DAWN PUTNAM SESSION

Putnam 1970 — Papers A and B

Putnam 120 Archive

Contents

Putnam 1970 — Paper A **5**
1970 A1 5
1970 A2 7
1970 A3 9
1970 A4 11
1970 A5 14
1970 A6 16

Putnam 1970 — Paper B **21**
1970 B1 21
1970 B2 23
1970 B3 26
1970 B4 28
1970 B5 31
1970 B6 33

Putnam 1970 — Paper A

1970 A1

1970 A1. Show that the power series for the function $e^{ax}\cos bx$ $(a > 0, b > 0)$ in powers of x has either no zero coefficients or infinitely many zero coefficients.

Caveat. The MAA 1965–1984 OCR renders the statement as "e'"" cos b:t (a> 0, b> 0)" with the exponent and cosine argument mangled by font substitution. The reconstruction above is standard and is confirmed by the book's own solution, which explicitly writes $e^{ax}\cos bx$ as $\mathrm{Re}\{(a+ib)^n\}$ inside the power series and reasons about the coefficient of x^n – there is no ambiguity about which function is meant. No mathematical content was altered.

Fields: Complex analysis, power series **Position:** A1 (opener) **Answer:** Either no coefficient vanishes, or infinitely many do; concretely, if the coefficient of x^{n_0} vanishes, so does the coefficient of x^{kn_0} for every odd integer k.

Layer 1 THE WHAT — surface analysis

The natural attempt. Write $e^{ax} = \sum_{m\ge 0} a^m x^m/m!$ and $\cos bx = \sum_{j\ge 0}(-1)^j b^{2j}x^{2j}/(2j)!$, and take the Cauchy product to get the coefficient of x^n directly:

$$c_n = \sum_{k=0}^{\lfloor n/2\rfloor} \frac{(-1)^k a^{n-2k} b^{2k}}{(n-2k)!\,(2k)!}.$$

This is a completely correct formula, and for small n you can chase it by hand: $c_0 = 1$, $c_1 = a$, $c_2 = (a^2 - b^2)/2$, and so on. But the moment you try to say something about *all* n at once – when does this alternating binomial-type sum vanish, and why would one such vanishing force infinitely many more – the sum offers no handle. It is a real trigonometric-polynomial identity in disguise, and staring at it as a sum of two-variable monomials does not reveal any structural reason for a zero at n to propagate to $3n, 5n, 7n, \ldots$. You can verify the claim numerically for the first several n, but a proof that covers every n at once refuses to fall out of this form.

The model chain. Let $z = a + ib$. Since $a > 0$ and $b > 0$, $z \neq 0$. The function e^{zx} is entire, and for real x,

$$e^{zx} = e^{ax}(\cos bx + i\sin bx), \qquad \text{so} \qquad e^{ax}\cos bx = \mathrm{Re}(e^{zx}).$$

Expanding the single exponential, $e^{zx} = \sum_{n\ge 0} \dfrac{z^n}{n!}x^n$, and since Re is $\mathbb{R}$-linear and x is real, the coefficient of x^n in $e^{ax}\cos bx$ is exactly

$$c_n = \frac{\mathrm{Re}(z^n)}{n!}.$$

So $c_n = 0$ if and only if z^n is purely imaginary. Now suppose $c_{n_0} = 0$ for some n_0, i.e. $z^{n_0} = iy$ for some real $y \neq 0$ (nonzero because $z \neq 0$). For any odd integer k,

$$z^{kn_0} = (z^{n_0})^k = (iy)^k = i^k y^k.$$

Since k is odd, $i^k = i \cdot (i^2)^{(k-1)/2} = i \cdot (-1)^{(k-1)/2}$, which is purely imaginary. Hence z^{kn_0} is purely imaginary too, so $c_{kn_0} = 0$. There are infinitely many odd k, so infinitely many indices $n_0, 3n_0, 5n_0, 7n_0, \ldots$ give a zero coefficient. Thus if any coefficient vanishes, infinitely many do; the only remaining possibility is that none vanish. This exhausts the dichotomy.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both readings begin from the same target: a formula for the coefficient c_n of x^n in $e^{ax}\cos bx$. They split before the formula is even finished being written down, at the choice of what object to expand. The natural attempt expands the product of two *real* series (an exponential and a cosine) and multiplies them out termwise. The model expands *one complex* series, e^{zx} with $z = a + ib$, and only afterwards takes the real part.

> *$e^{ax}\cos bx$ is not two independent functions glued together – it is the shadow, the real part, of a single complex exponential $e^{(a+ib)x}$. I should build that one clean object and read off its real part, instead of multiplying two real series and hoping a pattern appears in the product.*

▷ **Cue:** an exponential growth rate a and an oscillation rate b appear attached to the same real variable, inside a product like $e^{ax}\cos bx$ or $e^{ax}\sin bx$ $\longrightarrow$ **Move:** write $z = a + ib$ and recognise the expression as $\mathrm{Re}(e^{zx})$ or $\mathrm{Im}(e^{zx})$, then expand the one complex exponential instead of the two real factors separately

Layer 3 THE WHY

The recognition has to come from Euler's formula itself, not from pattern-matching against "problems that look like this." $\cos\theta = \mathrm{Re}(e^{i\theta})$ and $\sin\theta = \mathrm{Im}(e^{i\theta})$ are definitions, not tricks, and $e^{ax}\cos bx$ is built from exactly the two pieces – a real growth rate and a real angular rate, applied to the same variable x – that are the real and imaginary parts of one fixed complex number $z = a + ib$ acting by multiplication in the exponent. Once z is named, Re and "coefficient of x^n" commute for a free reason: both are $\mathbb{R}$-linear operations on the vector space of real sequences (or on the coefficients of a convergent power series with real argument), so $\mathrm{Re}\left(\sum z^n x^n/n!\right) = \sum \mathrm{Re}(z^n)x^n/n!$ termwise, with no approximation or limiting argument needed. This is what converts an analytic question (does a Taylor coefficient vanish?) into a purely algebraic one (does the argument of z^n land on $\pm\pi/2$?) – because multiplication of complex numbers adds arguments, $\arg(z^n) = n\arg(z) \pmod{2\pi}$, and "purely imaginary" is exactly the condition $\arg(z^n) \equiv \pi/2 \pmod{\pi}$. The archive's official solution does not spell out the argument-doubling geometry at all; it jumps straight to i^k for odd k, which is the same fact stated purely algebraically: multiplying a purely imaginary number by i^k (itself purely imaginary for odd k) keeps the product purely imaginary, because $i \cdot i = -1$ is real, so odd powers of i stay off the real axis while even powers land back on it. The reader does not need to know whether $\arg z$ is a rational multiple of π – that would only be needed to say exactly *which* n vanish, a question the problem never asks. It only asks for the dichotomy, and the dichotomy needs nothing more than: one zero, multiplied by any odd integer, is still a zero.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the specific pairing of e^{ax} with $\cos bx$ sharing the same real variable x, with a and b playing the role of a growth rate and an angular rate respectively. On seeing it, write down: this is Euler's formula wearing a disguise, so name $z = a + ib$.

2. That cue suggests expanding the single complex exponential $e^{zx} = \sum z^n x^n/n!$ and taking Re termwise, turning the coefficient condition $c_n = 0$ into the algebraic condition that z^n is purely imaginary.

3. The hesitation comes right after isolating $c_n = \mathrm{Re}(z^n)/n!$: it is tempting to think you now need to classify exactly which n satisfy $\mathrm{Re}(z^n) = 0$, which looks like it depends on whether $\arg(z)/\pi$ is rational. It is reinforced away by noticing the problem only asks for a dichotomy,

not a classification – so it suffices to show one zero propagates to infinitely many, which the odd-power trick does without ever discussing $\arg(z)$.

4. The trap is committing to case-splitting on the rationality of $\arg(z)/\pi$, a genuinely harder question that the problem does not require an answer to. Chasing it wastes the exam on a side question.
5. To recognise this faster: build the reflex that any product or quotient combining a real exponential rate and a real angular rate on the same variable is a complex exponential in a red coat, and that questions about a coefficient vanishing translate into questions about the argument of a single fixed complex number under integer powers.
6. Decision rule: when a real coefficient sequence is secretly Re or Im of the powers of one fixed nonzero complex number, do not try to characterise the full vanishing set directly – exploit that arguments add under multiplication, so any single vanishing index propagates to every one of its odd multiples, which is already enough to prove "finite or cofinite"-type dichotomies without any classification.

1970 A2

1970 A2. Consider the locus given by the real polynomial equation $Ax^2 + Bxy + Cy^2 + Dx^3 + Ex^2y + Fxy^2 + Gy^3 = 0$, where $B^2 - 4AC < 0$. Prove that there is a positive number δ such that there are no points of the locus in the punctured disk $0 < x^2 + y^2 < \delta^2$.

Caveat. The MAA 1965–1984 OCR is largely clean here; the only artefact is the punctured-disk bound rendered as "0< x1 + y1 < aŢ" with a standing in for δ and superscripts dropped. The reconstruction above restores the intended $0 < x^2 + y^2 < \delta^2$ and matches the book's own solution, which explicitly seeks "a positive number δ" and bounds r away from 0. No mathematical content was altered.

Fields: Algebraic curves, quadratic forms, compactness **Position:** A2 **Answer:** Yes; any δ with $0 < \delta \le m/M$ works, where $m = \min_\theta |A\cos^2\theta + B\sin\theta\cos\theta + C\sin^2\theta| > 0$ and $M = |D| + |E| + |F| + |G|$ (interpreting δ as arbitrary when $M = 0$).

Layer 1 THE WHAT — surface analysis

The natural attempt. Probe the curve one direction at a time: substitute $y = mx$ for a line through the origin of slope m, and see what the equation forces on x. This gives

$$x^2(A + Bm + Cm^2) + x^3(D + Em + Fm^2 + Gm^3) = 0,$$

i.e. $x^2\left[(A + Bm + Cm^2) + x(D + Em + Fm^2 + Gm^3)\right] = 0$. So on this line, either $x = 0$ (the origin itself) or

$$x = -\frac{A + Bm + Cm^2}{D + Em + Fm^2 + Gm^3}$$

whenever the denominator is nonzero – a single further point, not a whole neighbourhood. This is exactly the right idea (per direction, the curve meets the line in at most one nonzero point near the origin), and it is provable: you would go on to argue that $|x(m)|$ is bounded away from 0 as m ranges over $\mathbb{R}$, plus handle the vertical line $x = 0$ and the directions where the cubic denominator itself vanishes, separately. It works, but it costs you: you are re-deriving, slope by slope, a bound that has to hold uniformly over infinitely many lines, you must patch in the vertical direction by hand, and the denominator's zero set becomes its own case to rule out – three separate loose ends that all have to be tied off before the argument is complete.

The model chain. Write $Q(x,y) = Ax^2 + Bxy + Cy^2$ (the quadratic part) and $C_3(x,y) = Dx^3 + Ex^2y + Fxy^2 + Gy^3$ (the cubic part), so the locus is $Q + C_3 = 0$. Pass to polar coordinates $x = r\cos\theta$, $y = r\sin\theta$ with $r > 0$ (excluding the origin, which the punctured disk already excludes). By homogeneity,

$$Q(x,y) = r^2 q(\theta), \qquad C_3(x,y) = r^3 c(\theta),$$

where $q(\theta) = A\cos^2\theta + B\sin\theta\cos\theta + C\sin^2\theta$ and $c(\theta) = D\cos^3\theta + E\cos^2\theta\sin\theta + F\cos\theta\sin^2\theta + G\sin^3\theta$. On the locus, $r^2 q(\theta) + r^3 c(\theta) = 0$; dividing by $r^2 \neq 0$ gives $q(\theta) = -r\,c(\theta)$, so $|q(\theta)| \leq r\,|c(\theta)|$.

Since $B^2 - 4AC < 0$, the quadratic form Q is definite (of one fixed sign) and vanishes only at the origin; hence $q(\theta) \neq 0$ for every θ. As q is continuous on the compact set $[0, 2\pi]$, it attains a positive minimum $m = \min_\theta |q(\theta)| > 0$ by the extreme value theorem. Likewise c is continuous on a compact set, so $M = \max_\theta |c(\theta)| \leq |D| + |E| + |F| + |G|$ is finite. Combining, $m \leq |q(\theta)| \leq rM$, so any point of the locus with $r > 0$ satisfies $r \geq m/M$ (or, if $M = 0$, no such point can exist at all, since then $q(\theta) = 0$ is forced, contradicting $q(\theta) \neq 0$). Taking $\delta = m/M$ (any positive number if $M = 0$), no locus point has $0 < r < \delta$, i.e. $0 < x^2 + y^2 < \delta^2$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both readings agree that the equation splits into a quadratic part and a cubic part, $Q + C_3 = 0$, and that this difference in homogeneity degree is what must be exploited. They split on how to make that exploitation uniform. The natural attempt fixes a direction m, solves the resulting equation for the one nonzero intersection point on that line, and must then separately argue this holds "for every m, plus the vertical line, plus the degenerate lines." The model instead parametrises direction by the angle θ on the unit circle and divides the whole equation by r^2 once, extracting a single global bound over the entire compact parameter space at once.

> *I don't need to find the exact intersection point on each line and then show all of those bounds hold uniformly – I need one bound that already holds for every direction simultaneously. The unit circle is compact, so let θ do the job m was doing, and take the min and max of q and c once, globally, instead of solving line by line.*

▷ **Cue:** a curve's equation splits into homogeneous parts of two different degrees, and you need to show the origin is an isolated point of the locus $\longrightarrow$ **Move:** pass to polar coordinates, divide by the lowest power of r present, and bound the resulting angular functions using compactness of $\theta \in [0, 2\pi]$ rather than parametrising by slope and solving line by line

Layer 3 THE WHY

The hypothesis $B^2 - 4AC < 0$ is not merely "the quadratic doesn't factor over $\mathbb{R}$" – that weaker reading is where a student can stop and miss the point. Completing the square shows $Q(x,y) = A\left(x + \frac{B}{2A}y\right)^2 + \left(C - \frac{B^2}{4A}\right)y^2$ (for $A \neq 0$; the $B^2 < 4AC$ hypothesis forces A, C to be nonzero and of the same sign), a sum of one square with coefficient A and another with coefficient $(4AC - B^2)/4A$, both strictly positive or both strictly negative. That is exactly what "definite" means: $Q(x,y) = 0$ only at $(x,y) = (0,0)$. Definiteness is an algebraic fact about the discriminant, but the problem needs an *analytic* fact – that Q is bounded away from zero on every ray – and the bridge between the two is compactness: $q(\theta) = Q(\cos\theta, \sin\theta)$ is a continuous, nowhere-zero function on the compact circle $[0, 2\pi]$, so by the extreme value theorem it attains a strictly positive minimum. This is the only place in the whole argument where $B^2 - 4AC < 0$ is used, and it is used for exactly one purpose: to license the existence of $m > 0$. The cubic part needs no such hypothesis and gets none – it only needs to be *bounded above*, which any continuous function on a compact set is, for free. The asymmetry in the argument (a lower bound for the quadratic part, only an upper bound for the cubic part) mirrors the asymmetry in

the hypothesis: the problem gives you a sign-definiteness condition on the quadratic coefficients and nothing at all on D, E, F, G, because degree-3 terms are already suppressed by an extra factor of $r \to 0$ relative to the degree-2 terms, so they cannot threaten the bound no matter how large $|D| + |E| + |F| + |G|$ is – they only shrink the window $[0, \delta)$ in which the bound holds, they never destroy it.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the equation splitting into homogeneous pieces of different total degree, Q (degree 2) and C_3 (degree 3). On seeing it, write down: near the origin the lowest-degree part dominates, so switch to polar or scaling coordinates and factor out its power of r.

2. That cue suggests dividing $r^2 q(\theta) + r^3 c(\theta) = 0$ by r^2 (valid since $r > 0$ is guaranteed by the punctured disk) to isolate $q(\theta) = -r\, c(\theta)$, a relation between two bounded angular functions.

3. The hesitation is whether $q(\theta)$ can itself vanish for some bad direction, which would break the bound. It is reinforced away by recognising that $B^2 - 4AC < 0$ is precisely the algebraic statement that Q is definite, hence $q(\theta) \neq 0$ for every θ – combined with continuity and compactness, that already forces a strictly positive minimum, with no case analysis on θ required.

4. The trap is treating $B^2 - 4AC < 0$ as only "the quadratic doesn't factor," a true but weaker fact, and never pushing it to the stronger consequence – definiteness, hence a uniform positive lower bound – that the proof actually needs.

5. To recognise this faster: catalogue $B^2 - 4AC < 0$ as "definite quadratic form, hence bounded away from zero on the unit circle by compactness" as an immediate reflex, the same reflex used to recognise an ellipse from its defining quadratic form.

6. Decision rule: when a curve near the origin is a sum of homogeneous pieces of different degrees and you must show the origin is isolated on the locus, pass to polar coordinates, isolate the lowest-degree piece, and use compactness of the angular variable to get a uniform lower bound on that piece (when it is sign-definite) against an upper bound on the higher-degree pieces – never parametrise by slope and solve line by line when the angle variable already ranges over a compact set.

1970 A3

1970 A3. Find the length of the longest sequence of equal nonzero digits in which an integral square can terminate (in base 10) and find the smallest square which terminates in such a sequence.

Caveat. The MAA 1965–1984 OCR is clean here apart from the usual apostrophe and spacing artefacts ("x2=0 (mod 10)" for $x^2 \equiv 0 \pmod{10}$). The reconstruction above matches the book's own solution verbatim in content, which works entirely in modular arithmetic and ends with the explicit witness $38^2 = 1444$. No mathematical content was altered.

Fields: Number theory, quadratic residues **Position:** A3 **Answer:** The longest such run has length 3; the smallest square terminating in three equal nonzero digits is $38^2 = 1444$ (ending in 444).

Layer 1 THE WHAT — surface analysis

The natural attempt. List squares and watch the tail digits: $1, 4, 9, 16, 25, 36, 49, 64, 81, 100, 121, 144, \ldots, 1444, \ldots$. You notice $12^2 = 144$ ends in two 4's, and if you keep going you eventually find $38^2 = 1444$, ending in three 4's. This is genuinely useful – it produces the example the problem asks for – but it is search, not proof. Nothing in a finite table of computed squares tells you that no square

anywhere, of any size, ends in four equal nonzero digits, or in three equal copies of some digit other than 4. To close the problem this way you would need to search forever, which you cannot do in an exam, so the honest brute-force path finds the answer but cannot certify it is the answer.

The model chain. Work modulo powers of 10, in increasing steps. Squares mod 10 can only be $0, 1, 4, 5, 6, 9$ (checking $x = 0, \ldots, 9$), so a nonzero repeated terminal digit d must be one of $1, 4, 5, 6, 9$; call the candidate two-digit endings $11, 44, 55, 66, 99$.

Refine modulo 4: since $x \bmod 4 \in \{0, 1, 2, 3\}$ gives $x^2 \bmod 4 \in \{0, 1, 0, 1\}$, every square is $\equiv 0$ or $1 \pmod 4$. Now $11 \equiv 3, 55 \equiv 3, 99 \equiv 3 \pmod 4$ – all impossible – and $66 \equiv 2 \pmod 4$, also impossible. Only $44 \equiv 0 \pmod 4$ survives, so among two-digit repdigit endings, only $\ldots 44$ is possible for a perfect square.

Push to three digits: the only candidate is $\ldots 444$, and $38^2 = 1444$ exhibits it, so a run of length 3 is achieved.

Push to four digits: the only candidate for a longer run is $\ldots 4444$. Check modulo 16: since $x \bmod 8 \in \{0, \ldots, 7\}$ gives $x^2 \bmod 16 \in \{0, 1, 4, 9, 0, 9, 4, 1\}$, every square is $\equiv 0, 1, 4$, or $9 \pmod{16}$. But $4444 = 16 \cdot 277 + 12 \equiv 12 \pmod{16}$, which is none of those. So no square ends in 4444, and since 4 was the only surviving digit at length 3, no square can end in four equal nonzero digits at all. The maximum run length is 3.

Finally, find the smallest square ending in 444. Squares ending in $44 \pmod{100}$ come from $x \equiv 12, 38, 62, 88 \pmod{100}$ (direct check of two-digit endings whose square is $\equiv 44$). In increasing order, $x = 12$ gives $12^2 = 144$, whose last three digits are 144 – only two repeated 4's, not three, since it is preceded by 1. The next candidate is $x = 38$: $38^2 = 1444$, whose last three digits are 444. So 1444 is the smallest square terminating in three equal nonzero digits.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both approaches treat "the last k digits of x^2 are all equal to d" as a congruence condition, $x^2 \equiv d \cdot \underbrace{11\cdots1}_{k} \pmod{10^k}$. They split on how that congruence is attacked. The natural attempt tests it directly, empirically, by computing x^2 for consecutive x and reading off digits – treating 10^k as a single opaque modulus to be searched against. The model factors $10^k = 2^k \cdot 5^k$ and works one prime-power factor, 2^k, at a time, because quadratic residues modulo a power of 2 are a short, fully computable list.

> *10^k is not a prime power, so "$x^2 \bmod 10^k$" is secretly two independent conditions tangled together, mod 2^k and mod 5^k. I should isolate the power-of-2 condition and test it first, since squares modulo a power of 2 are unusually sparse and I can just write the whole residue list down – that is almost certainly where a fourth repeated digit gets killed, not the power-of-5 side.*

▷ **Cue:** a problem asks which terminal digit-patterns (mod 10^k) a perfect square can have
⟶ **Move:** factor $10^k = 2^k \cdot 5^k$ and test the residues of x^2 modulo the smallest power of 2 first, since $(\mathbb{Z}/2^k\mathbb{Z})^\times$-type residue counting makes squares mod 2^k far more restrictive, and escalate k one digit at a time rather than searching mod 10^k directly

Layer 3 THE WHY

The "digit" framing of this problem is a disguise; the actual content is a fact about how few elements of $\mathbb{Z}/2^k\mathbb{Z}$ are squares, and that fact has nothing to do with base 10 at all. Modulo 4, the map $x \mapsto x^2$ sends the four residues $0, 1, 2, 3$ to only two values, $0, 1$ – already a quarter of all inputs collapse onto half the possible outputs, so squares are structurally rare mod 4. Modulo 16, the same phenomenon is sharper: only 4 of the 16 residues, $\{0, 1, 4, 9\}$, are squares. This sparsity is a consequence of 2 being the unique even prime – squaring a unit modulo 2^k is a 2-to-1 (or worse) map because x and $-x$ (and, for $k \geq 3$, also $2^{k-1} \pm x$) give the same square, so the image is forced to be small. Base 5 carries no comparable obstruction: 5 is odd, squaring is closer to a genuine bijection on units modulo 5^k up to sign, and no analogous "only a quarter

survive" collapse occurs. So when a run of k equal terminal digits is translated into a congruence mod $10^k = 2^k \cdot 5^k$, the obstruction to making k large is entirely carried by the 2^k factor; the 5^k factor is essentially permissive and never the bottleneck. This is why testing mod 4 first, then mod 16, finds the wall (length 4 is impossible) with two of the smallest possible moduli, while the digit language of the problem statement is, mathematically, a red herring pointing at the wrong prime.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "the last k digits of x^2 are all the same digit d," which is a congruence $x^2 \equiv d \cdot (10^k - 1)/9 \pmod{10^k}$. On seeing it, write down: $10^k = 2^k \cdot 5^k$, so this is really two independent congruences glued together.

2. That cue suggests testing the power-of-2 side first, since the list of quadratic residues mod 2^k is short and fully computable by hand for small k (mod 4: $\{0,1\}$; mod 16: $\{0,1,4,9\}$), rather than attacking mod 10^k or mod 5^k directly.

3. The hesitation is deciding which modulus, 2^k or 5^k, actually carries the obstruction, before committing time to either. It is reinforced away by testing the cheapest case first: mod 4 alone already eliminates four of the five candidate two-digit endings $(11, 55, 66, 99)$, revealing that the 2-part is doing essentially all the work, so the same tool is escalated to mod 16 for the next digit.

4. The trap is trying to settle the problem by listing squares and hoping to eventually "see" that no square goes past three repeated digits – this finds the example 1444 but supplies no argument that a longer run is impossible, since exhaustive search never actually exhausts an infinite set.

5. To recognise this faster: memorise the short quadratic-residue tables for small powers of 2 – mod 4: $\{0,1\}$; mod 8: $\{0,1,4\}$; mod 16: $\{0,1,4,9\}$ – as an immediately available sieve for any "terminal digits of a square" question.

6. Decision rule: for any problem about which terminal digit-patterns a perfect square in base 10 can have, factor 10^k into its prime powers and test the smallest, tightest one – residues modulo a power of 2 are far more restrictive than modulo a power of 5 – before resorting to a direct search in base 10.

1970 A4

1970 A4. Given a sequence $\{x_n\}$, $n = 1, 2, \cdots$, such that $\lim_{n\to\infty}\{x_n - x_{n-2}\} = 0$. Prove that

$$\lim_{n\to\infty} \frac{x_n - x_{n-1}}{n} = 0.$$

Caveat. The MAA OCR renders the conclusion as "limit $x_n - x_{n-1} = 0$ / $n \to \infty$ n", with the fraction bar lost across two lines; the statement is restored as $\lim(x_n - x_{n-1})/n = 0$. Identity check: the book's solution under 1970 A-4 telescopes $x_n - x_{n-1}$ into alternating differences $x_k - x_{k-2}$, bounds it by $(n-N)\varepsilon + |x_N - x_{N-1}|$ and divides by n, which proves exactly this statement.

Fields: real analysis (sequences, limits) **Position:** A4 — mid-paper, but one of the shorter A4s; the difficulty is entirely in seeing which quantity to control **Answer:** The limit is 0. With $y_n = x_n - x_{n-1}$ the hypothesis says $y_n + y_{n-1} \to 0$; then $(-1)^n y_n$ has differences tending to 0, so it grows sublinearly, and $y_n/n \to 0$. The factor $1/n$ cannot be removed: $x_n = \sum_{k\le n}(-1)^k\sqrt{k}$ satisfies the hypothesis with $|x_n - x_{n-1}| = \sqrt{n} \to \infty$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The hypothesis talks about steps of two; the conclusion talks about steps of one. The first instinct is to split into even and odd subsequences. From $x_n - x_{n-2} \to 0$ you learn that $x_2, x_4, x_6, \ldots$ has consecutive differences tending to 0, and so does $x_1, x_3, x_5, \ldots$. A student who knows Cesàro's theorem then writes $x_{2k}/k \to 0$ and $x_{2k+1}/k \to 0$, so $x_n/n \to 0$.

Now the conclusion: $\dfrac{x_n - x_{n-1}}{n} = \dfrac{x_n}{n} - \dfrac{x_{n-1}}{n} \to 0 - 0 = 0$.

This is correct, and it is a perfectly good proof, provided you actually prove the lemma "if $a_k - a_{k-1} \to 0$ then $a_k/k \to 0$". Many students assert it and move on, which loses the marks, because it is the whole content of the problem. The weaker student does something else: tries to show $x_n - x_{n-1}$ itself tends to something, writes $x_n - x_{n-1} = (x_n - x_{n-2}) - (x_{n-1} - x_{n-2})$, and discovers that this relates a one-step difference to another one-step difference with no progress. Iterating that identity by hand is the right idea, but done without a plan it becomes a wall of indices.

The model chain. Put $y_n = x_n - x_{n-1}$ for $n \geq 2$ and $\varepsilon_n = x_n - x_{n-2} = y_n + y_{n-1}$ for $n \geq 3$. The hypothesis is $\varepsilon_n \to 0$.

Define $w_n = (-1)^n y_n$. Then

$$w_n - w_{n-1} = (-1)^n y_n - (-1)^{n-1} y_{n-1} = (-1)^n (y_n + y_{n-1}) = (-1)^n \varepsilon_n,$$

so $|w_n - w_{n-1}| = |\varepsilon_n| \to 0$.

Now fix $\varepsilon > 0$. Choose $N \geq 3$ with $|\varepsilon_k| < \varepsilon$ for all $k > N$. For $n > N$, telescoping,

$$|w_n| \leq |w_N| + \sum_{k=N+1}^{n} |w_k - w_{k-1}| < |y_N| + (n - N)\varepsilon.$$

Since $|w_n| = |y_n|$, dividing by n gives

$$\frac{|x_n - x_{n-1}|}{n} < \frac{|y_N|}{n} + \varepsilon.$$

N is fixed, so $|y_N|/n < \varepsilon$ once $n > |y_N|/\varepsilon$. Hence $|x_n - x_{n-1}|/n < 2\varepsilon$ for all large n, and since ε was arbitrary, the limit is 0.

Sharpness: take $y_n = (-1)^n \sqrt{n}$ and $x_n = \sum_{k=2}^{n} y_k$. Then $\varepsilon_n = y_n + y_{n-1} = \pm(\sqrt{n} - \sqrt{n-1}) \to 0$, but $|y_n| = \sqrt{n}$ is unbounded. So $x_n - x_{n-1}$ need not tend to 0, or even stay bounded; dividing by n is necessary.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains introduce the one-step difference $y_n = x_n - x_{n-1}$ and see $x_n - x_{n-2} = y_n + y_{n-1}$. The natural attempt then leaves y and returns to x, splitting into parities. The model stays with y and asks what it takes to turn a *sum* of consecutive terms into a *difference* of consecutive terms: an alternating sign.

> *The hypothesis is that $y_n + y_{n-1}$ is small, and every tool I own is about $y_n - y_{n-1}$ being small. A sum becomes a difference if I flip the sign of every other term, so look at $(-1)^n y_n$.*

> *A sequence whose steps eventually have size less than ε can grow by at most ε per step, so after n steps it is at most about εn. That is exactly why the problem divides by n.*

Layer 3 THE WHY

Why is the alternating sign the natural move and not a trick? Because of what the two sides of the problem are measuring. The conclusion is about y_n/n, and there is exactly one standard

reason a sequence divided by n goes to 0: its increments go to 0. That fact is the discrete version of "if $f'(t) \to 0$ then $f(t)/t \to 0$", and it comes straight from the triangle inequality: a sequence is its starting value plus the sum of its increments, and n increments each of size below ε add up to less than $n\varepsilon$. So the target is set before any cleverness: find a sequence with the same absolute value as y_n whose increments tend to 0.

The hypothesis does not hand you that. It hands you $y_n + y_{n-1} \to 0$, which says consecutive terms nearly *cancel*, not that they are nearly *equal*. Nearly cancelling means $y_n \approx -y_{n-1}$: the sequence is nearly flipping sign at every step. If a sequence nearly flips sign at every step, then multiplying by $(-1)^n$ un-flips it, and the result nearly stays put. That sentence is the whole of the model chain. The sign is not pulled from a hat; it is the correction for the behaviour the hypothesis describes.

The official solution in the MAA volume writes the telescoping out in full,

$$x_n - x_{n-1} = (x_n - x_{n-2}) - (x_{n-1} - x_{n-3}) + (x_{n-2} - x_{n-4}) - \cdots \pm (x_{N+1} - x_{N-1}) \mp (x_N - x_{N-1}),$$

with alternating signs and no comment. Here is where those alternating signs come from: they are the $(-1)^k$ in w_k. Writing w_n makes the telescoping sum one line, and it makes the structure visible: $w_n = w_N + \sum(w_k - w_{k-1})$ is the most ordinary telescope there is.

Now look again at the natural attempt. It is not wrong. It proves the stronger intermediate fact $x_n/n \to 0$, which is true here for the same reason: the even subsequence has increments $\varepsilon_{2k} \to 0$, so $x_{2k}/k \to 0$. It works because it applies the "increments to 0 implies sublinear" principle to a different sequence, x_{2k}, whose increments *are* the given small quantities. So both chains are the same principle applied to two different sequences. The model's choice is shorter because it lands on the quantity the conclusion names, y_n, and avoids the detour through x_n/n and the subtraction afterwards. The lesson is not "the parity split is bad"; it is "identify the principle first, then pick the sequence to apply it to that sits closest to the conclusion".

Finally, why is the $1/n$ there at all? The counterexample answers it. If $y_n = (-1)^n f(n)$ for any slowly growing f with $f(n) - f(n-1) \to 0$, the hypothesis holds and y_n is as large as f. The strongest statement available is therefore $y_n = o(n)$, and the problem asks for exactly that. Checking why a normalisation is present, by trying to break the statement without it, is the cheapest way to understand what a problem is really about.

Layer $+\alpha$ BACKTESTING

1. The cue is the mismatch of step sizes: $x_n - x_{n-2}$ in the hypothesis, $x_n - x_{n-1}$ in the conclusion. Write $y_n = x_n - x_{n-1}$ at once and translate the hypothesis into $y_n + y_{n-1} \to 0$.

2. A conclusion of the form "$(\cdot)/n \to 0$" suggests the telescoping bound: if increments tend to 0, the sequence is $o(n)$. The method is to find the sequence whose increments are the given small quantities.

3. Hesitation comes at $y_n + y_{n-1} \to 0$, because it is a sum and the familiar lemma needs a difference. Reinforce with: "a sum of neighbours becomes a difference of neighbours under $(-1)^n$".

4. The trap is trying to prove $x_n - x_{n-1} \to 0$, which is false (take $y_n = (-1)^n\sqrt{n}$). The secondary trap is citing Cesàro or Stolz without proof when that lemma is the entire problem.

5. Whenever a hypothesis says consecutive terms nearly cancel, multiply by $(-1)^n$. Whenever a conclusion divides by n, ask "which sequence has small increments?".

6. To prove $a_n/n \to 0$, exhibit a_n (up to sign or bounded error) as a telescoping sum of terms tending to 0, fix N from the ε-definition, bound by $|a_N| + (n-N)\varepsilon$, divide by n. Before starting, try to break the statement without the normalisation to see what it is protecting.

▷ **Cue:** hypothesis on $x_n - x_{n-2}$, conclusion on $x_n - x_{n-1}$ $\longrightarrow$ **Move:** set $y_n = x_n - x_{n-1}$; the hypothesis becomes $y_n + y_{n-1} \to 0$

▷ **Cue:** a sum of consecutive terms tends to 0 $\longrightarrow$ **Move:** pass to $(-1)^n y_n$, whose consecutive differences tend to 0

▷ **Cue:** prove $a_n/n \to 0$ ⟶ **Move:** telescope: $|a_n| \le |a_N| + (n-N)\varepsilon$, then divide by n
▷ **Cue:** a conclusion normalised by $1/n$ (or $1/\sqrt{n}$, etc.) ⟶ **Move:** try a counterexample without the normalisation to see what size the problem is protecting
▷ **Cue:** temptation to quote Cesàro or Stolz ⟶ **Move:** prove the one-line telescoping bound instead; it is shorter than the citation and cannot be marked down

1970 A5

1970 A5. Determine the radius of the largest circle which can lie on the ellipsoid

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \qquad (a > b > c).$$

Caveat. The MAA OCR garbles the displayed equation ("x• y• z• - +-+- = 1 (a > b >c). a1 b2 c1"); it is restored as the standard ellipsoid. Identity check: the book's solution under 1970 A-5 intersects circular sections with the yz- and xy-planes, obtains radius b, and exhibits the planes $a^2(b^2-c^2)z^2 = c^2(a^2-b^2)x^2$, which are circular sections of exactly this ellipsoid (verified symbolically below and numerically for $(a,b,c) = (3,2,1)$).

Fields: solid analytic geometry, quadratic forms **Position:** A5 — late in the paper, but a short argument once two geometric facts are in hand **Answer:** The radius is b, the middle semi-axis. The circles of radius b are the sections by the two planes $a\sqrt{b^2-c^2}\,z = \pm c\sqrt{a^2-b^2}\,x$ through the y-axis.

Layer 1 THE WHAT — surface analysis

The natural attempt. A circle in space is the intersection of a sphere with a plane. The natural move is to parametrise: centre $\mathbf{p}$, unit normal $\mathbf{n}$, radius r, and orthonormal $\mathbf{u}, \mathbf{v}$ spanning the plane; the circle is $\mathbf{p} + r(\cos t\,\mathbf{u} + \sin t\,\mathbf{v})$. Substitute into the ellipsoid equation, demand that it hold for all t, and expand in $1, \cos t, \sin t, \cos 2t, \sin 2t$. Five coefficient equations, about eight unknowns, and then you are asked to maximise r subject to them.

This is honest and can be pushed through, but it is a swamp. The coefficient of $\cos 2t$ and $\sin 2t$ says the quadratic form restricted to the plane is a multiple of the identity; the $\cos t$, $\sin t$ terms tie $\mathbf{p}$ to the plane; the constant term gives r. Most students who start this way never get to the maximisation, and those who do have no reason to believe the answer b is a maximum rather than a critical value.

The other natural move is to guess. The largest circle "should" be as big as possible, so try radius a: impossible, because the only points at distance a from the origin are $(\pm a, 0, 0)$. Try the plane $z = 0$: its section is an ellipse with semi-axes a and b, not a circle. At that point the guess stalls.

The model chain. *Step 1: every circle on the ellipsoid is a plane section.* A circle C lies in some plane Π, so $C \subseteq \Pi \cap \mathcal{E}$, where $\mathcal{E}$ is the ellipsoid. That intersection is a conic in Π (an ellipse, a point, or empty). A circle is also a conic, and five points of a circle, no three collinear, determine a unique conic through them; so the conic $\Pi \cap \mathcal{E}$ is the circle C.

Step 2: the central section is the largest in its direction. Write $Q(\mathbf{p}) = x^2/a^2 + y^2/b^2 + z^2/c^2$, a positive definite quadratic form with bilinear form B. Fix a unit normal $\mathbf{n}$ and let $\Pi_h = \{\mathbf{p} : \mathbf{n}\cdot\mathbf{p} = h\}$. Let $\mathbf{p}_h$ minimise Q on Π_h; by scaling, $\mathbf{p}_h = h\,\mathbf{p}_1$. Minimality gives $B(\mathbf{p}_h, \mathbf{u}) = 0$ for every $\mathbf{u}$ with $\mathbf{n}\cdot\mathbf{u} = 0$, so for such $\mathbf{u}$,

$$Q(\mathbf{p}_h + \mathbf{u}) = Q(\mathbf{p}_h) + Q(\mathbf{u}) = h^2 Q(\mathbf{p}_1) + Q(\mathbf{u}).$$

Hence $\Pi_h \cap \mathcal{E}$ is $\mathbf{p}_h + \{\mathbf{u} \perp \mathbf{n} :\ Q(\mathbf{u}) = 1 - h^2 Q(\mathbf{p}_1)\}$, a copy of the central section $\{\mathbf{u} \perp \mathbf{n} : Q(\mathbf{u}) = 1\}$ scaled by $\sqrt{1 - h^2 Q(\mathbf{p}_1)} \le 1$. So if some section in direction $\mathbf{n}$ is a circle of radius r, the central one is a circle of radius $\ge r$. It suffices to find the central circular sections.

Step 3: every central circular section has radius exactly b. Let Π pass through the origin O and meet $\mathcal{E}$ in a circle C. The section is symmetric under $\mathbf{p} \mapsto -\mathbf{p}$, so the centre of C is O (a circle's reflection through O is the circle with reflected centre, and it is the same circle, so the centre is fixed). Π is not the yz-plane, whose section $y^2/b^2 + z^2/c^2 = 1$ is not a circle as $b \neq c$. So Π meets the yz-plane in a line through O, which meets $\mathcal{E}$ at a point $\mathbf{q} = (0, y, z)$ on C. Since $c < b$,

$$|\mathbf{q}|^2 = y^2 + z^2 \leq b^2\Big(\frac{y^2}{b^2} + \frac{z^2}{c^2}\Big) = b^2,$$

so $r = |\mathbf{q}| \leq b$. In the same way Π is not the xy-plane (semi-axes $a \neq b$), so it meets that plane in a line through O, giving a point $(x, y, 0) \in C$ with $x^2 + y^2 \geq b^2(x^2/a^2 + y^2/b^2) = b^2$. So $r \geq b$. Hence $r = b$.

Step 4: such sections exist. Subtract the sphere $x^2 + y^2 + z^2 = b^2$, divided by b^2, from the ellipsoid equation:

$$Q(\mathbf{p}) - \frac{|\mathbf{p}|^2}{b^2} = x^2\Big(\frac{1}{a^2} - \frac{1}{b^2}\Big) + z^2\Big(\frac{1}{c^2} - \frac{1}{b^2}\Big).$$

The first bracket is negative and the second positive, so this vanishes exactly on the two planes $a^2(b^2 - c^2)z^2 = c^2(a^2 - b^2)x^2$. On either plane, $Q(\mathbf{p}) = 1 \iff |\mathbf{p}|^2 = b^2$, so the section is a great circle of the sphere of radius b. The maximum radius is b.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start from "a circle lies in a plane". The natural attempt then parametrises the circle and substitutes. The model instead asks where on the ellipsoid a circle could possibly reach, and uses the one fact every circle has: all its points are the same distance from its centre, and that distance is pinned between the distances of particular points on the ellipsoid.

> *Parallel slices of an ellipsoid are all the same shape, shrinking away from the middle. So I only need slices through the centre, and then the centre of the circle is the origin.*

> *A circle through the origin-centred ellipsoid has every point at distance r from O. Any plane through O must cut the yz-plane, where points are at most b from O, and the xy-plane, where they are at least b. So r is squeezed to b.*

Layer 3 THE WHY

A circle is the set of points at a fixed distance from a centre. That definition is the whole problem. Once the centre is known to be O, "radius" means "distance from O", and the ellipsoid comes with a ready-made map of distances from O: along the x-axis it reaches a, along the y-axis b, along the z-axis c, and in each coordinate plane the distance varies between the two semi-axes of that plane's ellipse. A circle of radius r is a curve on which the distance function is constant, so it must avoid every region of the surface where the distance is always larger than r or always smaller. The yz-ellipse is the "no further than b" region and the xy-ellipse the "no closer than b" region, and every central plane cuts both. The squeeze is forced.

So the real question is why the centre is O, and that is Step 2. The official solution states it in one sentence: "parallel cross sections of the ellipsoid are always similar ellipses". Here is where that comes from. An ellipsoid is a sphere stretched along three axes. Parallel slices of a sphere are circles that shrink away from the middle; stretching is linear, carries parallel planes to parallel planes and circles to ellipses of one fixed shape, so parallel slices of the ellipsoid are homothetic ellipses that shrink away from the middle. The completed square in Step 2 is the algebraic form of the same sentence. A student who thinks of an ellipsoid as a stretched sphere never needs the eight-unknown parametrisation.

Step 4 has its own "of course". Asking where a circle of radius b on the ellipsoid could be is asking where the ellipsoid meets the sphere of radius b. Two quadrics centred at O meet where their difference vanishes, and because b is the middle value, the difference has one negative and one positive coefficient and none for y^2: it factors into two planes. With $r = a$ or $r = c$ the difference would be semidefinite and vanish only on an axis. That is why the middle axis, and only the middle axis, gives circles. The hypothesis $a > b > c$ is doing exactly this work.

The natural attempt fails not because it is wrong but because it throws away the definition of a circle and replaces it with a parametrisation that has to rediscover the definition coefficient by coefficient. Zombie substitution with eight unknowns is the enemy this problem is set against.

Layer $+\alpha$ BACKTESTING

1. The cue is "largest circle" on a centrally symmetric surface. Write down: a circle is a plane section; parallel sections are similar; so take the section through the centre, whose circle is centred at O.

2. With the centre at O, radius equals distance from O. The method is to bound that distance using points of the circle forced to lie in the coordinate planes.

3. Hesitation comes in believing that the central section is the largest. Reinforce with the stretched-sphere picture, or the completed square $Q(\mathbf{p}_h + \mathbf{u}) = Q(\mathbf{p}_h) + Q(\mathbf{u})$.

4. The trap is parametrising the circle in general position and substituting. The second trap is proving only $r \le b$ and forgetting to show that a circle of radius b exists.

5. When a problem asks about circles or spheres, use "fixed distance from the centre" before any coordinates. When two quadrics with the same centre must meet, subtract them and factor.

6. For extremal problems on a symmetric body, first reduce to sections through the centre of symmetry, then bound the extremal quantity by evaluating it on the symmetry planes. For intersections of quadrics, form the difference; a difference with coefficients of opposite sign factors into planes.

▷ **Cue:** "largest circle" or "largest section" of an ellipsoid $\longrightarrow$ **Move:** take the parallel section through the centre; parallel sections are homothetic and shrink away from it

▷ **Cue:** a circle whose centre is known to be O $\longrightarrow$ **Move:** radius = distance from O; bound it using points where the circle crosses coordinate planes

▷ **Cue:** a plane through O in $\mathbb{R}^3$ $\longrightarrow$ **Move:** it meets every other plane through O in at least a line

▷ **Cue:** find where an ellipsoid and a sphere with the same centre meet $\longrightarrow$ **Move:** subtract the equations; the middle semi-axis kills one term and the rest factors into two planes

▷ **Cue:** an ellipsoid appears $\longrightarrow$ **Move:** think of it as a linearly stretched sphere; linear maps preserve parallelism and similarity of sections

1970 A6

1970 A6. Three numbers are chosen independently at random, one from each of the three intervals $(0, L_i]$ $(i = 1, 2, 3)$. If the distribution of each random number is uniform with respect to length in the interval it is chosen from, determine the expected value of the smallest of the three numbers chosen.

Caveat. The MAA volume's solution to 1970 A-6 sets up $L_1 L_2 L_3 E[X]$ as a triple integral and gives an inclusion–exclusion formula for $P(X \le a)$, then says "the answer follows easily" without stating the answer. The closed form below is derived here, checked against the equal-lengths case $(L/4)$, the limit $L_2, L_3 \to \infty$ $(L_1/2)$, and a Monte Carlo simulation with $(L_1, L_2, L_3) = (1, 2, 5)$ (0.3910 simulated against 0.39167 predicted). Identity check: the book's solution is about the minimum of three independent uniforms on intervals of

lengths L_i, which matches this statement.

Fields: probability (continuous distributions, expectation) **Position:** A6 — last of the morning paper, but one of the gentler A6s: the difficulty is choosing the formula for $E[X]$, after which it is a polynomial integral **Answer:** With the labels chosen so that $L_1 \le L_2 \le L_3$,

$$E[\min] = \frac{L_1}{2} - \frac{L_1^2}{6}\Big(\frac{1}{L_2} + \frac{1}{L_3}\Big) + \frac{L_1^3}{12\,L_2 L_3}.$$

Layer 1 THE WHAT — surface analysis

The natural attempt. The definition of expectation is $E[X] = \int x\, f_X(x)\, dx$, and for a function of three independent variables that becomes

$$E[\min(x,y,z)] = \frac{1}{L_1 L_2 L_3}\int_0^{L_1}\!\!\int_0^{L_2}\!\!\int_0^{L_3} \min(x,y,z)\, dz\, dy\, dx.$$

This is where the MAA solution starts too. Now $\min(x,y,z)$ is a different variable in each of six regions of the box, according to which coordinate is smallest, and the regions are cut off differently by the three unequal walls L_1, L_2, L_3. You split the inner integral at $z = \min(x,y)$:

$$\int_0^{L_3} \min(x,y,z)\, dz = \int_0^{\mu} z\, dz + \int_{\mu}^{L_3} \mu\, dz = \mu L_3 - \tfrac{1}{2}\mu^2, \qquad \mu = \min(x,y),$$

which is valid because $\mu \le L_1 \le L_3$. Then split again at $y = x$, being careful that y runs up to $L_2 \ge L_1$, so the region $y > x$ has a piece where y goes beyond L_1. It can be finished, and it gives the right answer, but every split is a chance to get a limit wrong, and a student who did not first order the L_i is splitting on conditions that change from case to case.

The model chain. Relabel so that $L_1 \le L_2 \le L_3$; the problem is symmetric in the three numbers, so this loses nothing. Let x, y, z be the three numbers and $X = \min(x,y,z)$, so $0 < X \le L_1$.

Tail formula. For a random variable X with $0 \le X \le L_1$,

$$E[X] = E\Big[\int_0^{L_1} \mathbf{1}_{\{a<X\}}\, da\Big] = \int_0^{L_1} P(X > a)\, da,$$

since $\int_0^{L_1} \mathbf{1}_{\{a<X\}}\, da = X$ pointwise and the order of expectation and integration may be exchanged (the integrand is non-negative and bounded).

Tail of the minimum. $X > a$ exactly when all three numbers exceed a, so by independence, for $0 \le a \le L_1$,

$$P(X > a) = \Big(1 - \frac{a}{L_1}\Big)\Big(1 - \frac{a}{L_2}\Big)\Big(1 - \frac{a}{L_3}\Big).$$

Integrate. Write $s_1 = \sum 1/L_i$, $s_2 = \sum_{i<j} 1/(L_i L_j)$, $s_3 = 1/(L_1 L_2 L_3)$. The product is $1 - s_1 a + s_2 a^2 - s_3 a^3$, so

$$E[X] = L_1 - \frac{s_1 L_1^2}{2} + \frac{s_2 L_1^3}{3} - \frac{s_3 L_1^4}{4}.$$

Expanding each term with L_1 cancelled where possible:

$$L_1 - \frac{L_1}{2} - \frac{L_1^2}{2L_2} - \frac{L_1^2}{2L_3} + \frac{L_1^2}{3L_2} + \frac{L_1^2}{3L_3} + \frac{L_1^3}{3L_2L_3} - \frac{L_1^3}{4L_2L_3},$$

which collects to

$$E[X] = \frac{L_1}{2} - \frac{L_1^2}{6}\Big(\frac{1}{L_2} + \frac{1}{L_3}\Big) + \frac{L_1^3}{12 L_2 L_3}.$$

Checks. All equal to L: $\frac{L}{2} - \frac{L}{3} + \frac{L}{12} = \frac{L}{4}$, the known mean of the minimum of three uniforms on $[0, L]$. $L_2, L_3 \to \infty$: the other two numbers are almost never the smallest, and the formula tends to $L_1/2$, the mean of the first number alone.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start by writing $E[X]$. The natural attempt writes it as the integral of X against the joint density and has to describe X region by region. The model writes it as the integral of the tail $P(X > a)$, which for a minimum is a product of three linear factors.

> *The minimum is ugly, but the event "the minimum is bigger than a" is the same as "all three are bigger than a", and independence turns that into a product. So I want a formula for $E[X]$ that only uses $P(X > a)$.*

> *Before splitting anything, I order the lengths. The minimum can never exceed the shortest interval, so L_1 is the only upper limit that matters.*

Layer 3 THE WHY

Why is the tail formula the natural choice rather than a trick from a probability course? Because of what a minimum is. The minimum of several numbers is defined by comparisons, and comparisons are awkward inside an integrand: $\min(x, y, z)$ is a piecewise function with six pieces. But the statement "$\min(x, y, z) > a$" is not piecewise at all; it is the conjunction "$x > a$ and $y > a$ and $z > a$". Independence is precisely the property that turns "and" into multiplication. So the event $\{X > a\}$ is where the problem's hypothesis (independence) and the problem's object (a minimum) fit together. Anything that expresses $E[X]$ through $P(X > a)$ will use both at full strength.

The tail formula itself is not deep. A non-negative number is the length of the interval $[0, X)$: $X = \int_0^\infty \mathbf{1}_{\{a<X\}}\, da$. Take expectations of both sides and swap: $E[X] = \int_0^\infty P(X > a)\, da$. For a non-negative integer-valued variable this is the familiar $E[N] = \sum_{k\geq 0} P(N > k)$. It is the same identity as computing an area under a graph by horizontal strips instead of vertical ones.

The MAA solution's alternative writes $P(X \leq a)$ by inclusion–exclusion, seven terms, and then says the answer "follows easily" from $\int a\, dP(X \leq a)$. Here is where that comes from, and why it is the long way round. $P(X \leq a)$ is $1 - P(X > a)$; inclusion–exclusion is the expansion of $1 - \prod(1 - p_i)$. Working with the complement directly means you never write the seven terms. And integrating a against the density requires differentiating first; the tail formula integrates the distribution function as it stands.

The ordering $L_1 \leq L_2 \leq L_3$ matters for one reason: the tail formula needs to know where $P(X > a)$ becomes 0, and it becomes 0 at the shortest length. Beyond $a = L_1$ the factor $1 - a/L_1$ would turn negative, so the product formula is only valid on $[0, L_1]$, and the answer is not symmetric in the three lengths. A student who integrates to $\max L_i$, or who writes an answer symmetric in L_1, L_2, L_3, has made this mistake. The answer's asymmetry is information, not ugliness.

The problem punishes zombie mathematics: the six-region triple integral is the most primal route and it works, but only a solver who first asks "is there a formula for $E[X]$ that suits a minimum?" does it in five lines.

Layer $+\alpha$ BACKTESTING

1. The cue is "expected value of the smallest" of independent variables. Write down $P(\min > a) = \prod P(x_i > a)$ and $E[X] = \int_0^\infty P(X > a)\, da$ immediately.

2. The tail-integral method: compute the survival function, which is a product for a minimum, and integrate it up to the point where it vanishes.

3. Hesitation comes at the choice between the joint-density integral and the tail formula; students default to the definition. Reinforce with: "a minimum is a conjunction of comparisons; conjunctions of independent events multiply".

4. The trap is integrating the product up to L_3 (or any length other than the smallest), where a factor goes negative. A second trap is an answer symmetric in the L_i: the correct answer depends on which interval is shortest.

5. Whenever the object is a maximum or minimum of independent variables, move to distribution or survival functions. For a maximum, $P(\max \le a) = \prod P(x_i \le a)$; for a minimum, $P(\min > a) = \prod P(x_i > a)$.

6. For E of a non-negative quantity whose tail is simple, use $E[X] = \int_0^\infty P(X > a)\, da$. Before integrating, identify where the tail vanishes and order any parameters so that one limit governs.

▷ **Cue:** expected value of the minimum of independent random variables $\longrightarrow$ **Move:** compute $P(\min > a) = \prod_i P(x_i > a)$, then $E = \int_0^\infty P(\min > a)\, da$

▷ **Cue:** expected value of the maximum of independent random variables $\longrightarrow$ **Move:** compute $P(\max \le a) = \prod_i P(x_i \le a)$, then $E = \int_0^\infty \big(1 - P(\max \le a)\big) da$

▷ **Cue:** a symmetric problem with unequal parameters L_1, L_2, L_3 $\longrightarrow$ **Move:** relabel so $L_1 \le L_2 \le L_3$ before any case split

▷ **Cue:** an integrand like $\min(x, y, z)$ or $\max(x, y)$ over a box $\longrightarrow$ **Move:** swap to the tail formula instead of splitting into ordering regions

▷ **Cue:** a closed form for an expectation $\longrightarrow$ **Move:** test it on equal parameters and on a limit where one variable dominates

Putnam 1970 — Paper B

1970 B1

1970 B1. Evaluate

$$\lim_{n\to\infty} \frac{1}{n^4}\left[\prod_{i=1}^{2n}(n^2+i^2)\right]^{1/n}.$$

Caveat. The MAA 1965–1984 OCR renders this problem as "Evaluate 1 ... limit- II (n1 + i 1)11". n' t-1" – the exponents on n and i, the upper limit of the product, and the power outside it are all lost to column-misreading, leaving only the skeleton "limit of $\frac{1}{n^?}\prod(n^2+i^2)^{1/n}$, i starting at 1." The statement above was reconstructed by working backward from the book's own (equally garbled but more informative) solution line, which reads, after correcting obvious OCR substitutions (L for Σ, "1 2n" for the summation's upper limit, "t1" for "$i=1$"): $\log a_n = \frac{1}{n}\sum_{i=1}^{2n}\log\left(1+\frac{i^2}{n^2}\right)$, and whose final line integrates $\log(1+x^2)$ from 0 to 2 to get $2\log 5 - 4 + 2\arctan 2$. Requiring $\log a_n$ to reduce to exactly that sum, with no leftover $\log n$ terms, pins down both the missing upper limit ($2n$, so that i/n ranges over $(0,2]$) and the missing normalising power (n^4, so that $-4\log n$ cancels the $4\log n$ produced by pulling n^2 out of all $2n$ factors). The reconstructed statement was then verified end to end by rederiving the book's own final line independently (see Layer 1); no mathematical content was invented beyond what the solution itself forces.

Fields: Infinite products, Riemann sums / definite integrals **Position:** B1 (opener) **Answer:** The limit equals $25\,e^{2\arctan 2-4}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The shape is a product of $2n$ factors, each within a bounded ratio of n^2, raised to the power $1/n$ and divided by n^4. A natural first move is to squeeze it: since i ranges from 1 to $2n$, every factor n^2+i^2 satisfies $n^2 < n^2+i^2 \le n^2+4n^2 = 5n^2$. Raising the whole product to the $1/n$ and dividing by n^4 then gives

$$\left(n^2\right)^{2n/n}/n^4 \;<\; \frac{1}{n^4}\left[\prod_{i=1}^{2n}(n^2+i^2)\right]^{1/n} \;\le\; \left(5n^2\right)^{2n/n}/n^4,$$

i.e. $1 < a_n \le 25$. This is correct, cheap, and instantly tells you the limit is a specific number between 1 and 25 – but it is a bound, not a value, and no amount of refining the squeeze (splitting the range into smaller sub-blocks and bounding each) will converge to the exact transcendental constant $25e^{2\arctan 2-4} \approx 4.19$, because the terms n^2+i^2 genuinely vary continuously across the range rather than clustering near one endpoint. You can spend the rest of the hour subdividing the product into finer and finer blocks, tightening the sandwich, and you will approach the answer only in the limit of infinitely many blocks – which is precisely the integral you were trying to avoid setting up.

The model chain. Take logarithms first: with a_n the quantity in question,

$$\log a_n = -4\log n + \frac{1}{n}\sum_{i=1}^{2n}\log(n^2+i^2).$$

Factor n^2 out of every term of the sum – legitimate because n^2 is the common, n-dependent part of each factor, while i^2/n^2 stays bounded (by 4, since $i \le 2n$) as $n \to \infty$:

$$\log(n^2+i^2) = \log n^2 + \log\left(1+\frac{i^2}{n^2}\right) = 2\log n + \log\left(1+\frac{i^2}{n^2}\right).$$

Substituting,

$$\log a_n = -4\log n + \frac{1}{n}\sum_{i=1}^{2n}\left[2\log n + \log\left(1+\frac{i^2}{n^2}\right)\right] = -4\log n + 4\log n + \frac{1}{n}\sum_{i=1}^{2n}\log\left(1+\frac{i^2}{n^2}\right).$$

The $\log n$ terms cancel exactly, leaving

$$\log a_n = \frac{1}{n}\sum_{i=1}^{2n}\log\left(1+\left(\frac{i}{n}\right)^2\right),$$

which is precisely a right-endpoint Riemann sum, with mesh $1/n$, for $f(x) = \log(1+x^2)$ sampled at $x_i = i/n$ for $i = 1, \ldots, 2n$, i.e. over the interval $(0, 2]$. Hence

$$\lim_{n\to\infty}\log a_n = \int_0^2 \log(1+x^2)\,dx.$$

Integrate by parts with $u = \log(1+x^2)$, $dv = dx$: $\int \log(1+x^2)\,dx = x\log(1+x^2) - \int \frac{2x^2}{1+x^2}\,dx = x\log(1+x^2) - 2x + 2\arctan x + C$. Evaluating from 0 to 2:

$$\int_0^2 \log(1+x^2)\,dx = 2\log 5 - 4 + 2\arctan 2.$$

Therefore $\lim_{n\to\infty} a_n = e^{2\log 5 - 4 + 2\arctan 2} = 25\,e^{2\arctan 2 - 4}$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree that a_n is trapped between 1 and 25, since both notice $n^2 < n^2 + i^2 \le 5n^2$. They split at the very next decision: what to do with that trapped quantity. The natural attempt tries to tighten the trap directly, treating $n^2 + i^2$ as "a number close to n^2" and refining the closeness. The model chain instead takes log and factors n^2 out of every term, converting the whole sum into a function of the single bounded ratio i/n alone – at which point the "trap" stops being a bound to tighten and becomes an exact Riemann sum to evaluate.

> *Both the divisor n^4 and every factor $n^2 + i^2$ are contaminated by n directly, not just by the ratio i/n – I need to factor n^2 out of each log term so that what's left is a fixed function evaluated only at i/n, because a sum only becomes an integral once every term is (mesh)×(a function of the sample point alone), with no leftover, unrescaled n hiding inside it.*

Layer 3 THE WHY

A Riemann sum is not merely "a sum that looks big"; it has a precise shape: $\frac{1}{n}\sum f(x_i)$ where f is a single fixed function and x_i are points of a fixed interval, independent of n. The raw expression $\log(n^2+i^2)$ fails this test on sight – it depends on n in two places, once through the explicit n^2 and once through the implicit bound on i (which runs up to $2n$). Squeezing treats this

failure as an obstacle to be bounded around; factoring treats it as a signal to be removed. The removal is forced by the algebra of logarithms themselves: $\log(n^2+i^2) = \log n^2 + \log(1+i^2/n^2)$ is not a trick, it is the unique way to split a sum of two quantities of different orders (n^2, which diverges, against $i^2/n^2 \leq 4$, which stays bounded) into an unbounded part and a bounded part using the one algebraic identity available for logarithms of sums. Once split, the unbounded part is summed exactly ($2\log n$, $2n$ times, giving $4n\log n$, divided by n giving $4\log n$) and must be checked against whatever normalising power of n the problem supplies outside the product – here n^4 in the denominator was not an arbitrary given constant to be taken on faith, it is exactly the power forced by this cancellation, which is also how the reconstruction in the

Caveat.

above was possible: the normalisation is not free information, it is determined by requiring the $\log n$ terms to vanish. What remains, $\log(1+i^2/n^2)$, is manifestly a fixed function of the single bounded ratio i/n, and the moment every term of a sum has that shape, $\frac{1}{n}\sum f(i/n)$ is by definition converging to $\int f$. The upper limit of integration, 2, is likewise not chosen but read off directly from where i/n stops: i runs to $2n$, so i/n runs to 2.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the pairing of a product of order-n many terms with an exponent $1/n$ outside it: any time a product's size is being controlled by an exponent that is the reciprocal of (roughly) its number of factors, write down "take logs, this becomes an averaged sum."
2. That cue suggests taking $\log a_n$ and rewriting it as $\frac{1}{n}\sum(\text{something})$, then actively hunting for a way to make "something" a fixed function of i/n alone, rather than accepting it as an expression that still contains n explicitly.
3. The hesitation happens right after the squeeze bound $1 < a_n \leq 25$ is found: it looks like real progress, and it is tempting to keep refining it. It is reinforced away by noticing the squeeze bound never gets better no matter how it is subdivided until the number of blocks goes to infinity – at which point you have re-derived the Riemann sum anyway, just without having named it, so naming it immediately is strictly cheaper.
4. The trap is mistaking a genuine (and correct) two-sided bound for a method of evaluation. A bound that is tight only in the limit of infinite subdivision is, definitionally, an integral in disguise; treating it as an algebra problem to be squeezed harder wastes the exam on reproducing, term by term and without machinery, what the Riemann-sum definition already hands you in one line.
5. To recognise this faster: whenever a limit has the outward shape "(product or sum of $\sim n$ terms, each depending on n and an index i running proportionally to n)$^{\text{power}\sim 1/n}$", check within the first thirty seconds whether the terms factor into an n-only part times a function of i/n alone – that check is the entire difference between a clean integral and a page of squeezing.
6. Decision rule: when evaluating $\left[\prod_{i=1}^{cn} F(n,i)\right]^{1/n}$ for some external normalisation, take logarithms immediately, then split each $\log F(n,i)$ into a part depending on n alone and a part depending only on the bounded ratio i/n; the n-only part must cancel against whatever power of n sits outside the product (if it does not cancel, the normalisation given in the problem is wrong or the limit diverges), and the surviving part is a Riemann sum whose integration bounds are read directly off the range of i/n.

1970 B2

1970 B2. The time-varying temperature of a certain body is given by a polynomial in the time of degree at most three. Show that the average temperature of the body between 9

A.M. and 3 P.M. can always be found by taking the average of the temperatures at two fixed times, which are independent of which polynomial occurs. Also, show that these two times are 10:16 A.M. and 1:44 P.M. to the nearest minute.

Fields: Integral calculus (definite integrals), polynomials, quadrature **Position:** B2 (middle) **Answer:** The two fixed times are noon $\mp T/\sqrt{3}$ with $T = 3$ hours, i.e. 10:16 A.M. and 1:44 P.M.

Layer 1 THE WHAT — surface analysis

The natural attempt. Set clock time $t = 0$ at 9 A.M., so the interval in question is $t \in [0, 6]$, and write the general cubic $P(t) = at^3 + bt^2 + ct + d$. The average temperature is

$$\overline{P} = \frac{1}{6}\int_0^6 P(t)\,dt = \frac{1}{6}\left[a\cdot\frac{6^4}{4} + b\cdot\frac{6^3}{3} + c\cdot\frac{6^2}{2} + d\cdot 6\right] = 54a + 12b + 3c + d.$$

You want two times t_1, t_2 with $\frac{1}{2}\big[P(t_1) + P(t_2)\big] = \overline{P}$ for every choice of a, b, c, d. Expanding $P(t_1) + P(t_2)$ and matching coefficients of a, b, c, d separately (since the identity must hold for arbitrary values of each), you get four simultaneous conditions:

$$\tfrac{1}{2}(t_1^3 + t_2^3) = 54, \qquad \tfrac{1}{2}(t_1^2 + t_2^2) = 12, \qquad \tfrac{1}{2}(t_1 + t_2) = 3, \qquad \tfrac{1}{2}(1 + 1) = 1.$$

The last is an automatic identity and gives no information; the first three are a genuinely nonlinear system in two unknowns t_1, t_2 – three equations, two unknowns, and you cannot yet see that they are consistent. It is solvable (elimination using $t_1 + t_2 = 6$ and $t_1 t_2$ as the two symmetric functions, then checking the cubic condition is automatically satisfied), but it costs a page of simultaneous-equation bookkeeping and a moment of real anxiety about whether an over-determined system that happens to work is a coincidence or a certainty, when nothing in the setup so far explains why it should work.

The model chain. Recentre time at the midpoint of the interval, noon, so that the two endpoints 9 A.M. and 3 P.M. become $-T$ and T with $T = 3$ hours, and write $P(\tau) = a\tau^3 + b\tau^2 + c\tau + d$ in this shifted variable. The average is

$$\overline{P} = \frac{1}{2T}\int_{-T}^{T} P(\tau)\,d\tau.$$

Because τ^3 and τ are odd functions, their integrals over the symmetric interval $[-T, T]$ vanish identically, for every a and every c:

$$\int_{-T}^{T}\big(a\tau^3 + c\tau\big)\,d\tau = 0 \quad \text{for all } a, c.$$

So only the even-degree terms survive:

$$\overline{P} = \frac{1}{2T}\left[b\cdot\frac{2T^3}{3} + d\cdot 2T\right] = \frac{bT^2}{3} + d.$$

Now try the two-point average at symmetric times $\tau = \pm s$: since odd powers of τ cancel between $P(s)$ and $P(-s)$ for exactly the same reason ($s^k + (-s)^k = 0$ for odd k),

$$\tfrac{1}{2}\big[P(s) + P(-s)\big] = \tfrac{1}{2}\big[2bs^2 + 2d\big] = bs^2 + d.$$

This already holds for every a and every c automatically – no condition on a, c is needed, because both sides killed those terms identically. Matching what remains, $bs^2 + d = \frac{bT^2}{3} + d$ for every b, d, forces exactly one equation in one unknown:

$$s^2 = \frac{T^2}{3} \quad\Longrightarrow\quad s = \frac{T}{\sqrt{3}}.$$

With $T = 3$ hours, $s = \sqrt{3} \approx 1.7320508$ hours $= 1$ hour 43.92 minutes ≈ 1 hour 44 minutes. The two times, symmetric about noon, are 10:16 A.M. and 1:44 P.M.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains write down the same cubic $P(t) = at^3 + bt^2 + ct + d$ and demand the two-point average match the true average for all a, b, c, d. They split at the choice of where to put $t = 0$. The natural attempt keeps the problem's given clock coordinates (9 A.M. $= 0$), which forces every one of the four coefficients into the matching conditions and produces a 3-equation, 2-unknown nonlinear system. The model chain relocates the origin to the interval's own midpoint before writing down a single equation.

> *The interval is symmetric if I put its centre – noon – at the origin; once I do, a cubic's odd-degree terms vanish under both the integral over a symmetric interval and the sum $P(s) + P(-s)$, for exactly the same reason, so recentring kills two of my four unknown coefficients for free before I have done any algebra at all.*

Layer 3 THE WHY

The problem states that the answer is independent of the polynomial, i.e. that some two-point rule reproduces $\overline{P}$ exactly for every cubic. This is, by definition, a request for a quadrature rule exact on cubics, and the standard structural fact about such rules is that odd-degree behaviour and even-degree behaviour decouple completely as soon as the interval and the sample points are both symmetric about a common centre: for any odd k, $\int_{-T}^{T} \tau^k \, d\tau = 0$ and $s^k + (-s)^k = 0$, because τ^k is itself an odd function and $s, -s$ are its reflection pair. Both facts are consequences of the single definition of an odd function, $f(-\tau) = -f(\tau)$, nothing more. Choosing clock time (9 A.M. as $t = 0$) breaks this symmetry for no reason connected to the mathematics – it is an artefact of how the problem happened to be phrased, not a constraint on the answer – and paying for that arbitrary choice is exactly the extra work the natural attempt does: matching a and c by brute force what recentring makes automatic. Once centred, the two-point average and the true average are both, structurally, functions only of b and d; the single remaining equation $s^2 = T^2/3$ is then not a coincidence rescued by careful elimination, it is the entire content of the problem, isolated because nothing else could have survived the symmetry. This is the general Gaussian-quadrature insight in miniature: symmetric nodes are exact for odd-degree contributions automatically, so a search for an exact-on-cubics rule reduces, the instant symmetry is imposed, to matching only the even part.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the phrase "independent of which polynomial occurs" attached to an interval with two named, fixed endpoints: on seeing it, write down "this is a request for a quadrature rule exact on polynomials up to the stated degree."
2. That cue suggests asking whether the interval has a natural centre, and if so, shifting the time variable there before writing a single matching equation, since quadrature rules exact for a given parity class of polynomials are almost always stated (and most cleanly derived) in coordinates symmetric about that centre.
3. The hesitation in the recentred version occurs at trusting that the a- and c-conditions really do vanish rather than merely appearing to; it is reinforced away by writing the two cancellations out explicitly – $\int_{-T}^{T} a\tau^3 \, d\tau = 0$ and $as^3 + a(-s)^3 = 0$ – as instances of the one definition of an odd function, rather than accepting them as a lucky pattern.
4. The trap is solving the problem in the coordinates it was literally given in (9 A.M. as time zero), which forces all four coefficients into an over-determined-looking system of three equations in two unknowns and manufactures doubt about consistency that the symmetric formulation never raises.

5. To recognise this faster: whenever a problem demands a property hold "for every polynomial of degree at most k" on a stated interval, check first whether recentring at the interval's midpoint is available, and if it is, always take it – the number of live unknowns typically halves.

6. Decision rule: for an exact-for-all-polynomials-of-degree-$\leq k$ identity on an interval $[p, q]$, work in the variable $\tau = t - \frac{p+q}{2}$ centred at the interval's midpoint; odd-degree coefficients then satisfy the identity automatically by parity, and only the even-degree coefficients impose real constraints on the nodes.

1970 B3

1970 B3. A closed subset S of $\mathbb{R}^2$ lies in $a < x < b$. Show that its projection on the y-axis is closed.

Caveat. The MAA 1965–1984 OCR renders the ambient space as "$R1$" rather than $\mathbb{R}^2$. This must be $\mathbb{R}^2$: the problem speaks of S having a projection on "the y-axis," which presupposes a second coordinate, and the book's own solution explicitly works with points $(x_n, y_n) \in S$ and concludes $(x, y) \in S$ – a two-coordinate argument throughout. The bound "$a < x < b$" is taken at face value; Bolzano–Weierstrass below only needs the x-coordinates to be bounded, so whether the strip is open or closed at a, b does not affect the proof, and no reconstruction of that clause was necessary.

Fields: Real analysis (metric topology, closed sets) **Position:** B3 (middle) **Answer:** The projection of S onto the y-axis is closed; the proof extracts a convergent x-subsequence via Bolzano–Weierstrass (using that S is confined to the bounded strip $a < x < b$) and then invokes closedness of S once.

Layer 1 THE WHAT — surface analysis

The natural attempt. "Closed" has a second, equally standard definition: the complement is open. Following that definition, fix y_0 not in the projection of S; you want to produce $\delta > 0$ so that $(y_0 - \delta, y_0 + \delta)$ meets no point of the projection, i.e. so that S contains no point with y-coordinate in that band. For each individual $x \in [a, b]$, the point (x, y_0) is not in S (since y_0 is not in the projection), and S is closed, so there is some $\varepsilon_x > 0$ with the open ball of radius ε_x around (x, y_0) missing S entirely. That gives you, for every single x, a safe vertical margin around y_0 – but ε_x depends on x, and you need one δ that works simultaneously for every $x \in [a, b]$ at once, since S could in principle come close to height y_0 at a different x for each attempted δ. Getting a uniform δ from the pointwise ε_x's is exactly the tube lemma: cover the compact segment $[a, b] \times \{y_0\}$ by these balls, extract a finite subcover, and take δ to be the minimum of the finitely many radii involved. It works, but it requires setting up an open cover of a compact set and extracting a finite subcover by name – real machinery for what turns out to be a short proof.

The model chain. Use the other standard definition of closed: a set is closed exactly when it contains the limit of every convergent sequence of its own points. Let y be any limit point of $\mathrm{proj}_y(S)$, witnessed by a sequence $y_n \to y$ with $y_n \in \mathrm{proj}_y(S)$ for every n. By definition of projection, for each n there is some x_n with $(x_n, y_n) \in S$. Since $S \subset \{a < x < b\}$, the sequence (x_n) is bounded, so by the Bolzano–Weierstrass theorem it has a convergent subsequence $x_{n_k} \to x$ for some $x \in [a, b]$. Passing to this same subsequence of the y_n's changes nothing about their limit, since a subsequence of a convergent sequence converges to the same limit: $y_{n_k} \to y$ still. Hence $(x_{n_k}, y_{n_k}) \to (x, y)$ in $\mathbb{R}^2$, and every one of these points lies in S. Because S is closed, it contains the limit of any of its own convergent sequences, so $(x, y) \in S$. Therefore

$y = \text{proj}_y(x, y) \in \text{proj}_y(S)$. Since $\text{proj}_y(S)$ contains every limit of its own convergent sequences, it is closed.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start from the same fact: closed has two equivalent definitions, and either one is a legitimate way to prove a set is closed. They split at the decision of which definition to deploy. The natural attempt reaches for "complement is open," which then needs a uniform bound built from pointwise data – exactly what compactness via open covers (the tube lemma) supplies, but only after real setup. The model chain reaches for "contains its own limit points," which needs a convergent sequence – exactly what compactness via Bolzano–Weierstrass supplies, in one line, directly from the boundedness hypothesis already stated in the problem.

> *Closed has two definitions I can attack with – 'contains its limit points' and 'complement is open.' The hypothesis I've actually been handed is that the x-coordinates are bounded, and Bolzano–Weierstrass is exactly the theorem that turns a bounded sequence into a convergent one; I should use the sequential definition of closed, because it is the one whose proof consumes precisely the compactness tool the problem's own hypothesis supplies, not the open-cover machinery a different but equally valid definition would require me to build from scratch.*

Layer 3 THE WHY

The problem hands over exactly two structural facts and nothing else: "S is closed" and "S lies in the bounded strip $a < x < b$." The task is to choose the formulation of "closed" whose proof template consumes those two facts directly. The sequential definition of closedness – every convergent sequence of points of the set converges to a point of the set – is built to accept exactly one kind of raw material: a sequence. Bolzano–Weierstrass is built to produce exactly one kind of output from exactly one kind of hypothesis: a convergent subsequence, from a bounded sequence. The match is total: the bound on S's x-coordinates is not incidental scenery, it is precisely the hypothesis Bolzano–Weierstrass needs, and nothing about the problem supplies a hypothesis suited to the open-cover route instead (that route needs to manufacture compactness of $[a, b] \times \{y_0\}$ as a covering argument from scratch, using the boundedness only indirectly, through the definition of compact via open covers, rather than through the definition of compact via sequences). This is the general principle behind choosing between two textbook-equivalent characterisations of the same property: they are equivalent in truth, not in convenience, and the one to use is whichever one's proof-template has an input slot shaped exactly like the hypothesis you were actually given. Here that slot is "bounded sequence in," and the problem statement is already handing you a bounded sequence the moment you pick a convergent sequence in the projection and lift it back to S.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that "closed" appears twice – once as a given property of S, once as the property to be proved of its projection – alongside an explicit boundedness clause on exactly one coordinate. On seeing this pairing, write down: "sequential characterisation of closed, plus Bolzano–Weierstrass on the bounded coordinate."
2. That cue suggests the concrete method: take an arbitrary convergent sequence witnessing a candidate limit point of the projection, lift each term back to a point of S, extract a convergent subsequence in the bounded coordinate via Bolzano–Weierstrass, and invoke closedness of S exactly once on the resulting convergent pair in the plane.
3. The hesitation is whether the full sequence (x_n, y_n) converges – it need not, since x_n alone is only known to be bounded, not convergent. It is reinforced away by recognising you do not need the whole sequence to converge, only a subsequence, and that passing to a subsequence

of x_n costs nothing on the y_n side, since a subsequence of an already-convergent sequence keeps the same limit automatically.

4. The trap is reaching first for the "complement is open" definition, since it is often taught as the default definition of closed; it is not wrong, but it forces a uniform-radius argument out of pointwise data via an open cover and finite subcover, machinery the problem's actual hypothesis (a bounded coordinate) does not hand you directly.
5. To recognise this faster: whenever "closed" is both a hypothesis on one set and a conclusion about a derived set (projection, image, closure of an operation), and a boundedness clause is present on some coordinate, default first to sequences plus Bolzano–Weierstrass; reserve the open-cover / tube-lemma route for problems that supply a covering-type hypothesis instead of a boundedness one.
6. Decision rule: to show that a projection (or any continuous image) of a closed, coordinate-bounded set is closed, pick a convergent sequence in the image, lift to the domain, extract convergence in the bounded coordinate via Bolzano–Weierstrass, and let closedness of the original set finish the argument in one invocation; do not reach for open covers unless no coordinate is stated to be bounded.

1970 B4

1970 B4. An automobile starts from rest and ends at rest, traversing a distance of one mile in one minute, along a straight road. If a governor prevents the speed of the car from exceeding ninety miles per hour, show that at some time of the traverse the acceleration or deceleration of the car was at least 6.6 ft./sec.^{2}

Caveat. The MAA 1965–1984 OCR ends the statement with "6.6 ft./sec. 1"; the trailing "1" is a dropped superscript and the unit is ft/sec^2, as the book's own solution (which compares $|v'(t)|$ with 6.6 in feet and seconds) confirms. The book's solution under the label B-4 bounds v by $\min\{6.6t,\ 6.6(60-t),\ 132\}$ and integrates to 5280, so it solves exactly this statement. The book writes $0 \le v(t) \le 132$, silently assuming the car never reverses; the proof below bounds $|v|$ instead, so it holds whether "distance" means net displacement or total path length.

Fields: Real analysis (mean value theorem, integral bounds); kinematics **Position:** B4 (middle of the paper, but one of the easiest problems on it) **Answer:** Suppose $|a(t)| < 6.6$ throughout. The mean value theorem traps the velocity graph under the trapezoid $\min\{6.6t,\ 6.6(60-t),\ 132\}$, whose area is exactly 5280 ft. So the car covers less than a mile, a contradiction. The constant 6.6 is sharp.

Layer 1 THE WHAT — surface analysis

The natural attempt. You're strong at mechanics, so you draw the journey you'd actually plan. Accelerate uniformly at rate a up to the speed limit, cruise, then brake at rate a. In feet and seconds: 1 mile $= 5280$ ft, 1 minute $= 60$ s, and 90 mph $= 90 \cdot 5280/3600 = 132$ ft/s. Speeding up takes $132/a$ seconds, and so does braking. The area under the velocity–time graph is the distance, so

$$5280 = 132\left(60 - \frac{132}{a}\right) \quad\Longrightarrow\quad 60 - \frac{132}{a} = 40 \quad\Longrightarrow\quad a = 6.6.$$

The number comes out exactly, and that feels like the end. It isn't. You've shown that one particular family of motions (constant acceleration, cruise, constant deceleration) needs $a = 6.6$. The problem is about every motion. A driver could use a smooth S-shaped speed-up, never quite reach 90 mph, or brake in pulses. Nothing in your calculation rules out one of those

managing the mile with a peak acceleration below 6.6. You have found the right candidate for the extreme case, but you haven't proved it is extreme. Writing "clearly the trapezoid is optimal" is exactly the gap a grader looks for.

The model chain. Work in feet and seconds. Let $v(t)$ be the velocity on $[0, 60]$. It is differentiable with acceleration $v'(t)$, and $v(0) = v(60) = 0$. The governor gives $|v(t)| \le 132$. The distance covered is $\int_0^{60} |v(t)|\,dt$, and this is at least the net displacement $\left|\int_0^{60} v(t)\,dt\right|$. So in either reading of "distance",

$$5280 \le \int_0^{60} |v(t)|\,dt. \qquad (*)$$

Suppose, for contradiction, that $|v'(t)| < 6.6$ for every $t \in [0, 60]$. For $0 < t \le 60$, the mean value theorem on $[0, t]$ gives some $\xi \in (0, t)$ with $v(t) = v(t) - v(0) = t\,v'(\xi)$, so $|v(t)| < 6.6\,t$. In the same way, on $[t, 60]$ for $0 \le t < 60$, $|v(t)| = |v(60) - v(t)| = (60 - t)|v'(\eta)| < 6.6(60 - t)$. Combining these with the governor,

$$|v(t)| \le g(t) := \min\{6.6t,\ 6.6(60 - t),\ 132\} \quad \text{for all } t \in [0, 60],$$

and the inequality is strict on $(0, 20)$, where $g(t) = 6.6t$. The function $g - |v|$ is continuous, non-negative, and positive on an interval, so $\int_0^{60} |v| < \int_0^{60} g$. The graph of g is a trapezoid. It rises to 132 at $t = 132/6.6 = 20$, stays level until $t = 40$, and falls back to 0 at $t = 60$. Its area is

$$\int_0^{60} g(t)\,dt = \tfrac{1}{2} \cdot 132 \cdot (60 + 20) = 5280.$$

So $\int_0^{60} |v| < 5280$, which contradicts $(*)$. Hence $|v'(t)| \ge 6.6$ for some t. The bound can't be improved: accelerating at exactly 6.6 for 20 s, cruising at 132 for 20 s and braking at 6.6 for 20 s covers exactly 5280 ft.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains draw the same trapezoid and get the same 6.6. They split at the question of what the trapezoid is. In the natural attempt it's a proposed motion: the car you'd drive, used as a sample. In the model chain it's a ceiling: a curve that every allowed velocity graph has to stay under. That switch from one example to a bound on all motions is the whole proof. Checking the ceiling takes one application of the mean value theorem for each sloping side.

> *I've found the motion that needs exactly* 6.6*, but "show that at some time" means I have to handle every motion. So I'll turn the question round: if the acceleration never reaches* 6.6*, how big can the velocity be at time* t*? Starting from rest, it's below* $6.6t$*. Ending at rest, it's below* $6.6(60 - t)$*. The governor caps it at* 132*. So the graph lives under the minimum of the three, and the area under that minimum is the most distance any such car can cover.*

> *My calculation gave exactly* 5280*. That can't be a coincidence. The setter chose* 6.6 *so that the best possible car only just manages the mile, so the strict inequality has to come from assuming* $|a| < 6.6$*, and I must keep it strict all the way through the integral.*

Layer 3 THE WHY

Why is "bound the graph" the natural move rather than a trick? Go back to the A-level fact the whole problem rests on. Distance is the area under the velocity–time graph. That turns "cover a mile" into "have area 5280". It also changes the question from "what did the car do?" to "how big can this area be?". The largest area under a graph comes from the tallest graph allowed. So the right question is: what is the highest the graph can go at each time t?

Each hypothesis is a limit on height, and you can read them off one at a time. The governor is already a height limit: $|v| \leq 132$. The acceleration bound limits slope, not height. But a slope limit becomes a height limit once you fix a point on the graph, and the problem fixes two: $v(0) = 0$ and $v(60) = 0$. Any line leaving $(0, 0)$ with slope under 6.6 in absolute value stays under the line $6.6t$. The same is true for any graph, and the mean value theorem is the precise statement of that. It says the average slope over $[0, t]$ equals the actual slope at some point, so a bound on every slope bounds the average slope, and that bounds the height. Seen that way, the three terms $6.6t$, $6.6(60 - t)$ and 132 aren't guessed. They are the three hypotheses written as heights. Taking the minimum is "all three at once". Integrating is "distance is area". None of it needs inspiration. Each line applies a definition you already had.

This also explains why the natural attempt reached the right number. The trapezoid you would drive is the ceiling itself, because a car that pushes every constraint to the limit at every moment traces out exactly that boundary. The optimal motion and the envelope are the same curve, so the SUVAT calculation gives the right constant. But in that attempt nothing shows it is the extreme case. The envelope argument does. The book's solution gives the three-way minimum without comment. The reason for it is the one above: to bound an area from above, bound the height at every point.

Finally, the exact equality $\int g = 5280$ is deliberate. In general, for distance d, time T and speed cap V, the ceiling argument gives: some $|a| \geq V^2/(VT - d)$ when the cap binds ($V < 2d/T$), and $|a| \geq 4d/T^2$ from the triangle when it doesn't. Here $2d/T = 176 > 132$, so the cap binds and $132^2/(7920 - 5280) = 6.6$. Without the governor the triangle gives only $4 \cdot 5280/3600 \approx 5.87$. So the governor does real work, and the answer being exact tells you that 6.6 is the true minimum.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that the problem pins the velocity at both ends ("starts from rest and ends at rest"), bounds its size (the governor) and asks about its slope. On seeing that, write down "distance $= \int v$, and bound v pointwise", then convert everything to feet and seconds at once: 5280, 60, 132.

2. Proof by contradiction with a pointwise envelope. Assume $|v'| < 6.6$, use the mean value theorem from each endpoint to get two sloping ceilings, add the governor as a flat ceiling, integrate the minimum and compare with 5280.

3. The hesitation comes after the SUVAT calculation gives 6.6: you feel finished, but the argument covers only one type of motion. What removes it is asking whether you've shown this for every car or only for the car you chose. A second, smaller hesitation is over strictness. The bound $|v| \leq 132$ isn't strict on the cruise segment, so find one interval where the inequality is strict (here $(0, 20)$) and use it.

4. The trap is proving the result for the natural candidate motion and treating it as optimal without proof. A smaller trap is using $0 \leq v \leq 132$ as the book does, which assumes the car never reverses. Bounding $|v|$ avoids this at no extra cost.

5. When a problem asks you to show that some rate must have reached a threshold, and gives the start and end values of the quantity plus a total (an area or integral), assume the rate stays below the threshold and turn that into a ceiling on the quantity. If a quick calculation with an idealised profile gives exactly the stated constant, that profile is the ceiling. Your job is to prove every other profile lies under it.

6. Decision rule: a bound on a derivative together with a known value at a point is a bound on the function, a cone from that point, via the mean value theorem. To show that an integral constraint forces the derivative to be large somewhere, take the minimum of all such cones and any explicit caps, integrate that envelope, and show it falls short. Never prove a "for every motion" statement by optimising within one family of motions.

1970 B5

1970 B5. Let u_n denote the "ramp" function

$$u_n(x) = \begin{cases} -n & \text{for } x \le -n, \\ x & \text{for } -n < x \le n, \\ n & \text{for } x > n, \end{cases}$$

and let F denote a real function of a real variable. Show that F is continuous if and only if $u_n \circ F$ is continuous for all n. (Note: $(u_n \circ F)(x) = u_n[F(x)]$.)

Caveat. The MAA 1965–1984 OCR drops the middle and right-hand values of the piecewise definition, printing only "$-n$" followed by the three conditions "for $x \le -n$, for $-n < x \le n$, for $x > n$". The values x and n in the second and third branches are reconstructed from the word "ramp" and, decisively, from the book's own solution, which uses exactly that $u_n^{-1}[(a,b)] = (a,b)$ whenever $n > \max(|a|,|b|)$ – true for the clipping function above and for nothing materially different. The identity check passes: the solution printed under 1970 B-5 solves this statement. The index n is read as ranging over the positive integers; nothing changes if it ranges over all positive reals.

Fields: Real analysis (continuity, preimages of open sets) **Position:** B5 (upper-middle; hard only if the quantifiers are handled carelessly) **Answer:** True as stated. The forward direction is composition of continuous functions; for the converse, at each point x_0 pick n larger than $|F(x_0)| + 1$, and near x_0 the function $u_n \circ F$ is literally equal to F.

Layer 1 THE WHAT — surface analysis

The natural attempt. The forward direction takes one line and every student gets it: u_n is continuous (three continuous pieces that agree at the joins, $u_n(-n) = -n$ and $u_n(n) = n$), so if F is continuous then $u_n \circ F$ is a composition of continuous functions. The work is the converse. The strong student's first instinct is global: "as $n \to \infty$ the ramp becomes the identity, so $u_n \circ F \to F$." And that is true pointwise – for fixed x, once $n > |F(x)|$ we have $u_n(F(x)) = F(x)$. So F is the pointwise limit of the continuous functions $g_n = u_n \circ F$, and the plan is "a limit of continuous functions is continuous". Here is the wall. That statement is false for pointwise limits: x^n on $[0,1]$ is continuous for every n and converges pointwise to a function that jumps at 1. What rescues limits is uniform convergence, so the student tries to prove $\sup_x |g_n(x) - F(x)| \to 0$. But $|g_n(x) - F(x)| = \max(|F(x)| - n, 0)$, and if F is unbounded (take $F(x) = x$) this supremum is $+\infty$ for every n. Restricting to a compact interval does not help either: we do not yet know F is continuous, so we do not know it is bounded on compact sets – that would be assuming the conclusion. The student is now pushing on uniform convergence, which is the wrong tool, and pushing hard.

The model chain. Fix a point x_0 and $\varepsilon > 0$; we prove F is continuous at x_0. Replace ε by $\varepsilon' = \min(\varepsilon, 1)$, and choose an integer $n \ge |F(x_0)| + 1$. Then $-n < F(x_0) \le n$, so $u_n(F(x_0)) = F(x_0)$. By hypothesis $u_n \circ F$ is continuous at x_0, so there is $\delta > 0$ with

$$|x - x_0| < \delta \implies |u_n(F(x)) - F(x_0)| < \varepsilon'.$$

For such x, the number $y = u_n(F(x))$ lies in $(F(x_0) - 1,\ F(x_0) + 1)$, and since $n \ge |F(x_0)| + 1$ this interval sits inside $(-n, n)$. Now read the definition of u_n backwards: the only outputs of u_n equal to $\pm n$ come from clipped inputs, and an output strictly inside $(-n, n)$ can only come from the middle branch, where the output equals the input. So $u_n(F(x)) \in (-n, n)$ forces

$F(x) = u_n(F(x))$. Hence $|F(x) - F(x_0)| < \varepsilon' \le \varepsilon$ whenever $|x - x_0| < \delta$. As x_0 was arbitrary, F is continuous.

The book's solution says the same thing in the open-set language: it suffices that $F^{-1}[(a,b)]$ is open for every bounded interval (a,b) (every open set is a union of such intervals and preimages respect unions); for $n > \max(|a|,|b|)$ one has $u_n^{-1}[(a,b)] = (a,b)$, so $F^{-1}[(a,b)] = (u_n \circ F)^{-1}[(a,b)]$, which is open.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree on the observation "for large n, $u_n(F(x)) = F(x)$". They split on the order of the quantifiers in that sentence. The natural attempt reads it as "for every x, eventually in n" and tries to make n work for all x at once – that is uniform convergence, and it fails. The model reads it as "for this one x_0, pick one n" and then asks the continuity of that single $u_n \circ F$ to show that nearby x are also unclipped. The split happens at the words "choose n": the model chooses n *after* fixing x_0, depending on $F(x_0)$.

> *Continuity is checked one point at a time. So I don't need one n that works everywhere – I fix x_0 first, and then I'm allowed to pick n depending on $F(x_0)$.*

> *If n is well above $|F(x_0)|$, the ramp u_n is the identity on a whole window around $F(x_0)$. If $u_n \circ F$ is continuous, nearby x land in that window – and inside the window, $u_n \circ F$ and F are the same function.*

> *Why the $\min(\varepsilon, 1)$? Because I need $u_n(F(x))$ to stay strictly inside $(-n,n)$ before I can say it equals $F(x)$. I have to keep a margin from the clipping level, and the 1 is that margin.*

Layer 3 THE WHY

Why is "fix the point first" the natural move rather than a trick? Go down to the definition. F is continuous at x_0 when for every $\varepsilon > 0$ there is $\delta > 0$ with $|F(x) - F(x_0)| < \varepsilon$ whenever $|x - x_0| < \delta$. Look at what that sentence actually uses: the value $F(x_0)$, and whether other values $F(x)$ fall inside the window $(F(x_0) - \varepsilon,\ F(x_0) + \varepsilon)$. It never asks what $F(x)$ is when it is far from $F(x_0)$ – only that it is not near. Continuity is local twice over: local in the domain (a small δ-neighbourhood of x_0), and local in the target (a small ε-window round $F(x_0)$). Everything outside those windows is invisible to the definition.

Now look at what u_n does. It changes nothing inside $(-n,n)$ and flattens everything outside to $\pm n$. That is the only information the hypothesis hides from you. So the question is whether flattening far-away values can hide a discontinuity. By the paragraph above it cannot. Continuity at x_0 only looks at the window round $F(x_0)$, and once n is big enough that window is untouched by the flattening. With this definition in front of you, picking n after x_0 is the natural move. The definition is pointwise, and the only thing n has to accommodate is the single number $F(x_0)$.

The one subtlety – and the reason a margin is needed – is that the window is judged through u_n. A value $u_n(F(x))$ close to $F(x_0)$ tells you $F(x)$ is close to $F(x_0)$ only if u_n is injective at the level being read. u_n is injective on $(-n,n)$ and badly non-injective at the endpoints: every input $\ge n$ produces the output n (strictly, $u_n(n) = n$ comes from the middle branch too). So you must stay strictly inside $(-n,n)$, with room to spare, which is what $n \ge |F(x_0)| + 1$ and $\varepsilon' \le 1$ buy you. In open-set language this is the book's identity $u_n^{-1}(W) = W$ for any open $W \subseteq (-n,n)$: open sets well inside the unflattened zone pull back to themselves.

Why does the natural attempt die, and where exactly? "F is the limit of $u_n \circ F$" is a global statement and it throws away the locality. Uniform convergence is how you carry continuity through a limit when the limit is taken everywhere at once. Here you never need a limit,

because at each point the sequence $u_n(F(x))$ is not merely convergent. It is *eventually constant*, and eventually constant on a neighbourhood. Once one term of the sequence equals F near x_0, taking a limit is wasted effort. That is zombie maths in its analytic form: reaching for the heaviest general theorem (limits of continuous functions) when the problem hands you exact equality.

Finally, note that the quantifier "for all n" is necessary. No single n works: let $F(x) = 1$ for $x \leq 0$ and $F(x) = 5$ for $x > 0$. Then $u_1 \circ F \equiv 1$ is continuous, but F is not. The hypothesis is strong because the clipping level can be raised above any value F takes, and the proof spends exactly that: one n per point.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "if and only if $u_n \circ F$ is continuous for all n", where u_n is the identity on a growing interval. On seeing it, write down: "$u_n = \mathrm{id}$ on $(-n, n)$; continuity is local; fix x_0, then choose $n > |F(x_0)| + 1$."
2. The cue suggests a pointwise ε–δ argument, or equivalently the preimage criterion on bounded open intervals, in which the auxiliary parameter n is chosen after the point (or the interval) is fixed. No limit theorems are needed.
3. You hesitate at "$u_n(F(x))$ is close to $F(x_0)$, but is $F(x)$?". What settles it: an output of u_n lying strictly inside $(-n, n)$ equals its input. Leave a margin of 1 between $|F(x_0)|$ and n, and cap ε at 1, and that fact applies.
4. The trap is the global reading: $F = \lim_n u_n \circ F$, so try to pass continuity through the limit. Pointwise limits do not preserve continuity, and the convergence is not uniform when F is unbounded. Nor can you restrict to a compact set, since boundedness of F there would need the continuity you are trying to prove.
5. Recognise it faster by asking of any hypothesis of the form "$g_n \circ F$ has property P for all n": is P local (continuity, differentiability, measurability of preimages), and is g_n the identity on a region that grows to cover everything? If both, fix the point first and choose n from it.
6. Decision rule: to prove a local property from a family of approximations, first fix the location, then pick one member of the family that is exactly equal to the target near that location. Use limits and uniform convergence only if no member is exactly equal anywhere. When a property is local, look for exact local agreement before reaching for approximation.

1970 B6

1970 B6. A quadrilateral which can be inscribed in a circle is said to be inscribable or cyclic. A quadrilateral which can be circumscribed to a circle is said to be circumscribable. Show that if a circumscribable quadrilateral of sides a, b, c, d has area $A = \sqrt{abcd}$, then it is also inscribable.

Caveat. The MAA 1965–1984 OCR renders the area condition as "$A = yabcd$"; the "y" is a mis-scanned radical sign. The reading $A = \sqrt{abcd}$ is confirmed by the book's own solution under the label B-6 of the Thirty-First Competition, which opens "Since the quadrilateral is circumscribable, $a + c = b + d$" and uses $\frac{1}{2}ab\sin\alpha + \frac{1}{2}cd\sin\beta = \sqrt{abcd}$. That solution solves exactly this statement, so the identity of the problem is secure. The sides a, b, c, d are taken in cyclic order around the quadrilateral, as the book's solution assumes.

Fields: Euclidean geometry; trigonometry (cosine rule, area formula) **Position:** B6 (last of Paper B)
Answer: In every circumscribable quadrilateral $A^2 = abcd\,\sin^2\frac{B+D}{2}$, where B, D are a pair of opposite angles; so $A = \sqrt{abcd}$ forces $B + D = \pi$, which is the cyclic condition.

Layer 1 THE WHAT — surface analysis

The natural attempt. You know two facts about circles and quadrilaterals by heart: a quadrilateral has an incircle when (and, for a convex one, only when) $a + c = b + d$ (Pitot), and it is cyclic exactly when a pair of opposite angles sums to π. So the target is clear: show $B + D = \pi$. The instinct is then to go looking for a circle. Put the incircle at the origin with radius r, let the tangent lengths from the four vertices be w, x, y, z, so the sides are $w + x$, $x + y$, $y + z$, $z + w$ and the area is $r(w + x + y + z) = rs$. The half-angles at the vertices satisfy $\tan(\frac{1}{2}\angle V) = r/(\text{tangent length from } V)$, and the four half-angles sum to π. Now the condition reads

$$r^2(w + x + y + z)^2 = (w + x)(x + y)(y + z)(z + w),$$

and you need to eliminate r using the constraint that the four arctangents $\arctan(r/w) + \cdots + \arctan(r/z) = \pi$. That constraint, expanded through the tangent addition formula, is a polynomial relation of the form $r^2(\text{sum of pairwise products}) = wxyz \cdot (\ldots)$, and you are now in five variables with two polynomial constraints, trying to extract an angle condition you have not even written down yet. This is the wall. It is not a false route – everything is true – but the variables chosen (tangent lengths, inradius) are native to the *incircle*, while the conclusion is about the *circumcircle*, i.e. about opposite angles. Nothing in this parametrisation ever mentions $B + D$.

The model chain. Let the quadrilateral be $PQRS$ with $PQ = a$, $QR = b$, $RS = c$, $SP = d$. A circumscribable quadrilateral is the intersection of the four half-planes bounded by its sides that contain the circle, so it is convex. Let $\alpha = \angle PQR$ (between a and b) and $\beta = \angle RSP$ (between c and d): a pair of opposite angles, each in $(0, \pi)$. The diagonal $PR = k$ splits the quadrilateral into triangles PQR and RSP, which by convexity do not overlap.

Step 1 (Pitot). The two tangent segments from any external point to a circle are equal. Calling the tangent lengths from P, Q, R, S respectively w, x, y, z gives $a = w + x$, $b = x + y$, $c = y + z$, $d = z + w$, so $a + c = b + d$, i.e. $a - b = d - c$.

Step 2 (two expressions for the shared data). The cosine rule in both triangles gives one equation for the diagonal, and the area splits:

$$a^2 + b^2 - 2ab\cos\alpha = k^2 = c^2 + d^2 - 2cd\cos\beta, \qquad 4A = 2ab\sin\alpha + 2cd\sin\beta.$$

Rearranged, the first reads $2ab\cos\alpha - 2cd\cos\beta = a^2 + b^2 - c^2 - d^2$.

Step 3 (use Pitot on the left-over sides). Since $(a - b)^2 = (d - c)^2$, we have $a^2 + b^2 - 2ab = c^2 + d^2 - 2cd$, so $a^2 + b^2 - c^2 - d^2 = 2ab - 2cd$. Hence

$$2ab\cos\alpha - 2cd\cos\beta = 2ab - 2cd, \qquad 2ab\sin\alpha + 2cd\sin\beta = 4A.$$

Step 4 (square and add). Squaring both and adding, the $\cos^2 + \sin^2$ pairs collapse and the cross terms combine by the addition formula:

$$4a^2b^2 + 4c^2d^2 - 8abcd(\cos\alpha\cos\beta - \sin\alpha\sin\beta) = 4(ab - cd)^2 + 16A^2,$$

that is, $8abcd - 8abcd\cos(\alpha + \beta) = 16A^2$, or

$$A^2 = abcd \cdot \frac{1 - \cos(\alpha + \beta)}{2} = abcd\,\sin^2\frac{\alpha + \beta}{2}.$$

Step 5 (read off the equality case). If $A^2 = abcd$ then $\sin^2\frac{\alpha+\beta}{2} = 1$. Since $0 < \alpha + \beta < 2\pi$, the half-sum lies in $(0, \pi)$, where $\sin = 1$ only at $\pi/2$. So $\alpha + \beta = \pi$.

Step 6 (cyclic). Draw the circumcircle Γ of PQR. Points X on the opposite side of line PR from Q with $\angle PXR = \pi - \alpha$ are exactly the points of the arc of Γ on that side (inscribed-angle theorem and its converse: the locus of points seeing a fixed segment under a fixed angle on one

side is a circular arc). S is on that side by convexity and $\angle PSR = \beta = \pi - \alpha$, so $S \in \Gamma$ and the quadrilateral is inscribable. ■

The book's solution is the same computation dressed differently: from Step 3 it extracts $ab(1-\cos\alpha) = cd(1-\cos\beta)$, substitutes it twice into $4A^2$ expanded with $\sin^2 = (1-\cos)(1+\cos)$, and lands on $4 = 2 - 2\cos(\alpha+\beta)$. Square-and-add reaches the same line with less bookkeeping, and as a bonus it produces the general formula rather than only the equality case.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree on the target ("show $B + D = \pi$") and on Pitot. They split at the choice of *coordinates for the quadrilateral*. The natural attempt parametrises by the incircle (tangent lengths w, x, y, z and r), because the hypothesis mentions the incircle. The model parametrises by two opposite angles α, β and one diagonal (†), because the conclusion is a statement about $\alpha + \beta$. At (†) the natural attempt has already lost: $\alpha + \beta$ is not a variable of its system.

> *The thing I have to prove is $\alpha + \beta = \pi$ – so α and β must be my variables. Cut along the diagonal between them, write the cosine rule and the area in both triangles, and the only way to get $\cos(\alpha+\beta)$ out of $\cos\alpha, \cos\beta, \sin\alpha, \sin\beta$ is to square and add.*

> *Pitot gives me $a + c = b + d$ but the cosine rule gives me $a^2 + b^2 - c^2 - d^2$. Those are linked through $(a-b)^2 = (c-d)^2$ – that is the one place the incircle is allowed to enter, and it enters by killing the side-length clutter.*

> *$A = \sqrt{abcd}$ looks like the equality case of something. Before trying to prove the implication, I should find the formula for A^2 in which abcd is the maximum.*

Layer 3 THE WHY

Why is "make the opposite angles the variables" not a trick? Go back to the definitions. A quadrilateral with fixed side lengths is not rigid: it has exactly one degree of freedom (hinge it at the vertices). Everything about its shape – its area, its diagonals, whether it is cyclic – is a function of that one parameter. The cyclic condition is, by the inscribed-angle theorem, a condition on the angle sum $\alpha + \beta$ of opposite vertices. So the natural single parameter is $\alpha + \beta$ itself, and the question becomes: *write the area as a function of $\alpha + \beta$*. Once the problem is phrased that way the cosine rule on the shared diagonal is forced – it is the only equation that ties the two triangles together – and square-and-add is forced because $\cos(\alpha+\beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta$ is a cross term, and cross terms come from squaring sums.

The general formula this produces for any quadrilateral (Bretschneider's formula) is

$$A^2 = (s-a)(s-b)(s-c)(s-d) - abcd\cos^2\frac{\alpha+\beta}{2}, \qquad s = \tfrac{1}{2}(a+b+c+d).$$

Read it slowly: area is largest, for given sides, exactly when the quadrilateral is cyclic, and then it equals Brahmagupta's $\sqrt{(s-a)(s-b)(s-c)(s-d)}$. Now impose Pitot: $s = a + c = b + d$, so $s - a = c$, $s - b = d$, $s - c = a$, $s - d = b$, and the Brahmagupta product becomes $abcd$. That is where $\sqrt{abcd}$ comes from. The problem's number is not arbitrary: it is the *maximum possible area* of a quadrilateral with those four sides, and the maximum of $-abcd\cos^2(\cdot)$ is attained only when the cosine is zero. The whole problem is "the equality case of a maximum is attained only at the maximiser", wearing a disguise.

This also explains why the natural attempt stalled. The incircle hypothesis contributes one thing – the linear relation $a + c = b + d$ – and after that the incircle is irrelevant. Parametrising by incircle data spends all your variables on the hypothesis that carries the least information.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the conclusion, "inscribable": write down immediately "cyclic $\iff$ opposite angles sum to π" and name those angles α, β. The second cue is $A = \sqrt{abcd}$: a closed-form area from sides alone should make you write down Heron/Brahmagupta/Bretschneider.

2. The method is: split along the diagonal separating α from β; cosine rule twice, area twice; square and add to create $\cos(\alpha + \beta)$; feed in Pitot through $(a-b)^2 = (c-d)^2$.

3. Hesitation occurs at "which diagonal?" and at "how do I use $a + c = b + d$ when everything else is quadratic?". The first is settled by the target: the diagonal whose two sides contain α and β. The second is settled by squaring the Pitot relation in its difference form $a - b = d - c$, which produces exactly the combination $a^2 + b^2 - c^2 - d^2$ the cosine rule left behind.

4. The trap is the incircle: the hypothesis names a circle, so you set up the incircle (tangent lengths, inradius, half-angle tangents) and drown in a five-variable system that never mentions the angle sum. A second, smaller trap: concluding $\alpha + \beta = \pi$ from $\sin^2 \frac{\alpha+\beta}{2} = 1$ without saying why the half-sum lies in $(0, \pi)$.

5. When a problem gives an extremal-looking closed form (an area, a product, a sum equal to some clean expression in the data) and asks you to deduce a rigid property, suspect that the closed form is the maximum or minimum of a known formula, and that the property is the equality case. Find the formula with a slack term (here $-abcd\cos^2\frac{\alpha+\beta}{2}$) and set the slack to zero.

6. Parametrise by the quantity the conclusion talks about, not the one the hypothesis talks about. Use each hypothesis at the single point where it simplifies the computation, and nowhere else. And whenever an equality is asked of something that is usually only an inequality, prove the inequality with an explicit non-negative remainder; the equality case falls out as "remainder $= 0$".

▷ **Cue:** The conclusion is "the quadrilateral is cyclic" $\longrightarrow$ **Move:** Name a pair of opposite angles α, β and aim to show $\alpha + \beta = \pi$; make these the working variables.

▷ **Cue:** Two triangles share a side, and you know an angle opposite the shared side in each $\longrightarrow$ **Move:** Write the cosine rule for the shared side in both, pair it with the area formula in both, and square-and-add to produce $\cos(\alpha + \beta)$.

▷ **Cue:** A tangential quadrilateral with $a + c = b + d$ meets a quadratic expression in the sides $\longrightarrow$ **Move:** Rewrite Pitot as $a - b = d - c$ and square: $a^2 + b^2 - c^2 - d^2 = 2(ab - cd)$.

▷ **Cue:** A quantity is stated to equal a clean closed form such as $\sqrt{abcd}$ and a rigid property is to be deduced $\longrightarrow$ **Move:** Recognise the closed form as the extremal value of a general formula with a squared slack term; the property is the vanishing of the slack.