DAWN PUTNAM SESSION

Putnam 1978 — Papers A and B

Putnam 120 Archive

Contents

Putnam 1978 — Paper A **5**
1978 A1 5
1978 A2 7
1978 A3 9
1978 A4 12
1978 A5 15
1978 A6 17

Putnam 1978 — Paper B **21**
1978 B1 21
1978 B2 23
1978 B3 26
1978 B4 29
1978 B5 31
1978 B6 34

Putnam 1978 — Paper A

1978 A1

1978 A1. Let A be any set of 20 distinct integers chosen from the arithmetic progression $1, 4, 7, \ldots, 100$. Prove that there must be two distinct integers in A whose sum is 104.

Fields: combinatorics, pigeonhole principle **Position:** A1 (opener) **Answer:** Any such A must contain two integers summing to 104, because the progression splits into 18 classes (16 pairs and 2 singletons) that no valid A-avoiding-the-sum can meet more than once each, and $20 > 18$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You've seen "arithmetic progression, sum of two terms" before, and the reflex it triggers is the Gauss trick: write the progression out, pair the first term with the last, the second with the second-last, and watch the sums stay constant. Here the progression is $1, 4, 7, \ldots, 100$, which has $\frac{100-1}{3} + 1 = 34$ terms, so you pair $(1, 100)$, $(4, 97)$, $(7, 94)$, and so on, down to $(49, 52)$. Each of these pairs sums to 101. Seventeen pairs, sum 101 each.

This is a real result and it's tempting to declare victory: 17 pigeonholes, so 18 elements of A would force two into the same hole, hence two integers summing to 101. Except the problem doesn't ask about 101. It asks about 104. You have built, correctly, an exhaustive and elegant partition of the progression into pairs of constant sum – and it is the wrong constant. Nothing in the pairing $(1, 100), (4, 97), \ldots$ produces a pair summing to 104; the closest you get is noticing that $104 = 101 + 3$, which tells you the target sum is offset from the "natural" endpoint-pairing sum by exactly one step of the common difference. That observation is progress, but it means every one of your 17 pairs is aimed at the wrong invariant, and you cannot patch the argument by relabelling – you have to rebuild the partition around 104, not around the endpoints 1 and 100.

The model chain. Don't pair by position in the list (first with last). Pair by the target value itself: for each term a in the progression, ask whether $104 - a$ is also a term of the progression, and if so, put a and $104 - a$ in one class.

A term $a = 1 + 3m$ (for $m = 0, 1, \ldots, 33$) satisfies $a \equiv 1 \pmod 3$. Then $104 - a \equiv 104 - 1 \equiv 103 \equiv 1 \pmod 3$ as well (since $104 \equiv 2 \pmod 3$), so $104 - a$ automatically lies in the same residue class as every term of the progression – the arithmetic progression's defining congruence is preserved by $a \mapsto 104 - a$. The only extra condition is the range: you need $1 \leq 104 - a \leq 100$, i.e. $4 \leq a \leq 103$, which combined with $a \leq 100$ gives $4 \leq a \leq 100$.

So every term a with $4 \leq a \leq 100$ has a genuine partner $104 - a$ in the progression – except when $a = 104 - a$, i.e. $a = 52$, which pairs with itself. And the one term outside this range, $a = 1$, has no partner ($104 - 1 = 103$ is not a term, since the progression stops at 100).

This produces exactly:

- one singleton $\{1\}$ (no valid partner),
- one singleton $\{52\}$ (self-paired, i.e. $52 + 52 = 104$, but you can't use the same integer twice, so 52 can appear in A without forcing anything),

- sixteen genuine pairs $\{4,100\},\{7,97\},\{10,94\},\ldots,\{49,55\}$, each summing to 104.

Check the count: 16 pairs contribute 32 terms, plus the two singletons is 34 – exactly the size of the progression. And these 18 classes are disjoint, since a term can belong to at most one class (its class is determined by pairing it with 104 minus itself).

Now pigeonhole directly: if A contains both elements of any one of the 16 pairs, those two elements sum to 104 and you're done. If not, A contains at most one element from each pair (at most 16 elements) plus possibly both singletons (at most 2 more), for a maximum of $16+2=18$ elements avoiding the sum entirely. Since $|A|=20>18$, A cannot avoid every pair, so at least one pair $\{a,104-a\}\subset A$, giving the required two integers.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start from the same instinct – pair up terms of the progression to build pigeonholes. They diverge at the moment of choosing *which* pairing rule to use: the natural attempt pairs by position (first with last, symmetric about the middle of the list), producing pairs of sum 101. The model chain pairs by target value (each a with $104-a$), producing pairs of sum 104.

> *The pairing I need isn't "symmetric about the middle of the list" – it's "symmetric about the number 104." I should be asking, for each term a, whether $104-a$ is also in the progression, not whether a sits opposite some other term in the printed order.*

Layer 3 THE WHY

The Gauss pairing is seductive here because it is the pairing that always exists for any finite arithmetic progression, independent of what sum you're chasing – it drops out of the list structure alone. But the problem hands you a specific target, 104, and a partition built without reference to that target is solving a different problem (find the guaranteed sum for *some* constant) rather than this one (find two integers summing to *this* constant). The fix is to let the target dictate the pairing rule from the start: given a required sum s, the only sensible equivalence relation on the progression is $a\sim a'$ iff $a+a'=s$, i.e. $a'=s-a$. This relation is well defined on the whole progression (not just on the specific 34 terms you can see) because it only depends on s and a; whether $s-a$ happens to land back inside the progression is then a separate, checkable condition – here, the congruence $a\equiv 1 \pmod 3$ forcing $s-a\equiv 1 \pmod 3$ whenever $s\equiv 2 \pmod 3$, plus the boundary check $1\le s-a\le 100$.

The deeper structural fact worth keeping: a subset of size k of an n-element set can avoid every pair of a fixed pairing on at most (number of classes) elements, one per class. So the whole problem reduces to counting classes, and the pairing rule that counts them correctly is the one anchored to the actual quantity in the problem statement – the sum 104 – not to whatever pairing the progression would suggest on its own. The endpoint values 1 and 52 are not arbitrary leftovers; 1 is the term with no partner in-range, and $52=104/2$ is the unique fixed point of $a\mapsto 104-a$, and both facts follow directly from writing the relation $a'=104-a$ down and imposing the progression's own membership conditions on it.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "prove two [distinct chosen elements] sum to a fixed constant s" applied to a set built by choosing many elements from a larger structured set. On seeing it, write down the relation $a'=s-a$ immediately, before reaching for any pairing that ignores s.

2. That cue suggests partitioning the ambient set (here, the progression) into classes under $a\sim(s-a)$, then pigeonholing the count of classes against the given subset size.

3. Hesitation occurs right after the progression is written out, because a different pairing – first-with-last – is already sitting in memory from other problems and produces a clean-looking table almost for free. Reinforce against it by asking: "does this pairing's constant sum match the number in the problem, or did I just default to the pairing I already knew?"

4. The trap is that the natural (first-with-last) pairing is not wrong, elegant-looking, or even hard to build – it's simply aimed at the wrong target sum, and its internal consistency (17 clean pairs of sum 101) can mask the mismatch until you try to conclude something about 104 specifically.

5. Recognise this faster next time by naming the target sum s before doing any pairing, and defining the equivalence relation symbolically ($a \sim s - a$) before substituting any specific values from the progression.

6. Decision rule: when asked to force a pair summing to a specific constant s from a structured set, partition that set by the relation $a \mapsto s - a$ (checking membership via the set's defining congruence and range, and isolating any self-paired or unpaired elements), then pigeonhole the subset size against the resulting class count. Never substitute a generic "reverse the list" pairing for the target-specific one.

1978 A2

1978 A2. Let $a, b, p_1, p_2, \ldots, p_n$ be real numbers with $a \neq b$. Define $f(x) = (p_1 - x)(p_2 - x)(p_3 - x)\cdots(p_n - x)$. Prove that

$$\det \begin{pmatrix} p_1 & a & a & a & \cdots & a & a \\ b & p_2 & a & a & \cdots & a & a \\ b & b & p_3 & a & \cdots & a & a \\ b & b & b & p_4 & \cdots & a & a \\ \vdots & & & & \ddots & & \vdots \\ b & b & b & b & \cdots & p_{n-1} & a \\ b & b & b & b & \cdots & b & p_n \end{pmatrix} = \frac{bf(a) - af(b)}{b - a}.$$

Caveat. The MAA OCR drops the definition of f and mangles the matrix into unreadable fragments (it retains only the answer formula $\frac{bf(a)-af(b)}{b-a}$ embedded inside the garbled problem text, plus the solution's use of $G(a) = f(a)$, $G(b) = f(b)$ for triangular determinants). The statement above is reconstructed to match: (i) the answer formula visible in the OCR, (ii) the solution's claim that setting $t = a$ or $t = b$ in a uniformly-shifted matrix yields a triangular matrix with determinant $f(a)$ or $f(b)$ respectively (forcing $f(x) = \prod(p_i - x)$, not $\prod(x - p_i)$), and (iii) an independent citation: Andreescu and Feng's *Putnam and Beyond* explicitly labels this identity "Problem 1978A2 [PutnamII, p. 31]" and reproduces it in exactly this form. I did not substitute a problem from an adjacent year – the OCR's own A-1 and A-3 solutions on the surrounding pages confirm this is 1978 Paper A.

Fields: linear algebra, determinants **Position:** A2 **Answer:** The determinant equals $\frac{bf(a) - af(b)}{b - a}$, where $f(x) = \prod_{i=1}^{n}(p_i - x)$.

Layer 1 THE WHAT — surface analysis

The natural attempt. An $n \times n$ determinant with a visible recursive block structure (each leading $k \times k$ principal submatrix looks like a smaller copy of the same problem) says "induction on n" to a well-trained eye, and that reflex is not wrong – it's the official alternate solution. Subtract the second column from the first: every entry below the diagonal in column 1 becomes $b - b = 0$ except the very last one, and every entry in the new column 1 above row 2 becomes $p_1 - a$ or $a - a = 0$ depending on position. Expanding by cofactors down this column splits the determinant into two smaller determinants of the same shape, one built from $p_2, \ldots, p_n$ with a playing its old role, one from $a, p_3, \ldots, p_n$. You now have a two-term recursion, and

you have to track two different auxiliary functions ($F(x) = (p_2 - x)\cdots(p_n - x)$ and $G(x) = (a - x)(p_3 - x)\cdots(p_n - x)$) through the recursion, verify the base case, and reassemble. It works – it's the official alternate solution – but it costs a genuine induction with two nontrivial families of determinants in flight simultaneously, and there's no point in the computation where you can see the answer coming before you've finished it.

The model chain. Don't induct on the size of the matrix. Exploit the fact that the matrix has exactly *two* off-diagonal values, and turn the whole determinant into a function of a single real parameter.

Let M_t be the matrix obtained by subtracting t from every entry of the given matrix (diagonal included), and let $G(t) = \det M_t$. Two things are true about G.

First, G is an affine (degree ≤ 1) function of t. Write the original matrix as $M_0 = M - tJ$ evaluated at $t = 0$ isn't quite the framing you want; instead see directly: $M_t = M_0 - tJ$, where J is the all-ones $n \times n$ matrix. Subtracting row $i+1$ of M_t from row i, for each i, cancels the t-dependence in every shared column except where the diagonal entries $p_i - t$ and $p_{i+1} - t$ sit, since all the other entries in rows i and $i+1$ are equal to the same constant (a or b) and difference away. After these row operations at most two entries per row remain t-dependent, linearly, so $\det M_t$ – a sum of products picking one entry per row and column – is a polynomial in t of degree at most 1. (Equivalently: J has rank 1, and the determinant of $A - tJ$ is always affine in t for a rank-one perturbation J, by the matrix determinant lemma.)

Second, G is easy to evaluate at $t = a$ and $t = b$ specifically, because those are exactly the two values that kill one of the two off-diagonal constants. At $t = a$: every entry above the diagonal becomes $a - a = 0$, so M_a is lower triangular, and $\det M_a$ is the product of its diagonal entries, $\prod_{i=1}^{n}(p_i - a) = f(a)$. At $t = b$: every entry below the diagonal becomes $b - b = 0$, so M_b is upper triangular, and $\det M_b = \prod_{i=1}^{n}(p_i - b) = f(b)$.

So G is an affine function of t with $G(a) = f(a)$ and $G(b) = f(b)$, known at two points. Solve for $G(0)$, the original determinant, by linear interpolation: writing $G(t) = \alpha t + \beta$,

$$\alpha = \frac{f(a) - f(b)}{a - b}, \qquad \beta = f(a) - \alpha a = \frac{bf(a) - af(b)}{b - a}.$$

Since $G(0) = \beta$, the original determinant is exactly $\dfrac{bf(a) - af(b)}{b - a}$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains recognise the same structural fact – the matrix's off-diagonal entries take only two values, a above and b below. They diverge at what to do with that fact: the natural attempt uses it to set up a column-expansion recursion in the matrix *size* n. The model chain uses it to build a one-parameter family of matrices M_t and studies the determinant as a function of the shift t, never inducting on n at all.

> *I don't need to shrink the matrix – I need to find a scalar I can vary that makes the matrix degenerate into something I already know how to handle. Shifting every entry by t and asking "which value of t makes this triangular?" answers itself twice, once for each off-diagonal constant.*

Layer 3 THE WHY

The uniform-shift move is available here for a specific structural reason that's worth isolating: the matrix's t-dependence, once you write $M_t = M_0 - tJ$, is a *rank-one* perturbation, because $J = \mathbf{1}\mathbf{1}^{\mathsf{T}}$ is rank one. A determinant is a multilinear, alternating function of the rows of a matrix; perturbing a matrix by a rank-one term changes only one "direction" in row-space at a time, and multilinearity of the determinant in that single direction is exactly what forces $\det(M_0 - tJ)$ to be affine in t (degree at most 1, since t only ever enters to the first power inside the multilinear expansion once the rank-one structure is accounted for). This is the same fact underlying the

matrix determinant lemma $\det(A - uv^\mathsf{T}) = \det A - v^\mathsf{T} \operatorname{adj}(A)u$. Once you know $G(t)$ is affine, evaluating it at *two* points determines it completely – and the problem itself supplies two natural evaluation points, a and b, because those are the two constants already sitting in the matrix. Set $t = a$ and the entries built from a vanish; set $t = b$ and the entries built from b vanish. The matrix wasn't given two off-diagonal values by accident: that's the datum telling you which two points to interpolate through.

The official solution's terse justification ("$G(t)$ is linear in t" via generic row subtraction) is correct but doesn't say why row subtraction is the right move to try; the rank-one framing above answers that directly – you subtract rows because the t-dependence lives in a single rank-one direction, and eliminating it from all but two rows is precisely how you certify a degree bound on a multilinear function.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that the matrix has a visible parameter pattern in its off-diagonal entries – exactly two constants, a above and b below, with the diagonal free. On seeing it, ask whether there's a single scalar t whose value collapses the matrix into a shape (triangular, diagonal, rank-deficient) you already know how to evaluate.
2. That cue suggests treating the determinant as a function $G(t) = \det(M_0 - tJ)$ of the shift and bounding its degree in t before evaluating it anywhere.
3. Hesitation sets in right after noticing the recursive block structure of the matrix, because column-expansion induction is the default response to "matrix with a repeating pattern" and it visibly makes progress immediately. Reinforce against it by asking: "is there a parameter I can vary across the whole matrix at once, rather than shrinking it one row and column at a time?"
4. The trap is that induction on n genuinely works here (it's the book's own alternate solution), so there's no failure to signal you should have looked elsewhere – the only way to find the cheaper route is to ask the structural question before committing to the recursion.
5. Recognise this faster next time by treating "matrix with only k distinct off-diagonal values" as a cue for a k-point interpolation on a shift parameter, independent of the matrix's size n.
6. Decision rule: if a determinant's entries are built from a small number of free constants $(a, b, \ldots)$ plus a diagonal, introduce a shift $M_t = M_0 - tJ$ (rank-one perturbation, hence $\det M_t$ affine in t), evaluate at the values of t that zero out each constant in turn (making the matrix triangular), and interpolate. Reserve size-induction for problems where no such parameter is visible.

1978 A3

1978 A3. Let $p(x) = 2 + 4x + 3x^2 + 5x^3 + 3x^4 + 4x^5 + 2x^6$. For k with $0 < k < 5$, define

$$I_k = \int_0^\infty \frac{x^k}{p(x)}\,dx.$$

For which k is I_k smallest?

Caveat. The MAA OCR preserves "$p(x) = 2 + 4x + 3x^2 + 5x^3 + 3x^4 + 4x^5 + 2x^6$. For k with $0 < k < 5$, define" and then drops the actual definition of I_k before resuming with "For which k is I_k smallest?" I reconstructed $I_k = \int_0^\infty x^k/p(x)\,dx$ from the solution page, which is intact and performs the substitution $x = 1/t$ on exactly this integrand: "$I_k = \int_\infty^0 t^{-k}/p(1/t)\cdot(-dt/t^2) = \int_0^\infty t^{4-k}/p(t)\,dt = I_{4-k}$", explicitly computed with $p(t)/t^6$ in place of $p(1/t)$, and concludes "I_k is smallest for $k = 2$" via AM-GM on $x^k + x^{4-k} \geq 2x^2$. I

independently verified that this reconstructed integral converges precisely for $-1 < k < 5$ (matching the solution's stated domain) and reproduces every step of the given solution, so the reconstruction is not a guess but the unique integrand consistent with the intact solution text.

Fields: analysis, definite integrals, inequalities **Position:** A3 **Answer:** I_k is smallest at $k = 2$, by the substitution symmetry $I_k = I_{4-k}$ together with AM-GM.

Layer 1 THE WHAT — surface analysis

The natural attempt. "For which k is I_k smallest" is an optimisation question, and the reflex for optimisation is calculus: treat I_k as a differentiable function of the real parameter k, differentiate under the integral sign, and set the derivative to zero.

$$\frac{dI_k}{dk} = \int_0^\infty \frac{x^k \ln x}{p(x)}\,dx.$$

You'd want to show this vanishes at some specific k and is negative before it, positive after – that would pin down a unique minimiser. The trouble starts immediately: $\int_0^\infty \frac{x^k \ln x}{p(x)}\,dx$ is not an elementary integral for a generic sextic p, and there is no algebraic manipulation that turns it into something you can set to zero and solve for k in closed form. You can sometimes argue convexity in k abstractly (Hölder's inequality does show $\ln I_k$ is convex in k, which would at least guarantee a unique minimum exists), but that only tells you a minimiser exists – it doesn't tell you which k it is, and the problem wants a specific numerical answer. You're stuck differentiating a family of transcendental integrals with no route to an explicit critical point, and you haven't yet used a single one of the specific coefficients $2, 4, 3, 5, 3, 4, 2$ that the problem bothered to give you.

The model chain. Don't differentiate yet. Look at $p(x)$ itself first: its coefficients, read in order, are $2, 4, 3, 5, 3, 4, 2$ – the same sequence forwards and backwards. This is not decoration. A polynomial of degree $2n$ whose coefficient list is a palindrome is called self-reciprocal, and it satisfies $x^{2n}p(1/x) = p(x)$ identically. Check it directly for this p (degree 6, so $n = 3$):

$$p(1/x) = 2 + \frac{4}{x} + \frac{3}{x^2} + \frac{5}{x^3} + \frac{3}{x^4} + \frac{4}{x^5} + \frac{2}{x^6} = \frac{2x^6 + 4x^5 + 3x^4 + 5x^3 + 3x^2 + 4x + 2}{x^6} = \frac{p(x)}{x^6}.$$

That identity is exactly what the palindrome guarantees, and it turns the substitution $x = 1/t$ from a routine trick into a targeted one: it should relate I_k to some other $I_{k'}$, because p is invariant (up to a power of x) under $x \mapsto 1/x$.

Substitute $x = 1/t$, $dx = -dt/t^2$, with $x : 0 \to \infty$ corresponding to $t : \infty \to 0$:

$$I_k = \int_\infty^0 \frac{(1/t)^k}{p(1/t)}\left(-\frac{dt}{t^2}\right) = \int_0^\infty \frac{t^{-k-2}}{p(1/t)}\,dt = \int_0^\infty t^{-k-2}\cdot\frac{t^6}{p(t)}\,dt = \int_0^\infty \frac{t^{4-k}}{p(t)}\,dt = I_{4-k}.$$

So $I_k = I_{4-k}$ for every k in the domain of convergence – the map $k \mapsto I_k$ is symmetric about $k = 2$, with no calculus involved yet.

Now use that symmetry to locate the minimum. For any k with $0 < k < 5$ (so $0 < 4 - k < 5$ too, keeping both integrals inside the given range),

$$2I_k = I_k + I_{4-k} = \int_0^\infty \frac{x^k + x^{4-k}}{p(x)}\,dx.$$

By AM-GM, for every $x > 0$,

$$\frac{x^k + x^{4-k}}{2} \geq \sqrt{x^k \cdot x^{4-k}} = x^2,$$

with equality exactly when $x^k = x^{4-k}$, i.e. $x = 1$ (for any $k \neq 2$) – so equality holds only on a single point, not on an interval, whenever $k \neq 2$. Since $p(x) > 0$ for all $x \geq 0$ (every coefficient of p is positive), dividing by $p(x)$ preserves the inequality, and integrating a strict-almost-everywhere inequality over $(0, \infty)$ gives a strict inequality between the integrals whenever $k \neq 2$:

$$2I_k = \int_0^\infty \frac{x^k + x^{4-k}}{p(x)}\,dx > \int_0^\infty \frac{2x^2}{p(x)}\,dx = 2I_2 \qquad (k \neq 2).$$

So $I_k > I_2$ for every $k \neq 2$ in range, and I_k is smallest exactly at $k = 2$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains treat "for which k is I_k smallest" as the question it literally is – find the minimiser of a one-parameter family of integrals. They split at the very first move: the natural attempt reaches for calculus (differentiate I_k with respect to k) without first looking at what makes this particular $p(x)$ special. The model chain looks at $p(x)$'s coefficient sequence before doing anything else, notices it's a palindrome, and builds the whole argument (substitution symmetry plus AM-GM) around that fact instead.

> *The coefficients of p read the same forwards and backwards – before I differentiate anything, I should ask what $x \mapsto 1/x$ does to p, because a palindromic coefficient list is exactly the signature of a polynomial invariant under that substitution.*

Layer 3 THE WHY

The instinct to differentiate is not unreasonable in general – it's the correct tool for "find the extremum of a smooth family" when nothing else is visible. What makes it the wrong first move *here* is that the problem supplied a very specific, suspiciously symmetric polynomial, and "suspiciously symmetric coefficients" is itself mathematical content, not flavour text: a palindrome in the coefficients of a degree-$2n$ polynomial is definitionally equivalent to the identity $x^{2n}p(1/x) = p(x)$, and that identity is precisely the hypothesis that makes the substitution $x = 1/t$ do something other than relabel the same integral. Differentiating under the integral throws this structure away before you've used it; the derivative $\int x^k \ln x / p(x)\,dx$ knows nothing about where p's coefficients came from, because differentiation in k doesn't interact with symmetry in x at all. The self-reciprocal property lives in the x-variable, so the tool that can see it has to act on the x-variable – a substitution, not a k-derivative.

Once you have $I_k = I_{4-k}$, the rest is forced by the shape of the problem: an integral that's invariant under $k \mapsto 4 - k$ and is being minimised is a textbook AM-GM setup, because AM-GM is exactly the inequality that compares an average of two symmetric quantities ($\frac{x^k + x^{4-k}}{2}$) to their geometric mean ($x^{\frac{k+(4-k)}{2}} = x^2$, which is k-independent). The fixed point of the symmetry, $k = 2$, is forced to be the extremiser because it's the unique value where the two things being averaged in $I_k = \frac{1}{2}(I_k + I_{4-k})$ coincide identically in x, collapsing the AM-GM inequality to equality. This is the general pattern behind "the answer is a fixed point of a symmetry you found": symmetry alone tells you the extremum is *at* $k = 2$ or split symmetrically around it; it takes one convexity-type inequality (here AM-GM, applied pointwise before integrating) to tell you which of those two possibilities holds and that the fixed point is a minimum rather than a maximum or saddle.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the coefficient list of $p(x)$ itself – $2, 4, 3, 5, 3, 4, 2$ – read before anything else is attempted. On seeing a palindromic coefficient sequence in the denominator of a ratio being optimised, write down the identity $x^{2n}p(1/x) = p(x)$ immediately.
2. That cue suggests the substitution $x = 1/t$ inside the integral defining I_k, aimed specifically at producing a relation between I_k and $I_{k'}$ for some other exponent k'.

3. Hesitation occurs at the very first line, because "for which k is [a quantity] smallest" reads as a calculus prompt before you've looked at the integrand's structure. Reinforce against it by asking: "did the problem hand me a generic function, or a suspiciously specific one – and if specific, what property does that specific choice have?"

4. The trap is that differentiation under the integral sign is technically valid and looks like the direct route to an answer; it only fails silently, by leading into an intractable transcendental equation with no closed-form root, rather than failing visibly at the first step.

5. Recognise this faster next time by treating any "coefficients read the same forwards and backwards" polynomial as an immediate cue for the substitution $x \mapsto 1/x$, before any optimisation technique is attempted on the surrounding problem.

6. Decision rule: when asked to optimise $I_k = \int f(x,k)/p(x)\,dx$ over a parameter k, check the coefficient structure of p (and of f) for a hidden symmetry $x \mapsto 1/x$ (or another natural substitution) before differentiating in k; a symmetry $I_k = I_{g(k)}$ reduces the problem to locating and classifying its fixed point via a pointwise inequality (AM-GM, Cauchy-Schwarz, convexity), which is almost always cheaper than solving a transcendental critical-point equation.

1978 A4

1978 A4. A "bypass" operation on a set S is a mapping from $S \times S$ to S with the property

$$B(B(w,x),B(y,z)) = B(w,z) \quad \text{for all } w,x,y,z \text{ in } S.$$

(a) Prove that $B(a,b) = c$ implies $B(c,c) = c$ when B is a bypass.
(b) Prove that $B(a,b) = c$ implies $B(a,x) = B(c,x)$ for all x in S when B is a bypass.
(c) Construct a table for a bypass operation B on a finite set S with the following three properties:
(i) $B(x,x) = x$ for all x in S.
(ii) There exist d and e in S with $B(d,e) = d \neq e$.
(iii) There exist f and g in S with $B(f,g) \neq f$.

Caveat. The MAA OCR renders (ii) as "$B(d,e) \bullet d + e$" and (iii) as "$B(f,g) + f$". Both are read as $B(d,e) = d \neq e$ and $B(f,g) \neq f$: this is the standard wording, it is the only reading under which (ii) and (iii) are non-trivial and compatible with (i), and the book's solution ("(ii) and (iii) will hold if I and J, respectively, have more than one element") confirms it. The book's solution under A-4 solves exactly this problem, so the identity of the problem is not in doubt.

Fields: abstract algebra (binary operations, rectangular bands) **Position:** A4 — mid-paper; (a) and (b) are substitution exercises, (c) is where a strong student stalls **Answer:** (a) $B(c,c) = B(B(a,b),B(a,b)) = B(a,b) = c$. (b) Compute $B(c,B(x,x))$ in two ways. (c) Take $S = \{1,2\} \times \{1,2\}$ with $B((i,j),(h,k)) = (i,k)$; four elements is the minimum possible.

Layer 1 THE WHAT — surface analysis

The natural attempt. Parts (a) and (b) are one identity and some unknowns, so you start plugging in. For (a) you want $B(c,c)$, and $c = B(a,b)$, so you write $B(c,c) = B(B(a,b),B(a,b))$ and the axiom with $(w,x,y,z) = (a,b,a,b)$ gives $B(a,b) = c$. Done in one line; nobody is stuck here.

Part (b) is where the plugging gets aimless. You want $B(c,x) = B(a,x)$. The left side contains $c = B(a,b)$, so you try $B(c,x) = B(B(a,b),x)$ — but the axiom needs *both* arguments

to be B-values, and x is bare. You try substituting $x = B(\cdot,\cdot)$ for something; you do not know that every x is a B-value. Many students circle here, trying (w,x,y,z) at random.

Part (c) is where the natural attempt really costs. You open a blank $n \times n$ table and fill it by trial. With $n = 2$, $S = \{p,q\}$: (i) fixes the diagonal, $B(p,p) = p$, $B(q,q) = q$. (ii) needs $B(d,e) = d$ with $d \neq e$, say $B(p,q) = p$. (iii) needs $B(f,g) \neq f$, so $f \neq g$ and $B(f,g) = g$; the only unused off-diagonal slot is $B(q,p)$, so $B(q,p) = p$. Now test the axiom with $(w,x,y,z) = (q,q,p,p)$: $B(B(q,q),B(p,p)) = B(q,p) = p$ fine; with (q,p,q,q): $B(B(q,p),B(q,q)) = B(p,q) = p$ but $B(q,q) = q$. Contradiction. So $n = 2$ fails, and at $n = 3$ there are $3^6 = 729$ ways to fill the off-diagonal cells, each needing 81 axiom checks. Blind search is the wall.

The model chain. *(a)* Put $(w,x,y,z) = (a,b,a,b)$: $B(c,c) = B(B(a,b),B(a,b)) = B(a,b) = c$.

(b) The axiom only speaks when both arguments are B-values, so manufacture one: $B(x,x)$ is a B-value for every x. With $(w,x,y,z) = (a,b,x,x)$,

$$B(c,B(x,x)) = B(B(a,b),B(x,x)) = B(a,x).$$

By (a), $c = B(c,c)$, so with $(w,x,y,z) = (c,c,x,x)$,

$$B(c,B(x,x)) = B(B(c,c),B(x,x)) = B(c,x).$$

The two displays have the same left side, so $B(a,x) = B(c,x)$.

(c) Read the axiom as a statement about what survives: in $B(B(w,x),B(y,z))$ only w and z — the outermost left and outermost right entries — matter. So let each element carry a "left label" and a "right label", and let B keep the left label of its left argument and the right label of its right argument. Concretely, $S = I \times J$ and

$$B\big((i,j),(h,k)\big) = (i,k).$$

Check the axiom: $B(B(w,x),B(y,z)) = B((w_1,x_2),(y_1,z_2)) = (w_1,z_2) = B(w,z)$. Check (i): $B((i,j),(i,j)) = (i,j)$. Check (ii): with $d = (1,1)$, $e = (2,1)$ we get $B(d,e) = (1,1) = d \neq e$, which needs $|I| \geq 2$. Check (iii): with $f = (1,1)$, $g = (1,2)$ we get $B(f,g) = (1,2) \neq f$, which needs $|J| \geq 2$. So $I = J = \{1,2\}$ works. Writing $p = (1,1)$, $q = (1,2)$, $r = (2,1)$, $s = (2,2)$, the table (row = first argument, column = second) is

B	p	q	r	s
p	p	q	p	q
q	p	q	p	q
r	r	s	r	s
s	r	s	r	s

with $d = p$, $e = r$, $f = p$, $g = q$.

Four is the minimum. Suppose B is a bypass with (i). Fix $o \in S$ and send $x \mapsto (B(x,o),B(o,x))$. Using (i) and the axiom, $B(B(x,o),B(o,x)) = B(x,x) = x$, so the map is injective, and $B(x,y)$ is sent to $(B(B(x,y),B(o,o)),B(B(o,o),B(x,y))) = (B(x,o),B(o,y))$. So S is a set of pairs $L \times R$ (every pair (l,r) is hit, by $B(l,r)$) with the operation of the model chain. If $|S| \leq 3$ then $|L| = 1$ or $|R| = 1$. If $|L| = 1$ then $B(x,y) = y$ always and (ii) fails; if $|R| = 1$ then $B(x,y) = x$ always and (iii) fails. That is also why the trial table above had to fail at $n = 2$ and would fail at $n = 3$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. In (b), both chains reach "I want to apply the axiom to $B(c,x)$". The natural attempt stalls because x is not a B-value; the model writes x's surrogate $B(x,x)$ and evaluates $B(c,B(x,x))$ twice. In (c), both chains open an empty table; the natural attempt fills cells, the model first asks what the axiom says the operation *remembers*.

> *The axiom only fires when both slots hold B-values. x is not one, but $B(x,x)$ is. If I compute $B(c,B(x,x))$ once through $c = B(a,b)$ and once through $c = B(c,c)$, I get $B(a,x)$ and $B(c,x)$.*

> *In $B(B(w,x),B(y,z)) = B(w,z)$ the inner letters x,y vanish and only the outer left w and outer right z survive. So B keeps "the left of the left and the right of the right". That is the projection $((i,j),(h,k)) \mapsto (i,k)$ on pairs. Build the table from that, not by trial.*

Layer 3 THE WHY

Why is (a) the setup for (b)? Because part (a) says every B-value c is idempotent, which is precisely the tool that lets you rewrite a bare c as a B-value $B(c,c)$. The axiom is an identity about B applied to two B-values; a problem that asks you to reason about bare letters is asking you to dress them up as B-values first. There are only two dressings available: $c = B(a,b)$ (given) and $c = B(c,c)$ (from (a)). For the other argument x the only dressing you have is $B(x,x)$, and you do not need to know whether $B(x,x) = x$ — you need only that it is a B-value. Once you see the axiom as "needs B-values in both slots", the substitutions in (b) are forced: there is nothing else to substitute.

Why pairs in (c)? Go back to what the axiom literally does to the four letters: w and z survive, x and y are deleted. So $B(w,x)$ must remember *something about* w (its "left part") and $B(y,z)$ must remember *something about* z (its "right part"); when they are combined, only the left part of the first and the right part of the second are kept. An operation whose output is built from one piece of each input is a coordinate projection, and coordinates mean a Cartesian product. The pair construction is not a trick; it is the axiom read aloud. Condition (i) says each element is determined by its two parts (nothing else is stored). Condition (ii), $B(d,e) = d \neq e$, says two elements can share a right part while differing in their left part, so $|I| \geq 2$; condition (iii) says the right part can change, so $|J| \geq 2$. Each condition is one coordinate being non-trivial.

The book's solution states the construction and adds "except for notation, every bypass is obtained this way". That remark needs (i): the constant map $B \equiv s_0$ on any set is a bypass and is not of the product form. With (i), the injection $x \mapsto (B(x,o),B(o,x))$ in the model chain is the proof. Algebraists call such an S a *rectangular band*: with (i), B is associative, since $B(B(x,y),z) = B(B(x,y),B(z,z)) = B(x,z)$ and $B(x,B(y,z)) = B(B(x,x),B(y,z)) = B(x,z)$, and it satisfies $xyz = xz$. Knowing the name is not needed; knowing that "inner letters vanish" means "outer coordinates survive" is.

The deeper lesson for (c): a construction problem attached to an identity is almost always solved by *modelling the identity*, not by searching tables. You found in the natural attempt that $n = 2$ is impossible; the structure explains why, and tells you the least n is 4 before you write a single cell.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that the axiom has B-values in both argument slots on the left. Write down: "to use the axiom I must present both arguments as $B(\cdot,\cdot)$".

2. For (a) and (b): substitution into the axiom, dressing bare letters as $B(x,x)$ or $B(c,c)$ and evaluating one expression two ways. For (c): model the identity with a concrete structure (coordinates on a product).

3. Hesitation in (b) occurs at $B(c,x)$ with x bare. It is reinforced away by recalling that (a) is there for a reason, and that $B(x,x)$ is always a legal B-value. Hesitation in (c) occurs at the blank table; ask what the operation keeps.

4. The trap is trial-filling the table: $n = 2$ provably fails, $n = 3$ provably fails, and a student can burn twenty minutes discovering that by hand.

5. In any identity where some variables disappear from one side, the survivors tell you what the operation remembers; build the model from the survivors. In any multi-part problem, part (a) is the lemma for part (b).

6. Reusable rule: for a functional or operational identity, (1) evaluate one cleverly chosen expression in two ways, using only the identity and earlier parts; (2) for a construction, read off which inputs survive the identity and realise the operation as a projection onto those coordinates, then check the conditions coordinate by coordinate.

- ▷ **Cue:** an identity whose left side needs B-values in every slot, applied to a bare element x $\longrightarrow$ **Move:** replace x by the B-value $B(x,x)$ (or any available B-value) and evaluate one expression in two ways
- ▷ **Cue:** part (a) proves idempotence of B-values $\longrightarrow$ **Move:** use it in (b) to rewrite c as $B(c,c)$
- ▷ **Cue:** an identity in which inner variables vanish and only the outermost survive $\longrightarrow$ **Move:** model elements as pairs and the operation as $((i,j),(h,k)) \mapsto (i,k)$
- ▷ **Cue:** "construct a table" for an operation satisfying an identity $\longrightarrow$ **Move:** find the structure the identity describes first; never fill cells by trial
- ▷ **Cue:** conditions of the form "some output differs from the left input" and "some output equals the left input but not the right" $\longrightarrow$ **Move:** each forces one coordinate of the product to have at least two values

1978 A5

1978 A5. Let $0 < x_i < \pi$ for $i = 1, 2, \ldots, n$ and set

$$x = \frac{x_1 + x_2 + \cdots + x_n}{n}.$$

Prove that

$$\prod_{i=1}^{n} \frac{\sin x_i}{x_i} \leq \left(\frac{\sin x}{x}\right)^n.$$

Caveat. The MAA OCR drops both displayed formulas of this problem (it reads only "Let $0 < x_i < \pi$ for $i = 1, 2, \ldots, n$ and set … Prove that"). The statement above is reconstructed from the book's own solution under A-5, which proves exactly $\prod(\sin x_i)/x_i \leq (\sin x/x)^n$ with x the mean, and agrees with the wording of this problem in *Putnam and Beyond* (Problem 442, credited to the 39th Putnam, 1978).

Fields: real analysis (inequalities, convexity) **Position:** A5 — late-paper by position, but mid-difficulty once the right function is chosen **Answer:** The function $g(t) = \ln \frac{\sin t}{t}$ is concave on $(0, \pi)$, since $g''(t) = \frac{1}{t^2} - \frac{1}{\sin^2 t} < 0$; Jensen gives $\sum g(x_i) \leq n\, g(x)$, and exponentiating gives the inequality, with equality iff all x_i are equal.

Layer 1 THE WHAT — surface analysis

The natural attempt. A product on the left, an n-th power of the "average" on the right: this looks like AM–GM, and you know two relevant facts. First, $\prod \sin x_i \leq \sin^n x$ — the product of sines is largest when the angles are equal (it follows from concavity of $\ln \sin$ on $(0, \pi)$, or from repeated use of $\sin a \sin b = \frac{1}{2}[\cos(a-b) - \cos(a+b)] \leq \sin^2 \frac{a+b}{2}$). Second, $\prod x_i \leq x^n$, by AM–GM. So you split the ratio:

$$\prod \frac{\sin x_i}{x_i} = \frac{\prod \sin x_i}{\prod x_i}.$$

The numerator is at most $\sin^n x$. The denominator is at most x^n — and a *smaller* denominator makes the fraction *larger*. So all you get is

$$\frac{\prod \sin x_i}{\prod x_i} \leq \frac{\sin^n x}{\prod x_i} \quad \text{and} \quad \frac{\sin^n x}{\prod x_i} \geq \frac{\sin^n x}{x^n},$$

which point in opposite directions and prove nothing. That is the wall: two correct inequalities, pulling against each other. The numerator gains from equalising and so does the denominator; the question is which gains more, and splitting cannot answer it.

A second natural attempt is smoothing: replace x_1, x_2 by their mean and show the product does not decrease. With $s = x_1 + x_2$ fixed you must show $\dfrac{\sin x_1 \sin x_2}{x_1 x_2} \le \dfrac{\sin^2(s/2)}{(s/2)^2}$. That is a genuine two-variable inequality in which the same numerator-versus-denominator competition reappears. It can be done, but only by essentially proving the concavity below in disguise.

The model chain. Both sides are positive, because $0 < x_i < \pi$ gives $\sin x_i > 0$, and x also lies in $(0, \pi)$. Take logarithms; the claim becomes

$$\sum_{i=1}^{n} g(x_i) \le n\, g(x), \qquad g(t) = \ln \frac{\sin t}{t} = \ln \sin t - \ln t \quad (0 < t < \pi).$$

Differentiate:

$$g'(t) = \cot t - \frac{1}{t}, \qquad g''(t) = -\frac{1}{\sin^2 t} + \frac{1}{t^2} = \frac{\sin^2 t - t^2}{t^2 \sin^2 t}.$$

For $t > 0$ we have $0 < \sin t < t$, so $\sin^2 t < t^2$ and $g''(t) < 0$ on $(0, \pi)$: g is strictly concave.

Jensen's inequality for a concave function now finishes. To keep it self-contained: since $g'' < 0$, g' is strictly decreasing, so by the mean value theorem the graph lies below each tangent line,

$$g(t) \le g(x) + g'(x)(t - x) \qquad (0 < t < \pi),$$

with equality only at $t = x$. (For $t > x$, $g(t) - g(x) = g'(\xi)(t - x)$ with $\xi \in (x, t)$ and $g'(\xi) < g'(x)$; for $t < x$ similarly.) Put $t = x_i$ and add over i:

$$\sum g(x_i) \le n\, g(x) + g'(x) \sum (x_i - x) = n\, g(x),$$

since $\sum (x_i - x) = 0$ by the definition of x. Exponentiating (the exponential is increasing) gives

$$\prod_{i=1}^{n} \frac{\sin x_i}{x_i} = e^{\sum g(x_i)} \le e^{n g(x)} = \left(\frac{\sin x}{x}\right)^n,$$

with equality iff every $x_i = x$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start by recognising "product $\le$ (value at the mean)^n" as a Jensen/AM–GM shape. At the next line the natural attempt splits $\dfrac{\prod \sin x_i}{\prod x_i}$ into numerator and denominator; the model keeps $\dfrac{\sin t}{t}$ as one function and takes its logarithm.

> *The numerator and the denominator both get bigger when I equalise, so bounding them separately is a tug-of-war I cannot referee. The problem gives me the ratio $\frac{\sin t}{t}$ as a single object, so I treat it as one function $f(t) = \frac{\sin t}{t}$ and ask whether $\ln f$ is concave.*

> *"Product of $f(x_i)$ against $f(\text{mean})^n$" is Jensen for $\ln f$. The only thing to check is the sign of $(\ln f)''$, and $\ln$ turns the quotient into a difference, so the second derivative is two easy pieces: $-\csc^2 t + t^{-2}$.*

Layer 3 THE WHY

Why the logarithm? The inequality compares a product of values $\prod f(x_i)$ with the n-th power of the value at the mean. Jensen's inequality is a statement about *sums* of values against

n times the value at the mean. The logarithm is exactly the dictionary between the two: it turns $\prod$ into $\sum$ and $(\cdot)^n$ into $n\cdot$. So the problem is "Jensen for $\ln f$" before you have done anything. AM–GM itself is the special case $f(t) = t$: $\prod x_i \le x^n$ is concavity of $\ln t$. Seeing this problem as "AM–GM with t replaced by $\frac{\sin t}{t}$" tells you that the proof of AM–GM (concavity of ln) is the proof to imitate — with $\ln \frac{\sin t}{t}$ in place of $\ln t$.

Why must the ratio stay together? At axiom level: $\ln \frac{\sin t}{t} = \ln \sin t - \ln t$ is a *difference* of two concave functions, and a difference of concave functions has no sign of concavity in general. The natural attempt proves that each piece is concave separately and then has nothing, because concavity does not survive subtraction. The only way forward is to compute the second derivative of the difference itself and see whether the concavity of $\ln \sin$ ($-\csc^2 t$) beats the concavity of $\ln t$ ($-t^{-2}$, entering with a minus sign as $+t^{-2}$). It does, and the reason is the most basic inequality of trigonometric calculus, $\sin t < t$. That is the whole content of the problem: $|\sin t| < |t|$, squared and reciprocated, says the curvature of $\ln \sin$ is stronger than the curvature of $\ln t$.

Why the hypothesis $0 < x_i < \pi$? It makes every $\sin x_i$ positive so that logarithms exist, and it places everything in an interval where g is defined and concave. (For x_i allowed up to 2π the left side can be negative and the statement becomes a different question.)

The book's solution is the model chain in four lines: "Let $g(x) = \ln[(\sin x)/x]$. Then $g''(x) = -\csc^2 x + 1/x^2 < 0$ since $x > \sin x$. Thus the graph of g is concave down and hence $\frac{1}{n}\sum g(x_i) \le g(\bar{x})$." Here is what it leaves out: why anyone would take ln of the ratio (the product-to-sum dictionary above), why the ratio must not be split (concavity is not preserved by differences), and the tangent-line proof of Jensen that turns "concave down" into the averaged inequality.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "$\prod f(x_i) \le f(\text{mean})^n$". Write down $\sum \ln f(x_i) \le n \ln f(x)$ and ask whether $\ln f$ is concave.

2. Jensen's inequality (equivalently, the tangent-line trick at the mean) for $g = \ln f$.

3. Hesitation occurs after splitting into $\prod \sin x_i$ and $\prod x_i$ and finding the two bounds point opposite ways. Reinforce: if the problem hands you a ratio as a unit, analyse it as a unit.

4. The trap is that both $\prod \sin x_i \le \sin^n x$ and $\prod x_i \le x^n$ are true and familiar; they invite a split that cannot close, because a smaller denominator enlarges the fraction.

5. Any inequality of the form "product of values against the value at the mean to the power n" is Jensen on a logarithm. AM–GM is the case $f(t) = t$.

6. Reusable rule: to compare $\prod f(x_i)$ with $f(\bar{x})^n$, take logs and test the sign of $(\ln f)''$ on the interval; never bound the factors of f separately when they move in the same direction.

▷ **Cue:** $\prod f(x_i)$ compared with $f\left(\frac{\sum x_i}{n}\right)^n$ $\longrightarrow$ **Move:** take logarithms and apply Jensen to $\ln f$

▷ **Cue:** a ratio whose numerator and denominator both increase under equalising $\longrightarrow$ **Move:** do not split it; treat the ratio as one function

▷ **Cue:** $\ln \frac{\sin t}{t}$ or any $\ln \frac{u(t)}{v(t)}$ to be tested for concavity $\longrightarrow$ **Move:** write it as $\ln u - \ln v$ and differentiate twice; the sign reduces to comparing $\frac{u'^2 - uu''}{u^2}$ with $\frac{v'^2 - vv''}{v^2}$

▷ **Cue:** a second derivative $\frac{1}{t^2} - \frac{1}{\sin^2 t}$ $\longrightarrow$ **Move:** sign from $\sin t < t$ for $t > 0$

▷ **Cue:** Jensen needed in an exam where citing it feels risky $\longrightarrow$ **Move:** prove it in two lines: tangent line at the mean, sum, the linear term vanishes

1978 A6

1978 A6. Let n distinct points in the plane be given. Prove that fewer than $2n^{3/2}$ pairs of them are unit distance apart.

Fields: combinatorics (extremal graph theory, double counting), plane geometry **Position:** A6 — last problem of the morning paper; hard because the geometry must be reduced to a single combinatorial fact before any counting can start **Answer:** Two distinct unit circles meet in at most two points, so no two points have three common unit-distance neighbours; double counting gives $\sum_i \binom{e_i}{2} \le n(n-1)$ where e_i is the number of points at unit distance from p_i, and Cauchy–Schwarz then gives at most $\frac{n}{4}\left(1+\sqrt{8n-7}\right) < 2n^{3/2}$ unit pairs.

Layer 1 THE WHAT — surface analysis

The natural attempt. You start geometrically. Draw points, draw unit segments, and try to build configurations with many unit pairs: an equilateral triangle gives 3 pairs from 3 points, two triangles sharing an edge (a rhombus) give 5 from 4, a triangular lattice patch gives about $3n$ pairs, and the grid $\{0,\dots,k\}^2$ gives about $2n$. So you suspect the truth is linear and try to prove something like "each point has at most 6 unit neighbours" — which is false: all points on the unit circle about p are unit neighbours of p, so one point can have $n-1$ of them.

So you switch to a per-point bound and try to argue by induction: remove a point of smallest degree k and bound k. You need to show that a configuration in which every point has at least k unit neighbours must have many points. Take x_0 with neighbours $x_1,\dots,x_k$ on the unit circle C_0 about x_0. Each x_i has at least k neighbours of its own; you try to show many of these are new points. This is the book's alternate solution, and it works — but it needs a careful accounting of which neighbours can coincide, then a recursion $f(n) \le f(n-1) + k_n$ with $k_n \approx \sqrt{2n}$, then an integral comparison to sum the recursion. It is long, and every step of it is powered by one geometric fact that the attempt uses without isolating: two different unit circles share at most two points.

The model chain. Let the points be $p_1,\dots,p_n$, let e_i be the number of p_j with $|p_ip_j| = 1$, and let E be the number of unit pairs. Each unit pair is counted once from each end, so

$$\sum_{i=1}^{n} e_i = 2E. \tag{1}$$

The geometric fact. If $j \ne k$, the unit circles centred at p_j and p_k are distinct (different centres), so they meet in at most two points. Hence at most two of the p_i are at unit distance from both p_j and p_k.

Double counting. Count the configurations $\left(p_i, \{p_j, p_k\}\right)$ in which $p_j \ne p_k$ and both are at unit distance from p_i. Counting by the apex p_i: p_i has e_i unit neighbours, which form $\binom{e_i}{2}$ pairs. Counting by the base $\{p_j, p_k\}$: there are $\binom{n}{2}$ bases and each has at most 2 apexes by the geometric fact. So

$$\sum_{i=1}^{n} \binom{e_i}{2} \le 2\binom{n}{2} = n(n-1), \qquad \text{i.e.} \qquad \sum e_i^2 \le 2n(n-1) + \sum e_i = 2n(n-1) + 2E. \tag{2}$$

Cauchy–Schwarz. $\left(\sum e_i\right)^2 \le n \sum e_i^2$. With (1) and (2),

$$4E^2 \le n\big(2n(n-1) + 2E\big), \qquad \text{so} \qquad 2E^2 - nE - n^2(n-1) \le 0.$$

This quadratic in E has positive root $\frac{n}{4}\left(1+\sqrt{8n-7}\right)$ (discriminant $n^2 + 8n^2(n-1) = n^2(8n-7)$), so

$$E \le \frac{n}{4}\left(1+\sqrt{8n-7}\right) < \frac{n}{4}\left(\sqrt{n} + \sqrt{8}\,\sqrt{n}\right) = \frac{1+2\sqrt{2}}{4}\, n^{3/2} < n^{3/2} < 2n^{3/2},$$

using $1 \le \sqrt{n}$ and $\sqrt{8n-7} < \sqrt{8n}$ for $n \ge 1$. (Check at small n: $n=3$ gives $E \le \frac{3}{4}(1+\sqrt{17}) \approx 3.84$, and the equilateral triangle has 3; $n=4$ gives $E \le 1+\sqrt{25} = 6$, and the rhombus has 5, which is the true maximum.)

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains draw unit circles around the points. The natural attempt then follows one point and its neighbours; the model steps back and asks what *any two* circles can share, isolating "two centres have at most two common unit neighbours" as the one fact, and counts cherries $(p_i, \{p_j, p_k\})$ globally.

> *The answer has $n^{3/2}$ in it, not n and not n^2. A $3/2$ exponent comes from a sum of squares of degrees controlled by n^2, then Cauchy–Schwarz. So I need a bound on $\sum e_i^2$, and $\sum \binom{e_i}{2}$ counts pairs of neighbours, which is a question about what two points can have in common.*

> *What can two points p_j, p_k have in common? A common unit neighbour lies on both unit circles about them, and two different circles cross at most twice. So each pair is "shared" at most twice. That is the only geometry in the problem; the rest is counting.*

Layer 3 THE WHY

Why the exponent points to the method. The problem does not ask for the true maximum; it asks you to beat $2n^{3/2}$. An exponent $3/2$ is the signature of one argument in combinatorics: bound a quadratic quantity in the degrees ($\sum e_i^2$ or $\sum \binom{e_i}{2}$) by $O(n^2)$, then pass from $\sum e_i^2$ to $\sum e_i$ by Cauchy–Schwarz, which costs a factor $\sqrt{n}$: $\sum e_i \le \sqrt{n \sum e_i^2} \le \sqrt{n \cdot O(n^2)} = O(n^{3/2})$. Reading the target this way tells you what to look for before you look at the geometry: something that makes $\sum \binom{e_i}{2} \le Cn^2$.

Why $\binom{e_i}{2}$ is the right quantity. $\binom{e_i}{2}$ counts pairs of neighbours of p_i — "cherries" with apex p_i. Turning the count round, each cherry is a pair $\{p_j, p_k\}$ plus a common neighbour. There are only $\binom{n}{2}$ pairs, so if each pair has at most a bounded number of common neighbours, the total number of cherries is $O(n^2)$. In graph language: the unit-distance graph contains no $K_{2,3}$ (two vertices joined to the same three), and the argument above is the standard Kővári–Sós–Turán count for $K_{2,t}$-free graphs.

Why the geometry is "at most two". Drive it to the definition: a point at unit distance from p_j lies on the circle of radius 1 centred at p_j — that is what a circle *is*. A common unit neighbour of p_j and p_k is therefore a common point of two circles with different centres. Two distinct circles meet in at most two points (subtracting their equations leaves a line, which meets a circle at most twice). So the centres are the objects of interest, just as in the Blueprint's circle example, and "draw the unit circles" is not a trick but a restatement of "unit distance".

What the book does. The MAA solution is the same double count, organised slightly differently: it notes the n unit circles meet in at most $2\binom{n}{2}$ points, bounds $\sum \binom{e_i}{2} \le n(n-1)$, reduces to the case $e_i \ge 1$, and applies Cauchy–Schwarz to $\sum (e_i - 1)$. The reduction is needed there because $(e_i - 1)^2 \le e_i(e_i - 1)$ fails when $e_i = 0$; deleting isolated points lowers n without changing E, and the bound is increasing in n. The version above applies Cauchy–Schwarz to $\sum e_i$ and solves a quadratic, which avoids the reduction and gives the sharper constant $\frac{1+2\sqrt{2}}{4} \approx 0.96$. The book's alternate solution is the natural attempt carried to completion.

Where the truth lies. The bound $n^{3/2}$ is far from sharp. Spencer, Szemerédi and Trotter (1984) proved the number of unit distances is $O(n^{4/3})$, and Erdős's lattice construction gives $n^{1+c/\log\log n}$; closing that gap is a famous open problem. The Putnam asks only for what the "no $K_{2,3}$" count yields, and that is the right expectation to set: an A6 that names $n^{3/2}$ is telling you which tool it wants.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the bound $2n^{3/2}$. Write down: "count $\sum \binom{e_i}{2}$, bound it by $O(n^2)$, then Cauchy–Schwarz".

2. Double counting of cherries (a point together with two of its neighbours), followed by Cauchy–Schwarz on the degree sequence.

3. Hesitation occurs when you realise one point can have $n-1$ unit neighbours, so no per-point bound exists. Reinforce: per-point degree is unbounded, but per-*pair* common degree is at most 2; bound pairs, not points.

4. The trap is constructing examples and trying to prove the linear bound the examples suggest; it is false (Erdős's lattice construction gives $n^{1+c/\log\log n}$ unit pairs, which is superlinear), and it sends you hunting for a local degree bound that does not exist.

5. An exponent $3/2$ in a bound on the number of edges means a forbidden $K_{2,t}$ and Cauchy–Schwarz. In geometry, "two circles meet at most twice" and "two lines meet at most once" are the usual sources of the forbidden configuration.

6. Reusable rule: to bound incidences or edges, find a small configuration that cannot occur (here, two points with three common neighbours), double count the pairs-with-a-common-neighbour, and convert the quadratic bound into a linear one with Cauchy–Schwarz.

▷ **Cue:** a bound of the form $Cn^{3/2}$ on a number of pairs $\longrightarrow$ **Move:** bound $\sum \binom{e_i}{2}$ by $O(n^2)$, then apply Cauchy–Schwarz to $\sum e_i$

▷ **Cue:** points at a fixed distance r from a point $\longrightarrow$ **Move:** they lie on the circle of radius r about it; draw the circles

▷ **Cue:** two distinct circles $\longrightarrow$ **Move:** at most two common points; so two centres have at most two common neighbours at that distance

▷ **Cue:** no per-point degree bound is available $\longrightarrow$ **Move:** bound the number of common neighbours of each pair instead and double count cherries

▷ **Cue:** $\sum e_i^2$ bounded, $\sum e_i$ wanted $\longrightarrow$ **Move:** $(\sum e_i)^2 \le n \sum e_i^2$, and if $\sum e_i^2$ involves $\sum e_i$, solve the resulting quadratic

Putnam 1978 — Paper B

1978 B1

1978 B1. Find the area of a convex octagon that is inscribed in a circle and has four consecutive sides of length 3 units and the remaining four sides of length 2 units. Give the answer in the form $r + s\sqrt{t}$ with r, s, and t positive integers.

Fields: geometry (cyclic polygons, central angles) **Position:** B1 — opening problem of the second session **Answer:** $13 + 12\sqrt{2}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. Eight sides, eight unknown central angles, one circle: you set up coordinates. Place the centre at the origin, let r be the radius, and let the eight vertices $P_0, \ldots, P_7$ go around in order, with $P_0P_1, \ldots, P_3P_4$ the four sides of length 3 and $P_4P_5, \ldots, P_7P_0$ the four of length 2, exactly as given. Each side of length 3 subtends a central angle 2α with $2r\sin\alpha = 3$, each side of length 2 subtends 2β with $2r\sin\beta = 2$, and going all the way around forces $4(2\alpha) + 4(2\beta) = 2\pi$, i.e. $4\alpha + 4\beta = \pi$. That much is clean and matches the model. The trouble starts when you try to get the actual *area* out of this arrangement while respecting the stated order: the octagon $P_0 \ldots P_7$ with four long sides bunched together and four short sides bunched together has only a single mirror symmetry (through the midpoints of the two "seams" where a 3-side meets a 2-side), not the four-fold symmetry of an alternating pattern. Summing the eight vertex coordinates or shoelacing the polygon directly means carrying eight different vertex angles $0, 2\alpha, 4\alpha, 6\alpha, 8\alpha, 8\alpha + 2\beta, \ldots$ through sines and cosines with no simplification available, and you are still stuck first solving the transcendental-looking system $4\alpha + 4\beta = \pi$, $r\sin\alpha = 3/2$, $r\sin\beta = 1$ for r before any of that arithmetic can even start. The problem is doable this way but it is long, symmetry-starved, and easy to arithmetic-error your way out of a clean closed form.

The model chain. The area of a polygon inscribed in a circle of radius r is the sum of the areas of the isosceles triangles cut off by the centre, and each such triangle's area, $\frac{1}{2}r^2\sin\theta_i$ where θ_i is that side's central angle, depends only on that side's *length* and on r – not on where the side sits in the cyclic order. So the total area $\frac{1}{2}r^2\sum_i \sin\theta_i$, and the constraint pinning down r (namely $\sum_i \theta_i = 2\pi$), both depend only on the *multiset* of side lengths, four 3's and four 2's, never on their arrangement. This licenses replacing the given (awkward) order with the friendliest one: alternating $3, 2, 3, 2, 3, 2, 3, 2$.

An alternating octagon inscribed in a circle has full 4-fold symmetry, and in fact it embeds cleanly inside a square. All its interior angles are equal (by the alternating symmetry each vertex looks the same up to rotation), and since the angles of a convex cyclic octagon sum to 6π, each interior angle is $\frac{6\pi}{8} = \frac{3\pi}{4}$. At each vertex where a side of length 2 ends, the exterior turn is $\pi - \frac{3\pi}{4} = \frac{\pi}{4}$, exactly the turning angle of the leg of a right isosceles triangle with legs ℓ and hypotenuse $\ell\sqrt{2}$; placing such a triangle (legs x, hypotenuse 2, so $x = \sqrt{2}$) on each of the four sides of length 2, with the right-angle vertex pointing outward, completes the octagon into a square. The square's side is $3 + 2\sqrt{2}$ (a side-3 edge of the octagon, plus a leg $\sqrt{2}$ of each adjacent right-isosceles triangle on either end).

Hence

$$\text{Area of octagon} = (3+2\sqrt{2})^2 - 4\cdot\frac{(\sqrt{2})^2}{2} = (9+12\sqrt{2}+8) - 4 = 13+12\sqrt{2}.$$

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree on the governing relations: central angle θ for a chord of length ℓ satisfies $\ell = 2r\sin(\theta/2)$, and the octagon's area is $\frac{1}{2}r^2\sum\sin\theta_i$ with $\sum\theta_i = 2\pi$. They split at the decision of *which cyclic arrangement of the eight chords to actually draw and compute with*. The natural attempt keeps the order the problem states – four 3's then four 2's – because that is the shape that was asked for. The model discards that order and substitutes the alternating arrangement of the same eight chord lengths, because the two governing relations above never referenced the order at all.

> *The area formula is a sum over sides, and a sum doesn't know what order its terms were listed in – $\frac{1}{2}r^2\sum_i\sin\theta_i$ and the constraint $\sum_i\theta_i = 2\pi$ are both symmetric functions of the multiset $\{\theta_i\}$, so if I can find* any *valid convex cyclic arrangement of four chords of length 3 and four of length 2, its area is automatically the area of every other arrangement of the same multiset, including the one actually named in the problem. I should therefore build whichever arrangement is easiest to compute, not the one that was drawn.*

Layer 3 THE WHY

Why is it legitimate to swap the arrangement at all – doesn't the problem specify a particular octagon? It specifies a particular *combinatorial* octagon (this side, then this side, ...), but the object being computed, the area, factors entirely through two symmetric functions of the central angles: their sum (which pins r) and the sum of their sines (which gives the area). Central angle is itself a function purely of chord length and r, by the isosceles-triangle relation $\ell = 2r\sin(\theta/2)$ – it carries no memory of which neighbour a side has. So the map from (ordered list of eight chord lengths) to (area) factors through (multiset of eight chord lengths), and reordering is a symmetry of the problem, not a change of problem, in exactly the way that $2+3+5$ and $5+2+3$ are the same computation in a different order. This is the general fact underneath *every* cyclic-polygon area problem stated with a specific but otherwise-unconstrained vertex order: if the area functional you are about to compute only reads off (length, radius) pairs and sums, permuting the sides is free.

Why does the alternating arrangement specifically embed into a square, rather than some other convenient shape? Because alternation is what buys the four-fold rotational symmetry that forces all eight interior angles to be equal, and a convex cyclic octagon has interior angles summing to $(8-2)\pi = 6\pi$, so equal angles pins each one at $\frac{3\pi}{4}$ – exactly the angle you get by cutting the corner off a square with a right-isosceles triangle. The corner-cutting is not an ad hoc image; it is what an interior angle of $\frac{3\pi}{4}$, sitting between two of the octagon's straight sides, always looks like when the two sides it's not cutting extend to meet at a right angle outside the octagon – and a right isosceles triangle is the cheapest possible right-angled shape to compute with, since its area is half the square of its leg.

Layer $+\alpha$ BACKTESTING

1. The cue is a cyclic polygon inscribed in a circle, area asked for, with a specific but seemingly arbitrary order of given side lengths. On seeing it, write down: area and radius depend only on the multiset of side lengths, via $\ell = 2r\sin(\theta/2)$ and $\sum\theta_i = 2\pi$, Area $= \frac{1}{2}r^2\sum\sin\theta_i$ – both symmetric in the θ_i.
2. That cue suggests re-arranging the same multiset of chords into the most symmetric cyclic order before computing anything, rather than working with the order as literally stated.
3. Hesitation occurs right at the reordering step – it feels like solving a different problem. It is reinforced away by checking explicitly that both quantities the answer depends on (the

radius equation and the area sum) are manifestly symmetric functions of the chord multiset, so any valid arrangement gives the same numbers.

4. The trap is coordinate- or angle-bashing the octagon in the exact order given, which has only one axis of symmetry and forces you to carry eight distinct angles through the whole computation instead of one.

5. Recognise instantly: any time a cyclic-polygon problem states an order for the side lengths that isn't needed to determine the area (i.e. convexity and inscribed-ness are all that's used), that order is a red herring and can be replaced by the most symmetric one.

6. Reusable rule: for a convex polygon inscribed in a circle, area and circumradius are determined by the multiset of side lengths alone, never by their cyclic order; when a problem hands you an inconvenient order, substitute the most symmetric arrangement of the same multiset and solve that instead.

▷ **Cue:** a cyclic polygon's area is asked for and a specific (inconvenient) order of side lengths is given $\longrightarrow$ **Move:** recognise that area and radius depend only on the multiset of side lengths, not the order; substitute the most symmetric arrangement of the same lengths

▷ **Cue:** chord of length ℓ in a circle of radius r $\longrightarrow$ **Move:** central angle θ satisfies $\ell = 2r\sin(\theta/2)$; triangle it cuts off from the centre has area $\frac{1}{2}r^2\sin\theta$

▷ **Cue:** a convex cyclic polygon with an alternating pattern of two side lengths $\longrightarrow$ **Move:** use the resulting equal-interior-angle symmetry to embed it in a regular figure (here a square, by attaching right-isosceles corner triangles to the shorter sides)

1978 B2

1978 B2. Express

$$\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\frac{1}{mn(m+n+2)}$$

as a rational number.

Fields: analysis (double series, integral representations) **Position:** B2 **Answer:** $\frac{7}{4}$.

Caveat. The MAA book OCR renders the denominator as "mln+mn2+2mn". Reading this as $m^2n+mn^2+2mn = mn(m+n+2)$ makes the source's own solution – which manipulates $\frac{1}{n(n+2)}$ and partial-fraction pairs $\left(\frac{1}{m}-\frac{1}{m+n+2}\right)$ and arrives at the value $7/4$ – fully consistent line by line, and $7/4$ is independently confirmed below by a second, unrelated method. This is taken as the correctly reconstructed statement.

Layer 1 THE WHAT — surface analysis

The natural attempt. The obstruction is immediate: the denominator $mn(m+n+2)$ couples m and n through their sum, so the sum does not split into a product of two one-variable series. The standard first move is partial fractions in one variable at a time. Fix n and decompose in m:

$$\frac{1}{m(m+n+2)} = \frac{1}{n+2}\left(\frac{1}{m}-\frac{1}{m+n+2}\right),$$

so

$$\sum_{m=1}^{\infty}\frac{1}{mn(m+n+2)} = \frac{1}{n(n+2)}\sum_{m=1}^{\infty}\left(\frac{1}{m}-\frac{1}{m+n+2}\right) = \frac{H_{n+2}}{n(n+2)},$$

where the sum telescopes to the finite harmonic number $H_{n+2} = 1+\frac{1}{2}+\cdots+\frac{1}{n+2}$ (the first $n+2$ terms of $1/m$ survive; everything past that cancels against the shifted series). This is

correct and it is exactly what the official solution does. But now you are staring at

$$S = \sum_{n=1}^{\infty} \frac{H_{n+2}}{n(n+2)},$$

a sum of a harmonic number against a second rational function of n. Partial-fractioning $\frac{1}{n(n+2)} = \frac{1}{2}\left(\frac{1}{n} - \frac{1}{n+2}\right)$ gives another telescoping shape, but H_{n+2} itself changes with n, so the terms do not cancel outright – you need a second, more delicate telescoping argument (regrouping a double limit, as the source solution does over several lines) to extract the finite value $7/4$. It works, but it costs a second nontrivial telescoping argument stacked on the first.

The model chain. Replace the coupling $m + n + 2$ in the denominator with an integral, using $\dfrac{1}{k+1} = \displaystyle\int_0^1 x^k\,dx$ for integer $k \geq 0$: with $k = m + n + 1$,

$$\frac{1}{m+n+2} = \int_0^1 x^{m+n+1}\,dx.$$

Substituting and exchanging the (nonnegative-term) sum and integral,

$$S = \sum_{m,n\geq 1} \frac{1}{mn} \int_0^1 x^{m+n+1}\,dx = \int_0^1 x \left(\sum_{m=1}^{\infty} \frac{x^m}{m}\right)\left(\sum_{n=1}^{\infty} \frac{x^n}{n}\right) dx = \int_0^1 x\left(-\ln(1-x)\right)^2 dx,$$

using $\sum_{m\geq 1} x^m/m = -\ln(1-x)$ for $0 \leq x < 1$, and the sum-integral exchange is justified by Tonelli's theorem since every term is nonnegative.

Substitute $u = 1 - x$, $du = -dx$:

$$S = \int_0^1 (1-u)\,(\ln u)^2\,du = \int_0^1 (\ln u)^2\,du - \int_0^1 u(\ln u)^2\,du.$$

Both are standard: $\int_0^1 x^a(\ln x)^2\,dx = \dfrac{2}{(a+1)^3}$ for $a > -1$ (differentiate $\int_0^1 x^a\,dx = \frac{1}{a+1}$ twice with respect to a). With $a = 0$: $\int_0^1 (\ln u)^2\,du = 2$. With $a = 1$: $\int_0^1 u(\ln u)^2\,du = \frac{2}{8} = \frac{1}{4}$. Hence

$$S = 2 - \frac{1}{4} = \frac{7}{4}.$$

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree on the very first move: the denominator's real obstruction is that $m + n + 2$ couples the two indices additively, and the standard partial-fractions tool decouples an additive expression only one variable at a time. The natural attempt applies partial fractions in m (the only tool that decouples an additive coupling directly), which succeeds locally but leaves the harmonic number H_{n+2} – itself an n-dependent object built from the first decoupling – entangled in the outer sum over n, forcing a second telescoping argument. The model instead abandons partial fractions on the coupling entirely and converts $\frac{1}{m+n+2}$ into an integral $\int_0^1 x^{m+n+1}dx$ *before* touching m or n individually.

> *The coupling in the denominator is $m + n$, a* sum *in the exponent-like position, but $x^{m+n} = x^m \cdot x^n$ is a* product *– so if I can turn $\frac{1}{m+n+2}$ into something that carries $m + n$ as a power of a dummy variable rather than as a sum, the double sum will separate into a product of two independent single sums, and each of those is just $-\ln(1-x)$, a function I already know how to integrate against x afterward.*

Layer 3 THE WHY

Why does the integral trick succeed where partial fractions in m alone leaves residue work behind? Partial fractions on $\frac{1}{m(m+n+2)}$ decouples m from the *product* structure of the denominator, but $m+n+2$ is still an additively shifted function of m for each fixed n – the decoupling produces H_{n+2}, an object whose definition (a partial sum up to $n+2$) still carries n inside it, so the entanglement has moved, not vanished. The representation $\frac{1}{k+1} = \int_0^1 x^k\,dx$ is different in kind: it does not decouple an additive expression, it *converts* the additive coupling $m+n$ into the exponent of a single variable x, and exponents of a common base turn addition into multiplication, $x^{m+n} = x^m x^n$. Multiplication is exactly the operation a double sum $\sum_m \sum_n$ already factors through when the summand itself is multiplicative in m and n (here $\frac{1}{m}\cdot\frac{1}{n}\cdot x^{m+n}$ is manifestly of the form $f(m)g(n)$ for fixed x), so the double sum collapses into a product $\left(\sum_m f(m)\right)\left(\sum_n g(n)\right)$ with no residual object connecting m and n at all. This is the reason the integral route needs only one telescoping-free evaluation ($\sum x^m/m = -\ln(1-x)$, itself just the Taylor series of a logarithm) followed by one elementary integral, while the partial-fractions route needs two rounds of telescoping because the additive coupling keeps reappearing in a new guise each time it is only partially resolved.

Why is $1/(k+1) = \int_0^1 x^k dx$ the right integral to reach for, rather than some other representation? Because it is the unique elementary way to turn a bare reciprocal of a linear expression in the summation index into a power of a single auxiliary variable – it is the base case of the Beta function, $B(p,q) = \int_0^1 x^{p-1}(1-x)^{q-1}dx$, restricted to $q=1$ – and any time a sum's index appears *only* in the exponent-like position $k+1$ in a denominator, this is the standard device for exchanging "index in an exponent" for "index as an exponent of a fixed dummy variable", which is precisely what makes a doubly-indexed additive coupling become a singly-indexed multiplicative one.

Layer $+\alpha$ BACKTESTING

1. The cue is a double sum whose summand's denominator contains an additive coupling of the two indices, here $m+n+2$. On seeing it, write down: $\dfrac{1}{m+n+2} = \displaystyle\int_0^1 x^{m+n+1}\,dx.$

2. That cue suggests converting the double sum into an integral of a product of two independent single-variable power series, rather than partial-fractioning one index at a time.

3. Hesitation occurs at trusting the interchange of $\sum$ and $\int$; it is reinforced away by noting every term is nonnegative, so Tonelli's theorem licenses the swap without further justification.

4. The trap is partial-fractioning in one variable only, which decouples that variable from the product structure but reintroduces the other variable's dependence through a new object (here H_{n+2}), doubling the work rather than finishing it.

5. Recognise instantly: any denominator that is linear in a *sum* of two summation indices is a candidate for the integral representation, because it converts additive coupling into multiplicative separability via $x^{m+n} = x^m x^n$.

6. Reusable rule: when a double sum's denominator couples its indices additively as $(m+n+c)$, rewrite $\frac{1}{m+n+c}$ as $\int_0^1 x^{m+n+c-1}dx$ before doing anything else; this separates the sum into a product of two independent series (each often summable in closed form, frequently as a logarithm or a power), and finishes with one elementary integral instead of iterated telescoping.

▷ **Cue:** a double sum's denominator is linear in the sum of the two indices, $m+n+c$ $\longrightarrow$ **Move:** rewrite $\frac{1}{m+n+c} = \int_0^1 x^{m+n+c-1}dx$ to convert additive index-coupling into a multiplicative $x^m \cdot x^n$, separating the double sum into a product of two single sums under an integral

▷ **Cue:** a series of the form $\sum_{m\geq 1} x^m/m$ appears $\longrightarrow$ **Move:** recognise it as $-\ln(1-x)$, the Taylor series of the logarithm

▷ **Cue:** need $\int_0^1 x^a(\ln x)^2 dx$ $\longrightarrow$ **Move:** differentiate $\int_0^1 x^a dx = \frac{1}{a+1}$ twice with respect to a to get $\frac{2}{(a+1)^3}$, rather than integrating by parts twice from scratch

1978 B3

1978 B3. The sequence $\{Q_n(x)\}$ of polynomials is defined by $Q_1(x)$ and $Q_2(x)$ as base cases, and, for $m \geq 1$, by

$$Q_{2m+1}(x) = Q_{2m}(x) + (m+1)xQ_{2m-1}(x), \qquad Q_{2m+2}(x) = Q_{2m+1}(x) + (m+1)xQ_{2m}(x).$$

Let x_n be the largest real solution of $Q_n(x) = 0$. Prove that $\{x_n\}$ is an increasing sequence and that $\lim_{n\to\infty} x_n = 0$.

Fields: analysis/algebra (recursively defined polynomials, real roots, sign arguments) **Position:** B3
Answer: $\{x_n\}$ is increasing and $x_n \to 0$ (every $x_n < 0$; no closed form is needed or found).

Caveat. The MAA book OCR drops the explicit base-case definitions of $Q_1(x)$ and $Q_2(x)$ entirely – the problem text reads "…is defined by" followed immediately by "and, for $m > 1$, by" with a gap where the initial polynomials must have stood as a displayed formula the OCR did not capture at all (not a garbled string, an outright omission). No reconstruction of Q_1, Q_2 was found in *Putnam and Beyond* or elsewhere in the project's sources. The two recursion formulas themselves, however, are intact and are confirmed correct: substituting the test point $a = -1/(m+1)$ into them independently reproduces the source solution's stated identity $Q_{2m+2}(a) = -Q_{2m-1}(a)$ exactly, which is strong evidence this is the right recursion for the right problem. The source solution also asserts, without displaying the polynomials, that "Clearly, $x_1 = -1$, $x_2 = -\frac{1}{2}$" (this fraction is itself a best-effort OCR reading) as the base case for the monotonicity induction; this claim is reproduced below as an unverified starting fact rather than re-derived, since Q_1, Q_2 are not recoverable from any available source. Every other step of the analysis below – positivity for $x \geq 0$, the monotonicity induction, and the limit argument – is carried out independently from the recursion alone and does not depend on the specific base case.

Layer 1 THE WHAT — surface analysis

The natural attempt. A recursively defined sequence of polynomials with a question about the location of its roots invites the most direct tool: find a closed form. You try small cases (using whatever Q_1, Q_2 turn out to be) and look for a pattern – perhaps a product formula, perhaps something Chebyshev-flavoured, since $(m+1)x$ multiplying the two-back term looks structurally like a three-term recursion for orthogonal polynomials. But the coefficient $(m+1)$ is not constant; it grows with the step, so unlike a genuine Chebyshev or continuant recursion (constant coefficients, or coefficients depending only on parity), the generating function or explicit sum you'd normally write down does not close up cleanly – the growing coefficient means each Q_n's coefficients involve a product of increasing integers with no obvious closed form, and even if you push through and get an explicit polynomial for Q_3, Q_4, Q_5, extracting general monotonicity of the *roots* from those explicit (and increasingly unwieldy) polynomials directly is not something a formula alone hands you – you'd still need calculus or a separate argument per polynomial. You have spent real effort and are no closer to a general proof for all n.

The model chain. Abandon the search for an explicit formula. The two things to be proved – $\{x_n\}$ increasing, $x_n \to 0$ – are both statements about *where the roots sit relative to specific numbers*, and the Intermediate Value Theorem lets you answer exactly that kind of question from sign information alone, without ever solving $Q_n(x) = 0$.

Step 1 (positivity for $x \geq 0$). By induction using both recursion formulas: if $Q_{2m-1}(x) > 0$ and $Q_{2m}(x) > 0$ for all $x \geq 0$, then for $x \geq 0$, $(m+1)x \geq 0$, so $Q_{2m+1}(x) = Q_{2m}(x) + (m+$

$1)xQ_{2m-1}(x) > 0$, and likewise $Q_{2m+2}(x) > 0$. So every $Q_n(x) > 0$ for $x \geq 0$ once the base case is positive there, which forces every real root x_n to be negative.

Step 2 (monotonicity). Suppose inductively $x_1 < x_2 < \cdots < x_{2m-1} < x_{2m}$. Since Q_{2m-1} has x_{2m-1} as its *largest* real root and $Q_{2m-1}(x) \to +\infty$ (its leading term dominates for large x), Q_{2m-1} keeps one constant sign for all $x > x_{2m-1}$, namely positive; since $x_{2m} > x_{2m-1}$ by hypothesis, $Q_{2m-1}(x_{2m}) > 0$. Now evaluate the recursion *at* x_{2m}, using that $Q_{2m}(x_{2m}) = 0$ by definition:

$$Q_{2m+1}(x_{2m}) = Q_{2m}(x_{2m}) + (m+1)x_{2m}Q_{2m-1}(x_{2m}) = (m+1)x_{2m}Q_{2m-1}(x_{2m}).$$

Since $x_{2m} < 0$ (Step 1) and $Q_{2m-1}(x_{2m}) > 0$, this product is negative: $Q_{2m+1}(x_{2m}) < 0$. But $Q_{2m+1}(x) \to +\infty$, so by the Intermediate Value Theorem Q_{2m+1} has a root beyond x_{2m}, and since x_{2m+1} is its *largest* root, $x_{2m+1} > x_{2m}$. The identical argument, one step later (evaluating Q_{2m+2} at x_{2m+1}), gives $x_{2m+2} > x_{2m+1}$. The induction closes: $\{x_n\}$ is increasing.

Step 3 (limit). Fix m and evaluate both recursion formulas at the single point $a = -\frac{1}{m+1}$, chosen because it makes $(m+1)a = -1$ exactly, collapsing each recursion from three terms to two:

$$Q_{2m+1}(a) = Q_{2m}(a) + (m+1)a\,Q_{2m-1}(a) = Q_{2m}(a) - Q_{2m-1}(a),$$

$$Q_{2m+2}(a) = Q_{2m+1}(a) + (m+1)a\,Q_{2m}(a) = Q_{2m+1}(a) - Q_{2m}(a) = \big[Q_{2m}(a) - Q_{2m-1}(a)\big] - Q_{2m}(a) = -Q_{2m-1}($$

So $Q_{2m+2}(a) = -Q_{2m-1}(a)$: these two values have opposite sign (or are both zero), so at least one of $Q_{2m+2}(a)$, $Q_{2m-1}(a)$ is ≤ 0. Whichever one is ≤ 0, that polynomial is positive for $x \geq 0$ (Step 1) and ≤ 0 at $a < 0$, so by the Intermediate Value Theorem it has a root in $[a, 0)$; since $\{x_n\}$ is increasing (Step 2), the corresponding index's largest root – and hence every later x_n – is $\geq a$. Either way, both $x_{2m-1} \geq a$ and $x_{2m+2} \geq a$, i.e. $x_n \geq -\frac{1}{m+1}$ for all the relevant n near $2m$. Combined with $x_n < 0$ from Step 1 and letting $m \to \infty$ (so $n \to \infty$ and $-\frac{1}{m+1} \to 0$), the squeeze gives $x_n \to 0$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start from the same recursion and the same goal, but they split at the very first strategic decision: what kind of object to try to control. The natural attempt tries to pin down $Q_n(x)$ itself – an explicit formula, a closed-form polynomial, something you could differentiate or factor. The model chain never tries to know $Q_n(x)$ at all; it only ever evaluates the recursion at specific, chosen test points (x_{2m}, x_{2m+1}, and $a = -1/(m+1)$) and reads off a *sign*, letting the Intermediate Value Theorem convert that sign into root-location information.

> *I don't need to know where the root x_{2m+1} actually is – I only need one point where Q_{2m+1} is negative, and I already know it eventually goes to $+\infty$, so a sign change (hence a root) beyond that point is guaranteed by the Intermediate Value Theorem. And evaluating at x_{2m} is free: I already know $Q_{2m}(x_{2m}) = 0$ by definition, so the recursion collapses to a single term I can sign-check.*

> *The coefficient $(m+1)x$ in the recursion is begging to be cancelled: plugging in $x = -1/(m+1)$ makes $(m+1)x = -1$ exactly, and the three-term recursion collapses into a bare subtraction, $Q_{2m+2}(a) = Q_{2m+1}(a) - Q_{2m}(a)$, which telescopes one more step into $-Q_{2m-1}(a)$ – turning an otherwise-unevaluable expression into a sign relation between two OLD terms I already have sign control over.*

Layer 3 THE WHY

Why does testing at chosen points beat trying to solve $Q_n(x) = 0$ explicitly? The two claims to be proved are claims about *location* of roots relative to other numbers (is x_{2m+1} bigger than x_{2m}? is x_n eventually within ε of 0?), not claims about the roots' exact values – and the

Intermediate Value Theorem is precisely the tool that answers "is there a root between these two points" from nothing but the sign of the function at those two points, without ever inverting the function. Whenever a proof's target is comparative or asymptotic rather than exact, sign information at well-chosen points is strictly what is needed; solving for the root exactly is solving a harder problem than the one that was asked. This is the general escape from an intractable recursion: when a recursively defined family has no tractable closed form but the question only concerns its zero set's order or limit, evaluate the recursion at points you control (previous roots, or points forced by the recursion's own coefficients) rather than trying to invert it.

Why is $a = -1/(m+1)$ specifically the right point to test, rather than any other number? Because the recursion's only term connecting three consecutive polynomials is $(m+1)x$ – a coefficient tied to the very index m that indexes the induction already in progress – and setting x so that $(m+1)x = -1$ is the unique choice that annihilates the coefficient down to a fixed constant, collapsing a three-term relation into a two-term one. This is not a guess pulled from nowhere: any time a recursion carries a coefficient that scales with the step index, testing the value of x that makes that coefficient equal to a clean constant (most often ± 1, since that is what turns "+ (coefficient)$\cdot x \cdot$ (term)" into "$\pm$(term)") is the systematic move, because it is the only way to remove the recursion's index-dependence from the arithmetic while keeping the relation between the polynomials themselves intact.

Layer $+\alpha$ BACKTESTING

1. The cue is a sequence of polynomials defined by a coupled recursion, with a question about the *order or limit* of their real roots – not their exact values. On seeing it, write down: don't solve for the roots; get sign information at chosen points and invoke the Intermediate Value Theorem.

2. That cue suggests (i) an induction establishing a sign region (here, positivity for $x \geq 0$, forcing all roots negative), then (ii) evaluating the recursion at the *previous* root to get monotonicity, then (iii) evaluating at a specially chosen point to pin down a limit.

3. Hesitation occurs in choosing which point to test at step (iii); it is reinforced away by scanning the recursion for a coefficient that depends on the induction index (here $(m+1)x$) and solving for the x that collapses it to ± 1.

4. The trap is hunting for an explicit formula for $Q_n(x)$ or its roots – the coefficient $(m+1)$ growing with the step makes any such formula intractable to extract clean monotonicity from, even if you do find it.

5. Recognise faster: whenever a recursion contains a term (coefficient depending on the step)$\times x \times$(older term), test the x-value that makes that coefficient collapse to a clean constant before doing anything else with the recursion.

6. Reusable rule: to control the real roots of a recursively defined sequence of functions without a closed form, (i) establish a sign region by induction directly on the recursion, (ii) prove monotonicity of the root sequence by evaluating the recursion at the previous root and reading the resulting sign via the Intermediate Value Theorem, (iii) pin down limits by testing at a point chosen to make the recursion's index-dependent coefficient collapse to a constant, turning a three-term relation into a two-term (often telescoping) one.

▷ **Cue:** a sequence of polynomials/functions defined by a coupled recursion; the question concerns the order or limit of their real roots, not their values $\longrightarrow$ **Move:** do not seek a closed form; establish a sign region by induction, then get root-order information by evaluating the recursion at previously-known roots and applying the Intermediate Value Theorem

▷ **Cue:** need to show a root x_{n+1} exceeds a known point p $\longrightarrow$ **Move:** evaluate the defining relation for the corresponding function at $x = p$ (using any known zero, e.g. $Q_{2m}(x_{2m}) = 0$, to kill a term); if the result is negative and the function tends to $+\infty$, the Intermediate Value Theorem forces a root beyond p, hence (since it's the largest root) $x_{n+1} > p$

▷ **Cue:** a recursion contains a term (index-dependent coefficient)·x·(older term) $\longrightarrow$ **Move:** test the value of x that makes the coefficient collapse to a clean constant (typically ± 1), reducing a three-term recursion to a two-term subtraction relation between older terms

1978 B4

1978 B4. Prove that for every real number N, the equation

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 = x_1x_2x_3 + x_1x_2x_4 + x_1x_3x_4 + x_2x_3x_4$$

has a solution for which x_1, x_2, x_3, x_4 are all integers larger than N.

Fields: number theory; Diophantine equations; Vieta jumping **Position:** B4 — middle of the paper; an existence problem that looks like number theory and is really about a quadratic in one variable **Answer:** Start from $(1,1,1,1)$. The equation is quadratic in each variable, so the smallest coordinate x_4 can be replaced by the other root $x_4' = x_1x_2 + x_1x_3 + x_2x_3 - x_4$. The new value is an integer and exceeds every other coordinate, so repeated jumping gives $(1,1,1,2)$, $(1,1,2,4)$, $(1,2,4,13)$, $(2,4,13,85)$, $(4,13,85,1495)$, …, whose smallest entry tends to infinity.

Layer 1 THE WHAT — surface analysis

The natural attempt. You need infinitely many integer solutions, arbitrarily large. The equation is symmetric, so the first thing a strong student tries is to make it *more* symmetric: put all four variables equal to t. Then $4t^2 = 4t^3$, so $t = 0$ or $t = 1$. That gives $(1,1,1,1)$, one small solution and nothing bigger.

Next attempt: break the symmetry as little as possible. Take $x_1 = x_2 = x_3 = a$ and $x_4 = b$. Then

$$3a^2 + b^2 = a^3 + 3a^2b, \qquad \text{i.e.} \qquad b^2 - 3a^2b + (3a^2 - a^3) = 0.$$

For $a = 1$ this is $b^2 - 3b + 2 = 0$, so $b = 1$ or $b = 2$. Good: $(1,1,1,2)$ is a new solution. For $a = 2$ it is $b^2 - 12b + 4 = 0$, whose discriminant 128 is not a square. For $a = 3$ it is $b^2 - 27b = 0$, so $b = 27$ — but $b = 0$ is the other root and $(3,3,3,27)$ still contains 3s, which will not go above a large N. A one-parameter family with three equal entries keeps three small entries, so it cannot work. Trying $x_1 = x_2 = a$, $x_3 = x_4 = b$ gives $2a^2 + 2b^2 = 2a^2b + 2ab^2$, i.e. $a^2 + b^2 = ab(a + b)$, which for positive integers forces $a = b = 1$. The search for a parametric family is hitting a wall: every family you guess either has no integer points or keeps a coordinate fixed.

The model chain. *Base solution.* $(1,1,1,1)$ works: both sides equal 4.

The jump. Fix x_1, x_2, x_3 and read the equation as a quadratic in x_4:

$$x_4^2 - (x_1x_2 + x_1x_3 + x_2x_3)\, x_4 + \left(x_1^2 + x_2^2 + x_3^2 - x_1x_2x_3\right) = 0.$$

If x_4 is one root, the sum of the roots is $x_1x_2 + x_1x_3 + x_2x_3$, so the other root is

$$x_4' = x_1x_2 + x_1x_3 + x_2x_3 - x_4.$$

So (x_1, x_2, x_3, x_4') is again a solution, and it is an integer solution when the old one was. Because the equation is symmetric, the same move can be made on any coordinate.

The jump increases. Suppose all four coordinates are positive integers and x_4 is a smallest one. Write the other three as $a \le b \le c$, so $1 \le x_4 \le a$. Then

$$x_4' = ab + ac + bc - x_4 \ge ab + ac + bc - a = a(b-1) + ac + bc \ge c + c = 2c > c,$$

using $b \ge 1$, $a \ge 1$ and $c \ge 1$. So the new coordinate is a positive integer strictly larger than each of the three coordinates it sits beside.

Iterate. From a positive integer solution, replace a smallest coordinate by its partner root, and repeat. Call the new values $y_1, y_2, \ldots$. Each y_{k+1} exceeds every coordinate present when it is created, including y_k, so $y_1 < y_2 < y_3 < \cdots$. Starting from $(1,1,1,1)$:

$$(1,1,1,1) \to (1,1,1,2) \to (1,1,2,4) \to (1,2,4,13) \to (2,4,13,85) \to (4,13,85,1495) \to \cdots$$

(Check $(2,4,13,85)$: the left side is $4+16+169+7225 = 7414$ and the right side is $104+680+2210+4420 = 7414$.) After step 4 the tuple is $\{y_1, y_2, y_3, y_4\}$: each of the first four steps removes an entry that is not a y, because while a 1 remains it is the minimum. From then on, if the tuple is $\{y_{k-3}, y_{k-2}, y_{k-1}, y_k\}$, the minimum is y_{k-3}; it is replaced by y_{k+1} and the tuple becomes $\{y_{k-2}, \ldots, y_{k+1}\}$. So after step $k \geq 4$ the smallest coordinate is y_{k-3}. The y_k are strictly increasing integers, so $y_k \geq k+1$ and hence $y_{k-3} \geq k-2$. Take $k > N+2$ and every coordinate exceeds N. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains write down $(1,1,1,1)$, and both look at the equation with some variables fixed. They split at the equation $b^2 - 3a^2b + (3a^2 - a^3) = 0$. The natural attempt sees a quadratic and *solves* it, asking when the discriminant is a square. The model chain sees a quadratic and asks what the *other* root is, since one root is already known.

> *This equation has degree two in each variable separately. If I already have a solution, then x_4 is one root of a quadratic whose coefficients I know. The other root is (sum of roots) $- x_4$ — an integer, free of charge. I do not need to find a family. I need to jump.*

> *Which coordinate do I jump? The smallest one. It is the one holding the tuple down, and replacing it by the partner root throws it far above the rest.*

- ▷ **Cue:** a Diophantine equation that is quadratic in each variable separately (every monomial has degree at most 2 in each x_i), and one small solution is visible $\longrightarrow$ **Move:** fix all but one variable, use Vieta's sum of roots to replace that variable by the other root; the result is again an integer solution
- ▷ **Cue:** "prove there are solutions larger than any N" for a symmetric equation $\longrightarrow$ **Move:** jump the smallest coordinate each time and prove the new value exceeds the old maximum; then the minimum is forced to infinity
- ▷ **Cue:** you keep guessing parametric families and each has a fixed coordinate or no integer points $\longrightarrow$ **Move:** stop parametrising; generate solutions from solutions with a move that preserves the equation

Layer 3 THE WHY

Why is the other root the right object, and not a trick you have to have seen? Look at what the problem gives you. It gives an equation and no formula. An existence problem with "for every N" wants an infinite supply of solutions, and there are only two ways to get one: write down a formula (a parametric family), or find an operation that turns one solution into another. The failed families show that the first route is closed. Every guess either kills the integrality or leaves a coordinate stuck at a small value. So look for an operation.

What operations preserve the equation? Look at its degrees. Each variable appears squared on the left and linearly on the right. So with three variables fixed the equation is a monic quadratic in the fourth, with integer coefficients. A monic quadratic with integer coefficients and one integer root has its other root an integer too, because the two roots add to an integer. That is Vieta's formula, which you learned as "sum of roots is $-b/a$". Here $a = 1$, so it is an integer statement. Nothing more is used. Monic in x_4 and integer coefficients: those two facts, read straight off the equation, are the whole reason the jump works.

Why jump the smallest coordinate? Because what you have to prove is about the minimum: *all* four entries exceed N. The only way to raise the minimum is to change the entry that achieves it. The inequality $x_4' \geq 2c$ then says the jump overshoots everything, and so the entries leave in the order they came in. The oldest, smallest entry always goes next, like a queue.

The official solution compresses this to "one can start with the solution $(1,1,1,1)$ and through repeated use of the procedures stated above obtain a solution with each x_i an integer greater than N". It proves the new coordinate exceeds the minimum, but that alone does not show the minimum grows. It would not rule out cycling among small values. The queue argument above is the missing step: every new value beats the current maximum, so the new values increase and the minimum is always the oldest of the last four.

This is the same mechanism as Markov's equation $x^2+y^2+z^2 = 3xyz$ and IMO 1988 Problem 6 ("Vieta jumping"). There the jump is used downwards, to reach a contradiction by descent. Here it is used upwards, to build. It is the same move in both.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the shape of the equation: squares on the left, products with each variable to the first power on the right. On seeing it, write the equation as a quadratic in x_4 with the other three fixed, and write the two Vieta relations.

2. Vieta jumping: from a known integer solution, replace one variable by the other root of the quadratic in that variable.

3. Hesitation comes after $(1,1,1,1)$, when the parametric guesses fail and the student concludes solutions are rare. What reinforces it away is noticing that the two-root structure already gave a second solution: $b=1$ and $b=2$ in the $a=1$ case. The second root is the jump. You found it and did not use it.

4. The trap is proving too little at the iteration. Showing that $x_4' > x_4$, or that x_4' exceeds the minimum, does not show the minimum goes to infinity. You need the new value to beat the *maximum*, so that the entries cycle out in order.

5. Whenever an equation is quadratic in each variable separately and you are handed (or can spot) one solution, the equation generates its own solutions. Look for the partner root before you look for a formula.

6. Reusable decision rule: *for a Diophantine equation of degree two in each variable, fix all but one variable and use the sum of the roots to jump. To build large solutions, jump the smallest coordinate and prove it lands above the largest. To prove no solutions, jump downwards and descend.*

1978 B5

1978 B5. Find the largest A for which there exists a polynomial

$$P(x) = Ax^4 + Bx^3 + Cx^2 + Dx + E,$$

with real coefficients, which satisfies

$$0 \leq P(x) \leq 1 \quad \text{for} \quad -1 \leq x \leq 1.$$

Fields: algebra; polynomials; extremal problems; Chebyshev polynomials **Position:** B5 — late in the paper; the answer is guessable in minutes, and the marks are all in the upper bound **Answer:** $A=4$, attained by $P(x) = 4x^4 - 4x^2 + 1 = (2x^2-1)^2$. The upper bound comes from the identity $A = P(1) + P(-1) + 2P(0) - 2P(\frac{1}{\sqrt{2}}) - 2P(-\frac{1}{\sqrt{2}})$, valid for every quartic, which gives $A \leq 1+1+2 = 4$.

Layer 1 THE WHAT — surface analysis

The natural attempt. A strong student first looks for a good example. The condition is symmetric under $x \mapsto -x$, so try an even polynomial. Something that stays between 0 and 1 and has a

large x^4 coefficient: a square is automatically ≥ 0, so try $P(x) = A(x^2 - c)^2$. On $[-1, 1]$ the quantity $x^2 - c$ runs over $[-c, 1 - c]$, so the largest value of P is $A \max(c^2, (1 - c)^2)$. That is smallest at $c = \frac{1}{2}$, where it equals $A/4$. So $A = 4$ works: $P(x) = 4(x^2 - \frac{1}{2})^2 = 4x^4 - 4x^2 + 1$.

Now prove nothing larger works. The instinct is calculus: P is at most 1, so at an interior maximum $P' = 0 \ldots$ and five unknown coefficients appear, with no clear equation to set up. Or Lagrange-style: maximise A subject to infinitely many constraints $0 \leq P(x) \leq 1$. You can write down the conditions at the points you suspect are extremal, but it is not clear which constraints are active, or why checking finitely many of them should be enough. Students who do not know Chebyshev polynomials often stop here with the right answer and a proof only for the family $A(x^2 - c)^2$. That proof is worth very little: it shows 4 is best *among squares of even quadratics*, not among all quartics.

The model chain. Example. $P(x) = (2x^2 - 1)^2 = 4x^4 - 4x^2 + 1$. For $-1 \leq x \leq 1$ we have $0 \leq x^2 \leq 1$, so $-1 \leq 2x^2 - 1 \leq 1$ and $0 \leq P(x) \leq 1$. So $A = 4$ is attained.

Upper bound. Let $s = 1/\sqrt{2}$. I claim that every polynomial P of degree at most 4 with x^4-coefficient A satisfies

$$A = P(1) + P(-1) + 2P(0) - 2P(s) - 2P(-s). \qquad (*)$$

Both sides are linear in P, so it is enough to check $(*)$ on $1, x, x^2, x^3, x^4$.

- $P = 1$: $1 + 1 + 2 - 2 - 2 = 0$, and $A = 0$.
- $P = x$ and $P = x^3$: every term pairs x with $-x$ with equal weights, so the right side is 0; and $A = 0$.
- $P = x^2$: $1 + 1 + 0 - 2 \cdot \frac{1}{2} - 2 \cdot \frac{1}{2} = 0$, and $A = 0$.
- $P = x^4$: $1 + 1 + 0 - 2 \cdot \frac{1}{4} - 2 \cdot \frac{1}{4} = 1$, and $A = 1$.

So $(*)$ holds. Now if $0 \leq P \leq 1$ on $[-1, 1]$, the positive-weight terms are each at most their weight and the negative-weight terms are at least 0:

$$A \leq 1 + 1 + 2 \cdot 1 - 0 - 0 = 4.$$

So the largest A is 4. □

Where the weights come from. They are the coefficients in the divided difference $P[x_0, \ldots, x_4] = \sum_i P(x_i) / \prod_{j \neq i} (x_i - x_j)$ at the nodes $-1, -s, 0, s, 1$, which equals the leading coefficient of a quartic. For example at $x = 0$ the product is $(0 + 1)(0 + s)(0 - s)(0 - 1) = s^2 = \frac{1}{2}$, giving weight 2; at $x = s$ it is $(s + 1)(2s)(s)(s - 1) = 2s^2(s^2 - 1) = -\frac{1}{2}$, giving weight -2. The equality case forces $P(\pm 1) = P(0) = 1$ and $P(\pm s) = 0$, which is exactly $(2x^2 - 1)^2$, so the extremiser is unique.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains find $(2x^2 - 1)^2$. They split at the upper bound, at the sentence "maximise A subject to $0 \leq P(x) \leq 1$ for all x". The natural attempt treats the constraint as a continuum and reaches for calculus. The model chain says that A is a *linear* function of P, and a linear function of a quartic is determined by five values. So it writes A as a combination of P at five points and bounds each point separately.

> *The leading coefficient is a linear function of P. A quartic is pinned down by its values at five points, so A is some fixed combination of five values of P. If I choose the five points where my example touches* 0 *and* 1*, the signs of the weights should match: plus where P = 1, minus where P = 0. Then the bound is one line.*

> *Which five points? The places where the extremal example is pressed against the constraints:*

> $x = \pm 1, 0$ *where it equals* 1, *and* $x = \pm\frac{1}{\sqrt{2}}$ *where it equals* 0.

- ▷ **Cue:** "find the largest leading coefficient of a polynomial bounded on an interval" $\longrightarrow$ **Move:** guess the extremiser (a Chebyshev polynomial, shifted and scaled), then express the leading coefficient as a weighted sum of values at the points where the extremiser touches its bounds
- ▷ **Cue:** an optimisation over polynomials with infinitely many pointwise constraints $\longrightarrow$ **Move:** the objective is linear in the coefficients, so find finitely many constraints that already force the bound: the points where the conjectured extremiser is tight
- ▷ **Cue:** a constraint on $[-1, 1]$ invariant under $x \mapsto -x$ $\longrightarrow$ **Move:** replace P by $\frac{1}{2}(P(x) + P(-x))$: it satisfies the same constraint and has the same even coefficients, so without loss P is even

Layer 3 THE WHY

Why does the interpolation identity exist, and why is it the natural bound and not a rabbit from a hat? Start from what A is. It is a coefficient, and coefficients depend linearly on the polynomial. The space of polynomials of degree at most 4 has dimension 5, and evaluation at five distinct points is a bijection onto $\mathbb{R}^5$ (that is Lagrange interpolation, or equivalently a Vandermonde determinant being nonzero). So *any* linear quantity, A included, is a fixed combination $\sum w_i P(x_i)$ of five values. This is not a choice. It is forced by dimension counting.

The only freedom is where to put the five points. Now look at the bound you want. You want $\sum w_i P(x_i) \le 4$ using only $0 \le P(x_i) \le 1$. The best that pointwise bounds can give is $\sum_{w_i > 0} w_i$. For that to equal 4 with equality at the example, the example must satisfy $P(x_i) = 1$ where $w_i > 0$ and $P(x_i) = 0$ where $w_i < 0$. So the points have to be where the example touches its bounds. The example $(2x^2 - 1)^2$ touches 1 at $0, \pm 1$ and touches 0 at $\pm 1/\sqrt{2}$. That is five points, which is exactly enough. Then the sign pattern $+, -, +, -, +$ of the weights is automatic, because divided-difference weights alternate in sign along ordered nodes. The extremiser tells you where to look. This is the "circle has a centre, so draw the centre" step: the thing you are bounding is linear, and a linear bound is tight only where the constraints are active.

The official solution takes a different route: it symmetrises twice. Averaging $P(x)$ with $P(-x)$ kills the odd terms. Then $y = x^2$ turns the even quartic into a quadratic on $[0, 1]$. A shift $y = (z + 1)/2$ moves it to $[-1, 1]$ with leading coefficient $A/4$. Averaging again and setting $w = z^2$ leaves a *linear* function $\frac{A}{4}w + G$ on $[0, 1]$ with values in $[0, 1]$, whose slope is at most 1. Each step preserves the constraint and loses no information about A. That is a valid and tidy proof. But it only works because this problem has a double symmetry, and it gives you no hint what to do for a quintic. The five-point identity works for every degree: the Chebyshev polynomial T_n has $n + 1$ extremal points on $[-1, 1]$, and the same argument shows 2^{n-1} is its largest leading coefficient among polynomials bounded by 1 there. Here $P = (T_4 + 1)/2$ with $T_4 = 8x^4 - 8x^2 + 1$, so the answer is $8/2 = 4$.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "largest A" with A the leading coefficient and a two-sided bound on an interval. On seeing it, write down a candidate extremiser and mark where it touches 0 and 1.

2. Guess the extremiser, then prove optimality by expressing A as a linear combination of values of P at the touching points (interpolation, or divided differences).

3. Hesitation comes once $A = 4$ is found and the upper bound resists calculus. What reinforces it away is the observation that A is linear in P, so the proof should be one linear inequality, not an optimisation.

4. The trap is proving the bound only inside the family you guessed, such as $A(x^2 - c)^2$, and presenting that as the answer. A second trap is assuming P is even without saying why; symmetrising by $\frac{1}{2}(P(x) + P(-x))$ is what licenses it.

5. Whenever you bound a linear quantity (a coefficient, a derivative at a point, an integral) of a polynomial subject to pointwise constraints, the proof is a finite weighted sum of values. The points are where the extremal example is tight.

6. Reusable decision rule: *to maximise a linear functional under pointwise constraints, find the extremiser first, list the points where it meets its constraints, and write the functional as a combination of values at exactly those points. The signs of the weights should match which constraint is tight at each point.*

1978 B6

1978 B6. Let p and n be positive integers. Suppose that the numbers c_{hk} $(h = 1, 2, \ldots, n;\ k = 1, 2, \ldots, ph)$ satisfy $0 \le c_{hk} \le 1$. Prove that

$$\Big(\sum \frac{c_{hk}}{h}\Big)^2 \le 2p \sum c_{hk},$$

where each summation is over all admissible ordered pairs (h, k).

Caveat. The MAA OCR drops the displayed inequality of this problem. It has been restored as $\big(\sum c_{hk}/h\big)^2 \le 2p \sum c_{hk}$. This is the standard wording, and it matches the book's solution under the label B-6, which sets $a_h = \big(\sum_k c_{hk}\big)/h$ and proves $\big(\sum_h a_h\big)^2 \le 2p \sum_h h a_h$. Since $\sum_{h,k} c_{hk}/h = \sum_h a_h$ and $\sum_{h,k} c_{hk} = \sum_h h a_h$, that is the same statement.

Fields: inequalities; algebra; induction; summation **Position:** B6 — last problem of the paper, but the hardest part is seeing that the c_{hk} are a distraction **Answer:** Put $a_h = \frac{1}{h}\sum_{k=1}^{ph} c_{hk}$, so $0 \le a_h \le p$. The claim becomes $\big(\sum a_h\big)^2 \le 2p \sum h a_h$. Expand the square as $\sum_j a_j\big(a_j + 2\sum_{i<j} a_i\big)$ and bound the bracket by $(2j-1)p$. That gives $p\sum(2j-1)a_j \le 2p\sum j a_j$.

Layer 1 THE WHAT — surface analysis

The natural attempt. A square of a sum on the left, a sum on the right: every strong student reaches for Cauchy–Schwarz. Write $\frac{c_{hk}}{h} = \sqrt{c_{hk}} \cdot \frac{\sqrt{c_{hk}}}{h}$. Then

$$\Big(\sum \frac{c_{hk}}{h}\Big)^2 \le \Big(\sum c_{hk}\Big)\Big(\sum \frac{c_{hk}}{h^2}\Big) \le \Big(\sum c_{hk}\Big) \sum_{h=1}^{n} \frac{ph}{h^2} = p\, H_n \sum c_{hk},$$

where $H_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n}$ and the second step used $c_{hk} \le 1$ and the fact that row h has ph entries. So the constant is pH_n, not $2p$. For $n \le 3$ that is fine ($H_3 = \frac{11}{6} < 2$), but $H_4 = \frac{25}{12} > 2$ and $H_n \to \infty$. The wall is real: the inequality has a constant independent of n, and Cauchy–Schwarz produces a logarithm. Other splittings (weights $\sqrt{h}$, and so on) produce the same $\log n$ or worse.

The reason is that Cauchy–Schwarz used $c_{hk} \le 1$ only *after* squaring, in the crude bound $\sum c_{hk}/h^2 \le pH_n$. It never paired that information with the size of the left-hand side.

The model chain. Compress the data. For each h let

$$a_h = \frac{1}{h} \sum_{k=1}^{ph} c_{hk}.$$

Row h has ph entries, each in $[0, 1]$, so $0 \le a_h \le p$. Also

$$\sum_{h,k} \frac{c_{hk}}{h} = \sum_{h=1}^{n} a_h, \qquad \sum_{h,k} c_{hk} = \sum_{h=1}^{n} h\, a_h.$$

So it suffices to prove: *if* $0 \le a_1, \dots, a_n \le p$ *then* $\left(\sum_{h=1}^{n} a_h\right)^2 \le 2p \sum_{h=1}^{n} h\, a_h$.

Expand the square in order. Writing out the square and grouping each cross term $2a_ia_j$ $(i < j)$ with the later index j,

$$\Big(\sum_{j=1}^{n} a_j\Big)^2 = \sum_{j=1}^{n} a_j^2 + 2\sum_{i<j} a_i a_j = \sum_{j=1}^{n} a_j\Big(a_j + 2\sum_{i=1}^{j-1} a_i\Big).$$

Bound each bracket. Each $a_i \le p$, so the bracket is at most $p + 2(j-1)p = (2j-1)p$. Since $a_j \ge 0$, we can multiply:

$$\Big(\sum_{j=1}^{n} a_j\Big)^2 \le p \sum_{j=1}^{n} (2j-1)\, a_j \le 2p \sum_{j=1}^{n} j\, a_j. \qquad \square$$

Sharpness. Take every $c_{hk} = 1$. Then $a_h = p$, the left side is n^2p^2 and the right side is $2p \cdot p \cdot \frac{n(n+1)}{2} = n(n+1)p^2$. The ratio $n/(n+1) \to 1$, so the constant 2 cannot be lowered. The intermediate inequality $\left(\sum a_j\right)^2 \le p \sum (2j-1) a_j$ is an *equality* in this case, because $\sum_{j=1}^{n} (2j-1) = n^2$.

The official version is the same computation as an induction on n. With $S_m = \sum_{h \le m} a_h \le mp$, $(S_m + a_{m+1})^2 \le 2p \sum_{h \le m} h a_h + 2mp\, a_{m+1} + p\, a_{m+1} \le 2p \sum_{h \le m+1} h a_h$, using $a_{m+1}^2 \le p a_{m+1}$ and $2m + 1 \le 2(m+1)$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains face $\left(\sum c_{hk}/h\right)^2$. The natural attempt applies Cauchy–Schwarz to the c_{hk} directly. The model chain first sums along each row, replacing all $p\,n(n+1)/2$ numbers c_{hk} by n numbers $a_h \in [0, p]$, and then expands the square *ordered by* h. The split is at the first line: whether you keep the c_{hk} or throw them away.

> *The inner index k appears nowhere except in the count of terms. Both sides only see the row sums. Let me replace each row by one number,* $a_h = \frac{1}{h}\sum_k c_{hk}$*, and see what the inequality says about the* a_h*.*

> *Now the right side has weight h on* a_h*, and the left is a square. If I expand the square and charge each cross term* $2a_ia_j$ *to the larger index, term j picks up* $j-1$ *partners, each at most p. Weight j on the right against about* $2j-1$ *copies of p on the left — that is exactly the balance I need.*

- ▷ **Cue:** a double-indexed sum where the inner index only affects the number of terms in each row $\longrightarrow$ **Move:** sum out the inner index first and reduce to a single sequence with a bound inherited from the row length
- ▷ **Cue:** $(\sum a_j)^2$ against $\sum w_j a_j$ with increasing weights w_j, and a uniform bound $0 \le a_j \le M$ $\longrightarrow$ **Move:** write $(\sum a_j)^2 = \sum_j a_j\big(a_j + 2\sum_{i<j} a_i\big)$ and bound the bracket by $(2j-1)M$
- ▷ **Cue:** Cauchy–Schwarz gives the right shape but a constant that grows like $\log n$ $\longrightarrow$ **Move:** the problem uses an upper bound on the terms, which Cauchy–Schwarz only uses crudely; switch to an expansion that uses the upper bound on every factor

Layer 3 THE WHY

Why reduce to row sums? Because the problem does. Read the constraint: k runs from 1 to ph. The inner index is there only to make row h have ph entries. On the left, c_{hk} is divided by h, which depends only on the row. On the right, the c_{hk} appear unweighted. So both sides are functions of the row sums $R_h = \sum_k c_{hk}$ alone, and the only thing the individual bounds $0 \le c_{hk} \le 1$ tell you about R_h is $0 \le R_h \le ph$. The problem is a one-index problem written with two indices. Normalising by h to get $a_h = R_h/h \in [0, p]$ makes the bound the same for every row, which is the cleanest form to carry.

Why expand the square that way? Look at the reduced inequality: a square of a sum on the left and a *weighted linear* sum on the right, with weight h on a_h. To compare a quadratic with a linear expression, one factor of each quadratic term must be converted into a constant, and the only constant available is p (from $a_i \le p$). So write the square as $\sum_j a_j \times (\text{something})$ and bound the something by a constant. The question is which "something". The symmetric choice $a_j \cdot \sum_i a_i$ bounds every bracket by np. That gives $np \sum a_j$, with weight n on every term, too much for small j. The right side gives small weight to small j. So the terms with small j must take few partners, which means charging each pair $\{i, j\}$ to the larger index. Then a_j carries $j - 1$ partners and the bracket is at most $(2j - 1)p$. The ordering is dictated by the weights on the right-hand side. It is not a choice.

And why does 2 come out exactly? Because $1 + 3 + 5 + \cdots + (2n - 1) = n^2$. The square on the left, at the extremal point $a_j \equiv p$, equals p^2 times the sum of the first n odd numbers. The right side replaces each odd number $2j - 1$ by the even number $2j$. That is where the factor 2 in $2p$ comes from, and it is why the bound is asymptotically sharp.

The official solution writes this as an induction on n with no comment on where the step comes from. The step is the telescoping identity $(S + a)^2 - S^2 = a(a + 2S)$. That identity is the ordered expansion above, one term at a time.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "$k = 1, 2, \ldots, ph$": an inner index whose only role is to fix row lengths. On seeing it, define the row sums and rewrite both sides in terms of them.

2. Reduction to a one-index inequality, then an ordered expansion of the square (equivalently, induction on n with $(S + a)^2 = S^2 + a(a + 2S)$).

3. Hesitation comes after Cauchy–Schwarz yields pH_n. The student either tries to shave the harmonic sum or gives up. What reinforces it away is the check $H_4 > 2$: a logarithmic constant cannot be tuned to 2, so the method itself is wrong and something must use $c_{hk} \le 1$ more fully.

4. The trap is the double index. It makes the problem look like a matrix or combinatorial inequality and invites heavy tools. A second trap is expanding the square symmetrically, which gives the constant n instead of the needed weight j.

5. When a square of a sum is compared with a linear sum carrying increasing weights, and the terms are bounded above, expand the square as "each term times itself plus twice everything before it". The weights on the right tell you the order.

6. Reusable decision rule: *first collapse every index that the inequality does not really see. Then, to bound a quadratic form by a linear one using $0 \le a_i \le M$, write the quadratic as $\sum_j a_j \cdot L_j$ with each L_j bounded by M times the weight you need, and choose the order of the charging to match the weights.*