor})$. On seeing it, write down the recurrence $(z - x^j)P_j = (1 - zx^{j-1})P_{j-1}$ and the partial sum S_1 as a single fraction.

2. Telescoping: find a closed form for the partial sum as (last term)$\times$(simple factor), prove it by induction using the recurrence, then take the limit.

3. Hesitation comes after S_1, when you have a fraction but no pattern. It goes away when you divide by the last term: S_1/P_1 is visibly $\frac{1-zx}{1-z}$, and $S_0/P_0 = 1$ fits $\frac{1-zx^0}{1-z}$.

4. The trap is expansion in powers of x: it feels systematic but needs a closed form for every coefficient. A quieter trap is forgetting that "the sum equals 0" includes "the series converges": the limit step with $P_N \to 0$ is not optional.

5. Recognise faster by asking of any series with product-shaped terms: "what is S_N divided by the last term?" Compute it for $N = 0, 1, 2$ before anything else.

6. Decision rule: when the terms of a sum have rational ratios, guess that the partial sum is the last term times a simple factor, fit the factor from small cases, and verify by one induction step against the ratio. Then the infinite sum is controlled entirely by the size of the last term.

▷ **Cue:** a series whose terms are running products $P_j = P_{j-1} \cdot r_j$ with r_j rational $\longrightarrow$ **Move:** compute S_N/P_N for small N and guess the partial sum as P_N times a simple factor

▷ **Cue:** a sum is claimed to be exactly 0 (or a constant) on a whole region $\longrightarrow$ **Move:** look for a telescope, not a coefficient-by-coefficient check

▷ **Cue:** a hypothesis like $|x| < 1$, $|z| > 1$ appears only in a limit $\longrightarrow$ **Move:** prove the finite identity for all parameters, then use the hypothesis for the ratio of consecutive factors

▷ **Cue:** a product of factors each tending to a limit λ with $|\lambda| < 1$ $\longrightarrow$ **Move:** pick r between $|\lambda|$ and 1; the tail is bounded by a geometric sequence

1990 B3

1990 B3. Let S be a set of 2×2 integer matrices whose entries a_{ij} (1) are all squares of integers and, (2) satisfy $a_{ij} \leq 200$. Show that if S has more than 50387 $(= 15^4 - 15^2 - 15 + 2)$ elements, then it has two elements that commute.

Fields: combinatorics (pigeonhole), linear algebra (2×2 matrices) **Position:** B3 – middle of the session; the difficulty is entirely in reading the number **Answer:** Proved: there are 15^4 admissible matrices, and they contain disjoint pairwise-commuting families whose sizes force any pairwise non-commuting set to have at most 50387 (indeed at most 50192) elements.

Layer 1 THE WHAT — surface analysis

The natural attempt. You start with the matrices. The squares of integers that are at most 200 are $0, 1, 4, \ldots, 196$, fifteen of them, so there are $15^4 = 50625$ admissible matrices. Then you ask when two matrices commute. Writing $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, $B = \begin{pmatrix} e & f \\ g & h \end{pmatrix}$ and comparing AB with BA entry by entry, you get

$$bg = cf, \qquad (a-d)f = b(e-h), \qquad (a-d)g = c(e-h),$$

that is, the vectors $(b, c, a-d)$ and $(f, g, e-h)$ are parallel (or one is zero). Now the problem looks like: the largest set of matrices with pairwise non-parallel vectors $(b, c, a-d)$. You start to classify the directions that occur, count how many matrices share each direction, and try to find the maximum. That is a genuinely hard counting problem – the directions $(b, c, a-d)$ with b, c squares and $a-d$ a difference of squares are messy – and it is not what is being asked. You have set out to compute the exact extremal number when the problem only asks you to beat a given one.

The model chain. Let M be the set of all 2×2 matrices whose entries lie in $Q = \{0, 1, 4, \ldots, 196\} = \{k^2 : 0 \le k \le 14\}$ (these are exactly the squares ≤ 200, since $14^2 = 196$ and $15^2 = 225$). Then $|M| = 15^4 = 50625$, and $S \subseteq M$. "Two elements" means two distinct elements.

Two commuting families. Let

$$D = \left\{ \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix} : a, d \in Q \right\}, \qquad T = \left\{ \begin{pmatrix} a & b \\ b & a \end{pmatrix} : a, b \in Q,\ b \neq 0 \right\}.$$

Diagonal matrices commute with one another, so any two elements of D commute. Every element of T equals $aI + bJ$ with $J = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, and any two polynomials in the same matrix J commute, so any two elements of T commute. The families are disjoint (elements of T have a nonzero off-diagonal entry), with $|D| = 15^2 = 225$ and $|T| = 15 \cdot 14 = 210$.

Pigeonhole. Suppose S has no two distinct commuting elements. Then S contains at most one element of D and at most one element of T, so it omits at least 224 elements of D and at least 209 elements of T. Hence

$$|S| \le 15^4 - 224 - 209 = 50192 < 50388.$$

So a set with more than 50387 elements has two distinct elements that commute.

Where the constant comes from. Using D together with the fifteen matrices $aI + J$ ($a \in Q$), which commute pairwise and are disjoint from D, the same argument gives $|S| \le 15^4 - (15^2 - 1) - (15 - 1) = 15^4 - 15^2 - 15 + 2 = 50387$. That is exactly the problem's number: the setters' bound is what one diagonal family plus one fifteen-element line of commuting matrices yields.

Caveat. The argument above proves the stated result and a stronger one (50192 in place of 50387), and both families were checked by machine to be pairwise commuting. It does not determine the true largest size of a pairwise non-commuting subset of M; more commuting families (for instance other directions of $(b, c, a-d)$) would lower the bound further. The problem does not ask for the sharp value, and none is claimed here.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths count $|M| = 15^4$. The natural attempt then characterises commuting pairs in general. The model reads the constant instead: $15^4 - 50387 = 238 = (15^2 - 1) + (15 - 1)$, and asks which families of pairwise-commuting matrices of sizes 15^2 and 15 are sitting inside M.

> *The number* 50387 *is* 15^4 *minus something small. So the problem is: show that S must miss at least* 238 *matrices, and "S has no commuting pair" can only force misses through families that commute among themselves. I need commuting families, not a classification.*

> $238 = (225 - 1) + (15 - 1)$. *A family of* $225 = 15^2$ *pairwise-commuting matrices is two free entries: the diagonal ones. A family of* 15 *is one free entry: aI plus a fixed non-diagonal matrix.*

> *Diagonal plus $\{aI + J\}$ gives exactly* 238.

Layer 3 THE WHY

Why read the constant first? Because of what a statement of the form "if $|S| > N$ then S contains a pair with property P" is. Its contrapositive is "a set with no P-pair has at most N elements", which is an upper bound on an independent set. The universal tool for upper-bounding independent sets is pigeonhole: cover the universe by groups inside which every pair has property P; an independent set takes at most one from each group, so its size is at most the number of groups. With $|M|$ fixed, the bound is $|M|$ minus the total "excess" $\sum(|\text{group}| - 1)$. The problem hands you N, so it hands you the excess you need: $15^4 - 50387 = 238$. Once the problem is read this way, the remaining question is only "which matrices visibly commute?", and a student who has done any linear algebra answers at once: diagonal matrices, and polynomials in a single matrix.

Why are those the natural families? A matrix commutes with every polynomial in itself: A commutes with $aI + bA$ for all scalars a, b. So for any fixed non-scalar matrix A, the set $\{aI + bA\}$ is a commuting family, and it has two free parameters. The diagonal matrices are the case $A = \text{diag}(1, 0)$, since $\text{diag}(a, d) = dI + (a - d)\,\text{diag}(1, 0)$; parametrised by their two diagonal entries, both entries range freely over Q, giving 15^2 matrices. The family $aI + bJ$ is the case $A = J$, and here the parameters are the entries themselves, so again both range over Q (with $b \neq 0$ to stay disjoint from D). That is where the size-225 families come from, and the shape of the constant 238 tells you the setters used one such family plus a one-parameter slice of another.

The general commutation criterion – $(b, c, a - d)$ parallel to $(f, g, e - h)$ – is the same fact seen from the other end: for non-scalar A, the matrices commuting with A are exactly the $aI + bA$. It is correct, and it is what you would need to find the sharp bound. But the problem asks for a threshold, not the extremum, and the cheap route is to exhibit enough commuting mass, not to classify it. This is "zombie maths" in its combinatorial form: computing the exact answer to a question that only needed an inequality.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the strange explicit constant $50387 = 15^4 - 15^2 - 15 + 2$ next to "more than". On seeing it, write down $|M| = 15^4$ and $15^4 - 50387 = 238 = (15^2 - 1) + (15 - 1)$.

2. Pigeonhole via a partial cover by commuting families: a set with no commuting pair takes at most one from each family.

3. Hesitation comes after deriving the commutation equations, when the classification of directions $(b, c, a - d)$ looks endless. It is reinforced away by rereading the goal: an inequality, so exhibit families; do not classify.

4. The trap is trying to compute the maximum size of a non-commuting set exactly. A smaller trap is miscounting the admissible entries: 0 is a square, and $14^2 = 196 \leq 200 < 225$, so there are 15 values, not 14.

5. When a problem gives a threshold of the form (total) − (small correction), compute the correction first and factor it as a sum of (family size − 1) terms; the families it suggests are the proof.

6. Decision rule: to prove "any large enough set contains a related pair", group the universe into blocks inside which all pairs are related and count the excess; the given threshold usually tells you the block sizes.

▷ **Cue:** "if $|S|$ exceeds an explicit constant N then S contains two related elements" $\longrightarrow$ **Move:** compute (universe size) $-N$ and interpret it as $\sum(|\text{block}| - 1)$ over blocks of pairwise-related elements

▷ **Cue:** matrices must be shown to commute $\longrightarrow$ **Move:** look for families of polynomials in one fixed matrix, $aI + bA$, including the diagonal matrices

▷ **Cue:** a count of matrices or tuples with entries restricted to squares up to a bound ⟶ **Move:** list the allowed values, including 0, and count them before anything else
▷ **Cue:** the question asks only to beat a threshold, not to find an extremum ⟶ **Move:** exhibit enough structure to force the inequality; do not classify

1990 B4

1990 B4. Let G be a finite group of order n generated by a and b. Prove or disprove: there is a sequence

$$g_1, g_2, g_3, \ldots, g_{2n}$$

such that

(1) every element of G occurs exactly twice, and

(2) g_{i+1} equals $g_i a$ or $g_i b$ for $i = 1, 2, \ldots, 2n$. (Interpret g_{2n+1} as g_1.)

Fields: combinatorics (directed graphs, Eulerian circuits), group theory **Position:** B4 – opens the hard half of the afternoon; in the room it was hard (of 201 top papers tallied by Kedlaya–Poonen–Vakil, 17 scored 7 or more and 179 scored 0 or left it blank), although the whole difficulty is one reframing **Answer:** Prove: such a sequence always exists. List, in order, the starting points of the arrows of an Eulerian circuit in the digraph on G with an arrow $g \to ga$ and an arrow $g \to gb$ for every g.

Layer 1 THE WHAT — surface analysis

The natural attempt. You read "prove or disprove" and start building the sequence by hand. Begin at e and multiply by a: $e, a, a^2, \ldots$, round the cyclic subgroup $\langle a \rangle$ and back to e. Now you want to jump with b into a new coset, go round that, and eventually come round everything a second time. Small cases first. $G = \mathbb{Z}_n$ with $a = b = 1$: the sequence $0, 1, \ldots, n-1, 0, 1, \ldots, n-1$ works. S_3 with $a = (12)$, $b = (123)$: after a few minutes of trial and error you find one. But every general recipe ("go round the a-cycles, jump with b when you have to") collapses in the bookkeeping: a b-jump lands you in the middle of an a-cycle, some of whose elements you have already used twice, and you have no rule for what to do next.

Then you do the sensible thing and try a smaller version of the question: can you visit every element *once*? Take $G = \mathbb{Z}_6$ with $a = 3$, $b = 2$ (these generate, since $\gcd(2, 3) = 1$). A closed walk of 6 steps with x steps of $+3$ and y steps of $+2$ returns to its start, so $3x + 2y \equiv 0 \pmod 6$ with $x + y = 6$, giving $x + 12 \equiv 0$, so $x \in \{0, 6\}$. All $+2$ steps visit only the even residues; all $+3$ steps visit only 0 and 3. So the "once" version is false. That makes "disprove" feel likely for "twice" too, and you spend the rest of the time hunting a counterexample among non-abelian groups. There is none.

The model chain. Build a directed multigraph D. Its vertices are the n elements of G. Its arrows are indexed by the pairs $(g, x) \in G \times \{a, b\}$: the arrow (g, x) runs from g to gx. So D has exactly $2n$ arrows. (If $a = b$ the two arrows out of g are parallel; if $a = e$ the arrow (g, a) is a loop. Both are allowed.)

Out-degree. Exactly two arrows start at g, namely (g, a) and (g, b).

In-degree. The arrow (k, a) ends at h if and only if $ka = h$, that is $k = ha^{-1}$: exactly one arrow labelled a ends at h. Likewise exactly one labelled b, from hb^{-1}. So every vertex has in-degree 2 as well.

Connectivity. Every $g \in G$ is a product $y_1 y_2 \cdots y_r$ with each $y_j \in \{a, a^{-1}, b, b^{-1}\}$. The walk $e,\ y_1,\ y_1 y_2,\ \ldots,\ g$ moves at each step along an arrow of D (if $y_j \in \{a, b\}$) or backwards along one (if $y_j \in \{a^{-1}, b^{-1}\}$: the step from k to ka^{-1} reverses the arrow (ka^{-1}, a), and likewise for b^{-1}). So D is weakly connected. (It is in fact strongly connected: G finite gives $a^{-1} = a^{\mathrm{ord}(a)-1}$, a forward word.)

Euler's theorem, directed form. A finite directed multigraph with at least one arrow, weakly connected, with in-degree equal to out-degree at every vertex, has a closed walk using every arrow exactly once. Proof. Call a walk using no arrow twice a *trail*. A trail started at v along unused arrows can only get stuck at v: each time it enters $u \neq v$ it has used one more arrow into u than out of u, and in-degree equals out-degree, so an unused arrow leaves u. By finiteness it stops, so closed trails exist; let W be one of maximal length. The arrows not on W are still balanced at every vertex, because W enters and leaves each vertex equally often. If some arrow is unused, weak connectivity gives a vertex u on W touching an unused arrow (otherwise the vertices of W with their arrows would form a whole connected piece, and every arrow would lie on W). Balance among unused arrows means u has an unused arrow *leaving* it; a trail of unused arrows from u returns to u, and splicing it into W at u gives a longer closed trail. Contradiction, so W uses every arrow.

Reading off the sequence. Let the Eulerian circuit be $\varepsilon_1, \ldots, \varepsilon_{2n}$ and let g_i be the start of ε_i. The end of ε_i is the start of ε_{i+1} (indices mod $2n$), and ε_i is (g_i, a) or (g_i, b), so $g_{i+1} \in \{g_i a, g_i b\}$, including $g_{2n+1} = g_1$: condition (2). Each g starts exactly two arrows, each used exactly once, so g occurs exactly twice: condition (1). The statement is true.

For $\mathbb{Z}_6$, $a = 3$, $b = 2$, where the "once" version failed, this procedure gives $0, 3, 0, 2, 5, 2, 4, 1, 3, 5, 1, 4$.

Caveat. Constructed here before the official solution was consulted; it then matched Kedlaya–Poonen–Vakil's (Cayley digraph, in- and out-degree 2, weak connectivity from generation, Eulerian circuit, startpoints of arcs). The construction was also run by machine (Hierholzer's algorithm plus a direct check of (1) and (2)) on all 8172 ordered generating pairs of subgroups of $S_2, \ldots, S_5$ of order at most 60, with no failures; the $\mathbb{Z}_6$ sequence above was checked the same way.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths read condition (2) the same way: consecutive terms differ by right multiplication by a or b. At that point the natural attempt treats the sequence as a list of *elements* to be arranged – a vertex-visiting problem. The model treats it as a list of *steps*: $2n$ steps, each leaving the current element by a or by b – and there are exactly $2n$ such possible steps in the whole group. The split is the moment you count steps instead of elements.

> *Every element appears twice, and out of every element there are exactly two allowed moves. That's $2n$ slots and $2n$ moves $g \to ga$, $g \to gb$. If I use every move exactly once, condition (1) is automatic. So I'll look for an Euler circuit, not a Hamilton-style arrangement.*

> *Euler needs in equals out. Out is 2. Who steps into h? Only ha^{-1} and hb^{-1}. That's 2. Connected because a, b generate. Done.*

> *The "once" version is false for $\mathbb{Z}_6$ with steps 3 and 2. So the "twice" is not decoration – it is exactly what makes the visit count match the out-degree. I shouldn't let the failed smaller version push me towards "disprove".*

Layer 3 THE WHY

Why is the edge view natural and not a trick? Go back to what a Hamilton cycle and an Euler circuit are. A Hamilton cycle visits every *vertex* once; an Euler circuit traverses every *arrow* once, and therefore passes through each vertex as many times as that vertex has arrows leaving it. Here every vertex has two arrows leaving it, and the problem asks for every vertex to be passed through twice. The numbers in the statement are the out-degree of the move

graph. When a problem prescribes a visit count per vertex, the first question is "what is the out-degree?", and when the answers agree the problem is almost certainly about edges.

Be precise about the direction of the implication, because it matters in a written proof. An Euler circuit *gives* a valid sequence. A valid sequence need not be an Euler circuit: in $G = \mathbb{Z}_2$ with $a = 1$, $b = 0$, the sequence $0, 1, 0, 1$ satisfies (1) and (2) while using the move $0 \to 0 + 1$ twice and the loop $0 \to 0 + 0$ never. The counting does not force Euler; it suggests it as the candidate to try, and the candidate then works.

Why do the degree conditions hold? This is the only place "group" is used. Right multiplication $g \mapsto ga$ is a bijection of G, because it has the inverse $g \mapsto ga^{-1}$. A bijection sends exactly one arrow into each vertex. Two bijections, two arrows in. Take away inverses – two arbitrary maps on a finite set – and in-degrees can be anything, so the claim fails. The group axiom that inverses exist is precisely what balances the graph.

Why is connectivity supplied by "generated by"? Generation means every element is a word in $a^{\pm 1}, b^{\pm 1}$, which is a walk that may go against arrows: weak connectivity. The directed Euler theorem needs only that, which is why Kedlaya–Poonen–Vakil write "D is weakly connected, since a and b generate G" and stop. If you only remember the undirected version you will want forward words, and then finiteness gives them, since a^{-1} is a positive power of a.

Why does the hand construction fail and the "once" version mislead? Hamilton problems have no local criterion: the $\mathbb{Z}_6$ obstruction is a global parity-type count, and there is no general degree test. Euler problems are decided entirely by degrees and connectivity, which are local and checkable in two lines. The hand construction is an attempt to solve an Euler problem with Hamilton tools, and the "once" test case is a Hamilton problem that answers the wrong question. Choosing which of the two you are facing is the whole problem.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the pairing of "occurs exactly twice" with "$g_i a$ or $g_i b$": two visits, two choices. Write down: "$2n$ terms; $2n$ possible moves $g \to gx$; draw the digraph of moves".

2. Eulerian circuit in the Cayley digraph of G with respect to $\{a, b\}$: check in-degree = out-degree and connectivity.

3. Hesitation comes at "Hamilton or Euler?" and at the in-degree. Reinforce the first by comparing visit count with out-degree; reinforce the second by asking who can precede h – only ha^{-1} and hb^{-1}.

4. The trap is the phrase "prove or disprove" combined with a hand search that gets messy: you guess "disprove". A deeper trap is testing the "once" version, which really is false ($\mathbb{Z}_6$, steps 3 and 2), and extrapolating. A rigour trap is claiming that every valid sequence is an Euler circuit; only the converse is needed and only the converse is true.

5. Any time a closed walk must pass through each object exactly d times and each object has exactly d allowed successors, suspect an Euler circuit in the graph of moves. De Bruijn sequences are the same pattern.

6. Decision rule: before searching for an arrangement, decide whether the constraint is on vertices (Hamilton: no local test, look for structure or a counterexample) or on edges (Euler: check degrees and connectivity, then stop). If the per-vertex visit count equals the degree, it is on edges.

▷ **Cue:** a cyclic sequence in which every element appears exactly d times and each step is one of d allowed moves $\longrightarrow$ **Move:** reinterpret as an Eulerian circuit in the digraph of moves and read off the starting points of its arrows

▷ **Cue:** steps of the form $g \to gx$ in a group $\longrightarrow$ **Move:** right multiplication by x is a bijection, so each generator contributes in-degree 1 and out-degree 1 at every vertex

▷ **Cue:** "generated by" and a directed walk is needed $\longrightarrow$ **Move:** words in $x^{\pm 1}$ give weak connectivity, which the directed Euler theorem accepts; if forward words are wanted, use $x^{-1} = x^{\mathrm{ord}(x)-1}$ in a finite group

▷ **Cue:** a covering-walk problem whose smaller version (visit once) fails ⟶ **Move:** check whether the larger version is an Euler problem in disguise before extrapolating the failure

1990 B5

1990 B5. Is there an infinite sequence $a_0, a_1, a_2, \ldots$ of nonzero real numbers such that for $n = 1, 2, 3, \ldots$ the polynomial

$$p_n(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

has exactly n distinct real roots?

Fields: algebra (real polynomials), real analysis (intermediate value theorem, perturbation) **Position:** B5 – the hard end of the paper by position, but a two-idea problem once the right question is asked **Answer:** Yes. Start from $p_1(x) = 1 - x$ and at each stage add a term $a_{n+1}x^{n+1}$ so tiny that it cannot disturb the sign pattern already present, with its sign chosen so that one new root appears far out to the right. An explicit example is $a_n = (-1)^n 10^{-n^2}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You read "infinite sequence" and "for every n" and look for a formula: a known power series whose partial sums all have real roots. The first candidate is $e^{-x} = \sum(-1)^k x^k/k!$, since its coefficients are nonzero and alternate. Then $p_1 = 1 - x$ has one root, good; but

$$p_2(x) = 1 - x + \tfrac{1}{2}x^2$$

has discriminant $1 - 4 \cdot \frac{1}{2} = -1 < 0$, so no real roots at all. The wall is immediate. You try $\cos x$ – coefficients vanish at odd powers, forbidden. You try $\prod(1 - x/k)$ – but the coefficient of x in $(1 - x)(1 - x/2)$ is $-\frac{3}{2}$, not -1, so p_2 would change a_1, and the problem demands that p_{n+1} *extend* p_n, keeping every earlier coefficient.

So you build by hand. $p_1 = 1 - x$. For $p_2 = 1 - x + a_2 x^2$ you need $1 - 4a_2 > 0$, so $0 \neq a_2 < \frac{1}{4}$; take $a_2 = \frac{1}{10}$. For $p_3 = 1 - x + \frac{1}{10}x^2 + a_3 x^3$ you need three distinct real roots, and you reach for the cubic discriminant

$$\Delta = 18abcd - 4b^3 d + b^2 c^2 - 4ac^3 - 27a^2 d^2$$

with $(a, b, c, d) = (a_3, \frac{1}{10}, -1, 1)$. It is a quadratic inequality in a_3; you solve it; you get an interval. Then p_4 needs the quartic discriminant, which is a polynomial of degree 6 in the coefficients with sixteen terms, and positivity of the discriminant is not even sufficient for four real roots of a quartic. Here the attempt dies – not from a mistake, from a method that gets strictly worse at every step and has no induction in it.

The model chain. *Claim.* There is such a sequence. Put $a_0 = 1$, $a_1 = -1$, so $p_1(x) = 1 - x$ has exactly one real root. We construct $a_{n+1} \neq 0$ from p_n, for $n \geq 1$, keeping the induction hypothesis

$$(\mathrm{H}_n) \qquad p_n \text{ has degree } n \text{ and } n \text{ distinct real roots } x_1 < x_2 < \cdots < x_n.$$

Step 1: sample points with alternating signs. Since p_n has degree n and n distinct roots, $p_n(x) = a_n \prod_{j=1}^{n}(x - x_j)$, each root is simple, and p_n has no other zeros. Choose

$$\alpha_0 < x_1 < \alpha_1 < x_2 < \cdots < \alpha_{n-1} < x_n < \alpha_n.$$

Passing from α_{i-1} to α_i changes the sign of exactly one factor, $x - x_i$, so $p_n(\alpha_0), p_n(\alpha_1), \ldots, p_n(\alpha_n)$ are all nonzero and alternate in sign.

Step 2: a perturbation too small to matter at the sample points. Let

$$M = \max_{0\le i\le n} |\alpha_i|^{n+1} + 1, \qquad m = \min_{0\le i\le n} |p_n(\alpha_i)| > 0,$$

and choose $0 < \varepsilon < m/M$. Define

$$a_{n+1} = -\varepsilon \operatorname{sgn} p_n(\alpha_n) \neq 0, \qquad p_{n+1}(x) = p_n(x) + a_{n+1}x^{n+1}.$$

For each i, $|a_{n+1}\alpha_i^{n+1}| \le \varepsilon M < m \le |p_n(\alpha_i)|$, so $p_{n+1}(\alpha_i)$ has the same (nonzero) sign as $p_n(\alpha_i)$. Hence the values $p_{n+1}(\alpha_0), \dots, p_{n+1}(\alpha_n)$ still alternate.

Step 3: count roots. By the intermediate value theorem p_{n+1} has a root in each open interval (α_{i-1}, α_i), $i = 1, \dots, n$: that is n roots. Next, p_{n+1} has degree $n+1$ with leading coefficient a_{n+1}, so for large x its sign is $\operatorname{sgn} a_{n+1} = -\operatorname{sgn} p_n(\alpha_n) = -\operatorname{sgn} p_{n+1}(\alpha_n)$. So there is another root in (α_n, ∞). The $n+1$ intervals are disjoint, so these $n+1$ roots are distinct. A polynomial of degree $n+1$ has at most $n+1$ roots, so it has exactly $n+1$, all real and distinct: (H_{n+1}) holds.

Every a_k is nonzero, every p_n extends the previous one, and induction gives the sequence.

Explicit version. Let $a_k = (-1)^k 10^{-k^2}$. For $0 \le k \le n$,

$$(-1)^k 10^{-k^2} p_n(10^{2k}) = \sum_{i=0}^{n} (-1)^{i-k} 10^{-(i-k)^2} = 1 + \sum_{\substack{j=-k \\ j\neq 0}}^{n-k} (-1)^j 10^{-j^2},$$

using $-i^2 + 2ik = k^2 - (i-k)^2$. The tail has absolute value at most $2\sum_{j\ge 1} 10^{-j^2} < 2(0.1 + 0.001 + 0.001) < 1$, so $\operatorname{sgn} p_n(10^{2k}) = (-1)^k$. The $n+1$ values at $1, 10^2, \dots, 10^{2n}$ alternate, giving n roots in disjoint intervals; degree n caps it at n.

Caveat. Kedlaya–Poonen–Vakil give exactly these two solutions (inductive perturbation; $a_n = (-1)^n 10^{-n^2}$). Both were constructed here first and then compared. Both were also checked by machine with exact rational arithmetic: the explicit sequence gives n distinct real roots for $n \le 8$ (Sturm count), and the inductive recipe, run with α_i at gap midpoints, produced $a_2 = \frac{1}{10}$, $a_3 = -10^{-4}$, $a_4 = 10^{-11}$, …, each p_n having n real roots for $n \le 8$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree on $p_1 = 1 - x$ and on wanting p_2 to have two roots. They split at the question asked about a_{n+1}. The natural attempt asks "for exactly which a_{n+1} does p_{n+1} have $n+1$ real roots?" – an exact characterisation, hence a discriminant. The model asks "can I choose a_{n+1} so small that it breaks nothing, and still gain a root?" – a sufficiency question, hence signs and the intermediate value theorem.

> *I do not need the set of all good a_{n+1}. I need one. What is the cheapest one? A tiny one: if a_{n+1} is tiny, p_{n+1} looks like p_n wherever x is not huge, so the n old roots survive.*

> *Roots are fragile but sign changes are not. If p_n is positive here and negative there, a small enough nudge keeps it positive here and negative there. So I should record the roots as sign changes at fixed points α_i, not as the roots themselves.*

> *Where does the extra root come from? Far away, where x^{n+1} beats everything else however small its coefficient. So choose the sign of a_{n+1} to disagree with what p_n is doing at the right end.*

Layer 3 THE WHY

Why ask for sufficiency, not characterisation? Read the quantifiers. "Is there a sequence" is an existence question, and the sequence is built term by term with the earlier terms frozen. That is an induction whose step needs *one* valid a_{n+1}, not all of them. A discriminant answers "all of them"; it is the wrong tool for the question asked, and it is zombie maths in its purest form – the most primal manipulation, undertaken because it is the thing you know for "how many real roots". The moment you see that the step only needs existence, the freedom to make a_{n+1} as small as you like is the obvious free parameter to spend.

Why sign changes, not roots? This is where the argument reaches its axioms. A root is the solution of an equation, $p(x) = 0$ – a closed, knife-edge condition that any perturbation can destroy (a double root can vanish into the complex plane). A strict sign, $p(\alpha) > 0$, is an open condition: by the definition of continuity (here, of $a \mapsto p_n(\alpha) + a\alpha^{n+1}$ in a), it survives every small enough change. The intermediate value theorem then converts "strict sign change between α_{i-1} and α_i" back into "a root in between". So the only way to carry roots through a perturbation is to translate them into strict inequalities, perturb, and translate back. This is exactly why distinctness matters in the hypothesis: simple roots are exactly the ones where p_n genuinely changes sign, so n distinct roots of a degree-n polynomial give n robust sign changes. Once you have said "roots are fragile, strict inequalities are robust", the choice of sample points α_i is not a trick. It is the only way to write the hypothesis in a form that survives the step.

Why is there a new root at all? Degree. The one thing a tiny $a_{n+1}x^{n+1}$ does change is the behaviour at infinity: however small ε is, εx^{n+1} eventually dominates $a_n x^n$ (for $x > |a_n|/\varepsilon$ roughly). So the sign at $+\infty$ is ours to choose, independently of everything near the old roots. Choose it to disagree with $p_n(\alpha_n)$ and the intermediate value theorem hands you root number $n+1$. The machine check shows where it lands: near $-a_n/a_{n+1}$, the point where the two top terms balance – 8.87, then 990, then 10^7, then 10^{15}. The roots march off to infinity, which is exactly why the coefficients have to shrink so fast.

Where the explicit example comes from. Kedlaya–Poonen–Vakil's second solution sets $a_k = (-1)^k 10^{-k^2}$ without comment. Here is where that comes from. The inductive construction says each new root lives where two consecutive terms balance, and between two such balance points one single term should dominate the whole sum. At $x = 10^{2k}$ the i-th term has size $10^{k^2-(i-k)^2}$: a Gaussian bump in i centred at $i = k$, falling off so fast that the k-th term outweighs all others combined. So $p_n(10^{2k})$ has the sign of its k-th term, $(-1)^k$. The quadratic exponent k^2 is chosen precisely because $\log|a_k|$ must be concave with gaps between consecutive slopes big enough for one term to win – the "Newton polygon" picture of their remark. The explicit solution is the inductive one with the ε's frozen into a formula.

Caveat. A sharper known criterion (D. C. Kurtz, *Amer. Math. Monthly* 1992): if all $a_k > 0$ and $a_k^2 > 4a_{k-1}a_{k+1}$ for $1 \le k \le n-1$, then $\sum a_k x^k$ has n distinct real (negative) roots. It is quoted as context only and was not proved or re-verified here; nothing in the solution above depends on it.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that the sequence is *infinite* and each p_{n+1} is p_n plus one term. Write down: "induction; the step is $p_{n+1} = p_n + a_{n+1}x^{n+1}$ with p_n frozen".

2. A perturbation argument: make a_{n+1} small, keep the old sign changes via the intermediate value theorem, and gain one root at infinity from the new leading term.

3. Hesitation occurs at "how do I know the old roots survive?". It is reinforced away by replacing roots with sample points α_i where p_n has strict, alternating signs – strict inequalities survive small changes.

4. Two traps. First, hunting for a closed-form power series (e^{-x} fails already at p_2; $\cos x$ has zero coefficients; products of linear factors change earlier coefficients). Second, the discriminant:

exact root-counting criteria explode in degree and are the wrong kind of statement for an existence induction.

5. Recognise it by the shape "build an infinite object term by term, each finite stage must have property P". If P is defined by strict inequalities or can be witnessed by them, the step is a small-perturbation step.

6. Decision rule: to preserve a property through a perturbation, first rewrite it as finitely many strict inequalities (open conditions); then any small enough change preserves it, and you are free to spend the remaining freedom on the new requirement.

▷ **Cue:** an infinite sequence where every finite truncation must have a property, with earlier terms frozen $\longrightarrow$ **Move:** induct and ask for one good next term, not all of them; make it as small as needed

▷ **Cue:** need to preserve the real roots of a polynomial under a change $\longrightarrow$ **Move:** replace roots by points α_i with strict alternating signs; a small perturbation keeps the signs and the intermediate value theorem returns the roots

▷ **Cue:** adding a tiny top-degree term εx^{n+1} $\longrightarrow$ **Move:** it is negligible on any bounded set but controls the sign at infinity; use that to create exactly one new root beyond the old ones

▷ **Cue:** "exactly n distinct real roots" for a degree-n polynomial $\longrightarrow$ **Move:** only a lower bound is needed: find n sign changes in disjoint intervals; degree gives the upper bound

▷ **Cue:** a polynomial whose coefficients shrink like 10^{-k^2} $\longrightarrow$ **Move:** at $x = 10^{2k}$ the k-th term dominates all others; evaluate there to read off signs

1990 B6

1990 B6. Let S be a nonempty closed bounded convex set in the plane. Let K be a line and t a positive number. Let L_1 and L_2 be support lines for S parallel to K, and let $\overline{L}$ be the line parallel to K and midway between L_1 and L_2. Let $B_S(K,t)$ be the band of points whose distance from $\overline{L}$ is at most $(t/2)w$, where w is the distance between L_1 and L_2. What is the smallest t such that

$$S \cap \bigcap_K B_S(K,t) \neq \emptyset$$

for all S? (K runs over all lines in the plane.)

Fields: convex geometry, integral calculus (centroids, concave functions) **Position:** B6 – the last problem of the paper, and genuinely hard: very few contestants scored on it **Answer:** $t = \frac{1}{3}$. A triangle shows nothing smaller works, and for $t = \frac{1}{3}$ the centroid of S lies in every band, because the centroid of a plane convex set always sits between $\frac{1}{3}$ and $\frac{2}{3}$ of the way across the set, in every direction.

Caveat. The project copy `putnam_1990_problems.tex` reads "parallel to K_1"; this is a typo, and the Kedlaya–Poonen–Vakil book reads "parallel to K", as reproduced above.

Layer 1 THE WHAT — surface analysis

The natural attempt. You unpack the definition. Put coordinates so that K is horizontal; S then fills the strip $0 \le y \le w$ exactly, touching both edges, and the band $B_S(K,t)$ is $\frac{1-t}{2}w \le y \le \frac{1+t}{2}w$ – the middle fraction t of the strip. So the question is: how thin a middle slice, in *every* direction at once, is still guaranteed to catch a common point of S?

You try shapes. A disc, a square, any centrally symmetric set: the centre lies on every midline, so every $t > 0$ works. That tells you the answer is decided by lopsided sets. You try a triangle ABC and take K parallel to BC: the band is the middle slice between BC and the apex A. Take the three sides in turn and ask where the three slices meet. It is not obvious from a picture, and here many students stall – they draw it, it looks like the slices meet in a small

hexagon or triangle near the middle, and they guess $t = \frac{1}{2}$ or try to push the picture further with coordinates for general K, a direction at a time, which leads nowhere: there are infinitely many K and no candidate point to test them on.

The model chain. Notation. For a direction K, measure the signed distance y perpendicular to K, normalised so that $S \subseteq \{0 \le y \le w\}$ with both lines $y = 0$, $y = w$ touching S. Then $B_S(K,t) = \{\frac{1-t}{2}w \le y \le \frac{1+t}{2}w\}$. The bands grow with t, so the set of good t is an up-interval; we show it is $[\frac{1}{3}, \infty)$.

Lower bound: $t < \frac{1}{3}$ fails. Let S be a triangle ABC and write every point as $P = \alpha A + \beta B + \gamma C$ with $\alpha + \beta + \gamma = 1$ (barycentric coordinates). If h_A is the height from A, the distance from P to line BC is αh_A for P on the same side as A (the distance is affine in P, 0 on BC, h_A at A). For $K \parallel BC$ the support lines are BC and the parallel through A, so $w = h_A$ and $P \in B_S(K,t)$ iff $\frac{1-t}{2} \le \alpha \le \frac{1+t}{2}$. Doing the same for the other two sides, a point in all three bands (and on the A-side of BC etc.) has

$$1 = \alpha + \beta + \gamma \;\ge\; 3 \cdot \tfrac{1-t}{2}, \qquad \text{so } t \ge \tfrac{1}{3}.$$

A point on the far side of BC has distance $\ge h_A$ from the far line, so is outside the band anyway. Hence for $t < \frac{1}{3}$ the three bands alone have empty intersection; at $t = \frac{1}{3}$ they meet only at $\alpha = \beta = \gamma = \frac{1}{3}$, the centroid.

Upper bound: $t = \frac{1}{3}$ works. Degenerate cases first. If S is a point, it lies on every $\overline{L}$. If S is a segment (zero area), its midpoint projects onto the midpoint of the projection of S in every direction, so it lies on every $\overline{L}$. Otherwise S has positive area; let G be its centroid.

(i) $G \in S$. If not, since S is closed and convex there is a linear function f with $f(G) > \max_S f$ (separation). But $f(G)$ is the average of f over S, which is at most $\max_S f$. Contradiction.

(ii) For each K, G lies in $\frac{1}{3}w \le y \le \frac{2}{3}w$, i.e. in $B_S(K, \frac{1}{3})$. Let $\ell(y)$ be the length of the cross-section $S \cap \{\text{height } y\}$, for $0 \le y \le w$. By Cavalieri, the height of G is

$$\bar{y} = \frac{\int_0^w y\,\ell(y)\,dy}{\int_0^w \ell(y)\,dy}.$$

ℓ is concave. If $[a_1, b_1]$ is the section at height y_1 and $[a_2, b_2]$ at y_2, then by convexity S contains, at height $sy_1 + (1-s)y_2$, both $sa_1 + (1-s)a_2$ and $sb_1 + (1-s)b_2$, hence $\ell(sy_1 + (1-s)y_2) \ge s\ell(y_1) + (1-s)\ell(y_2)$.

Comparison with a triangle. Let $g(y) = c(w - y)$ with c chosen so that $g(\frac{w}{3}) = \ell(\frac{w}{3})$, and put $h = \ell - g$: concave (concave minus linear), $h(\frac{w}{3}) = 0$, $h(w) = \ell(w) \ge 0$. For $\frac{w}{3} \le y \le w$, h lies above its chord between $\frac{w}{3}$ and w, so $h(y) \ge 0$. For $0 \le y < \frac{w}{3}$, the point $\frac{w}{3}$ lies between y and w, so $0 = h(\frac{w}{3}) \ge s\,h(y) + (1-s)h(w)$ with $s \in (0,1]$, forcing $h(y) \le 0$. Thus $(y - \frac{w}{3})h(y) \ge 0$ on all of $[0, w]$. Also $\int_0^w (y - \frac{w}{3})(w - y)\,dy = 0$, because $\int_0^w y(w-y)\,dy = \frac{w^3}{6}$ and $\frac{w}{3}\int_0^w (w-y)\,dy = \frac{w}{3} \cdot \frac{w^2}{2} = \frac{w^3}{6}$. Therefore

$$\int_0^w (y - \tfrac{w}{3})\,\ell(y)\,dy = \underbrace{\int_0^w (y - \tfrac{w}{3})\,g(y)\,dy}_{=0} + \int_0^w (y - \tfrac{w}{3})\,h(y)\,dy \;\ge\; 0,$$

i.e. $\bar{y} \ge \frac{w}{3}$. Applying the same argument after reflecting $y \mapsto w - y$ gives $\bar{y} \le \frac{2w}{3}$.

So $G \in S \cap \bigcap_K B_S(K, \frac{1}{3})$, and the smallest t is $\frac{1}{3}$.

Caveat. Verified here by hand as written and by machine: for 3000 random convex polygons and 60 directions each, the centroid's relative position across the width stayed in $[0.33333, 0.66666]$, approaching both ends (near-triangles). Kedlaya–Poonen–Vakil prove (ii) differently, by a continuity/area-exchange argument that pushes mass of S into a triangle

QP_1P_2 without raising the centroid; the concave-section argument above is an alternative, constructed independently and checked line by line.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree on the translation "$B_S(K,t)$ is the middle fraction t of the strip" and on trying a triangle. They split at the question asked of the triangle. The natural attempt asks "what does the intersection of the three slices look like?" and tries to read it off a picture. The model asks "which point would I put in every slice, and what number does that point force?" – and writes barycentric coordinates, in which "distance to side BC as a fraction of the height" is literally the coordinate α.

> *A triangle's width in the direction of a side is just its height, and a point's distance to that side, as a fraction of the height, is its barycentric coordinate. The three coordinates add to* 1*, so they can't all be bigger than* $\frac{1}{3}$*. That's the answer:* $\frac{1}{3}$*, and the only survivor is the centroid.*

> *Infinitely many directions, so I can't hunt for the point direction by direction. I need one canonical point of S that doesn't depend on K and behaves well in every direction. The triangle just told me which: the centroid.*

> *Centroid height is a weighted average of y with weight the section length* $\ell(y)$*. Convexity makes* ℓ *concave. The worst concave weight is a triangle pointing the other way. So I compare* ℓ *with the linear function that agrees with it at the critical height* $\frac{w}{3}$*.*

Layer 3 THE WHY

Why a triangle? The problem asks for a bound valid for *all* S, so the answer is set by the worst S. Centrally symmetric sets are the best case (the centre lies on every midline), so the worst case is the most lopsided convex set; in the plane that is the triangle, the set with the fewest vertices, where every direction has a vertex on one side and an edge on the other. Testing extremes is not a hunch; it is what "for all S" demands.

Why barycentric coordinates? Read the definition of the band: distance from a line, measured as a fraction of the width. Distance from a line is an affine function of the point, 0 on the line; for a triangle and $K \parallel BC$ that affine function, normalised by the width, takes the value 1 at A and 0 at B, C – which is the definition of the barycentric coordinate α. So the problem, restricted to the three side-directions, is literally "find α, β, γ in $[\frac{1-t}{2}, \frac{1+t}{2}]$ with sum 1". The number $\frac{1}{3}$ is forced by $\alpha + \beta + \gamma = 1$, nothing more.

Why the centroid in general? The problem is a "one point, infinitely many conditions" problem. The only way to satisfy infinitely many direction-dependent conditions with a single point is to use a point defined without reference to any direction – an affine-invariant "centre" of S. The band condition is itself affine in flavour (ratios of distances along parallel lines), so the centre must be one that respects affine maps: the centroid does (averaging commutes with affine maps), and it is exactly the survivor in the triangle. Kedlaya's motivation says "the triangle suggests the centroid"; this is why it is not a guess.

Why concavity of sections, and why compare at $\frac{w}{3}$*?* The centroid's height is an average of y weighted by $\ell(y)$. Convexity of S is the only hypothesis, and its one-dimensional shadow is exactly concavity of ℓ – that is the whole content of convexity in this direction. Among nonnegative concave weights on $[0, w]$, the one pulling the average lowest puts as much weight as possible near $y = 0$ and none at $y = w$: the linear weight $c(w - y)$, a triangle with apex on top, whose average is $\frac{w}{3}$. To show every other concave ℓ is no worse, compare it with that linear one at the single height where signs change, $y = \frac{w}{3}$: concavity then forces $\ell - g$ to have the

same sign as $y - \frac{w}{3}$, so the difference can only push the average up. Kedlaya–Poonen–Vakil's area-swapping argument does the same thing geometrically: it trades the part of S that is "too high" for a triangle whose centroid is known. The analytic version shows why it works: a concave function crosses a line at most twice, and we have pinned one crossing.

Why is $t = \frac{1}{3}$ attained, so "smallest" makes sense? The bands are closed and the centroid argument gives non-strict inequalities, so $t = \frac{1}{3}$ itself works. The triangle shows equality: its centroid is exactly at $\frac{1}{3}$ in the three side-directions.

Layer $+\alpha$ BACKTESTING

1. The first cue is "for all S" with "smallest t": an extremal problem over shapes. Write down: "answer = worst shape; best shape is symmetric; try a triangle". The second cue is the band definition in fractions of the width: write $y \in [\frac{1-t}{2}w, \frac{1+t}{2}w]$.

2. Two halves: a construction (triangle, barycentric sum) for the lower bound, and one canonical point (the centroid) for the upper bound, proved by a one-dimensional weighted-average estimate.

3. Hesitation occurs at "infinitely many K – how can one point work for all?". It is reinforced away by naming a point defined without reference to K. Second hesitation: "how do I bound the centroid?". Reduce to the section-length function and use concavity.

4. The trap is analysing the intersection of bands geometrically, direction by direction, instead of choosing a candidate point. A second trap: guessing $t = \frac{1}{2}$ from a rough triangle picture, or proving the centroid bound only for polygons.

5. Any "one point must satisfy a condition for every direction/line" problem wants an intrinsic centre (centroid, circumcentre, incentre, centre of symmetry), chosen by testing the extremal shape; the centroid is the one for conditions stated as ratios along parallel lines.

6. Decision rule: to bound a weighted average whose weight is only known to be concave, compare the weight with the linear function that agrees with it at the target point; concavity fixes the sign of the difference on each side, and the linear case is computed directly.

▷ **Cue:** "smallest constant that works for every convex set" $\longrightarrow$ **Move:** test the symmetric case (best) and the triangle (worst); the triangle usually gives the answer

▷ **Cue:** distance to a side of a triangle as a fraction of the height $\longrightarrow$ **Move:** that ratio is a barycentric coordinate; use $\alpha + \beta + \gamma = 1$

▷ **Cue:** one point must lie in a set for every direction K $\longrightarrow$ **Move:** choose a point defined independently of K – the centroid for affine, ratio-type conditions

▷ **Cue:** position of the centroid of a convex region along a direction $\longrightarrow$ **Move:** write $\bar{y} = \int y\ell(y)\,dy / \int \ell$ with ℓ the section length, which convexity makes concave

▷ **Cue:** an integral inequality with a concave weight and an extremal linear weight $\longrightarrow$ **Move:** subtract the linear function matching at the sign-change point; concavity makes the product of signs nonnegative

▷ **Cue:** need a centroid to lie inside a convex set $\longrightarrow$ **Move:** if not, separate by a line; the average of a linear function cannot exceed its maximum