DAWN PUTNAM SESSION

Putnam 1966 — Papers A and B

Putnam 120 Archive

Contents

Putnam 1966 — Paper A **5**
1966 A1 5
1966 A2 7
1966 A3 9
1966 A4 12
1966 A5 14
1966 A6 17

Putnam 1966 — Paper B **21**
1966 B1 21
1966 B2 23
1966 B3 26
1966 B4 29
1966 B5 31
1966 B6 33

Putnam 1966 — Paper A

1966 A1

1966 A1. Let $f(n)$ be the sum of the first n terms of the sequence $0, 1, 1, 2, 2, 3, 3, 4, \ldots$, where the nth term is given by

$$a_n = \begin{cases} n/2 & \text{if } n \text{ is even,} \\ (n-1)/2 & \text{if } n \text{ is odd.} \end{cases}$$

Show that if x and y are positive integers and $x > y$ then $xy = f(x+y) - f(x-y)$.

Caveat. The MAA 1965–1984 OCR renders the case definition of a_n as "S n/2 if n is even, 1 (n - 1)/2 if n is odd", where the "S" and stray "1" are font-substitution artifacts of a large left brace in the original typeset cases-bracket, not additional terms. The sequence $0, 1, 1, 2, 2, 3, 3, 4, \ldots$ printed just before it fixes the reading beyond doubt: $a_1 = 0, a_2 = 1, a_3 = 1, a_4 = 2, a_5 = 2, \ldots$, exactly $a_n = \lfloor n/2 \rfloor$, which is what the reconstructed cases give. The book's own solution confirms this reading, deriving $f(n) = n^2/4$ for even n and $f(n) = (n^2-1)/4$ for odd n – consistent with summing $\lfloor k/2 \rfloor$ for $k = 1, \ldots, n$. No mathematical content was altered.

Fields: Number theory, sequences (summation of a floor-type sequence) **Position:** A1 (opener) **Answer:** The identity holds identically; $f(n) = \lfloor n^2/4 \rfloor$, and $x+y$, $x-y$ always share parity, which is exactly what makes the subtraction collapse to xy.

Layer 1 THE WHAT — surface analysis

The natural attempt. The first job is a closed form for $f(n) = a_1 + a_2 + \cdots + a_n$ where $a_k = \lfloor k/2 \rfloor$. Splitting into even and odd k and summing each arithmetic run gives, after a short induction,

$$f(n) = \begin{cases} n^2/4 & n \text{ even} \\ (n^2-1)/4 & n \text{ odd.} \end{cases}$$

This much is unavoidable and correct. The trouble starts at the next step. The problem asks for $f(x+y) - f(x-y)$, and the closed form for f has *two* branches, so the honest first instinct is to case-split on the parities of $x+y$ and $x-y$ separately: (even, even), (even, odd), (odd, even), (odd, odd) – four cases, each requiring its own substitution into the piecewise formula and its own algebraic simplification. Two of the four cases are in fact vacuous, but nothing in the four-way split tells you that in advance; you find out only by grinding through each one and discovering that two produce a contradiction with x, y both being integers. The case (even, odd) forces $x+y$ and $x-y$ to have different parities, i.e. $(x+y) - (x-y) = 2y$ to be odd – which is false for every integer y – so that branch dies, but only after you have written it down and attempted the algebra. The same happens to (odd, even). You are left re-deriving, case by case and after the fact, a fact you could have used from the very first line.

The model chain. Before touching the formula for f, note that

$$(x+y)+(x-y)=2x,$$

which is even for every integer x. Two integers whose sum is even have the same parity. So $x+y$ and $x-y$ are automatically either both even or both odd – there are only two cases, not four, and which one occurs never needs to be determined, because the algebra is identical in both:

$$f(x+y)-f(x-y)=\frac{(x+y)^2-c}{4}-\frac{(x-y)^2-c}{4}=\frac{(x+y)^2-(x-y)^2}{4},$$

where $c=0$ in the even–even case and $c=1$ in the odd–odd case, and c cancels either way. Expanding,

$$(x+y)^2-(x-y)^2=4xy,$$

so $f(x+y)-f(x-y)=4xy/4=xy$, exactly as claimed, for every pair of positive integers $x>y$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains reach the same closed form $f(n)=\lfloor n^2/4\rfloor$ (with its two branches) before doing anything else with it. They split at the very next line: on what basis to decide which branch(es) of f apply to $f(x+y)$ and to $f(x-y)$. The natural attempt treats the two branch-choices as independent unknowns and enumerates all four combinations. The model chain instead first asks whether the two choices are related, notices $(x+y)+(x-y)=2x$ is even, and collapses the four combinations to two matched ones before writing a single case.

> *$x+y$ and $x-y$ are not two independent numbers whose parities I need to determine separately – they differ by $2y$, an even number, so whatever parity one has, the other has too. I should check that relationship first, before I let the piecewise formula for f tempt me into a four-way split.*

▷ **Cue:** a formula splits by parity of n (or any residue class), and you need to evaluate it at two related arguments such as $x+y$ and $x-y$ $\longrightarrow$ **Move:** first compute the difference (or sum) of the two arguments – here $(x+y)-(x-y)=2y$ – and check whether that difference is a multiple of the modulus; if so, the two arguments necessarily lie in the *same* residue class, and no case split is needed at all

Layer 3 THE WHY

The move is forced by what parity actually is: a partition of the integers into two classes under the equivalence $m\sim n \iff 2\mid(m-n)$. Two integers are automatically in the same class the instant their difference is shown to be even – that is the definition, not a coincidence to be verified after the fact by trying both branches and seeing which one is consistent. The problem itself manufactures this situation: it asks about f evaluated at $x+y$ and $x-y$, and these are not two arbitrary integers but two numbers built from the *same* pair (x,y) by addition and subtraction, so their difference, $2y$, is transparently even before any arithmetic about f is done. The four-case enumeration is not wrong, but it is solving a harder question than the one being asked: it treats "what is the parity of $x+y$?" and "what is the parity of $x-y$?" as two separate one-bit unknowns, when the problem's own construction has already correlated those two bits perfectly. Once you see that correlation, the piecewise formula for f stops being an obstacle – it becomes convenient, because whichever single case holds, the correction term c (0 or $-1/4$ against n^2) is identical on both sides of the subtraction and cancels regardless. The book's own solution states this in one line ("since $x+y$ and $x-y$ always have the same parity") and moves directly to the algebra; the archive's job is to say explicitly why that line is true and why it is the first thing to check, rather than a fact discovered by exhaustion.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that f is defined piecewise by parity, and the quantity being asked about, $f(x+y)-f(x-y)$, involves the same formula evaluated at two related arguments. On seeing it, write down the difference of those two arguments, $(x+y)-(x-y)=2y$, before writing down any cases.
2. That cue suggests checking whether $2y$ is a multiple of the modulus (here 2, since f is defined mod 2). It is, so the method is: conclude $x+y\equiv x-y \pmod 2$ immediately, reducing four potential cases to at most two, both algebraically identical after cancellation.
3. The hesitation happens right after deriving the closed form for $f(n)$: the piecewise structure visually invites a full case split on both arguments independently, because that is the mechanical way to "unpack" a cases-defined function. It is reinforced away by remembering that the two arguments were not chosen independently – they are both built from x and y – so their relationship should be checked before, not after, expanding cases.
4. The trap is the four-way case split: two of the four cases are self-contradictory (they would require $2y$ to be odd), and discovering this by brute enumeration wastes time and obscures the one-line reason the identity holds at all.
5. To recognise this faster: whenever a piecewise/parity-dependent object is evaluated at two arguments that are sums or differences of the same underlying variables, compute that sum or difference first and read off the parity relationship directly, before touching the piecewise formula.
6. Decision rule: before case-splitting on a residue-class-dependent formula evaluated at two related arguments p, q, always check $p-q$ (or $p+q$) against the modulus first. If it is divisible by the modulus, p and q lie in the same class automatically, and the apparent case split collapses to one case – turning a piecewise identity into a single clean computation.

1966 A2

1966 A2. Let a, b, c be the lengths of the sides of a triangle, let $p=(a+b+c)/2$, and let r be the radius of the inscribed circle. Show that

$$\frac{1}{(p-a)^2}+\frac{1}{(p-b)^2}+\frac{1}{(p-c)^2}\geq\frac{1}{r^2}.$$

Caveat. The MAA 1965–1984 OCR renders the exponents on the left-hand side as stray letters and marks – "(p - a) 2 (p - b)l (p - c)' " – with the superscript 2's dropped or turned into "l" and prime-like glyphs by font substitution. The reading above is standard and is confirmed unambiguously by the book's own solution, which explicitly sets $x=1/(p-a)$, $y=1/(p-b)$, $z=1/(p-c)$ and reduces the claim to $x^2+y^2+z^2\geq xy+yz+zx$ – a reduction that only makes sense if the original inequality has squares on every term, on both sides. No mathematical content was altered.

Fields: Geometry (triangle inequalities), algebra (sum of squares) **Position:** A2 **Answer:** The inequality holds for every triangle, with equality iff the triangle is equilateral; it reduces to the trivial fact that $x^2+y^2+z^2\geq xy+yz+zx$ for any reals x, y, z.

Layer 1 THE WHAT — surface analysis

The natural attempt. The statement is dressed entirely in triangle language – sides, semiperimeter, inradius – so the natural instinct is to attack it with triangle tools directly. Write $r=(p-a)\tan(A/2)=(p-b)\tan(B/2)=(p-c)\tan(C/2)$ (a standard incircle identity, since the tangent length from vertex A to the incircle is $p-a$ and the half-angle at A sees the inradius

across it). Substituting into the claim turns it into a statement purely about the half-angles $A/2, B/2, C/2$ with $A+B+C=\pi$:

$$\cot^2\frac{A}{2}\tan^2\frac{A}{2}\cdot(\text{something})\ \ldots$$

and at this point the algebra becomes a genuine trigonometric identity chase: three half-angle tangents constrained by $A/2+B/2+C/2=\pi/2$, an inequality to prove among their reciprocals-squared, no obvious symmetric simplification, and no clean stopping point. You can convince yourself numerically that it is probably true and probably reduces to something like the tangent half-angle product identity $\tan\frac{A}{2}\tan\frac{B}{2}+\tan\frac{B}{2}\tan\frac{C}{2}+\tan\frac{C}{2}\tan\frac{A}{2}=1$, but turning that identity into the stated inequality by trigonometric manipulation alone is slow, and it is easy to lose the thread of which half-angle identity is even the relevant one to reach for.

The model chain. The area of the triangle has two standard formulas: $A=pr$ (semiperimeter times inradius) and Heron's formula $A=\sqrt{p(p-a)(p-b)(p-c)}$. Equate and square:

$$p^2r^2=p(p-a)(p-b)(p-c)\implies pr^2=(p-a)(p-b)(p-c).$$

Now substitute $x=\dfrac{1}{p-a}, y=\dfrac{1}{p-b}, z=\dfrac{1}{p-c}$, all positive since $p-a, p-b, p-c>0$ in any triangle. Then $(p-a)(p-b)(p-c)=\dfrac{1}{xyz}$, and $p=(p-a)+(p-b)+(p-c)=\dfrac{1}{x}+\dfrac{1}{y}+\dfrac{1}{z}$. Substituting into $pr^2=(p-a)(p-b)(p-c)$ gives

$$r^2\left(\frac{1}{x}+\frac{1}{y}+\frac{1}{z}\right)=\frac{1}{xyz}\implies\frac{1}{r^2}=xyz\left(\frac{1}{x}+\frac{1}{y}+\frac{1}{z}\right)=yz+xz+xy.$$

So the original inequality is now exactly

$$x^2+y^2+z^2\geq xy+yz+zx,$$

which follows immediately from the three trivial inequalities $x^2+y^2\geq 2xy$, $y^2+z^2\geq 2yz$, $z^2+x^2\geq 2zx$, added and halved. Equality holds iff $x=y=z$, i.e. $p-a=p-b=p-c$, i.e. $a=b=c$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains begin by recognising that $p-a, p-b, p-c$ and r are tied together by *some* triangle identity, and both need one such identity to make progress. They split at the choice of *which* identity to reach for: the natural attempt reaches for the half-angle tangent identities that describe r vertex-by-vertex ($r=(p-a)\tan\frac{A}{2}$, etc.), keeping the angles A, B, C in play throughout. The model reaches instead for the two formulas for the *area* of the whole triangle ($A=pr$ and Heron's formula), which involve no angles at all and directly relate r, p, and the three quantities $p-a, p-b, p-c$ as a single algebraic package.

> *I don't need to know anything about the angles to relate r to $(p-a),(p-b),(p-c)$ – the area of the triangle is a single number computable two different ways, once via r and once via Heron's formula, and setting those two expressions equal gives me exactly the relation I need, with no trigonometry at all.*

▷ **Cue:** an inequality or identity mixes a triangle's inradius r with its semiperimeter p and the three quantities $p-a$, $p-b$, $p-c$ ⟶ **Move:** equate the two standard formulas for the triangle's area, $A=pr$ and $A=\sqrt{p(p-a)(p-b)(p-c)}$, to eliminate A and produce a purely algebraic relation among $r, p, p-a, p-b, p-c$ – then treat that relation as the entire problem, discarding the geometry

Layer 3 THE WHY

The reason the area-formula route works and the half-angle route doesn't collapse as easily is structural: area is the one quantity in a triangle that has *two independent, exact* formulas built from different data – one from r and p (which measures the triangle "from the inside," via the incircle), and one from a, b, c alone (Heron, which measures it "from the outside," via the raw side l-engths). Whenever a single geometric quantity has two exact formulas in different variables, equating them is not a trick – it is the definition of that quantity being well-defined, and it is always available as a substitution engine to eliminate the quantity and relate its two sets of ingredients directly. The half-angle formulas $r = (p-a)\tan(A/2)$ are also true, but they route the relationship through a *third* kind of data (angles), which the final inequality never mentions – so using them means introducing variables you will later have to eliminate again, at real cost. Once the area is eliminated via the two area formulas, the resulting equation $pr^2 = (p-a)(p-b)(p-c)$ is symmetric in $p-a, p-b, p-c$ in exactly the way the desired inequality is symmetric in those same three quantities, which is why the substitution $x = 1/(p-a)$ etc. turns it into the single cleanest possible inequality: $x^2+y^2+z^2 \geq xy+yz+zx$, provable from nothing but $(x-y)^2 \geq 0$. The official solution states the two area formulas and the substitution in three lines and calls the rest "trivial" – correctly, but only after the substitution has already done all the real work of stripping the geometry away.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that the inequality involves r, p, and $p-a, p-b, p-c$ together but no angles and no side lengths on their own. On seeing it, write down: this is asking for a relation between r and the semiperimeter-differences, and the object that ties all of these together without introducing new variables is the triangle's area.

2. That cue suggests writing the area two ways – $A = pr$ and Heron's $A = \sqrt{p(p-a)(p-b)(p-c)}$ – and equating them to eliminate A, producing a purely algebraic identity in $p, r, p-a, p-b, p-c$.

3. The hesitation occurs right after seeing r appear with vertex-associated terms like $p-a$: it is tempting to recall $r = (p-a)\tan(A/2)$ and chase the problem through half-angles, because that identity also directly involves r and $p-a$ together. It is reinforced away by noticing that the target inequality mentions no angles at all, so any route through angles will require eliminating them again later at extra cost.

4. The trap is committing to the half-angle identities and trying to prove a trigonometric inequality among $\tan(A/2), \tan(B/2), \tan(C/2)$ constrained by $A+B+C = \pi$ – a real inequality, but one with no clean symmetric form and no obvious stopping point, chosen only because it also happens to relate r and $p-a$.

5. To recognise this faster: whenever a problem's variables are exactly the ingredients of two different standard formulas for the same geometric quantity (here, area), write both formulas down immediately and equate them before reaching for any other identity, even one that is individually true and relevant.

6. Decision rule: when a triangle identity or inequality involves a mix of quantities that are each individually expressible via two different named formulas (e.g. area via pr and via Heron; or, more generally, any quantity with a "from the incircle" formula and a "from the side lengths" formula), equate the two formulas first to eliminate the shared quantity and reduce to a relation purely among the given variables – before reaching for angle-based identities that introduce variables the final statement doesn't need.

1966 A3

1966 A3. Let $0 < x_1 < 1$ and $x_{n+1} = x_n(1 - x_n)$, $n = 1, 2, 3, \ldots$. Show that

$$\lim_{n\to\infty} nx_n = 1.$$

Fields: Real analysis (recurrence sequences, Stolz–Cesàro / discrete L'Hôpital) **Position:** A3 **Answer:** $\lim_{n\to\infty} nx_n = 1$; equivalently $x_n \sim 1/n$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The recursion $x_{n+1} = x_n - x_n^2$ is easy to get a feel for: since $0 < x_1 < 1$, an easy induction shows $0 < x_n < 1$ for all n, and $x_{n+1} = x_n(1 - x_n) < x_n$, so $\{x_n\}$ is decreasing and bounded below by 0, hence converges to some $L \geq 0$ with $L = L(1 - L)$, forcing $L = 0$. So $x_n \to 0$ – that much is free. The problem wants a *rate*: not just that $x_n \to 0$, but that $nx_n \to 1$. The natural next move, with nx_n sitting right there in the statement, is to attack nx_n itself. Multiply the recursion by $(n + 1)$:

$$(n + 1)x_{n+1} = nx_n + x_n\big[1 - (n + 1)x_n\big].$$

To show nx_n is increasing you now need $1 - (n + 1)x_n \geq 0$ for every n, which is not obvious from the recursion alone – it requires a separate induction on the logistic map itself (using that $x_2 \leq \frac{1}{2}$, and that $a \leq \frac{1}{2}$ implies $a(1 - a)$ is still controlled, to bound $(n + 1)x_n \leq 1$ for all $n \geq 2$). Having fought through that, you now know nx_n is increasing and bounded above by 1, so it converges to some limit $a \in (0, 1]$ – but you still haven't shown $a = 1$ rather than some smaller value. That requires a *second*, independent argument: summing the bracketed terms $[1 - (n + 1)x_n]$ from 2 to n telescopes the left side, and a proof by contradiction (if $a < 1$, the bracketed terms stay bounded below by a positive constant infinitely often, forcing $\sum x_n$ to converge – but $nx_n \geq x_1$ forces $\sum x_n \geq x_1 \sum(1/n)$, which diverges) is needed to rule out $a < 1$. Three separate pieces of machinery – monotonicity of nx_n, boundedness via an auxiliary induction, and a contradiction argument via a divergent harmonic sum – stacked on top of each other, each individually correct, none of them signalled in advance by the recursion itself.

The model chain. The quantity nx_n is n divided by $1/x_n$ – an $n \to \infty$ limit of the form (index)/(reciprocal of the sequence), which is exactly the shape the Stolz–Cesàro theorem (the discrete analogue of L'Hôpital's rule for ∞/∞) is built to handle: if $b_n \to \infty$ strictly increasing and $\lim(a_{n+1} - a_n)/(b_{n+1} - b_n)$ exists, then $\lim a_n/b_n$ equals that same limit. Take $a_n = 1/x_n$ and $b_n = n$. Compute the increment directly from the recursion:

$$\frac{1}{x_{n+1}} - \frac{1}{x_n} = \frac{1}{x_n(1 - x_n)} - \frac{1}{x_n} = \frac{1 - (1 - x_n)}{x_n(1 - x_n)} = \frac{x_n}{x_n(1 - x_n)} = \frac{1}{1 - x_n}.$$

Since $x_n \to 0$ (shown above, for free), this increment tends to $\dfrac{1}{1 - 0} = 1$. By Stolz–Cesàro,

$$\lim_{n\to\infty} \frac{1/x_n}{n} = \lim_{n\to\infty} \left(\frac{1}{x_{n+1}} - \frac{1}{x_n}\right) = 1,$$

and since this limit is a nonzero finite number, $\lim_{n\to\infty} \dfrac{n}{1/x_n} = \lim_{n\to\infty} nx_n$ is its reciprocal, 1. Done, in four lines after establishing $x_n \to 0$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains use exactly the same two free facts – x_n decreasing to 0 – before doing anything harder. They split at the choice of what quantity to take the limit of next. The natural attempt keeps working with x_n itself, multiplying the recursion by $(n + 1)$ and chasing nx_n directly through induction. The model chain instead changes variables to $1/x_n$ before computing anything further, because the target nx_n is recognisable as n over a reciprocal sequence.

> *I want a limit of the form nx_n, which is the same shape as $n/(1/x_n)$ – and a ratio of an index to a transformed sequence is exactly what Stolz–Cesàro is for. Instead of fighting with x_n's own increments, which are quadratically small and awkward, I should look at the increments of $1/x_n$, because the recursion $x_{n+1} = x_n - x_n^2$ is precisely the kind of quadratic-decay recursion that reciprocal substitution linearises.*

▷ **Cue:** a recursion loses a term quadratically, $x_{n+1} = x_n - cx_n^2 + \cdots$, and the problem asks for the linear-in-n decay rate (a limit of the form $nx_n \to L$) $\longrightarrow$ **Move:** substitute $y_n = 1/x_n$ and compute $y_{n+1} - y_n$ directly from the recursion; this increment will tend to a finite limit, and Stolz–Cesàro then converts $y_n/n \to$ (that limit) into the desired rate for x_n, without any separate induction or contradiction argument

Layer 3 THE WHY

The recursion $x_{n+1} = x_n - x_n^2$ is the Euler step, with step size 1, for the differential equation $\frac{dx}{dn} = -x^2$ – and that ODE has the exact solution $x(n) = \frac{1}{n+C}$, whose reciprocal $1/x(n) = n + C$ is *linear* in n. The reciprocal is the natural variable for a quadratic-decay recursion for exactly the same structural reason it is the natural variable for the continuous equation: differentiating $1/x$ by the chain rule gives $-x'/x^2$, and when $x' = -x^2$ this is identically 1, a constant – the quadratic nonlinearity cancels perfectly against the reciprocal's own derivative. Stolz–Cesàro is the discrete shadow of that same cancellation: it exists precisely because a ratio a_n/b_n of two divergent sequences can fail to have a nice limit even when the *increments* $a_{n+1} - a_n$ and $b_{n+1} - b_n$ do, and it says that when b_n increases steadily to infinity, the increment ratio governs the whole limit. Choosing $a_n = 1/x_n$ is not a trick reached for by pattern memory; it is reached for because computing $1/x_{n+1} - 1/x_n$ from $x_{n+1} = x_n(1 - x_n)$ produces an exact algebraic simplification (the x_n in the numerator cancels one power of x_n in the denominator) rather than an approximation, which is the signal that the reciprocal is the right coordinate. The book's own solution never takes this route; it treats the recursion as opaque and manufactures monotonicity, a boundary bound via an auxiliary induction on the logistic map, and a divergence-of-the-harmonic-series contradiction, three separate constructions in place of one substitution. Every piece of that argument is correct, but none of it is motivated by anything visible in the recursion itself – which is exactly the kind of unmotivated correctness the archive exists to replace.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the shape of the recursion itself: $x_{n+1} = x_n - x_n^2$ loses a term *quadratically* in x_n, and the claim to prove is about n times x_n – a rate statement. On seeing a quadratic-decay recursion paired with a claimed linear-in-n rate, write down: try the reciprocal.

2. That cue suggests substituting $y_n = 1/x_n$ and computing $y_{n+1} - y_n$ directly; the method is then Stolz–Cesàro (discrete L'Hôpital), converting the increment's limit into the limit of y_n/n.

3. The hesitation happens immediately after writing $x_{n+1} = x_n(1 - x_n)$ and seeing nx_n named explicitly in the problem: the instinct is to multiply through by $(n + 1)$ and chase nx_n in its own terms, because that is the object the problem names. It is reinforced away by recognising $nx_n = n/(1/x_n)$ as a ratio-of-sequences shape, which should trigger Stolz–Cesàro before any direct induction is attempted.

4. The trap is committing to a direct induction on nx_n: it does eventually work (it is the published solution), but it requires two independently-motivated sub-arguments – an auxiliary bound $(n + 1)x_n \leq 1$ proved by its own induction, and a proof by contradiction using divergence of the harmonic series to pin the limit exactly at 1 rather than some smaller value – neither of which is suggested by the recursion until after you have committed to the direct approach.

5. To recognise this faster: build the reflex that any recursion of the form $x_{n+1} = x_n - cx_n^2 + \cdots$ (quadratic loss) paired with a question about the linear rate of decay should immediately

trigger "take reciprocals, then Stolz–Cesàro," in the same way $e^{ax}\cos bx$ should trigger "take the real part of a complex exponential."

6. Decision rule: when a recursion decays quadratically and the question is about its linear-in-n rate, do not analyse x_n's own increments by direct induction – substitute the reciprocal $y_n = 1/x_n$ first. If $y_{n+1} - y_n$ tends to a finite limit L, then by Stolz–Cesàro $y_n/n \to L$, and if $L \neq 0$ this immediately gives $nx_n \to 1/L$ with no separate monotonicity or boundedness argument required.

1966 A4

1966 A4. Prove that after deleting the perfect squares from the list of positive integers the number we find in the nth position is equal to $n + \{\sqrt{n}\}$, where $\{\sqrt{n}\}$ denotes the integer closest to $\sqrt{n}$.

Caveat. Identity check: the MAA volume's solution under 1966 A-4 proves by induction that $n + \{\sqrt{n}\} - (n - 1 + \{\sqrt{n-1}\})$ is 1 or 2, with 2 exactly when a square is skipped; that is a proof of this statement. The OCR text of the statement is clean. Note that $\{\cdot\}$ here is *not* the fractional part; it is the nearest integer, which is well defined because $\sqrt{n}$ is never a half-integer ($(k+\frac{1}{2})^2 = k^2 + k + \frac{1}{4}$ is not an integer).

Fields: number theory / combinatorics (counting, floor and nearest-integer functions) **Position:** A4 — mid-paper, but among the gentler A4s once the right direction of counting is chosen **Answer:** True for every $n \geq 1$. A non-square m with $k^2 < m < (k+1)^2$ sits in position $m - k$; as m runs over that gap, $n = m - k$ runs over $k^2 - k + 1, \ldots, k^2 + k$, which is exactly the set of n with nearest integer to $\sqrt{n}$ equal to k. So the nth non-square is $n + k = n + \{\sqrt{n}\}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. Write the list out: $2, 3, 5, 6, 7, 8, 10, 11, \ldots$. Position n and value a_n: $(1,2), (2,3), (3,5), (4,6), (5,7), (6,8), (7,10)$. The formula checks: $\{\sqrt{3}\} = 2$, $3 + 2 = 5$; $\{\sqrt{7}\} = 3$, $7 + 3 = 10$. Good.

Now prove it. The instinct, trained by years of sequence questions, is induction on n. Assume $a_{n-1} = n - 1 + \{\sqrt{n-1}\}$. The next non-square is $a_{n-1} + 1$ unless $a_{n-1} + 1$ is a square, in which case it is $a_{n-1} + 2$. So you must show

$$\{\sqrt{n}\} - \{\sqrt{n-1}\} = \begin{cases} 1 & \text{if } n - 1 + \{\sqrt{n-1}\} + 1 \text{ is a perfect square,} \\ 0 & \text{otherwise.} \end{cases}$$

Put $q = \{\sqrt{n-1}\}$, so $q - \frac{1}{2} < \sqrt{n-1} < q + \frac{1}{2}$, i.e. $q^2 - q + \frac{1}{4} < n - 1 < q^2 + q + \frac{1}{4}$. Then $n + q$ lies strictly between $q^2 + \frac{5}{4}$ and $(q+1)^2 + \frac{1}{4}$, so it is a square only if $n + q = (q+1)^2$, i.e. $n = q^2 + q + 1$. And separately you must check that $\{\sqrt{n}\}$ jumps from q to $q + 1$ exactly at $n = q^2 + q + 1$, which is where $\sqrt{n}$ crosses $q + \frac{1}{2}$.

This is the MAA solution, and it works. It costs: two separate facts to establish (when a square is skipped; when the nearest integer jumps), two chains of half-integer inequalities, and a coincidence of thresholds that has to be verified rather than seen. Most students who go this way lose the thread in the inequalities, or prove one direction of the "if and only if" and forget the other.

The model chain. Turn the question round. Instead of asking "what is the value at position n?", ask "what is the position of the value m?". That question has an immediate answer.

Let m be a non-square, and let $k = \lfloor \sqrt{m} \rfloor$, so $k^2 < m < (k+1)^2$. The positive integers up to m number m; the squares among them are $1^2, 2^2, \ldots, k^2$, which number k. So the non-squares up to m number $m - k$, and m is the $(m-k)$th non-square.

Now fix $k \geq 1$. As m runs through the $2k$ non-squares $k^2+1, \ldots, k^2+2k$ strictly between k^2 and $(k+1)^2$, its position $n = m - k$ runs through

$$B_k = \{k^2 - k + 1,\ k^2 - k + 2,\ \ldots,\ k^2 + k\}.$$

These blocks are consecutive: $B_1 = \{1, 2\}$, $B_2 = \{3, \ldots, 6\}$, $B_3 = \{7, \ldots, 12\}$, and in general the last element of B_k is $k^2 + k$ while the first element of B_{k+1} is $(k+1)^2 - (k+1) + 1 = k^2 + k + 1$. So every positive integer n lies in exactly one block B_k, and for $n \in B_k$ the nth non-square is $m = n + k$.

It remains to identify k from n. For $n \in B_k$,

$$\left(k - \tfrac{1}{2}\right)^2 = k^2 - k + \tfrac{1}{4} < k^2 - k + 1 \leq n \leq k^2 + k < k^2 + k + \tfrac{1}{4} = \left(k + \tfrac{1}{2}\right)^2,$$

so $k - \frac{1}{2} < \sqrt{n} < k + \frac{1}{2}$, which says precisely that the integer nearest to $\sqrt{n}$ is k. Therefore the nth non-square is $n + \{\sqrt{n}\}$. ■

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains write the list out and check the formula on small cases. At the moment of proving it, the natural attempt stays with the direction the problem is phrased in — position → value — and reaches for induction. The model reverses the map to value → position, where the count is one line, and then reads off which positions belong to which gap between squares.

> *The problem asks me for the nth non-square, which is hard. But the position of a given non-square m is easy: it is m minus the number of squares up to m, which is $m - \lfloor \sqrt{m} \rfloor$. I'll compute the inverse map and invert it.*

> *The positions coming from the gap between k^2 and $(k+1)^2$ run from $k^2 - k + 1$ to $k^2 + k$. Those endpoints are $(k \mp \frac{1}{2})^2$ rounded inward — that is where the "nearest integer to $\sqrt{n}$" comes from.*

Layer 3 THE WHY

Why is reversing the direction the natural move and not a trick? Because of what "the nth element of a set" means. The nth element of an increasing list S is defined *implicitly*: it is the $m \in S$ such that exactly n elements of S are at most m. The explicit function is the counting function

$$N(m) = \#\{s \in S : s \leq m\},$$

and "the nth element" is its inverse. Any time a question asks for the nth term of a sparse or thinned-out set, the counting function is the primary object and the nth-term function is secondary. For the non-squares the counting function is $N(m) = m - \lfloor \sqrt{m} \rfloor$: one subtraction, because squares are easy to count. The problem is disguised as a question about a_n but is really a question about inverting N.

Once you have N, the answer's shape is forced. Inverting $n = m - \lfloor \sqrt{m} \rfloor$ means solving for m, and the only obstacle is the floor. So hold the floor constant: on each gap $k^2 < m < (k+1)^2$ the floor is the constant k, the map is $m \mapsto m - k$, a shift, and a shift is inverted by $n \mapsto n + k$. All that is left is to express "which gap" in terms of n rather than m, which is where the block endpoints $k^2 - k + 1$ and $k^2 + k$ appear. Those numbers sit symmetrically around k^2, at distance about k, and $(k \pm \frac{1}{2})^2 = k^2 \pm k + \frac{1}{4}$. So the rounding in the problem, "integer closest to $\sqrt{n}$", is

not an arbitrary choice of the setter: it is exactly the rounding that the block structure produces. If the problem had asked about deleting squares and said "$n + \lfloor\sqrt{n}\rfloor$" you would find the formula fails at $n = 3$, and the block computation tells you why.

The MAA solution's induction is the same mathematics viewed through a keyhole. It checks, one step at a time, that consecutive values of $n + \{\sqrt{n}\}$ differ by 1 or 2 and that the 2 occurs exactly at the squares. That is proving the inverse relation locally, step by step. The block argument proves it globally, in one pass, by computing the counting function. Here is the cost comparison: the induction needs two coinciding thresholds ($n + q$ square iff $n = q^2 + q + 1$; $\{\sqrt{n}\}$ jumps iff $n = q^2 + q + 1$) and it is the coincidence that the reader must take on trust unless both inequalities are done carefully. The block argument never meets a coincidence, because both thresholds are the same number $k^2 + k + 1$ computed once.

There is also a sanity reason to prefer counting. The induction proves "the formula agrees with the list" but gives no feeling for why the formula has the form it does. The count shows it: removing squares shifts the nth integer up by the number of squares already passed, and the number of squares passed by position n is about $\sqrt{n}$. The precise rounding is fine print on top of a fact you could have guessed in your head.

Layer $+\alpha$ BACKTESTING

1. The cue is "the number we find in the nth position" of a set defined by deletion. Write down immediately: position of $m = m - \#\{\text{deleted numbers} \le m\} = m - \lfloor\sqrt{m}\rfloor$.

2. Counting function, then inversion. On each interval where the floor is constant the counting function is a shift; invert the shift and translate the interval's endpoints into the variable n.

3. Hesitation comes when the formula uses $\sqrt{n}$ (of the position) while the count uses $\sqrt{m}$ (of the value): they look incompatible. Reinforce with "compute the block of n's coming from one gap in m, then see which $\sqrt{n}$ that block corresponds to".

4. Two traps. First, reading $\{\sqrt{n}\}$ as the fractional part, which makes the statement nonsense. Second, an induction that establishes when $n + \{\sqrt{n}\}$ jumps by 2 but never proves the jump happens *only* at squares, which leaves the "if and only if" half done.

5. Whenever a problem asks for the nth element of a set, ask whether the number of elements below a given m is easy. If it is, the problem is an inversion problem, not a sequence problem.

6. To find the nth term of an increasing set S: write the counting function $N(m)$; split the range into pieces where N is affine; invert on each piece; express the piece boundaries in terms of n. Rounding functions in the answer come from those boundaries.

▷ **Cue:** "the nth term" of the integers with some set deleted $\longrightarrow$ **Move:** count instead: the position of m is m minus the number of deleted elements up to m

▷ **Cue:** a counting function involving $\lfloor f(m)\rfloor$ $\longrightarrow$ **Move:** hold the floor constant on each interval, where the map is a shift, and invert the shift

▷ **Cue:** an answer containing "the integer nearest to $\sqrt{n}$" $\longrightarrow$ **Move:** look for block endpoints near $k^2 \pm k$, since $(k \pm \frac{1}{2})^2 = k^2 \pm k + \frac{1}{4}$

▷ **Cue:** an induction whose step needs two thresholds to coincide $\longrightarrow$ **Move:** look for a global count that produces both thresholds as the same number

1966 A5

1966 A5. Let C denote the family of continuous functions on the real axis. Let T be a mapping of C into C which has the following properties:

1. T is linear, i.e. $T(c_1\psi_1 + c_2\psi_2) = c_1T\psi_1 + c_2T\psi_2$, for c_1 and c_2 real and ψ_1 and ψ_2 in C.
2. T is local, i.e. if $\psi_1 \equiv \psi_2$ in some interval I then also $T\psi_1 \equiv T\psi_2$ holds in I.

Show that T must necessarily be of the form $T\psi(x) = f(x)\psi(x)$ where $f(x)$ is a suitable continuous function.

Caveat. Identity check: the MAA solution under 1966 A-5 sets $f = T1$ and, for fixed x_0, compares ψ with the function equal to ψ on $x \le x_0$ and to the constant $\psi(x_0)$ on $x > x_0$; that proves exactly this statement. The OCR renders that auxiliary function's second branch as "$-\psi(x_0)$"; the minus sign is an OCR artefact (with it the function would be discontinuous and the argument would fail), and the constant is $+\psi(x_0)$. The book's comment is worth keeping: the interval I must be non-degenerate, since if I may be a single point the problem is empty. The proof below uses only open intervals, so it is valid under the strictest reading.

Fields: real analysis / functional analysis (linear operators) **Position:** A5 — late in the paper; short once seen, but the key step is invisible to anyone who has not asked which hypothesis is unused **Answer:** Take $f = T1$, the image of the constant function 1. For each x_0, split $\psi - \psi(x_0)$ into a piece vanishing on $(-\infty, x_0]$ and a piece vanishing on $[x_0, \infty)$; locality kills T of each piece on an open half-line, continuity of the image carries the zero to x_0, and so $T\psi(x_0) = \psi(x_0) f(x_0)$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The target form $T\psi = f\psi$ tells you what f must be: put $\psi \equiv 1$ and read off $f = T1$. That is continuous because T maps C into C. So the claim is $T\psi(x_0) = \psi(x_0)\,(T1)(x_0)$ for every x_0.

Fix x_0 and subtract the constant: $\psi = \psi(x_0) \cdot 1 + g$ with $g(x_0) = 0$. By linearity $T\psi = \psi(x_0)T1 + Tg$, so everything reduces to

$$g(x_0) = 0 \quad \Longrightarrow \quad (Tg)(x_0) = 0. \tag{$*$}$$

Now the wall. Locality speaks about functions that agree on an *interval*; g agrees with 0 only at the single point x_0. The instinct is to approximate: let g_ε be g modified to vanish on $(x_0 - \varepsilon, x_0 + \varepsilon)$, so that $Tg_\varepsilon(x_0) = T0(x_0) = 0$ by locality, and then let $\varepsilon \to 0$. But $Tg_\varepsilon \to Tg$ would need T to be continuous in some topology, and nothing in the problem says so. A linear map on C with no continuity assumption can do strange things under limits. The approximation argument is dead, and most students who reach this point either assert the limit (and lose the problem) or try to prove T is bounded (and get nowhere).

The model chain. Let $f = T1$; $f \in C$. Note $T0 = 0$, by linearity with $c_1 = c_2 = 0$.

Claim $(*)$. If $g \in C$ and $g(x_0) = 0$, then $(Tg)(x_0) = 0$.

Define

$$g_-(x) = \begin{cases} g(x), & x \le x_0, \\ 0, & x > x_0, \end{cases} \qquad g_+(x) = \begin{cases} 0, & x \le x_0, \\ g(x), & x > x_0. \end{cases}$$

Both are continuous: each is continuous on either side of x_0, and at x_0 both one-sided limits equal $0 = g(x_0)$. Also $g = g_- + g_+$.

On the open interval (x_0, ∞) we have $g_- \equiv 0$, so by locality $Tg_- \equiv T0 = 0$ there. Since Tg_- is continuous, $(Tg_-)(x_0) = \lim_{x \to x_0^+} (Tg_-)(x) = 0$. Symmetrically, $g_+ \equiv 0$ on $(-\infty, x_0)$, so $Tg_+ \equiv 0$ there, and continuity at x_0 from the left gives $(Tg_+)(x_0) = 0$. By linearity $(Tg)(x_0) = (Tg_-)(x_0) + (Tg_+)(x_0) = 0$.

Conclusion. For any $\psi \in C$ and any x_0, apply $(*)$ to $g = \psi - \psi(x_0) \cdot 1$:

$$0 = (Tg)(x_0) = (T\psi)(x_0) - \psi(x_0)\,(T1)(x_0),$$

so $T\psi(x_0) = f(x_0)\psi(x_0)$. Since x_0 was arbitrary, $T\psi = f\psi$. ■

(The MAA solution packs the same idea into one function: ψ' equals ψ on $x \le x_0$ and the constant $\psi(x_0)$ on $x > x_0$. Locality on $(-\infty, x_0)$ gives $T\psi' = T\psi$ there; locality on (x_0, ∞) gives $T\psi' = \psi(x_0) f$ there; continuity of $T\psi'$ and $T\psi$ at x_0 gives $T\psi(x_0) = \psi(x_0) f(x_0)$.)

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains reduce to $(*)$: T must kill, at x_0, any g vanishing at x_0. There the natural attempt tries to make g vanish on a small *two-sided* neighbourhood and take a limit,

which needs continuity of T. The model makes g vanish on a *one-sided* half-line, by cutting it at x_0, and uses continuity of Tg instead.

> *I've used linearity and locality, and I'm stuck. Which hypothesis haven't I touched? "T maps C into C" — the output is continuous. Continuity lets me learn a value at x_0 from values on one side of x_0. So I only need Tg to vanish on a half-line ending at x_0, not on a neighbourhood.*

> *g vanishes at x_0, so I can cut it there without breaking continuity: keep the left half and put 0 on the right, and vice versa. Each half vanishes on an open half-line, and they add back to g.*

Layer 3 THE WHY

Why is "cut at x_0" the natural move? Start from what the conclusion says: T is multiplication by a function. The one-line reason to believe that is the unused hypothesis. Linear and local operators on smooth functions include differentiation: $D\psi = \psi'$ is linear, and if $\psi_1 = \psi_2$ on an interval then $\psi_1' = \psi_2'$ there. (It is a theorem of Peetre that linear local operators are, locally, differential operators.) So linearity and locality alone cannot force multiplication. Differentiation is excluded here only because T must be defined on every continuous function and return a continuous one, and a continuous function need not have a derivative. That tells you where the proof lives. Any argument that uses only linearity and locality would also "prove" that d/dx is a multiplication operator, which is false; so the continuity of $T\psi$ must enter, and it must enter decisively.

Now ask what continuity of $T\psi$ can do. It is a statement about limits of *values of the output*, not about limits of inputs. So it cannot rescue the approximation argument, which needs T to commute with a limit of inputs. What it can do is transfer information along the real line: if $Tg = 0$ on (x_0, ∞) then $Tg(x_0) = 0$. So the job is to make g agree with 0 on an open half-line with endpoint x_0. A general g does not, but a continuous g with $g(x_0) = 0$ can be cut at x_0 into two continuous pieces, each of which does. The condition $g(x_0) = 0$ is exactly what makes the cut continuous; that is why the reduction to $(*)$ subtracts the constant $\psi(x_0)$ first. Every step is forced by the question "what does this hypothesis let me do?"

Seen this way the MAA solution's ψ' is not a rabbit from a hat. It is ψ with its right-hand half replaced by the constant that keeps it continuous. On the left half-line it looks like ψ; on the right half-line it looks like a multiple of 1, whose image you know. Continuity at the join then forces the two descriptions to agree at x_0. The two-piece version in the model chain is the same move made symmetric.

The book's comment about the interval I also has a clean explanation from here. If I is allowed to be a single point, locality says $\psi_1(x_0) = \psi_2(x_0) \Rightarrow T\psi_1(x_0) = T\psi_2(x_0)$, which is $(*)$ outright; the problem would be over after one line. The content of the problem is the passage from "agree on an interval" to "agree at a point", and the bridge is continuity.

Layer $+\alpha$ BACKTESTING

1. The cue is the form of the conclusion, $T\psi = f\psi$. Write down $f = T1$ at once, and reduce to showing $(T\psi)(x_0)$ depends only on $\psi(x_0)$, i.e. $(*)$.

2. A pointwise statement from interval hypotheses suggests: make the relevant function vanish on an interval with x_0 as endpoint, then pass to x_0 using continuity.

3. Hesitation comes at $(*)$, where the only visible tool is an approximation that T is not known to respect. Reinforce with the audit: list the hypotheses, tick those used, and ask what the unticked one can do.

4. The trap is taking a limit through T, e.g. $Tg_\varepsilon \to Tg$, without any continuity of T. A second trap is proving $(*)$ only for g vanishing near x_0 and forgetting that the general g vanishes only at x_0.

5. Test the claim against the most natural operator that satisfies the algebraic hypotheses but should fail the conclusion (here d/dx). Whatever hypothesis excludes it is the one the proof must use.

6. When a function vanishes at a point and a hypothesis applies only to functions vanishing on intervals, cut the function at the point into one-sided pieces, each vanishing on a half-line, and recover the point by continuity.

▷ **Cue:** a conclusion of the form $T\psi = f\psi$ $\longrightarrow$ **Move:** set $f = T1$ and reduce to: $\psi(x_0) = 0 \Rightarrow T\psi(x_0) = 0$

▷ **Cue:** linear operator with no continuity assumption $\longrightarrow$ **Move:** never pass a limit through T; look for exact algebraic decompositions instead

▷ **Cue:** a hypothesis on intervals, a conclusion at a point $\longrightarrow$ **Move:** produce agreement on a half-line ending at the point, then use continuity of the output

▷ **Cue:** a continuous function vanishing at x_0 $\longrightarrow$ **Move:** cut it at x_0 into a left piece and a right piece, both still continuous

▷ **Cue:** stuck after using some hypotheses $\longrightarrow$ **Move:** find the natural counterexample the unused hypothesis excludes (e.g. differentiation); that hypothesis is the key

1966 A6

1966 A6. Justify the statement that

$$3 = \sqrt{1+2\sqrt{1+3\sqrt{1+4\sqrt{1+5\sqrt{1+\cdots}}}}}.$$

Caveat. Identity check: the OCR of the problem statement is unreadable (the radical is rendered as a string of symbols in which only the coefficients $2, 3, 4, 5$ survive). The statement above is the standard wording of 1966 A-6 (Ramanujan's nested radical). The MAA solution printed under A-6 proves exactly this: it reads the radical as $\lim_{n\to\infty}\sqrt{1+2\sqrt{1+3\sqrt{\cdots\sqrt{1+(n-1)\sqrt{1+n}}}}}$, derives $3 = \sqrt{1+2\sqrt{1+\cdots+n\sqrt{(n+2)^2}}}$ from $(n+2)^2 = 1+(n+1)(n+3)$, and bounds the other side by $3 \le (n+2)^{2^{1-n}} \cdot$ (truncation). The OCR drops the auxiliary inequality it calls "for any $a > 1$"; the version used below ($\sqrt{1+\lambda ky} \le \sqrt{\lambda}\sqrt{1+ky}$ for $\lambda \ge 1$) is reconstructed and proved in full, and it reproduces the book's exponent 2^{1-n} exactly. Every numerical claim in this analysis was checked by direct computation.

Fields: real analysis (limits, monotone sequences, nested radicals); algebraic identities **Position:** A6 — the last problem of the paper; the identity is a five-minute find, the real content is that "justify" asks you to say what the infinite expression means and then prove a two-sided limit **Answer:** Define $a_n = \sqrt{1+2\sqrt{1+3\sqrt{\cdots\sqrt{1+n}}}}$ (the radical truncated by putting 1 at depth n). The identity $k+1 = \sqrt{1+k(k+2)}$ telescopes to give exactly 3 when the depth-n slot holds $n+2$, so $a_n < 3$ and a_n increases; the scaling bound $\sqrt{1+\lambda ky} \le \sqrt{\lambda}\sqrt{1+ky}$ ($\lambda \ge 1$) gives $3 \le (n+2)^{2^{1-n}} a_n$, and $(n+2)^{2^{1-n}} \to 1$. Hence $a_n \to 3$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The reflex for an infinite nested radical is self-similarity: call the whole

thing x, square, and find x again inside. Try it.

$$x^2 = 1 + 2\sqrt{1 + 3\sqrt{1 + 4\sqrt{\cdots}}}.$$

The tail is not x: its coefficients start at 3, not 2. The trick that works for $\sqrt{2 + \sqrt{2 + \cdots}}$ fails here because the radical is not periodic. The better students repair this by naming the whole family: let $f(n) = \sqrt{1 + n\sqrt{1 + (n+1)\sqrt{\cdots}}}$, so that

$$f(n)^2 = 1 + n\, f(n+1), \tag{†}$$

and the target is $f(2) = 3$. Guessing from $f(2) = 3$ that $f(n) = n + 1$, you check $(n + 1)^2 = n^2 + 2n + 1 = 1 + n(n + 2)$. It fits. Most scripts stop here and write "therefore $3 = \sqrt{1 + 2\sqrt{1 + 3\sqrt{\cdots}}}$".

That is not a justification, and here is the wall. Equation (†) runs forwards: pick $f(2)$, then $f(3) = (f(2)^2 - 1)/2$, $f(4) = (f(3)^2 - 1)/3$, and so on. Every starting value produces a sequence satisfying (†) for as long as it stays non-negative. Start at $3 + 10^{-6}$ and the sequence is still a perfectly good solution; it passes 10^{100} by $n = 30$. Start at $3 - 10^{-6}$ and it goes negative at $n = 25$. So (†) has solutions other than $n + 1$, and "$n + 1$ fits" proves nothing about which one the radical is. You have verified a candidate, not identified the value.

The model chain. *Meaning.* For $k \geq 2$ let $g_k(y) = \sqrt{1 + ky}$ for $y \geq 0$, and let $G_n = g_2 \circ g_3 \circ \cdots \circ g_n$ ($n \geq 2$), a composition of $n - 1$ maps. Then $G_n(t) = \sqrt{1 + 2\sqrt{1 + 3\sqrt{\cdots\sqrt{1 + nt}}}}$. The infinite radical means $\lim_{n\to\infty} a_n$, where

$$a_n = G_n(1) = \sqrt{1 + 2\sqrt{1 + 3\sqrt{\cdots\sqrt{1 + (n-1)\sqrt{1 + n}}}}}.$$

Each g_k is strictly increasing, so each G_n is.

Exact identity. Since $1 + k(k + 2) = (k + 1)^2$, $g_k(k + 2) = k + 1$. Hence

$$G_n(n + 2) = g_2 \circ \cdots \circ g_{n-1}(n + 1) = \cdots = g_2(4) = 3 \qquad \text{for every } n \geq 2.$$

Upper bound and monotonicity. $1 < n + 2$ and G_n is increasing, so $a_n < 3$. Also $a_{n+1} = G_n(g_{n+1}(1)) = G_n(\sqrt{n + 2}) > G_n(1) = a_n$. So (a_n) is increasing and bounded above by 3, and converges to some $L \leq 3$.

Scaling lemma. If $\lambda \geq 1$ and $y \geq 0$, then $g_k(\lambda y) = \sqrt{1 + \lambda k y} \leq \sqrt{\lambda + \lambda k y} = \lambda^{1/2} g_k(y)$. By induction on the number of layers,

$$G_n(\lambda y) \leq \lambda^{2^{-(n-1)}} G_n(y) \qquad (\lambda \geq 1,\ y \geq 0).$$

Indeed, for the innermost layer $g_n(\lambda y) \leq \lambda^{1/2} g_n(y)$ with $\lambda^{1/2} \geq 1$; apply g_{n-1} (increasing) and the lemma again with $\lambda^{1/2}$ in place of λ, and so on through $n - 1$ layers.

Lower bound. With $y = 1$, $\lambda = n + 2$:

$$3 = G_n(n + 2) \leq (n + 2)^{2^{1-n}} a_n, \qquad \text{so} \qquad 3(n + 2)^{-2^{1-n}} \leq a_n < 3.$$

Finally $\ln\left((n + 2)^{2^{1-n}}\right) = 2^{1-n} \ln(n + 2) \to 0$, so the left side tends to 3 and $L = 3$. ■

(The MAA solution is this argument, with the scaling inequality stated for "any $a > 1$". Numerically $a_{10} \approx 2.98992$ against the lower bound 2.98548; $a_{20} \approx 3 - 1.2 \times 10^{-5}$.)

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains find the identity $(k + 1)^2 = 1 + k(k + 2)$. At that point the natural attempt declares victory. The model asks what the infinite expression *is*, defines it as a limit of finite truncations, and uses the identity not as an answer but as an exact finite anchor, $G_n(n + 2) = 3$, against which the truncation $G_n(1)$ is compared from both sides.

> *"Justify" means the examiners know I can guess* 3*. An infinite radical has no value until I say which finite things it is the limit of. So: truncate at depth* n*, and prove the truncations tend to* 3*.*

> *My identity gives exactly* 3 *if the slot at depth* n *holds* $n+2$ *instead of* 1*. So I'm comparing the same finite radical at two inputs. One side is monotonicity. For the other side I need to know how much a big change at depth* n *can matter at the top, and each square root halves the exponent of a multiplicative error.*

> *Pulling a factor* $\lambda \geq 1$ *out:* $\sqrt{1+\lambda k y} \leq \sqrt{\lambda}\sqrt{1+ky}$*, because* $1 \leq \lambda$*. After* $n-1$ *roots, the error* $n+2$ *has become* $(n+2)^{2^{1-n}}$*, which goes to* 1*.*

Layer 3 THE WHY

Why must the thought "define it first" be thought? Because of (†) read forwards. The self-similar relation $f(n)^2 = 1 + nf(n+1)$ is the only structural fact the symbols give you, and it does not pin down $f(2)$: the forward recursion accepts every starting value. For $f(2) > 3$ the solution grows like a tower of squarings, $f(n) \approx C^{2^n}$; for $f(2) < 3$ it dies. So the equation has a one-parameter family of solutions, and the value 3 is singled out only by an extra condition — the growth rate. That is not a technicality. It is the entire mathematical content of the problem, and the word "justify" is the examiners telling you so.

Now the growth condition has to come from somewhere, and the only place it can come from is the definition of the infinite expression. Truncating at depth n with a fixed number (here 1) is a choice that says "the tail is small". The scaling lemma quantifies exactly how small is small enough. Replace 1 by any tail $t_n \geq 0$ with $\ln t_n = o(2^n)$: the lower bound $G_n(t_n) \geq G_n(0) = a_{n-1}$ and the upper bound $G_n(t_n) \leq 3\big(t_n/(n+2)\big)^{2^{1-n}}$ (when $t_n \geq n+2$) both still squeeze to 3. But take $t_n = e^{2^n}$ and every truncation is at least $t_n^{2^{1-n}} = e^2$, so the limit is not 3; numerically it is about 20.4. The symbols alone do not determine the number. The definition does. That is why the model begins by writing $a_n = G_n(1)$: it is not pedantry, it is the step on which the answer depends.

Next, why is the identity $(k+1)^2 = 1 + k(k+2)$ the natural anchor, rather than a lucky find? Work backwards from the outermost layer. $3 = \sqrt{9} = \sqrt{1+2\cdot 4}$, so the tail after the 2 must be 4. $4 = \sqrt{16} = \sqrt{1+3\cdot 5}$, so the next tail is 5. The pattern $k+1 = \sqrt{1+k(k+2)}$ is forced by peeling one layer at a time; it is the $f(n) = n+1$ solution of (†). What the model adds is the observation that this peeling, done $n-1$ times, is a finite exact statement: $G_n(n+2) = 3$. A finite identity is something you can compare with; an infinite one is not.

Last, why is multiplicative scaling the right comparison tool? The two inputs you must compare, 1 and $n+2$, are far apart in absolute terms and get further apart as n grows, so any additive estimate is hopeless. But the square root is the map that turns a factor λ into $\lambda^{1/2}$: nested m times, a factor λ at the bottom is at most $\lambda^{2^{-m}}$ at the top. The "1+" inside each root is the only obstacle, and it goes the right way because $1 \leq \lambda \cdot 1$. Double-exponential decay of the error beats the polynomial growth $n+2$ of the discrepancy, and that is the whole proof. Once you see that each layer takes a square root, "how does an error at the bottom propagate to the top?" has only one answer.

Layer $+\alpha$ BACKTESTING

1. The first cue is the verb "justify" applied to an infinite expression. Write down at once: $a_n = \sqrt{1+2\sqrt{\cdots\sqrt{1+n}}}$, and the claim $\lim a_n = 3$.

2. A nested expression with a guessable value suggests peeling layers to find the exact tail that makes the value come out right $(4, 5, 6, \ldots)$, then comparing the actual truncation with that exact finite version from both sides.

3. Hesitation comes after the identity is found: it feels like the problem is over. Reinforce it away by asking whether the recurrence (†) has other solutions — run it forwards from $3+\varepsilon$ and watch it explode. Then the need for a limit argument is obvious.

4. The trap is to verify that $f(n) = n+1$ satisfies $f(n)^2 = 1+nf(n+1)$ and call that a proof. It proves only that 3 is a consistent value, not that the radical converges to it; super-exponentially large tails give other values.

5. Any time a problem says "justify", "make sense of" or "show that this infinite expression equals", the grader wants the definition as a limit of finite objects and a two-sided estimate. Recognise the verb as an instruction to define.

6. To evaluate an infinite nested expression: find the exact finite version (the tail that makes the value exact), then bound the error from replacing that tail by the real truncation, tracking how an error at the bottom is damped by each layer. For square roots the damping is $\lambda \mapsto \lambda^{1/2}$ per layer.

▷ **Cue:** "justify" or "prove" an equality involving an infinite expression $\longrightarrow$ **Move:** define the expression as a limit of explicit finite truncations before touching the algebra

▷ **Cue:** a non-periodic nested radical with a guessable value $\longrightarrow$ **Move:** peel the outer layer to find the required tail; the pattern of tails gives an exact identity such as $(k+1)^2 = 1+k(k+2)$

▷ **Cue:** a candidate satisfies a self-similar relation $f(n)^2 = 1+nf(n+1)$ $\longrightarrow$ **Move:** run the relation forwards from nearby starting values; if they also satisfy it, the candidate needs a growth condition as well as a check

▷ **Cue:** comparing two inputs to a deep composition of square roots $\longrightarrow$ **Move:** use multiplicative scaling: $\sqrt{1+\lambda u} \le \sqrt{\lambda}\sqrt{1+u}$ for $\lambda \ge 1$, so a factor λ at depth m is at most $\lambda^{2^{-m}}$ at the top

▷ **Cue:** an increasing sequence bounded by an exact value $\longrightarrow$ **Move:** get the matching lower bound by comparing with the exact finite version and showing the comparison factor tends to 1

Putnam 1966 — Paper B

1966 B1

1966 B1. Let a convex polygon P be contained in a square of side one. Show that the sum of the squares of the sides of P is less than or equal to 4.

Fields: Plane geometry, geometric inequalities **Position:** B1 (opener) **Answer:** The bound 4 holds, and it is tight – the unit square itself achieves equality.

Layer 1 THE WHAT — surface analysis

The natural attempt. You want a bound on $\sum |e_i|^2$, where $e_1, \ldots, e_n$ are the side vectors of P, traversed once around, so $\sum e_i = 0$. The most immediately available fact is that every side of P lies inside a unit square, so no side can be longer than the square's diagonal: $|e_i| \le \sqrt{2}$ for every i. You also know a genuinely true fact about convex regions: a convex region contained in a unit square has perimeter at most the perimeter of the square, so $L = \sum |e_i| \le 4$. Put these together and you get

$$\sum |e_i|^2 \le \Big(\max_i |e_i|\Big) \sum_i |e_i| \le \sqrt{2} \cdot 4 = 4\sqrt{2} \approx 5.66.$$

This is a completely legitimate chain – both inequalities used are true and provable – and it produces a finite bound of the right general shape. But $4\sqrt{2}$ is not 4. You have proved something true and weaker than what was asked. Tightening it from here is not obvious: the diagonal bound $|e_i| \le \sqrt{2}$ is only ever achieved by a side running corner-to-corner, and a real polygon cannot have every side doing that simultaneously, but turning "not every side is that long" into a numerically sharper constant, by staying inside this one-dimensional bookkeeping of "total length times worst side," is a dead end. The length of each side and the perimeter total are simply the wrong quantities to be squaring and multiplying together.

The model chain. Set up coordinates so the square is $[0,1] \times [0,1]$. Write each side vector as $e_i = (a_i, b_i)$, so the side length satisfies $|e_i|^2 = a_i^2 + b_i^2$, and it suffices to bound $\sum a_i^2$ and $\sum b_i^2$ separately by 2 each.

Consider only the x-components a_i. As you traverse the boundary of a convex polygon once, in a fixed direction, the x-coordinate of the moving vertex is a *unimodal* function of the step index: it increases monotonically from the leftmost vertex to the rightmost vertex, then decreases monotonically back to the start. (This is exactly convexity: the two arcs of the boundary between the leftmost and rightmost points project onto the x-axis without folding back, because a convex curve can cross any vertical line at most twice.) Consequently the positive a_i's sum to exactly the horizontal extent of P, and the negative a_i's sum to exactly the same extent with the opposite sign – both are at most 1, since P sits inside the unit square. Hence

$$\sum_i |a_i| \;=\; \sum_{a_i>0} a_i \;+\; \sum_{a_i<0} (-a_i) \;\le\; 1 + 1 \;=\; 2.$$

At the same time, every individual $|a_i| \leq 1$, because it is the length of a horizontal projection confined to $[0,1]$. So

$$\sum_i a_i^2 \;\leq\; \Big(\max_i |a_i|\Big) \sum_i |a_i| \;\leq\; 1 \cdot 2 \;=\; 2.$$

The identical argument, run on the y-components, gives $\sum_i b_i^2 \leq 2$. Adding,

$$\sum_i |e_i|^2 = \sum_i a_i^2 + \sum_i b_i^2 \leq 2 + 2 = 4. \qquad \blacksquare$$

Equality is achieved by the unit square itself: four sides of length 1, sum of squares exactly 4.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths start from the same object – the side vectors e_i of a convex polygon confined to a unit square – and both correctly reach for the identity $|e_i|^2 = a_i^2 + b_i^2$ in spirit. They split at the choice of *which single number* to bound each side by before summing squares. The natural attempt bounds every side by the same global ceiling, the diagonal $\sqrt{2}$, applied uniformly regardless of a side's orientation. The model instead bounds each side's two *coordinates separately*, each by 1 (not $\sqrt{2}$), and only combines them into $a_i^2 + b_i^2$ at the very last step.

> *A side's length is capped by the diagonal only when it runs corner to corner – but I don't need a bound on the length of one side at all. I need a bound on the sum of two coordinate-wise sums, and each coordinate individually never exceeds the side of the square, which is* 1*, not* $\sqrt{2}$*. I should split the vector into its components before I square anything, not after.*

> *The polygon is convex and closed, so its boundary, read off against either axis alone, goes up once and down once – it cannot wiggle back and forth. That unimodality is what turns "confined to a unit interval" into "total positive displacement plus total negative displacement is at most* 2*," which is the only fact doing real work in the whole proof.*

- ▷ **Cue:** a closed convex polygon (or any closed convex curve) sits inside a bounding box and you need a bound on a sum of squares of its side lengths or displacements $\longrightarrow$ **Move:** decompose every side into its two axis components separately, use convexity to bound the total positive and total negative displacement along each axis by the box's width/height, and only recombine via $|e_i|^2 = a_i^2 + b_i^2$ after each coordinate sum is separately controlled
- ▷ **Cue:** you have a sum $\sum x_i^2$ where each $0 \leq x_i \leq M$ (a known uniform ceiling) and you also know $\sum x_i \leq S$ $\longrightarrow$ **Move:** bound $\sum x_i^2 \leq M \sum x_i \leq MS$, since $x_i^2 = x_i \cdot x_i \leq M \cdot x_i$ termwise – this converts a sum of squares into a scaled sum of the (linear, easier-to-control) quantity, and is sharp exactly when the x_i that matter sit at M

Layer 3 THE WHY

The natural attempt fails because it treats "side length" as the fundamental unit to be squared and summed, when the problem's real structure lives one level lower, in the two coordinate axes separately. Why is that the right level? Because the containing region is a *square*, not a disc: its constraint on the polygon is not "distance from a fixed centre is at most r" (which would make the diagonal-length bound natural, since a disc has one radius in every direction) but "the x-coordinate lies in $[0,1]$ *and* the y-coordinate lies in $[0,1]$" – two independent, axis-aligned constraints. A shape defined by two separate one-dimensional constraints should be attacked by two separate one-dimensional arguments, recombined only at the point where the geometry (Pythagoras) forces recombination. Squaring the diagonal bound first collapses the two constraints into one number too early, and that number, $\sqrt{2}$, is an artefact of the square's

shape at a 45° angle – it has nothing to do with how any *particular* side is oriented, so applying it to every side simultaneously is exactly where the slack of $4\sqrt{2}$ versus 4 comes from.

The convexity hypothesis is not decoration; it is what makes "total positive displacement ≤ 1" true at all. Without convexity a polygon's x-coordinate could oscillate many times while staying in $[0, 1]$, and the positive and negative parts of $\sum a_i$ could each be much larger than 1 (a zig-zag can retrace the same interval many times). Convexity is precisely the statement that the boundary crosses any line at most twice, which is what forces the x-projection to be unimodal – go right, then go left, no more reversals. This is the same "of course" as the METHOD.md circle example: once you know a convex curve cannot double back across a fixed line, projecting it onto that line and bounding the up-swing and down-swing separately is not a trick, it is reading off the definition of convexity in coordinates. The official archive solution states this projection step in one line ("the first side will be covered at most twice") without unpacking why convexity guarantees the "at most twice" – that guarantee *is* the definition of a convex boundary meeting a line in at most two points, restated for a bounding side instead of an arbitrary line, and it is worth saying explicitly rather than passing over.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "convex polygon contained in a square," which pairs a shape whose defining constraint is genuinely two separate one-dimensional bounds (an x-range and a y-range) with a quantity, sum of squared side lengths, that Pythagoras will split into x- and y-parts anyway. On seeing it, write down: bound each coordinate separately before recombining, don't bound the diagonal length first.
2. That cue suggests decomposing each side vector into components and bounding $\sum a_i^2$ and $\sum b_i^2$ independently, each via the elementary "sum of squares $\leq$ ceiling $\times$ sum" trick, using convexity to control the sum of $|a_i|$ (respectively $|b_i|$).
3. The hesitation happens right after writing $|e_i| \leq \sqrt{2}$: it feels like real progress, because it is true and it is the most obviously available fact about a segment in a unit square. It is reinforced away by actually finishing that computation to see the constant come out as $4\sqrt{2}$, not 4 – carrying a plausible-looking approach all the way to a number is how you discover it is the wrong one, rather than abandoning it on a hunch.
4. The trap is the diagonal bound $|e_i| \leq \sqrt{2}$: it is correct, tempting, and uniform (applies to every side identically without any case analysis), which makes it feel like the natural first move – and it is precisely uniform because it throws away the side's orientation, which is the one piece of information the problem's answer of exactly 4 depends on.
5. To recognise this faster: whenever the bounding region is a box (product of independent coordinate constraints) rather than a ball (one radial constraint), reflexively split any vector quantity into its coordinates before bounding it, and only reassemble with the Pythagorean identity at the final step.
6. Decision rule: to bound $\sum |v_i|^2$ for vectors v_i constrained by a coordinate-wise box (rather than a single radius), decompose into components, bound each component's sum-of-squares via "each term $\leq$ known ceiling, so sum-of-squares $\leq$ ceiling $\times$ sum-of-terms," and use any monotonicity or convexity in the data to bound that linear sum – then add the per-axis bounds via Pythagoras. Never let a single worst-case diagonal or norm bound stand in for a component-wise argument when the constraint itself is component-wise.

1966 B2

1966 B2. Prove that among any ten consecutive integers at least one is relatively prime to each of the others.

Fields: Number theory, divisibility and the pigeonhole principle **Position:** B2 (middle) **Answer:** True: in any block of ten consecutive integers, some element shares no common factor greater than 1 with any

of the other nine.

Layer 1 THE WHAT — surface analysis

The natural attempt. The instinctive first move is parity: half of any ten consecutive integers are odd, so surely one of those five odd numbers is the answer – an odd number can't share a factor of 2 with anything. Try to run with this. Take $2, 3, 4, \ldots, 11$. The odd numbers are $3, 5, 7, 9, 11$. Check $9 = 3^2$: it shares the factor 3 with 3 itself. Check 3: it shares a factor of 3 with 9. So neither 3 nor 9 works, even though both are odd. The odd-number idea is not false, exactly – it correctly rules out sharing a factor of 2 – but it was never enough on its own, because 3 and 5 and 7 can still collide with each other or with a multiple of themselves sitting nearby. You could try to patch this by also excluding multiples of 3 from the odd numbers, leaving $5, 7, 11$ in this example, and that patch happens to work here. But it's not obvious the patch always leaves something over: what if, in a different block of ten, the surviving "odd and not-a-multiple-of-3" numbers were still colliding with each other via 5 or 7? You are now stacking ad hoc exclusions (not 2, then not 3, then check 5, then check 7...) without a principle telling you when to stop, or why four primes are enough, or why you won't run out of numbers before you run out of primes to worry about.

The model chain. The key structural fact is: if two of the ten consecutive integers $n, n+1, \ldots, n+9$ share a common factor $d > 1$, then d divides their difference, and that difference is at most 9. So $d \leq 9$, which means every prime factor of d is at most 9 – and since the primes up to 9 are exactly $2, 3, 5, 7$, *any* shared factor between two numbers in this block must trace back to one of these four primes.

This converts the problem into: show that among the ten integers, at least one is divisible by none of $2, 3, 5, 7$. If such a number exists, it automatically cannot share a factor with any other number in the block, by the paragraph above – so it is relatively prime to all nine others, and we are done.

Now count. Exactly five of the ten integers are even. Among the remaining five odd integers, consider how many are multiples of 3. Multiples of 3 occur every third integer and alternate in parity (since 3 is odd, consecutive multiples of 3 differ by an odd number, so their parities alternate); hence among any run of ten consecutive integers, the multiples of 3 that are also odd number at most 2. So at least $5 - 2 = 3$ of the ten integers are simultaneously odd and not divisible by 3.

These three (or more) survivors still need checking against 5 and 7. Two multiples of 5 can appear in a window of length ten only if they differ by exactly 5, and since 5 is odd, one of that pair is even and one is odd – so at most *one* odd multiple of 5 survives into our set of "odd, not divisible by 3" numbers. The same argument, with 7 in place of 5, shows at most one odd multiple of 7 survives (two multiples of 7 within a length-ten window must differ by exactly 7, an odd number, so at most one of the pair is odd).

So starting from at least 3 survivors (odd, not divisible by 3), removing at most 1 for divisibility by 5 and at most 1 for divisibility by 7 leaves at least $3 - 1 - 1 = 1$ integer, divisible by none of $2, 3, 5, 7$. By the structural fact above, this integer is relatively prime to every other integer in the block. ■

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths begin at the same place: parity splits the ten integers in half, and an odd number cannot collide via the prime 2. They split at the very next question – *how many primes do you actually have to worry about, and why exactly that many?* The natural attempt answers this reactively, discovering by trial that 3 also matters, then presumably 5, then 7, patching the argument prime by prime with no assurance of termination. The model answers it structurally, in one stroke, before doing any casework: a shared factor must divide a difference that is at

most 9, so it is built entirely from primes at most 9 – there are exactly four of them, full stop, and no fifth prime will ever need checking.

> *If two numbers in this block of ten share a common factor d, that d has to divide their difference – and their difference is less than* 10*. So I don't need to test divisibility by every prime; I only need to worry about primes that can possibly divide a number less than* 10*. That's a finite, fixed, and small list:* $2, 3, 5, 7$*. I should find this list* before *I start sieving, not discover it one prime at a time by trial.*

> *Once I know the only primes that matter are* $2, 3, 5, 7$*, this is now a pure pigeonhole-and-counting problem: how many of the ten integers can I strike off (for being divisible by at least one of these four primes) before I'm forced to have at least one left over? I should count how many survive each sieve, not check individual numbers by hand.*

- ▷ **Cue:** two elements of a set of consecutive integers (or of any set with bounded spread) are said to share a common factor $\longrightarrow$ **Move:** that common factor must divide their difference; if the whole set has diameter (max minus min) at most N, only primes $p \leq N$ need ever be considered, giving a small fixed list of primes to sieve against instead of an open-ended search
- ▷ **Cue:** you need to show a sieve by finitely many primes, each removing roughly half (or a fixed fraction) of a finite set, cannot remove everything $\longrightarrow$ **Move:** count survivors after each prime's sieve explicitly (using that consecutive multiples of an odd prime alternate in parity to bound how many can already be excluded), and check the running total stays positive rather than guessing that "enough primes were used"

Layer 3 THE WHY

The natural attempt's real flaw is not that it is wrong about 2 and 3 – it is that it never asks the question that would tell it when to stop. "Odd numbers avoid a factor of 2" is true but says nothing about *why* there is a finite list of primes to worry about at all; without that, checking 3, then wondering about 5, then 7, feels like it could go on indefinitely, and there is no way to be confident the process terminates in your favour. The fact that resolves this is almost definitional: a common factor of two numbers must divide their difference (this is literally what "common factor" and "divides" mean, chained together via $\gcd(x, y) \mid (x - y)$), and the numbers in question are confined to a window of size 10, so their difference is bounded by 9. Boundedness of the window is therefore not incidental scaffolding – it is the single fact that turns "infinitely many primes could matter" into "only primes up to 9 can matter," because a prime larger than the window's diameter simply cannot divide any nonzero difference that small. This is the same move, one level down, as bounding a problem by its own size: the *ten* in "ten consecutive integers" is not just a count of how many numbers to sieve, it is also the ceiling on which primes are even eligible to cause a collision. Once that ceiling is named, the four primes $2, 3, 5, 7$ are not a guess refined by trial, they are the complete and provably exhaustive list, and the rest of the argument is bookkeeping: how much of the ten does sieving by four fixed primes remove, and is there slack left over. The archive's official solution reaches for exactly this sieve but states the parity-alternation facts about multiples of 3, 5, 7 tersely; spelling out *why* at most two odd multiples of 3 (respectively one each of 5 and 7) can appear is what turns "it works out" into a proof with no arithmetic left unaccounted for.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "ten consecutive integers" combined with "relatively prime to each of the others." On seeing it, write down: any shared factor divides a difference that is less than 10, so only primes up to 9 can ever be responsible for a collision.

2. That cue suggests turning the problem into a finite sieve: show at least one of the ten survives exclusion by each of $2, 3, 5, 7$, by counting how many integers each sieve removes and confirming the total removed is strictly less than ten.

3. The hesitation happens right after the parity observation ("half are odd"), when it is tempting to think the problem is basically solved and just needs "a bit more checking." It is reinforced away by actually testing a concrete example (like $2, \ldots, 11$, where 3 and 9 both fail despite being odd) – a single counterexample to the too-easy version forces you to ask what other primes are in play, and forces the question of *how many* before you trust any count.

4. The trap is stopping the sieve after 2 and 3 out of familiarity (they are the two smallest and most reflexive primes to check), without deriving that the list is exactly $\{2, 3, 5, 7\}$ and provably no larger – an argument that happens to survive on the first example you try is not yet a proof.

5. To recognise this faster: whenever a problem bounds the spread of a set of integers (here, "ten consecutive"), immediately translate that bound into a bound on which primes can possibly divide any pairwise difference – this converts an open-ended divisibility question into a finite, nameable list of primes before any casework begins.

6. Decision rule: to show some element of a bounded set of integers is coprime to all the others, first bound the primes that could possibly cause a collision (via $p \mid \gcd \Rightarrow p \mid$ difference $\leq$ diameter of the set), then run a counting or pigeonhole argument against exactly that finite list – never sieve prime-by-prime on faith that "enough" primes have been tried.

1966 B3

1966 B3. Show that if the series $\displaystyle\sum_{n=1}^{\infty} \frac{1}{p_n}$ is convergent, where $p_1, p_2, p_3, \ldots, p_n, \ldots$ are positive real numbers, then the series $\displaystyle\sum_{n=1}^{\infty} \left(\frac{n}{p_1 + p_2 + \cdots + p_n} \right)^2 p_n$ is also convergent.

Caveat. The MAA 1965–1984 OCR renders the problem statement's second display as a blank – the sentence reads "…then the series ___ is also convergent" with the formula itself dropped, evidently a typeset equation lost in scanning. The statement above is reconstructed from the book's own official solution, which is intact and unambiguous: it defines $q_n = p_1 + p_2 + \cdots + p_n$ and estimates $S_N = \sum_{n=1}^{N} (n/q_n)^2 (q_n - q_{n-1})$, i.e. exactly $\sum_{n=1}^{N} (n/q_n)^2 p_n$, against $T = \sum 1/p_n$. I independently re-derived a full, rigorous proof of the convergence of this exact series from $\sum 1/p_n < \infty$ (given in the model chain below) and it matches the shape and every intermediate quantity of the book's terse solution, which confirms the reconstruction rather than merely assuming it. No other mathematical content was altered.

Fields: Real analysis, series and inequalities (Cauchy–Schwarz, Abel summation) **Position:** B3 (middle-to-late) **Answer:** True: convergence of $\sum 1/p_n$ forces convergence of $\sum (n/q_n)^2 p_n$, where $q_n = p_1 + \cdots + p_n$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The instinct on seeing "one series converges, prove another converges" is comparison or ratio test, applied termwise. Try comparison first: does $(n/q_n)^2 p_n \leq C/p_n$ for some constant C, for all n? That would need $n^2 p_n^2 \leq C q_n^2$, i.e. $p_n/q_n \leq \sqrt{C}/n$ roughly – but p_n can be enormous compared to q_{n-1} (nothing stops p_n from being, say, q_{n-1} itself, doubling q_n at each step), so no such termwise bound holds in general. Termwise comparison against $1/p_n$ is a dead end.

Try Cauchy–Schwarz next, since it is the standard tool for turning one sum into a product of two others – but applied naively, at the level of the whole sum: split $(n/q_n)^2 p_n$ as $(n\sqrt{p_n}/q_n)\cdot(\sqrt{p_n})$ and try to bound $\sum(n/q_n)^2 p_n$ by $\left(\sum n^2 p_n/q_n^2\right)$ directly, which is circular – it is the same sum you started with. Or try pairing n/q_n with itself across the whole sum via $\left(\sum n/q_n\right)^2 \le N\sum n^2/q_n^2$, which introduces an N that grows and defeats the purpose. Every direct attempt to apply Cauchy–Schwarz *across the whole series at once*, rather than to a cleverly chosen finite decomposition inside each term, either reproduces the target sum unchanged or introduces an unwanted extra factor of N. You can verify the claim is at least plausible on an example – take $p_n = n^2$, so $\sum 1/p_n = \pi^2/6$ converges, and $q_n \sim n^3/3$, giving $(n/q_n)^2 p_n \sim 9/n^2$, which converges – but one convergent example built from a tidy formula for p_n does not touch the general case, where p_n can grow however it likes as long as $\sum 1/p_n < \infty$.

The model chain. Set $q_n = p_1 + \cdots + p_n$ (so $q_0 = 0$, q_n strictly increasing to ∞, since $\sum 1/p_n < \infty$ forces $p_n \to \infty$, hence $q_n \to \infty$) and $T = \sum_{n\ge 1} 1/p_n < \infty$. Write the partial sums $S_N = \sum_{n=1}^{N}(n/q_n)^2 p_n$; since every term is positive, it suffices to show S_N is bounded above independently of N.

The move that unlocks the problem is to weaken $1/q_n^2$ to $1/(q_n q_{n-1})$ for $n \ge 2$ (valid since $q_{n-1} \le q_n$, so $q_n^2 \ge q_n q_{n-1}$), because this makes the term *telescope*:

$$\frac{n^2}{q_n^2}p_n = \frac{n^2(q_n - q_{n-1})}{q_n^2} \le \frac{n^2(q_n - q_{n-1})}{q_n q_{n-1}} = n^2\left(\frac{1}{q_{n-1}} - \frac{1}{q_n}\right) \qquad (n \ge 2).$$

Summing this telescoping form by parts (Abel summation), for $n = 2, \ldots, N$:

$$\sum_{n=2}^{N} n^2\left(\frac{1}{q_{n-1}} - \frac{1}{q_n}\right) = \sum_{m=1}^{N-1}\frac{(m+1)^2}{q_m} - \sum_{n=2}^{N}\frac{n^2}{q_n} = \frac{4}{q_1} + \sum_{m=2}^{N-1}\frac{(m+1)^2 - m^2}{q_m} - \frac{N^2}{q_N}.$$

Since $(m+1)^2 - m^2 = 2m + 1 \le 3m$ and the dropped term $-N^2/q_N$ is negative, this is at most $\dfrac{4}{q_1} + 3\sum_{m=2}^{N-1}\dfrac{m}{q_m}$. Adding back the $n = 1$ term, $(1/q_1)^2 p_1 = 1/p_1$ (as $q_1 = p_1$), and using $4/q_1 = 4/p_1$:

$$S_N \le \frac{5}{p_1} + 3\sum_{m=2}^{N-1}\frac{m}{q_m}.$$

Now Cauchy–Schwarz is applied correctly – not to the whole target sum, but to the auxiliary sum $\sum m/q_m$, splitting each term as $(m/q_m)\sqrt{p_m}\cdot(1/\sqrt{p_m})$:

$$\sum_{m=1}^{N-1}\frac{m}{q_m} = \sum_{m=1}^{N-1}\frac{m\sqrt{p_m}}{q_m}\cdot\frac{1}{\sqrt{p_m}} \le \left(\sum_{m=1}^{N-1}\frac{m^2 p_m}{q_m^2}\right)^{1/2}\left(\sum_{m=1}^{N-1}\frac{1}{p_m}\right)^{1/2} \le \sqrt{S_N}\cdot\sqrt{T}.$$

Substituting back, $S_N \le 5/p_1 + 3\sqrt{S_N\,T}$. Writing $x = \sqrt{S_N} \ge 0$, this is $x^2 - 3\sqrt{T}\,x - 5/p_1 \le 0$, a quadratic inequality whose positive root gives

$$x \le \frac{3\sqrt{T} + \sqrt{9T + 20/p_1}}{2},$$

a bound depending only on T and p_1, not on N. So S_N is bounded above uniformly in N; being an increasing sequence of partial sums of a positive series, it therefore converges. ■

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths reach for Cauchy–Schwarz – it is the correct tool, and any strong student will name it almost immediately given the shape $(n/q_n)^2 p_n$, which looks built for it. They split at *what Cauchy–Schwarz is applied to.* The natural attempt applies it to the target sum

as a whole, treating $\sum(n/q_n)^2 p_n$ itself as one side of an inequality to be bounded – which is circular, since the sum you are trying to control appears (in disguise) on both sides. The model first performs an entirely different manoeuvre – weakening q_n^2 to $q_n q_{n-1}$ to make the summand telescope – and only *after* that reduces the problem to a strictly smaller and different quantity, $\sum m/q_m$, which is where Cauchy–Schwarz is then applied cleanly, with S_N appearing on only one side.

> *I keep getting the same sum back when I apply Cauchy–Schwarz directly to $\sum(n/q_n)^2 p_n$ – that's a sign I'm applying the inequality at the wrong level. I should first change the shape of the summand into something that is no longer self-referential, and only then bring in Cauchy–Schwarz on whatever smaller, different quantity is left over.*

> *The quantity q_n^2 in the denominator is an equality-forcing choice, not a necessary one – I only need an* upper *bound on the term, so I'm free to replace q_n^2 with anything smaller that makes the algebra behave, as long as it's still $\geq$ what I need it to dominate. Replacing it with $q_n q_{n-1}$ costs nothing (since $q_{n-1} \leq q_n$) and buys a telescoping difference $1/q_{n-1} - 1/q_n$, which is exactly the kind of object a partial-sum bound wants to see.*

- ▷ **Cue:** a sum of the form $\sum(n/q_n)^2 p_n$ or similar, where q_n is an increasing partial-sum sequence and $p_n = q_n - q_{n-1}$ appears both in the numerator's role and as the increment of the denominator $\longrightarrow$ **Move:** weaken the denominator q_n^2 to the smaller, telescoping-friendly product $q_n q_{n-1}$ (valid since $q_{n-1} \leq q_n$), turning p_n/q_n^2 into a genuine difference $1/q_{n-1} - 1/q_n$ that a partial sum will collapse via Abel summation, before reaching for any comparison or Cauchy–Schwarz
- ▷ **Cue:** after simplifying, Cauchy–Schwarz would need to be applied to the very sum you are trying to bound, reproducing it on the other side of the inequality $\longrightarrow$ **Move:** stop – this signals the inequality is being applied one level too high; first transform or reduce the target sum (via telescoping, substitution, or splitting off known-convergent pieces) into a genuinely smaller or different auxiliary quantity, and apply Cauchy–Schwarz to that instead, so the target sum S_N appears on only one side of the resulting inequality

Layer 3 THE WHY

The reason termwise comparison against $1/p_n$ fails is structural, not a failure of cleverness: q_n mixes together *all* of $p_1, \ldots, p_n$, so the size of one term of the target series depends on the entire history of the p_k's, not on p_n alone. Any proof strategy that tries to bound term n by a function of p_n alone is fighting the problem's actual dependency structure and will not work except in special cases. This is why the correct move has to operate on the *partial sum* q_n as a single evolving object, and why Abel summation – summation by parts, the discrete analogue of integration by parts – is the natural tool: it is specifically designed for sums where a running total (q_n) interacts with an index-dependent weight (n^2), converting a sum of products into a boundary term plus a sum of differences.

Why replace q_n^2 with $q_n q_{n-1}$ specifically? Because $q_n - q_{n-1} = p_n$ is already sitting in the numerator, and $\dfrac{q_n - q_{n-1}}{q_n q_{n-1}} = \dfrac{1}{q_{n-1}} - \dfrac{1}{q_n}$ is not an approximation or a coincidence – it is the algebraic identity for the derivative of $-1/x$ in discrete form, and it is the *only* way to make a difference of reciprocals appear exactly, rather than approximately, from a difference in the numerator. Once you see that p_n/q_n^2 is "almost" $-\Delta(1/q_n)$, weakening the denominator to make it exactly that difference is not a trick – it is recognising that the problem is begging for a telescoping sum, because the numerator is already phrased as an increment of the denominator's argument.

The second Cauchy–Schwarz application succeeds where the first failed because by that point $\sum m/q_m$ is a genuinely *different, smaller-looking* object than S_N – and Cauchy–Schwarz,

applied to $m/q_m = \left(m\sqrt{p_m}/q_m\right)(1/\sqrt{p_m})$, produces S_N on one side and T on the other, with no circularity, because S_N now sits inside a square root, one full inequality-application removed from where it started. This is the general shape of "self-bounding" arguments: you cannot close a loop with the same tool applied at the same level twice, but a self-referential inequality of the form $x^2 \leq a + bx$ (here $x = \sqrt{S_N}$) is perfectly closeable, because it is just a quadratic – the circularity is not avoided, it is embraced and solved algebraically once it takes that clean form.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the shape $(n/q_n)^2 p_n$ with q_n an increasing partial-sum sequence and $p_n = q_n - q_{n-1}$. On seeing it, write down: this numerator is an increment of the very quantity in the denominator, so the term is a disguised discrete derivative – look for a telescoping form before reaching for comparison or Cauchy–Schwarz.
2. That cue suggests weakening q_n^2 to $q_n q_{n-1}$ to force the exact identity $p_n/(q_n q_{n-1}) = 1/q_{n-1} - 1/q_n$, then summing by parts (Abel summation) to convert the weighted sum into a telescoped boundary term plus a smaller auxiliary sum.
3. The hesitation happens right after naming Cauchy–Schwarz as the tool (it is visually the obvious one for this shape) and then applying it directly to the whole target series, only to find the same sum reappearing on the other side of the inequality. It is reinforced away by noticing explicitly that a self-referential bound of this kind proves nothing new, and backing up to ask what smaller quantity Cauchy–Schwarz should be applied to instead.
4. The trap is treating Cauchy–Schwarz as a tool to be fired at the target sum as a whole, on the first pass, rather than saving it for a later, smaller auxiliary quantity produced by an earlier reduction (here, the telescoping step). Reaching for the strongest inequality immediately, before simplifying the object it will act on, is a specific case of zombie maths: applying real machinery without first asking what shape it needs to see.
5. To recognise this faster: whenever a series term is built from an increasing partial-sum sequence q_n and its own increment $p_n = q_n - q_{n-1}$, check first whether weakening q_n to q_{n-1} (or vice versa) in a denominator produces an exact telescoping difference of reciprocals – this is nearly always available whenever the numerator already looks like $q_n - q_{n-1}$.
6. Decision rule: when a bound would apply an inequality to the very quantity you are trying to bound, reproducing it unchanged (or with an unwanted growing factor) on the other side, stop and look for a prior transformation – typically a telescoping sum via a weakened denominator, or an Abel summation – that reduces the target to a genuinely smaller auxiliary quantity; only then apply the strong inequality (Cauchy–Schwarz, AM-GM, etc.), so that the original quantity appears on one side only, ideally inside a square root or other invertible wrapper that turns the self-reference into a solvable algebraic equation.

1966 B4

1966 B4. Let $0 < a_1 < a_2 < \cdots < a_{mn+1}$ be $mn+1$ integers. Prove that you can select either $m+1$ of them no one of which divides any other, or $n+1$ of them each dividing the following one.

Fields: Combinatorics (pigeonhole principle), number theory (divisibility), partial orders **Position:** B4 (upper middle) **Answer:** Always possible: label each a_i with the length of the longest divisibility chain that starts at a_i; either some label exceeds n, or $m+1$ numbers share a label, and numbers sharing a label cannot divide one another.

Layer 1 THE WHAT — surface analysis

The natural attempt. You try small cases. With $m = n = 1$ you have three integers $a_1 < a_2 < a_3$: either some pair has one dividing the other (a chain of length $2 = n+1$), or no pair does and

you have three, hence two, mutually non-dividing numbers. Fine. With $m = n = 2$ you have five integers, and you start building a chain greedily: begin at a_1, jump to the smallest later number it divides, and so on. If the chain reaches length 3 you're done. If it stalls at length 2, you try to argue that the numbers you skipped form an antichain (no one divides another) — but they need not. For example, take $2, 3, 4, 9, \ldots$: the greedy chain from 2 is $2 \mid 4$ and stops, the leftovers $3, 9$ divide each other, and now you have to restart from 3. You find yourself restarting chains from different starting points, keeping track of which numbers have been "used", and the case analysis grows with every new m, n. You can feel that some counting is lurking — $mn + 1$ is exactly one more than $m \cdot n$ — but the greedy construction gives you no single quantity to count. This is the wall: a construction that builds one chain at a time never produces $m + 1$ *objects of the same kind* for pigeonhole to act on.

The model chain. For each index i with $1 \le i \le mn + 1$, let n_i be the length (number of terms) of the longest chain

$$a_i = a_{j_1} \mid a_{j_2} \mid \cdots \mid a_{j_k}, \qquad i = j_1 < j_2 < \cdots < j_k,$$

that starts at a_i and uses only terms from $a_i, a_{i+1}, \ldots, a_{mn+1}$. Every $n_i \ge 1$ (the one-term chain a_i).

Case 1: some $n_i \ge n + 1$. Then the first $n + 1$ terms of that chain are $n + 1$ of the given integers, each dividing the following one. Done.

Case 2: every $n_i \le n$. Then the $mn + 1$ labels $n_1, \ldots, n_{mn+1}$ take values in $\{1, 2, \ldots, n\}$, a set of n values. If each value were taken at most m times, there would be at most mn labels; there are $mn + 1$. So some value ℓ is taken by at least $m + 1$ indices $i_1 < i_2 < \cdots < i_{m+1}$.

Claim: no one of $a_{i_1}, \ldots, a_{i_{m+1}}$ divides another. Take two of them, a_i and a_j with $i < j$. Since $0 < a_i < a_j$, the only possible divisibility is $a_i \mid a_j$ (a larger positive integer cannot divide a smaller one). Suppose $a_i \mid a_j$. Take a longest chain starting at a_j; it has $n_j = \ell$ terms, all with indices $\ge j > i$. Putting a_i in front gives a valid chain starting at a_i of length $\ell + 1$, so $n_i \ge \ell + 1 > \ell = n_i$, a contradiction. Hence the $m + 1$ numbers $a_{i_1}, \ldots, a_{i_{m+1}}$ form the required set in which none divides any other.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree that the answer is "chain or antichain" and that $mn + 1 = m \cdot n + 1$ smells of pigeonhole. They split at the question of *what to pigeonhole*. The natural attempt constructs one chain and inspects what's left over. The model chain constructs nothing: it attaches a number n_i to *every* element at once, and lets pigeonhole act on those numbers.

> *I need $m + 1$ things of the same kind, and $mn + 1$ is one more than m times n, so I want a label with only n possible values that I can hang on every element simultaneously; the natural label with n values is "how long a chain can I get from here", and if that label ever goes past n I have won the other way.*

> *If a_i divides a_j and j comes later, then anything that can be done starting from a_j can be done starting from a_i with one extra step — so equal labels forbid divisibility.*

Layer 3 THE WHY

Why is "longest chain from here" the label, and not something invented? Go back to what the two alternatives demand. A chain of length $n + 1$ is a statement about *length*. An antichain of size $m + 1$ is a statement about *how many*. Pigeonhole converts "how many" into "too few boxes": if you have $mn + 1$ objects in n boxes, one box holds $m + 1$. So the proof must supply n boxes, and the only quantity in the problem that is naturally bounded by n is chain length — precisely because if it isn't bounded by n, the first alternative holds. The boxes are

therefore forced: box ℓ contains the elements whose longest chain has length ℓ. The $mn+1$ in the statement is not arbitrary; it is exactly the count that makes n boxes overflow to $m+1$.

The second half — that one box is an antichain — follows from the definition of "longest". A longest-chain function is monotone along the relation it measures: if x precedes y in the order $(a_i \mid a_j, i<j)$, then $\text{long}(x) \geq \text{long}(y)+1$, because any chain from y extends backwards to x. Strictly increasing along the order means it can never be equal on two comparable elements. That is the whole content, and it's the definition of maximum, nothing more.

The ordering $a_1 < a_2 < \cdots$ does quiet work here. Divisibility among positive integers already respects size $(a \mid b, a \neq b \Rightarrow a < b)$, so "each dividing the following one" automatically means the chain runs left to right in the index order. That is why it's legitimate to define n_i using only later terms: no chain could ever step backwards.

This is the same mechanism as the Erdős–Szekeres theorem (any sequence of $mn+1$ distinct reals has an increasing subsequence of length $n+1$ or a decreasing one of length $m+1$), whose standard proof labels each term with the longest increasing subsequence starting there. It is also the easy direction of Mirsky's theorem: in any finite partial order, the elements can be split into h antichains, where h is the length of the longest chain — the "levels" are the boxes above. Divisibility is a partial order; the problem is Mirsky with numbers attached.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the count $mn+1$ set against two alternatives of sizes $m+1$ and $n+1$. On seeing it, write down: "$mn+1$ objects, n boxes, one box gets $m+1$ — what are the n boxes?"
2. The method is labelling plus pigeonhole: give every element a label bounded by n whenever the $n+1$ alternative fails, and show that equal labels produce the $m+1$ alternative.
3. Hesitation occurs at defining the label — "longest chain starting here" feels like it needs computing. It doesn't: you never compute it, you only use that it exists and is maximal. Reinforce this by writing the one-line extension argument $(a_i \mid a_j \Rightarrow n_i \geq n_j+1)$ before worrying about anything else.
4. The trap is greedy construction: building a single chain and hoping the leftovers form an antichain. Leftovers carry no structure, and the case analysis never closes.
5. Recognise faster: whenever a problem says "either A of one kind or B of another" and the total is $(A-1)(B-1)+1$, it is a pigeonhole on a label with $B-1$ values, and the label is "the best you can do from this element".
6. Decision rule: in an "either a long chain or a large antichain" problem for any order relation, label each element by the length of the longest chain starting (or ending) at it; comparable elements get different labels by maximality, so each label class is an antichain, and pigeonhole over the labels finishes the proof.

1966 B5

1966 B5. Given $n\,(\geq 3)$ distinct points in the plane, no three of which are on the same straight line, prove that there exists a simple closed polygon with these points as vertices.

Fields: Combinatorial geometry, extremal principle **Position:** B5 (hard half) **Answer:** Always exists: a closed polygon through all the points of minimum total length cannot cross itself, since swapping a crossing pair of edges strictly shortens it; alternatively, sort the points by angle around a convex-hull vertex.

Layer 1 THE WHAT — surface analysis

The natural attempt. You draw some points and start joining them up, trying to find a rule that never creates a crossing. The first rule you try is "go to the nearest unvisited point" — and

it fails: on a few points you can draw a nearest-neighbour tour whose closing edge back to the start cuts across an earlier edge. The second idea is repair: take any closed polygon, and whenever two edges cross, fix that crossing. You draw the fix (replace the two crossing edges by two non-crossing ones through the same four endpoints) and it works locally. Then you worry: the fix may create new crossings elsewhere, and fixing those may bring back the old ones. What stops the repairs going on forever? Without an answer, the repair procedure isn't a proof — it's a hope. You've found the right local move and have no reason to believe it terminates.

The model chain. Label the points $P_1, \dots, P_n$. Each ordering $(\sigma_1, \dots, \sigma_n)$ of the indices gives a closed polygon

$$P_{\sigma_1} P_{\sigma_2} \cdots P_{\sigma_n} P_{\sigma_1},$$

possibly self-intersecting. There are finitely many orderings, so there is one whose total length L (the sum of its n edge lengths) is as small as possible. Call this polygon Π. I claim Π is simple.

First, which intersections are even possible? Two edges sharing an endpoint meet only at that endpoint: if they overlapped further they would lie on one line, putting three of the points on a line. A vertex cannot lie on the interior of a non-incident edge, for the same reason. So the only way Π can fail to be simple is that two non-adjacent edges cross at a point X interior to both. Relabel so that these edges are $P_{\sigma_1}P_{\sigma_2}$ and $P_{\sigma_k}P_{\sigma_{k+1}}$ with $3 \le k \le n-1$ (index $n+1$ meaning 1). Write $A = P_{\sigma_1}$, $B = P_{\sigma_2}$, $C = P_{\sigma_k}$, $D = P_{\sigma_{k+1}}$.

Remove edges AB and CD and insert AC and BD, reversing the stretch in between:

$$\Pi' : \; A = P_{\sigma_1},\, P_{\sigma_k},\, P_{\sigma_{k-1}},\, \dots,\, P_{\sigma_2},\, P_{\sigma_{k+1}},\, \dots,\, P_{\sigma_n},\, P_{\sigma_1}.$$

This is again a closed polygon through all n points, with the same edges as Π except that AB, CD are replaced by AC, BD. Since X lies on segment AB and on segment CD,

$$|AB| + |CD| = |AX| + |XB| + |CX| + |XD| = \big(|AX| + |XC|\big) + \big(|BX| + |XD|\big).$$

By the triangle inequality $|AX| + |XC| \ge |AC|$, with equality only if X lies on segment AC. But X is on line AB and $X \ne A$, so X on line AC would force A, B, C onto one line — excluded. So $|AX| + |XC| > |AC|$, and likewise $|BX| + |XD| > |BD|$. Hence $L(\Pi') < L(\Pi)$, contradicting minimality. So Π has no crossings: it is a simple closed polygon with the given points as vertices. (For $n = 3$ the triangle is simple anyway.)

A second route: choose two points P_1, P_2 such that all other points lie strictly on one side of line P_1P_2 (an edge of the convex hull). Each other point P_i makes an angle $\theta_i \in (0, \pi)$ with ray P_1P_2 at P_1, and these angles are distinct because no three points are collinear. Visiting P_1, P_2 and then the rest in increasing order of θ_i and returning to P_1 gives a simple polygon: it is a "fan" around P_1.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains find the same local move: uncross two crossing edges. They split at what is done with the move. The natural attempt applies it as a procedure and must then prove the procedure halts. The model chain never runs the procedure; it asks which polygon the move could never improve, and starts there.

> *Uncrossing always makes the polygon strictly shorter, so instead of repairing crossings one by one, I'll take the shortest polygon from the start — it can't have a crossing, because uncrossing it would make it shorter still.*

> *There are only finitely many orderings of the points, so "the shortest one" exists; I don't need any limit argument.*

Layer 3 THE WHY

The repair procedure worries you because it has no measure of progress. The fix to that is to find a quantity that the local move strictly decreases. Once you have such a quantity, the procedure's termination and the extremal argument are the same fact: a strictly decreasing quantity on a finite set of states cannot decrease for ever, and the state where it is smallest is one where the move is unavailable. Writing the proof extremally ("take the minimum") is the same argument with the bookkeeping removed.

Why length? Because the geometry of a crossing is the geometry of the triangle inequality. Two crossing segments AB and CD are the diagonals of the quadrilateral $ACBD$, and the diagonals of a convex quadrilateral are together longer than either pair of opposite sides. At the crossing point X each diagonal splits into two pieces, and those four pieces can be regrouped into two broken paths $A \to X \to C$ and $B \to X \to D$ — each of which a straight segment beats. Length is the quantity the triangle inequality speaks about, so it is the quantity uncrossing reduces. Nothing else needs inventing.

The "no three collinear" hypothesis is used twice, and it's worth seeing exactly where. It rules out degenerate contacts (touching, overlapping edges), so "not simple" means "a proper crossing". And it makes the triangle inequality strict. Without it the minimum-length tour can be non-simple: four collinear points visited out and back retrace the same segment.

Choosing which new pair of edges to insert also has a reason. Replacing AB, CD by AD, BC would split the tour into two separate cycles; replacing by AC, BD with the middle stretch reversed keeps one cycle. On a sketch the reversal is visible: you walk from A to C, then run backwards along the old path to B, then jump to D and continue.

The alternative solution is the constructive version of the same idea: a fan from a hull vertex has no crossings because angle order around P_1 prevents two edges from overlapping in angle. It is shorter to write but needs the hull edge chosen with care; the extremal proof needs no geometry beyond the triangle inequality.

Layer $+\alpha$ BACKTESTING

1. First useful cue: "there exists" a polygon with a property (no crossings), with a finite set of candidates. Write down: "among all closed polygons through the points, take the extremal one — what is the right measure?"

2. The method is the extremal principle: pick a quantity the obvious repair move strictly decreases, take a minimiser, and show a defect would allow a further decrease.

3. Hesitation occurs at whether uncrossing might create new crossings. Reinforce it away by noticing the extremal argument doesn't care: it never repeats the move, it only needs one move to shorten the minimum, which is impossible.

4. The trap is treating the repair as an algorithm and chasing termination by hand, or trusting a greedy rule like nearest-neighbour, which fails on small examples.

5. Recognise faster: if you can describe a local fix for a defect, ask immediately which numerical quantity that fix improves. If one exists, the optimum has no defect.

6. Decision rule: to show some configuration without a defect exists among finitely many, find a local move that removes a defect while strictly decreasing a numerical measure, then take a configuration minimising that measure; for crossing segments the measure is total length, via the strict triangle inequality.

Caveat. The book's official solution (and its alternate) is used in substance; OCR garbled its subscripts, so the reversal of the tour between the crossing edges has been written out afresh here and checked directly.

1966 B6

1966 B6. Show that all solutions of the differential equation $y'' + e^x y = 0$ remain bounded as $x \to \infty$.

Fields: Differential equations (second-order linear, oscillation), energy methods **Position:** B6 (hardest slot) **Answer:** The quantity $E(x) = y^2 + e^{-x}(y')^2$ satisfies $E'(x) = -e^{-x}(y')^2 \le 0$, so $y(x)^2 \le E(0)$ for all $x \ge 0$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You recognise $y'' + ky = 0$: for a constant $k > 0$ the solutions are $A\cos\sqrt{k}\,x + B\sin\sqrt{k}\,x$, bounded. Here the "spring constant" e^x grows, so you try to solve the equation. No elementary substitution works directly, but a change of variable can: put $t = 2e^{x/2}$, so $dt/dx = t/2$. Then

$$\frac{dy}{dx} = \frac{t}{2}\frac{dy}{dt}, \qquad \frac{d^2y}{dx^2} = \frac{t^2}{4}\frac{d^2y}{dt^2} + \frac{t}{4}\frac{dy}{dt},$$

and the equation becomes $t^2 y_{tt} + t\,y_t + t^2 y = 0$ — Bessel's equation of order 0. So $y = c_1 J_0(2e^{x/2}) + c_2 Y_0(2e^{x/2})$, and since $J_0(t), Y_0(t) \to 0$ like $t^{-1/2}$ as $t \to \infty$, every solution is bounded (indeed $y \to 0$ like $e^{-x/4}$). This is correct, and in 1966 some contestants did it. But it rests entirely on quoting the large-t behaviour of Bessel functions, which is itself a theorem needing proof, and a student who hasn't met Bessel functions hits a wall at the transformed equation. The other natural move — "the frequency increases, so the oscillation speeds up but the amplitude shouldn't grow" — is a heuristic, not a proof: a speeding-up oscillation could in principle gain amplitude, and nothing you've written says it doesn't.

The model chain. Let y be any solution; it is defined and twice differentiable on all of $\mathbb{R}$ because the equation is linear with continuous coefficients. Define, for $x \ge 0$,

$$E(x) = y(x)^2 + e^{-x}\,y'(x)^2.$$

Differentiate, using $y'' = -e^x y$:

$$E'(x) = 2yy' + e^{-x}\cdot 2y'y'' - e^{-x}(y')^2 = 2yy' + 2e^{-x}y'\,(-e^x y) - e^{-x}(y')^2 = -e^{-x}\,(y')^2 \le 0.$$

So E is non-increasing on $[0,\infty)$, and for every $x \ge 0$

$$y(x)^2 \le y(x)^2 + e^{-x}y'(x)^2 = E(x) \le E(0) = y(0)^2 + y'(0)^2.$$

Hence $|y(x)| \le \sqrt{y(0)^2 + y'(0)^2}$ for all $x \ge 0$: every solution is bounded as $x \to \infty$.

Equivalently, in the integrated form the official solution uses: multiply the equation by $e^{-x}y'$ and integrate from 0 to T. The term $\int_0^T yy'\,dx = \frac{1}{2}\left[y^2\right]_0^T$, and integrating by parts

$$\int_0^T e^{-x}y'y''\,dx = \tfrac{1}{2}\left[e^{-x}(y')^2\right]_0^T + \tfrac{1}{2}\int_0^T e^{-x}(y')^2\,dx,$$

so $y(T)^2 + e^{-T}y'(T)^2 + \int_0^T e^{-x}(y')^2\,dx = y(0)^2 + y'(0)^2$, which gives the same bound.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start from the constant-coefficient picture $y'' + ky = 0$. The natural attempt tries to *solve* the variable equation and read off the size of y. The model chain gives up on a formula and instead asks what quantity the constant-coefficient oscillator conserves, and whether a version of it stays controlled here.

> *I don't need the solution, I need a bound on it; for $y'' + ky = 0$ the conserved quantity is $ky^2 + (y')^2$, so I'll divide by the spring constant to isolate y^2, try $y^2 + (y')^2/e^x$, and see whether differentiating it gives something with a sign.*

> *The only term that doesn't cancel comes from differentiating e^{-x}, and e^{-x} is decreasing — so that term is negative and the quantity can only go down.*

Layer 3 THE WHY

"Bounded" is a statement about size, not about form. A formula would tell you the size, but asking for a formula is asking for much more than the question needs. The cheap tool for size in a second-order equation is energy, and the reason is physical before it is mathematical: $y'' = -ky$ is a mass on a spring, and its total energy $\frac{1}{2}(y')^2 + \frac{1}{2}ky^2$ is constant (differentiate it and use the equation — the two terms cancel exactly). Energy controls position: $\frac{1}{2}ky^2 \le$ energy.

When k varies with x, write $k = q(x)$. The direct energy $qy^2 + (y')^2$ has derivative $q'y^2$, which is *positive* here ($q' = e^x > 0$): energy is pumped in. That looks bad, but it's the wrong quantity to watch. You don't care about the energy; you care about y^2, which is energy divided by q. So divide first:

$$E = y^2 + \frac{(y')^2}{q}, \qquad E' = 2yy' + \frac{2y'y''}{q} - \frac{q'}{q^2}(y')^2 = -\frac{q'}{q^2}(y')^2,$$

the first two terms cancelling by $y'' = -qy$. Now the growth of q helps rather than hurts: if $q > 0$ and q is increasing, E decreases, and $y^2 \le E \le E(0)$. The problem is the case $q = e^x$. The stiffer the spring becomes, the smaller the swing for a given energy — the energy rises, but more slowly than q, and the amplitude falls. That's why the Bessel answer decays like $e^{-x/4}$: amplitude squared times the square root of the frequency is roughly constant.

So the "trick" of multiplying by $e^{-x}y'$ in the official solution is not a trick. Multiplying by y' is how every energy identity is born (it turns y'' into $\frac{1}{2}\frac{d}{dx}(y')^2$ and y into $\frac{1}{2}\frac{d}{dx}y^2$); the factor $e^{-x} = 1/q$ is the normalisation that makes the y^2 term appear with coefficient 1. The book's solution states the multiplication without motivation; this is where it comes from.

Note what is not claimed: y' need not be bounded (the bound is only on $e^{-x}(y')^2$, and indeed y' grows like $e^{x/4}$). The problem asks only about y.

Layer $+\alpha$ BACKTESTING

1. First useful cue: "bounded", attached to a second-order linear equation with no y' term. Write down the constant-coefficient energy $ky^2 + (y')^2$ and ask how to adapt it.

2. The method is a Lyapunov / energy function: build a quantity that dominates y^2 and show its derivative has a sign.

3. Hesitation occurs when the direct energy $e^x y^2 + (y')^2$ turns out to increase. Reinforce it away by remembering that you want y^2, not energy: divide by the coefficient before differentiating.

4. The trap is trying to solve the equation. It can be solved (Bessel functions after $t = 2e^{x/2}$), but the bound then depends on asymptotics you'd have to quote, and without that background the attempt stalls.

5. Recognise faster: any question about boundedness, decay or growth of solutions to $y'' + q(x)y = 0$ is an energy question. Try $y^2 + (y')^2/q$ first.

6. Decision rule: to bound solutions of $y'' + q(x)y = 0$ with $q > 0$, differentiate $y^2 + (y')^2/q$; its derivative is $-(q'/q^2)(y')^2$, so solutions are bounded whenever q is positive and non-decreasing. More broadly: when a problem asks for size and not form, look for a monotone quantity instead of a formula.

Caveat. The book's solution to 1966 B-6 is damaged in the OCR: the displayed identity after "multiplying the equation by $e^{-x}y'$ and integrating from 0 to T" and the text after "for some $0 \le s \le T$" are lost. The method is identifiable from the surviving sentence. The proof given here is reconstructed independently and every step has been checked by direct differentiation; it does not depend on the missing text. The Bessel-function reduction in the natural attempt was checked by the chain rule; the claim that J_0, Y_0 decay like $t^{-1/2}$ is standard asymptotics, quoted rather than proved.