DAWN PUTNAM SESSION

Putnam 1976 — Papers A and B

Putnam 120 Archive

Contents

Putnam 1976 — Paper A **5**
1976 A1 5
1976 A2 7
1976 A3 9
1976 A4 11
1976 A5 14
1976 A6 17

Putnam 1976 — Paper B **23**
1976 B1 23
1976 B2 25
1976 B3 27
1976 B4 30
1976 B5 33
1976 B6 35

Putnam 1976 — Paper A

1976 A1

1976 A1. P is an interior point of the angle whose sides are the rays OA and OB. Locate X on OA and Y on OB so that the line segment XY contains P and so that the product of distances $(PX)(PY)$ is a minimum.

Caveat. The MAA 1965–1984 OCR renders this as "Pis an interior point ... 08" with the ray labels garbled by column-misreading. The statement above is the standard reconstruction, cross-checked against the book's own solution (which speaks of rays OA, OB, the bisector, and a circle tangent to both) and against the topic index, which lists this problem under "minimum problem 1976 A-1." No mathematical content was altered.

Fields: Euclidean geometry, optimisation **Position:** A1 (opener) **Answer:** X and Y are the points where the line through P perpendicular to the bisector of $\angle AOB$ meets OA and OB; equivalently, XY is the chord, through P, of the circle tangent to both rays.

Layer 1 THE WHAT — surface analysis

The natural attempt. Set up coordinates: O at the origin, the bisector of $\angle AOB$ along a fixed axis, P fixed at distance d from O. A line through P is determined by one parameter, its angle φ to the bisector. In the triangle OPX, the angle at O and the angle at P (namely φ, up to a fixed offset) are known, so the Law of Sines gives PX as $\dfrac{OP\sin(\angle POA)}{\sin(\angle OXP)}$, and $\angle OXP = \pi - \angle POA - \varphi$. The same construction on the other side gives PY in terms of φ. The product $PX \cdot PY$ is then a single-variable function of φ, built from ratios of sines. You differentiate it, using the quotient rule twice, clear denominators, and apply product-to-sum identities to isolate the critical φ. It can be done — the algebra converges on a clean condition after a page of trigonometric manipulation — but it costs the better part of the exam's allotted time for a single point, and a sign error anywhere in the quotient-rule chain is fatal and easy to miss, because the final expression is unrecognisable as anything you've seen before, so there is no way to sanity-check it by inspection.

The model chain. Fix a circle inscribed in the angle, tangent to OA at some point X and to OB at some point Y, chosen from the one-parameter family of such circles (all centred on the bisector) so that it passes through P; then XY, the chord joining the tangent points, is a segment through P meeting both rays. By symmetry about the bisector, $OX = OY$, and by a standard fact of Euclidean geometry, the product of the two segments a chord through P is cut into by P — here PX and PY — equals the power of P with respect to that circle, a number depending only on the circle and P, not on which chord through P is chosen. Now let X_1Y_1 be any other segment through P with X_1 on OA, Y_1 on OB, and let X_2, Y_2 be the two points where the line X_1Y_1 actually meets the fixed circle. Since the circle is tangent to (hence never crosses) OA and OB, and X_1 lies on OA while X_2 lies strictly inside the angle on the same ray from P,

we get $PX_1 \geq PX_2$, and likewise $PY_1 \geq PY_2$. The power-of-the-point identity applied to this second chord gives $PX_2 \cdot PY_2$ equal to the same constant power of P — that is, equal to $PX \cdot PY$. Hence $PX_1 \cdot PY_1 \geq PX_2 \cdot PY_2 = PX \cdot PY$, with equality exactly when X_1Y_1 coincides with the tangent chord XY. So $PX \cdot PY$ is the minimum, attained uniquely at the tangent-circle chord — which, since $OX = OY$, is precisely the chord perpendicular to the bisector at P's foot on it, or more directly: the line through P making equal angles with OA and OB measured from the perpendiculars, i.e. $(\pi - \angle AOB)/2$ with each ray.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both approaches start from the same reading of the problem: minimise $PX \cdot PY$ over segments through P. They split before either produces a single symbol of algebra, at the decision of what kind of object the product $PX \cdot PY$ is being treated as. The natural attempt treats it as *a function of a variable* (the angle φ) to be minimised by calculus. The model treats it as *an invariant of a fixed configuration* (the power of P with respect to a circle) and looks for the configuration in which that invariant equals the quantity in question.

> *This is a product of two lengths, both measured from the same fixed point, along a line that varies — that is exactly the shape of the power of a point with respect to a circle. I should be hunting for a circle through P tangent to both rays, not setting up a derivative.*

Layer 3 THE WHY

The recognition has to come from the phrase itself, not from familiarity with "this type of problem." Power of a point is defined as: for a fixed point Q and circle ω, and any line through Q meeting ω at two points M, N, the signed product $QM \cdot QN$ is the same for every such line. The defining feature of that invariant is precisely "two collinear distances from a fixed point, multiplied, with the line free to rotate" — which is a description of $(PX)(PY)$ itself, once you notice P is fixed and X, Y move together as the line through P rotates. A problem asking you to optimise a sum of two distances from a fixed point would not trigger this (that shape belongs to ellipses, via the reflection property); a product does, and only a product does, because power of a point is defined multiplicatively. This is the same instinct STYLE.md's own worked example insists on for the circle: a circle is the set of points equidistant from a centre, so of course the centre is the object of interest. Here the object is not a circle handed to you already, but a product-of-two-distances handed to you, and the product's algebraic shape is the pointer to which circle to construct: the one for which X and Y are not merely *a* chord through P, but happen to lie on OA and OB themselves. That last requirement — $X \in OA$, $Y \in OB$ — is exactly what tangency buys you: it's what pins X, Y to the rays rather than to an arbitrary secant, and it is what makes the inequality $PX_1 \cdot PY_1 \geq PX \cdot PY$ run in the right direction, since a circle tangent to a line touches it without crossing, so any other point of that same line lies strictly outside the circle, hence farther from P along that direction than the circle's own boundary point. Existence and uniqueness of the tangent circle through P follow from the one-parameter family of circles tangent to both rays sweeping continuously from radius 0 at O outward, so it passes through every interior point of the angle exactly once.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the phrase "product of distances $(PX)(PY)$," with P fixed and the line through it free to vary. On seeing it, write down: this is the algebraic shape of the power of a point.
2. That cue suggests constructing a circle through P for which X and Y — the actual intersections with OA and OB — are the two points of a chord, i.e. a circle tangent to both rays and passing through P, since tangency is what forces the chord's endpoints onto the rays themselves.

3. The hesitation comes after the circle is built: it is not obvious why the tangent chord gives the *minimum* rather than an arbitrary critical value. It is reinforced away by the direct comparison argument — any other secant through P meets the fixed circle at two points strictly closer to P than X_1, Y_1 are (because the circle cannot cross the tangent lines), and those two closer points already realise the same constant power. That converts "why is this extremal" into an explicit, checkable inequality.

4. The trap is committing to coordinates and calculus before asking what kind of object the target expression is. Once the derivative is set up, the problem no longer looks geometric, and a sign or algebra slip in the trigonometric identities is both easy to make and hard to catch, because the intermediate expressions carry no geometric meaning to check against.

5. To recognise this faster: build a private list of "two distances from one point, multiplied" → power of a point as a fixed reflex, the same way "distance to a fixed centre, constant" reads instantly as a circle.

6. Decision rule: when an optimisation problem's target quantity is a *product* of lengths sharing a common endpoint on a line that is otherwise free to rotate, look first for a circle through that endpoint realising the product as a power-of-a-point invariant, before reaching for calculus in a coordinate system.

1976 A2

1976 A2. Let $P(x,y) = x^2y + xy^2$ and $Q(x,y) = x^2 + xy + y^2$. For $n = 1, 2, 3, \ldots$, let $F_n(x,y) = (x+y)^n - x^n - y^n$ and $G_n(x,y) = (x+y)^n + x^n + y^n$. One observes that $G_2 = 2Q$, $F_3 = 3P$, $G_4 = 2Q^2$, $F_5 = 5PQ$, $G_6 = 2Q^3 + 3P^2$. Prove that, in fact, for each n either F_n or G_n is expressible as a polynomial in P and Q with integer coefficients.

Caveat. The MAA OCR renders Q as "O" throughout and drops several exponents and subscripts in the sample identities (e.g. "$0, = 20, F, = 3P, 0, = 202, F, = 5PO, G. = 20' + 3P'$"). The statement above reconstructs this as $G_2 = 2Q$, $F_3 = 3P$, $G_4 = 2Q^2$, $F_5 = 5PQ$, $G_6 = 2Q^3 + 3P^2$, which was verified independently in Step 3 below by direct expansion, and which is internally consistent with the recursion the book's own solution states. No other reconstruction of these five sample identities is consistent with both the algebra and the parity pattern (even n gives G, odd n gives F) that the induction requires.

Fields: algebra, symmetric functions **Position:** A2 **Answer:** F_n is a polynomial in P, Q with integer coefficients when n is odd; G_n is, when n is even. (Never both, in general — only one of the two families is polynomial in P, Q for a given n.)

Layer 1 THE WHAT — surface analysis

The natural attempt. The five given identities look like a pattern to be spotted and generalised by brute expansion. A natural first move is to write $s = x + y$, $p = xy$ — the two elementary symmetric functions everyone reaches for on sight of $x^n + y^n$ — since $F_n = s^n - x^n - y^n$ and $G_n = s^n + x^n + y^n$ are visibly symmetric, and power sums in x, y satisfy the standard Newton recursion $x^n + y^n = s(x^{n-1} + y^{n-1}) - p(x^{n-2} + y^{n-2})$. You could grind out F_n, G_n in terms of s and p this way, term by term, and then try to re-express s, p in terms of $P = ps$ and $Q = s^2 - p$ afterwards. This works in principle, but it detours through a second symmetric-function system (s, p) that the problem never mentions, and translating a polynomial identity in s, p into one in P, Q at the end requires solving $Q = s^2 - p$, $P = ps$ for p in terms of Q and s^2, reintroducing s — which is not itself a polynomial in P, Q (you'd need $s^3 = sQ + P$, a cubic in s, not a formula for s). The detour through s, p actively obstructs the very translation the problem is asking for.

The model chain. Work with P, Q directly, and find a recursion that stays inside the P, Q system without ever solving for s or p alone. Two elementary identities do this in one step: $s^n = s^{n-2} \cdot Q + s^{n-3} \cdot P$ and $x^n + y^n = (x^{n-2} + y^{n-2}) \cdot Q - (x^{n-3} + y^{n-3}) \cdot P$. Both reduce, after dividing through by the appropriate common factor, to the single degree-3 identity $s^3 = sQ + P$, which is immediate: $sQ + P = s(x^2 + xy + y^2) + xy \cdot s = s(x^2 + 2xy + y^2) = s \cdot s^2 = s^3$. Subtracting and adding the two displayed identities term by term gives the recursion

$$(R): \qquad F_n = Q\,F_{n-2} + P\,G_{n-3}, \qquad G_n = Q\,G_{n-2} + P\,F_{n-3}.$$

This recursion reaches back three steps at once (via the G_{n-3} or F_{n-3} term), so a single previous value is not enough to seed it; the problem statement hands over exactly the five values needed to seed it correctly: G_2, F_3, G_4, F_5, G_6. Now induct on $n \geq 7$: if n is odd, $F_n = QF_{n-2} + PG_{n-3}$ requires F_{n-2} (odd index, already polynomial in P, Q by the inductive hypothesis, since $n - 2 < n$) and G_{n-3} (even index, likewise already known); if n is even, symmetrically $G_n = QG_{n-2} + PF_{n-3}$ requires G_{n-2} (even) and F_{n-3} (odd), both already established. Since Q, P are themselves polynomials in x, y with integer coefficients, and multiplying and adding polynomials-in-P, Q stays inside the ring of polynomials in P, Q with integer coefficients, the induction closes: F_n is such a polynomial for every odd n, G_n for every even n.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree that the target is a recursion connecting F_n, G_n to lower-index terms. They split at the choice of coordinate system to carry that recursion in: the natural attempt reaches for $s = x + y$, $p = xy$ — the default pair for any symmetric expression in two variables — while the model insists on working in P, Q throughout, because P, Q are what the answer must be expressed in, and s, p are not obtainable from P, Q by a polynomial formula.

> *The problem already hands me the target alphabet, P and Q — I should build the recursion inside that alphabet from the start, not in s, p and translate afterwards, because translating back out of s, p means solving a cubic for s, which isn't a polynomial operation at all.*

Layer 3 THE WHY

The instinct to reach for s, p is not wrong in general — it is the correct move for almost any two-variable symmetric expression — but it is wrong *here* specifically because the problem is not asking "find a formula for F_n": it is asking "show F_n lies in a particular subring," namely $\mathbb{Z}[P, Q]$. A subring membership question has to be answered using operations available inside that subring: addition, multiplication, integer coefficients. Solving for s in terms of P, Q is not such an operation, because s satisfies $s^3 = sQ + P$, a genuinely cubic relation — s is algebraic over $\mathbb{Z}[P, Q]$ of degree 3, not a member of it. So any route through s, p that ends by "solving for s" has silently left the ring the problem lives in, and the fact that the final formula happens to simplify back down is not guaranteed and not efficient to arrange. The cubic identity $s^3 = sQ + P$ is exactly the seed of the correct approach in disguise: it says that multiplying by s (i.e. going from an F or G built from s^k to one built from s^{k+1}, or from a power sum of degree k to degree $k + 1$) can be re-expressed as a combination of $\times Q$ (drop two degrees in s, symmetric) and $\times P$ (drop three degrees in s, symmetric) — which is precisely a degree-2/degree-3 recursion in n, matching the F_{n-2}, G_{n-3} (respectively G_{n-2}, F_{n-3}) shape. The alternation between F and G as n runs through consecutive integers is likewise not a coincidence to be checked case by case: F_n is built from the odd-symmetric combination $s^n - x^n - y^n$, which changes sign under $x \leftrightarrow -x, y \leftrightarrow -y$ in a way tied to parity of n, and the recursion's index arithmetic ($n - 2$ preserves parity, $n - 3$ flips it) automatically routes F to F and G to G along parity-consistent chains — which is exactly why the five seed values alternate G, F, G, F, G for $n = 2, \ldots, 6$.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that the target alphabet (P, Q) is given by the problem, not left to be

chosen. On seeing that, write down: any recursion must be built directly in P, Q, because the answer needs to live inside $\mathbb{Z}[P, Q]$.

2. That cue suggests testing whether multiplication by $s = x + y$ — the operation that increases every power sum's degree by one — can itself be written using only P, Q, since that is the operation any degree-by-degree recursion needs to control.

3. Hesitation occurs at the point of choosing s, p out of habit, because that is the standard toolkit for symmetric functions in two variables. It is reinforced away by noticing that s is not a polynomial in P, Q at all (it satisfies a genuine cubic, $s^3 = sQ + P$) — so any route through s, p cannot close inside the target ring without extra work, and often none at all cleanly.

4. The trap is treating "find the pattern" as "expand more terms and eyeball it." The five given identities are not decoration; they are the exact seed values a index-3-back recursion requires, and skipping past them to search for a closed form directly (e.g. guessing F_n, G_n in terms of the roots of $t^2 - st + p$) wastes the one piece of scaffolding the problem supplies for free.

5. To recognise this faster: when a problem defines two target building blocks (here P, Q) and hands over several worked small cases, check first whether "multiplying by the natural symmetric generator" (here s) can be re-expressed purely in the target blocks — that single identity is usually the entire recursion in disguise.

6. Decision rule: to prove a family of symmetric expressions lies in a subring generated by two given symmetric building blocks, find how the ring's implicit generator (the elementary symmetric function that increases degree by one) multiplies in terms of those blocks, and let that identity drive an induction that never leaves the target ring.

1976 A3

1976 A3. Find all integral solutions of the equation $|p^r - q^s| = 1$, where p and q are prime numbers and r and s are positive integers larger than unity. Prove that there are no other solutions.

Caveat. The OCR renders the exponents as "$Y - 23 = 1$" and drops several inline superscripts in the solution text. The statement above is standard and unambiguous once the garbling is stripped; it is confirmed both by the topic index ("Number theory ... 1976 A-3") and by the book's own solution, which explicitly derives the unique answer $(p, r, q, s) = (3, 2, 2, 3)$ or $(2, 3, 3, 2)$, i.e. $3^2 - 2^3 = 1$.

Fields: number theory (elementary, a special case of Catalan's conjecture) **Position:** A3 **Answer:** The only solution, up to swapping the two prime powers, is $3^2 - 2^3 = 1$, i.e. $(p, r, q, s) = (3, 2, 2, 3)$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The equation says two prime powers, each with exponent at least 2, differ by exactly 1 — so they are consecutive integers. A natural first move is small-case hunting: list perfect prime powers with exponent ≥ 2 $(4, 8, 9, 16, 25, 27, 32, 49, \ldots)$ and look for adjacent pairs. This immediately finds $8, 9$ and nothing else in a long stretch, which is suggestive but proves nothing — the problem demands a proof that no further solution exists anywhere, and no amount of table-building settles that. The natural continuation is then to try to bound the search: since consecutive integers can't both be even, exactly one of p^r, q^s is even, forcing one of p, q to equal 2 (the only even prime, and a power of an odd prime is odd). That's real progress and is where both chains agree. But the tempting next step is to treat r, s generically and attempt modular arithmetic mod small numbers (mod 3, mod 4, mod 8...) hunting for a contradiction that rules out all $r, s > 2$ at once — a search without a clear endpoint, since no single modulus

rules out every case uniformly, and you can spend the rest of the session cycling through moduli without ever landing the argument that actually closes it, which is the parity of r itself.

The model chain. As above, one of p, q equals 2; say $q = 2$ (the case $p = 2$ is symmetric). Then p is an odd prime with $p^r = 2^s \pm 1$. **First split on the parity of r.** If r is odd, factor: $p^r \mp 1 = (p \mp 1)(p^{r-1} \pm p^{r-2} + \cdots \pm 1)$, and the second factor is a sum of r terms each ± 1, hence an odd integer, and it is strictly greater than 1 when $r > 1$ (since $p > 1$ makes the terms grow, or in the -1 case telescope to something exceeding 1 once $r > 1$). But $p^r \mp 1 = \pm 2^s$ (up to sign) has no odd factor greater than 1 — contradiction. So r must be even, $r = 2t$. **Now write** $p^t = 2n + 1$ (odd, since p is odd) and consider the two sign cases separately. If $p^r + 1 = 2^s$: $(2n+1)^2 + 1 = 4n^2 + 4n + 2 = 2(2n^2 + 2n + 1)$, and $2n^2 + 2n + 1$ is odd, so this is $2 \times (\text{odd})$ — forcing $s = 1$, but $s > 1$ is required, contradiction (equivalently: 4 does not divide $4n^2 + 4n + 2$, while $4 \mid 2^s$ for any $s > 1$). If $p^r - 1 = 2^s$: $(2n+1)^2 - 1 = 4n^2 + 4n = 4n(n+1)$, and since $n, n+1$ are consecutive, exactly one is odd; for $4n(n+1)$ to be a pure power of 2, that odd factor must itself equal 1. Either $n = 0$ (giving $p^t = 1$, impossible for a prime p) or $n = 1$ (giving $p^t = 3$, hence $p = 3, t = 1, r = 2$, and $n(n+1) = 2$ so $4n(n+1) = 8 = 2^s$, $s = 3$). This is the unique solution: $p = 3, r = 2, q = 2, s = 3$, i.e. $3^2 - 2^3 = 1$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree up through "one of p, q is 2." They split at the line "now bound r, s": the natural attempt reaches for modular arithmetic on the whole equation at once, searching for a modulus that kills every $r, s > 1$ simultaneously. The model instead asks a narrower, sequential question — first, what does the parity of the exponent r alone force? — and only after pinning r to be even does it look at the resulting expression mod small numbers.

> *I don't need one modulus to rule out every exponent at once — I should split on the parity of r first, because an odd exponent lets me factor $p^r \mp 1$ using the standard sum-of-a-geometric-series factorisation, which converts the exponent into an explicit odd cofactor, and 2^s has no odd factor greater than 1 to match it.*

Layer 3 THE WHY

The reason "split on parity of r first" is the natural next step, and not merely one option among many, is that the equation has already told you, structurally, that one side is a pure power of 2 — and powers of 2 have exactly one property that is easy to use against an arbitrary integer: they have no odd divisor greater than 1. That fact is only accessible if you can produce an explicit odd divisor of the other side, $p^r \mp 1$. The factorisation $a^r - b^r = (a-b)(a^{r-1} + a^{r-2}b + \cdots + b^{r-1})$ produces exactly such a divisor, but only when r is odd (so that the r-term alternating or non-alternating sum stays intact); for even r the same factorisation instead splits $a^r - b^r$ into $(a-b)(a+b)$ times more of the same, along a different route entirely (difference of squares, iterated). So the parity of r is not incidental — it determines which factorisation identity is even legal to reach for, which is why it must be the first case split, before any modular arithmetic: modular arithmetic mod a fixed small number is a blunt instrument that can rule out isolated residues, but it cannot, by itself, manufacture the divisor structure that "2^s has no odd factor > 1" needs to bite. Once r is forced even, p^t becomes the natural new variable (an odd number, since p is odd), and $(p^t)^2 \pm 1$ is now a difference or sum of two squares in disguise — the same $4n(n+1)$ versus $4n^2 + 4n + 2$ split that closes the two remaining cases is nothing more than expanding $(2n+1)^2 \pm 1$ and reading off the power of 2 dividing it, which is now tractable precisely because the exponent problem has been reduced to a parity-of-a-single-integer problem.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "$|p^r - q^s| = 1$ with p, q prime and $r, s > 1$" — two prime powers differing by exactly 1. On seeing it, write down: one of p, q must be 2, since consecutive integers can't both be even and a power of an odd prime is always odd.

2. That cue suggests reducing to $p^r = 2^s \pm 1$ for an odd prime p, and then asking what structural fact about powers of 2 can be brought to bear — specifically, that 2^s has no odd divisor exceeding 1.

3. Hesitation occurs right after the reduction, when the instinct is to search for a single modulus that eliminates all remaining (r, s) at once. It is reinforced away by noticing that the tool actually needed — an explicit odd factor of $p^r \mp 1$ — only exists when r is odd, via the geometric-series factorisation; so the case split on parity of r is not optional, it's what makes the "no odd factor" fact usable at all.

4. The trap is trying to handle r and s symmetrically, or bounding both at once with a blanket modular argument, rather than noticing that only one of the two exponents (r, attached to the odd prime p) is the one whose parity controls which factorisation is legal.

5. To recognise this faster: whenever an equation forces one side to be a pure prime power (here 2^s), immediately inventory that prime's one usable property (here: no odd factor > 1) and then ask which algebraic manipulation of the other side would produce a divisor testable against that property — factorisations keyed to the parity of an exponent are the first thing to check.

6. Decision rule: when a Diophantine equation pins one side to be a pure power of a single prime, do not search moduli blindly — split on the parity of the exponent on the other side first, since that parity determines which factorisation identity is available to expose a divisor the pure prime power cannot have.

1976 A4

1976 A4. Let r be a root of $P(x) = x^3 + ax^2 + bx - 1 = 0$ and $r + 1$ be a root of $y^3 + cy^2 + dy + 1 = 0$, where a, b, c, and d are integers. Also let $P(x)$ be irreducible over the rational numbers. Express another root s of $P(x) = 0$ as a function of r which does not explicitly involve a, b, c, or d.

Fields: algebra (polynomials, Vieta's formulas, irreducibility), with a Galois-theoretic shadow **Position:** A4 — mid-paper; one structural observation (the second cubic *is* the first one shifted) followed by a two-line Vieta computation **Answer:** $s = -\dfrac{1}{r+1}$; the third root is then $-\dfrac{r+1}{r}$, and either expression is a correct answer.

Caveat. The OCR renders the second cubic as "$y^3 + cy^2 + dy$ ..r $1 = 0$". The book's own solution writes it as $M(x) = x^3 + cx^2 + dx + 1$ and uses the constant term $+1$, so the statement above restores "$+1$". With that reading the solution under the label 1976 A-4 solves exactly this statement, and the answer has been checked numerically for $a = 0, 1, 2, -5, 7$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You have a cubic with unknown integer coefficients and a condition on a second cubic. The instinct is to turn the second condition into a condition on a, b. Substitute $x = y - 1$ into P: the number $r + 1$ is a root of

$$P(y-1) = (y-1)^3 + a(y-1)^2 + b(y-1) - 1 = y^3 + (a-3)y^2 + (3-2a+b)y + (a-b-2).$$

If you are willing to believe that this is the same polynomial as $y^3 + cy^2 + dy + 1$, comparing constant terms gives $a - b - 2 = 1$, so $b = a - 3$ and

$$P(x) = x^3 + ax^2 + (a-3)x - 1.$$

Now you want the other roots. Divide out $x - r$: by Vieta the other two roots s, t satisfy $st = 1/r$ (the product of all three is 1) and $s + t = -a - r$. The instructions forbid a in the answer, so eliminate it using $P(r) = 0$:

$$a(r^2 + r) = 1 + 3r - r^3, \qquad a = \frac{1 + 3r - r^3}{r(r+1)}, \qquad a + r = \frac{r^2 + 3r + 1}{r(r+1)}.$$

So s, t are the roots of $x^2 + \dfrac{r^2 + 3r + 1}{r(r+1)}x + \dfrac{1}{r} = 0$. Now the quadratic formula: the discriminant is $\big((r^2 + 3r + 1)^2 - 4r(r+1)^2\big) / \big(r^2(r+1)^2\big)$, and the numerator is $r^4 + 6r^3 + 11r^2 + 6r + 1 - (4r^3 + 8r^2 + 4r)$ — you are now expanding quartics, hoping for a perfect square. It is one, $r^4 + 2r^3 + 3r^2 + 2r + 1 = (r^2 + r + 1)^2$ after the dust settles, and the roots $\big(-(r^2 + 3r + 1) \pm (r^2 + r + 1)\big) / \big(2r(r+1)\big)$ come out. The route works. It costs a page of algebra, and it quietly relied on an identification ("this is the same polynomial") that it never justified.

The model chain. *Step 1: the two cubics are the same cubic, shifted.* P is irreducible over $\mathbb{Q}$ and has r as a root, so it is the minimal polynomial of r; hence $r + 1$ also has degree 3 over $\mathbb{Q}$, and its minimal polynomial is the monic cubic $P(y - 1)$. The polynomial $M(y) = y^3 + cy^2 + dy + 1$ has rational coefficients and $r + 1$ as a root, so the minimal polynomial divides it. Both are monic of degree 3, so $M(y) = P(y - 1)$. Consequently, if the roots of P are r, s, t, the roots of M are $r + 1, s + 1, t + 1$. Irreducible polynomials over $\mathbb{Q}$ have distinct roots, so r, s, t are distinct; and $r \neq 0, -1$ because P has no rational root.

Step 2: read the two constant terms with Vieta. For the monic cubic P with constant term -1, $rst = 1$. For the monic cubic M with constant term $+1$, $(r + 1)(s + 1)(t + 1) = -1$. Therefore

$$st = \frac{1}{r}, \qquad (s+1)(t+1) = -\frac{1}{r+1}.$$

Expanding the second, $st + (s + t) + 1 = -\frac{1}{r+1}$, so

$$s + t = -\frac{1}{r+1} - 1 - \frac{1}{r} = -\frac{r^2 + 3r + 1}{r(r+1)}.$$

Step 3: guess and check the factorisation. Two numbers with product $\frac{1}{r}$ and sum $-\frac{r^2+3r+1}{r(r+1)}$: try $-\frac{1}{r+1}$ and $-\frac{r+1}{r}$. Their product is $\frac{1}{r}$, and their sum is $-\frac{r+(r+1)^2}{r(r+1)} = -\frac{r^2+3r+1}{r(r+1)}$. Both match, and a pair of numbers is determined by its sum and product (they are the roots of one quadratic), so

$$\{s, t\} = \left\{-\frac{1}{r+1}, -\frac{r+1}{r}\right\}.$$

Either one is an admissible answer; take $s = -1/(r + 1)$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start from the same two facts, r is a root of P and $r + 1$ is a root of M. At that point the natural attempt converts the second fact into a condition on the *coefficients* ($b = a - 3$) and then has to eliminate a. The model converts it into a statement about the *roots* ($(r + 1)(s + 1)(t + 1) = -1$) and never meets a at all.

> *The answer is not allowed to contain a, b, c, d. So I should not be solving for them in the first place. What do I know about the roots that does not mention coefficients? The constant terms: $rst = 1$, and $(r + 1)(s + 1)(t + 1) = -1$ if the roots of the second cubic are the shifted roots of the first. That gives me st and $s + t$ in terms of r alone.*

> *Why does the problem bother telling me P is irreducible? Because that is exactly what I need to say the second cubic is $P(y-1)$ and not some other cubic that happens to share the root $r+1$.*

Layer 3 THE WHY

Start from the instruction "does not explicitly involve a, b, c, d". That sentence tells you what the answer is made of: r and nothing else. Vieta's formulas express symmetric functions of the roots through the coefficients; read in reverse, they tell you that any information about roots that holds *independently of the coefficients* must come from the parts of the polynomials the problem has fixed. What has the problem fixed? Only the leading coefficients (both 1) and the constant terms (-1 and $+1$). The constant term of a monic cubic is minus the product of its roots. So the problem has handed you exactly two products, rst and $(r+1)(s+1)(t+1)$, and nothing else. Two equations, two unknowns s, t: the whole computation is forced once you see which data are free of a, b, c, d.

The one thing that is not automatic is that the roots of M are $r+1, s+1, t+1$. A priori M is only known to share one root with $P(y-1)$. This is where irreducibility enters, and it enters through the definition of the minimal polynomial: the monic rational polynomial of least degree with r as a root divides every rational polynomial with r as a root. Irreducible and monic means P is that minimal polynomial. Shifting by the rational number 1 changes nothing about degree, so $P(y-1)$ is the minimal polynomial of $r+1$, and a monic cubic divisible by a monic cubic is equal to it. Drop irreducibility and the argument genuinely fails: if r were rational, M could be any cubic through $r+1$ with constant term 1, and s would not be determined by r.

Now the deeper reason the answer is a function of r at all. The map $\varphi(x) = -\dfrac{1}{x+1}$ satisfies

$$\varphi(\varphi(x)) = -\frac{1}{1-\frac{1}{x+1}} = -\frac{x+1}{x}, \qquad \varphi(\varphi(\varphi(x))) = x.$$

It has order 3, and it permutes the roots cyclically: $r \mapsto s \mapsto t \mapsto r$. So every root is a rational function of every other root. In the language beyond A-level, the splitting field is $\mathbb{Q}(r)$ itself and its Galois group is cyclic of order 3, generated by a Möbius transformation. With $b = a - 3$ and $a = -m$ the cubic is $x^3 - mx^2 - (m+3)x - 1$, Shanks' "simplest cubic"; with $a = 0$ you get $x^3 - 3x - 1$, whose roots are $2\cos(\pi/9), 2\cos(5\pi/9), 2\cos(7\pi/9)$, and the problem is the statement that one of them is a rational function of another. You do not need any of this to solve the problem, but it explains why a problem like this can exist: for a general cubic no rational expression in r gives another root.

Why does the natural attempt cost more? It routes the information through the coefficient a, then has to remove a again, and ends in a quadratic formula whose discriminant is a perfect square only because of the cyclic structure above — the solver meets that structure as an unexplained miracle. The model never lets a in.

Layer $+\alpha$ BACKTESTING

1. The cue is "as a function of r which does not explicitly involve a, b, c, d". Write down the Vieta relations that involve only fixed data: $rst = 1$ and $(r+1)(s+1)(t+1) = -1$.

2. Vieta on the constant terms of both cubics, after justifying that the second cubic's roots are the first's roots shifted by 1. Then solve for st and $s+t$ and factor the resulting quadratic by inspection.

3. Hesitation comes at "are the roots of M really $r+1, s+1, t+1$?" Reinforce with the minimal polynomial: irreducible and monic means it divides anything sharing a root, so equal degrees force equality. A second hesitation is at the quadratic; test the obvious candidates $-1/(r+1)$ and $-(r+1)/r$ (they already appeared in the computation of $s+t$) before reaching for the formula.

4. The trap is solving for a, b, c, d first. It works, but it adds an elimination and a quartic discriminant, and it invites you to skip the irreducibility step because the coefficient comparison "looks" automatic.

5. Whenever an answer must be expressed without the coefficients, look for relations among roots that use only the fixed parts of the polynomials (leading and constant terms, or known values). Whenever a problem states "irreducible", expect it to be used to identify a polynomial as a minimal polynomial.

6. Reusable rule: coefficients are the enemy when the answer must be in roots. Translate each piece of given data straight into a symmetric function of the roots; if two different polynomials share a root and one is irreducible, their root sets are linked by the substitution that relates them.

▷ **Cue:** "express another root as a function of r not involving the coefficients" $\longrightarrow$ **Move:** write Vieta only for the coefficients the problem fixes (usually leading and constant terms)

▷ **Cue:** "r is a root of P and $r+1$ (or $g(r)$) is a root of Q", with P irreducible $\longrightarrow$ **Move:** P is the minimal polynomial; Q is divisible by $P(y-1)$, so equal degree gives $Q(y) = P(y-1)$ and all roots shift together

▷ **Cue:** "irreducible over the rationals" in the hypotheses $\longrightarrow$ **Move:** it is there to invoke the minimal polynomial: distinct roots, no rational root, divides everything sharing a root

▷ **Cue:** a quadratic with known sum and product, both rational functions of a parameter $\longrightarrow$ **Move:** look at the partial fractions you produced while computing the sum; the roots are usually sitting there

▷ **Cue:** a cubic whose roots are rational functions of one another $\longrightarrow$ **Move:** look for a Möbius map of order 3 permuting the roots, such as $x \mapsto -1/(x+1)$

1976 A5

1976 A5. In the (x, y)-plane, if R is the set of points inside and on a convex polygon, let $D(x, y)$ be the distance from (x, y) to the nearest point of R. (a) Show that there exist constants a, b, and c, independent of R, such that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-D(x,y)}\,dx\,dy = a + bL + cA,$$

where L is the perimeter of R and A is the area of R. (b) Find the values of a, b, and c.

Fields: geometry (convex polygons, nearest-point projection), integral calculus (improper double integrals, polar coordinates) **Position:** A5 — late in the paper; the computation is short once the plane is cut up correctly, and the cutting is the whole problem **Answer:** $a = 2\pi$, $b = 1$, $c = 1$: the integral equals $2\pi + L + A$.

Caveat. The OCR of the displayed integral is garbled ("$J_{\sim}J_{\sim}e^{-D<x,>}$"). The reading $\iint_{\mathbb{R}^2} e^{-D(x,y)}\,dx\,dy$ is the only one consistent with the book's solution under 1976 A-5, which computes exactly this integral region by region and obtains $2\pi + L + A$. The answer was also checked numerically for the unit square ($2\pi + 5 \approx 11.2832$).

Layer 1 THE WHAT — surface analysis

The natural attempt. Part (a) says "show there exist constants". A strong student's first reaction is to test cases and let them determine a, b, c, then worry about (a) afterwards.

Try the simplest R you can think of. If R shrinks to a single point (a degenerate polygon, $L = A = 0$), then D is the distance ρ to that point and polar coordinates give

$$\int_0^{2\pi}\int_0^{\infty} e^{-\rho}\rho\, d\rho\, d\theta = 2\pi \cdot 1 = 2\pi,$$

so, if the formula holds at all, $a = 2\pi$. Next try a square of side s: outside the square, D depends on which of nine zones you are in. In the four edge zones D is the perpendicular distance, in the four corner zones it is the distance to the corner, and inside it is 0. You set up nine integrals, and they come out as s^2 for the inside, $s\int_0^{\infty} e^{-v}dv = s$ for each edge zone, and a quarter of 2π for each corner zone. Total $s^2 + 4s + 2\pi$. Consistent with $a = 2\pi$, $b = 1$, $c = 1$.

But you have not shown (a). The square is a lucky polygon: its corner zones are right angles and fit together into a full turn by symmetry. For a general triangle or pentagon the corner zones have different angles, and you do not yet know why they should always add up to the same 2π. Fitting constants from examples does not prove that a formula of the form $a + bL + cA$ exists. That is the wall.

The model chain. Write $I(S) = \iint_S e^{-D}\,dx\,dy$ for a region S. Let the vertices of R be $v_1, \dots, v_n$ in order and the edges $e_i = [v_i, v_{i+1}]$ of lengths ℓ_i (indices mod n), with outward unit normals $\mathbf{n}_i$.

The nearest-point map. R is closed, bounded and convex, so each point X has a unique nearest point $p(X) \in R$ (if p, q were two nearest points, their midpoint would lie in R and be strictly closer to X). For $X \notin R$ the vector $X - p(X)$ is perpendicular to any edge in whose relative interior $p(X)$ lies, since otherwise sliding p along the edge would decrease the distance.

Cut the plane. The plane is the union of the following sets, which overlap only along finitely many lines and rays (area zero):

- R itself, where $D = 0$;
- for each edge e_i, the half-strip $S_i = \{u + t\mathbf{n}_i : u \in e_i,\ t \geq 0\}$, where $D(u + t\mathbf{n}_i) = t$;
- for each vertex v_i, the wedge $W_i = \{v_i + \rho(\cos\theta, \sin\theta) : \rho \geq 0,\ \theta$ between the directions of $\mathbf{n}_{i-1}$ and $\mathbf{n}_i\}$, where $D = \rho$.

That these sets cover the plane: take $X \notin R$ and $p = p(X)$. If p is inside edge e_i, then $X = p + t\mathbf{n}_i$ with $t = D(X)$, so $X \in S_i$. If $p = v_i$, then $X - v_i$ makes a non-acute angle with both edges leaving v_i (else a point on that edge near v_i would be closer), and the set of such directions is precisely the angle between $\mathbf{n}_{i-1}$ and $\mathbf{n}_i$; so $X \in W_i$. Conversely on S_i and W_i the stated formula for D is correct, because the claimed nearest point satisfies the obtuse-angle condition that characterises the nearest point of a convex set.

The angle of W_i is the exterior angle $\alpha_i = \pi - \beta_i$, where β_i is the interior angle at v_i (the normals are each turned $\pi/2$ from their edges). The exterior angles of a convex polygon sum to 2π: $\sum(\pi - \beta_i) = n\pi - (n-2)\pi = 2\pi$.

Integrate piece by piece.

$$I(R) = A, \qquad I(S_i) = \int_0^{\ell_i}\int_0^{\infty} e^{-t}\,dt\,du = \ell_i, \qquad I(W_i) = \int_0^{\alpha_i}\int_0^{\infty} e^{-\rho}\rho\,d\rho\,d\theta = \alpha_i.$$

Adding (the overlaps have area zero, and all integrands are positive, so the improper integral is the sum of the pieces),

$$\iint_{\mathbb{R}^2} e^{-D} = A + \sum_i \ell_i + \sum_i \alpha_i = A + L + 2\pi.$$

This proves (a) with $a = 2\pi$, $b = 1$, $c = 1$, which is (b). The constants are unique: the point, a segment of length ℓ (with $L = 2\ell$, $A = 0$) and a unit square already pin them down.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains know how to integrate e^{-D} over a single zone. They split at the question "how many kinds of zone are there?" The natural attempt answers it separately for each shape; the model classifies points by *what their nearest point in R is* — interior, edge or vertex — which gives the same three kinds of zone for every convex polygon.

> *D is defined through "the nearest point of R". So I should sort the points of the plane by their nearest point. For a polygon that point is inside R, on an edge, or at a vertex. Three types of region, and I can integrate each type in general.*

> *The corner zones add up to 2π for the square. Is that the square's symmetry or something always true? The corner zone's angle is the angle between two outward normals, which is the exterior angle, and the exterior angles of any convex polygon add to one full turn.*

Layer 3 THE WHY

Why sort by the nearest point? Because the problem defined D that way. The only fact you are given about D is "distance to the nearest point of R", and the only way to make that computable is to know which point is nearest. Once you ask the question, convexity answers it: the nearest point is unique, and the direction from it to X is an outward normal direction at that point. A polygon's boundary has only two kinds of points — edge points, with one normal direction, and vertices, with a fan of normal directions — so the outside of R is swept out by normal rays from edges (half-strips) and fans from vertices (wedges). This is the circle example from the Blueprint in another form: the centre is the object of interest in a circle because the circle is defined by distance from it; here the nearest point is the object of interest because D is defined by distance to it.

Why is the answer of the form $a + bL + cA$ at all? Think about the sets $\{D \le t\}$. By the same decomposition, $\{D \le t\}$ is R together with a rectangle $\ell_i \times t$ on each edge and a circular sector of radius t and angle α_i at each vertex, so

$$\operatorname{area}\{D \le t\} = A + Lt + \pi t^2.$$

That is Steiner's formula for the parallel body of a convex set, and it is the real reason for the answer. The integral of a function of D depends only on how the area is distributed across levels of D: with $V(t) = \operatorname{area}\{D \le t\}$,

$$\iint e^{-D} = A + \int_0^\infty e^{-t} V'(t)\, dt = A + \int_0^\infty e^{-t}(L + 2\pi t)\, dt = A + L + 2\pi.$$

So the three constants are $c = e^{-0} = 1$, $b = \int_0^\infty e^{-t} dt = 1$, $a = 2\pi \int_0^\infty t e^{-t} dt = 2\pi$, and the same argument gives $\iint f(D) = f(0)A + L\int_0^\infty f + 2\pi \int_0^\infty t f(t)\, dt$ for any integrable f. The exponential is a decoy; the geometry is Steiner.

Why does convexity matter? For a non-convex polygon, at a reflex vertex the normal fans would overlap the half-strips of neighbouring edges and the zones would double-count, so $\{D \le t\}$ loses area and the formula fails. The hypothesis "convex" is the licence for the clean decomposition.

The book's solution states the half-strip and wedge decomposition without saying why it covers the plane or why the wedge angles sum to 2π; those are the two facts written out in the model chain, and they are exactly the facts that make the answer independent of R.

Layer $+\alpha$ BACKTESTING

1. The cue is "distance to the nearest point of R". Write down: for each X, its nearest point $p(X)$, and the three possible types of $p(X)$ (interior, edge, vertex).

2. Decompose the plane by the type of the nearest point, integrate e^{-D} on each piece in adapted coordinates (Cartesian along an edge, polar at a vertex), and add. Alternatively, compute area$\{D \le t\}$ and integrate against it.
3. Hesitation arises over whether the wedges have a controlled total angle. Reinforce: each wedge is bounded by the two outward normals at its vertex, so its angle is the exterior angle, and exterior angles of a convex polygon sum to 2π.
4. The trap is to find a, b, c from examples (point, square) and believe (a) has been proved. Examples determine the constants if the formula holds; they do not show that it holds.
5. When a function is defined by "nearest point" or "distance to a set", classify points by their nearest point first. When an integral of a function of a distance is asked for, compute the area of the level sets $\{D \le t\}$ and integrate against it.
6. Reusable rule: integrals of f(distance to K) over the plane depend on K only through the areas of its parallel bodies; for convex K these are $A + Lt + \pi t^2$, so the answer is always a combination of area, perimeter and a universal constant.

▷ **Cue:** "$D(x, y)$ = distance from (x, y) to the nearest point of a convex set" ⟶ **Move:** sort points by their nearest point: interior, edge (half-strip), vertex (wedge)
▷ **Cue:** integral over the plane of a function of the distance to a set ⟶ **Move:** find area$\{D \le t\}$ and integrate f against its derivative (layer-cake)
▷ **Cue:** wedges at the vertices of a convex polygon bounded by outward normals ⟶ **Move:** their angles are exterior angles and sum to 2π; they reassemble into a disc
▷ **Cue:** "show there exist constants independent of R" plus "find them" ⟶ **Move:** prove the general formula by decomposition; use degenerate cases (a point) only as a check
▷ **Cue:** area, perimeter and a constant appearing together for a convex shape ⟶ **Move:** Steiner's formula area$(K_t) = A + Lt + \pi t^2$ is behind it

1976 A6

1976 A6. Suppose $f(x)$ is a twice continuously differentiable real-valued function defined for all real numbers x and satisfying $|f(x)| \le 1$ for all x and $(f(0))^2 + (f'(0))^2 = 4$. Prove that there exists a real number x_0 such that $f(x_0) + f''(x_0) = 0$.

Caveat. The MAA OCR for this problem is unreliable and project search could not retrieve a clean, correctly-aligned block of book text for 1976 A-6 despite repeated targeted queries; per the task's authoritative override, the statement above is taken as given and the solution below is constructed and fully verified independently, not transcribed from the OCR. It agrees with the well-known published solution to this problem (build $G = f^2 + f'^2$, use the Mean Value Theorem twice, argue by an interior extremum), so no substantive reconstruction risk is believed to remain, but the exact wording/attribution in the 1965–1984 volume is not confirmed here.

Fields: real analysis (mean value theorem, extremal principle, second-order ODE intuition) **Position:** A6 — last problem of the morning session; the difficulty is entirely conceptual (one auxiliary function, one clean argument), not computational **Answer:** Let $G(x) = f(x)^2 + f'(x)^2$. Mean Value Theorem on $[0, 2]$ and $[-2, 0]$ gives points $c \in (0, 2)$, $d \in (-2, 0)$ with $|f'(c)|, |f'(d)| \le 1$, so $G(c), G(d) \le 2 < 4 = G(0)$. Hence G has an interior maximum $x_0 \in (d, c)$, where $G'(x_0) = 2f'(x_0)(f(x_0) + f''(x_0)) = 0$. If $f'(x_0) = 0$ then $G(x_0) = f(x_0)^2 \le 1$, contradicting $G(x_0) \ge G(0) = 4$. So $f(x_0) + f''(x_0) = 0$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The pair $(f(0), f'(0))$ satisfies $f(0)^2 + f'(0)^2 = 4$: it sits on a circle of radius 2 in the (y, y')-plane. A well-prepared student recognises that circle immediately, because

it is exactly the invariant of simple harmonic motion: if $y'' + y = 0$ then $y^2 + (y')^2$ is constant along any solution, and a solution with that constant equal to 4 is $y(x) = 2\cos(x - \theta)$ for some phase θ with $f(0) = 2\cos\theta$, $f'(0) = 2\sin\theta$. So the natural first move is to compare f to this model solution: if f actually satisfied $f'' + f = 0$ everywhere, it would have to be $2\cos(x - \theta)$, and that function takes the value 2 at $x = \theta$ — flatly contradicting $|f(x)| \le 1$. This is exactly the right instinct: the amplitude-4 energy is too large for a genuine SHM solution to stay inside the band $[-1, 1]$, so *something* must interrupt the identity $f'' + f = 0$ before f can grow that large, and the problem is asking you to locate the interruption.

The trouble is turning "f cannot behave like $2\cos(x - \theta)$ for long" into a proof about the actual f, which is not assumed to solve any differential equation at all — only $|f| \le 1$ and the value of $G(0)$. The natural next step is a comparison argument: suppose, for contradiction, that $f'' + f \neq 0$ everywhere, so $f'' + f$ has a constant sign (it is continuous, so if it changed sign it would vanish by the Intermediate Value Theorem, which is the answer already). Say $f''(x) > -f(x)$ for all x. A student now tries to convert this pointwise inequality on f'' into a growth bound on f or on G, typically by attempting a Sturm–type comparison with the true SHM solution $2\cos(x - \theta)$, or by trying to integrate the differential inequality $f'' > -f$ twice. Both routes demand machinery (comparison theorems for second-order ODEs, or careful sign-tracking of $f'' + f$ combined with the unknown, possibly changing, sign of f') that is far beyond what a single Putnam problem should need, and a student who pushes down this road typically produces a partial, case-heavy argument that never quite closes, because the inequality on f'' alone does not control f without also controlling f' at some starting point — which is precisely the missing ingredient the model chain supplies for free.

The model chain. *Step 1: name the energy function.* Let $G(x) = f(x)^2 + f'(x)^2$ for all real x. Since f is twice continuously differentiable, f' is differentiable, so G is differentiable, with

$$G'(x) = 2f(x)f'(x) + 2f'(x)f''(x) = 2f'(x)\big(f(x) + f''(x)\big).$$

This is the whole problem in one line: G' vanishes exactly where either $f' = 0$ or $f + f'' = 0$, and the second alternative is the conclusion wanted.

Step 2: bound G away from 0 using the Mean Value Theorem. f is differentiable on , so on $[0, 2]$ there is $c \in (0, 2)$ with

$$f'(c) = \frac{f(2) - f(0)}{2}, \qquad |f'(c)| \le \frac{|f(2)| + |f(0)|}{2} \le \frac{1 + 1}{2} = 1,$$

using $|f| \le 1$. Likewise on $[-2, 0]$ there is $d \in (-2, 0)$ with $|f'(d)| \le 1$. Since also $|f(c)|, |f(d)| \le 1$,

$$G(c) = f(c)^2 + f'(c)^2 \le 1 + 1 = 2, \qquad G(d) \le 2.$$

Meanwhile $G(0) = f(0)^2 + f'(0)^2 = 4$.

Step 3: force an interior extremum. G is continuous on the closed interval $[d, c]$, so it attains a maximum there. Since $G(0) = 4 > 2 \ge G(d), G(c)$ and $0 \in (d, c)$, that maximum is not attained at either endpoint (its value is $\ge G(0) = 4$, strictly more than the value at d or c). So the maximum is attained at some $x_0 \in (d, c)$, an interior point of the interval on which G is differentiable, and by Fermat's theorem $G'(x_0) = 0$.

Step 4: resolve the factorisation. By Step 1, $2f'(x_0)\big(f(x_0) + f''(x_0)\big) = 0$, so $f'(x_0) = 0$ or $f(x_0) + f''(x_0) = 0$. Suppose $f'(x_0) = 0$. Then $G(x_0) = f(x_0)^2 \le 1$, since $|f(x_0)| \le 1$. But $G(x_0)$ is the maximum of G on $[d, c]$, hence $G(x_0) \ge G(0) = 4$. This gives $4 \le G(x_0) \le 1$, absurd. So $f'(x_0) \neq 0$, and the only surviving alternative is

$$f(x_0) + f''(x_0) = 0,$$

which is exactly the statement to be proved. ■

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains reach the same recognition: $(f(0), f'(0))$ lies on the radius-2 "energy circle" of $y'' + y = 0$, and a genuine solution with that energy would leave $[-1, 1]$. The paths split on what to do with that recognition. The natural attempt tries to compare the unknown f directly to the model solution $2\cos(x - \theta)$, term by term, which needs a comparison theorem for the differential inequality $f'' > -f$. The model chain instead stops comparing functions and starts studying a single real number: it builds $G(x) = f(x)^2 + f'(x)^2$ — the same expression that makes the energy circle an energy circle — as an honest function of x on the whole line, and asks only where *that one function* has a critical point.

> *The pair $(f(0), f'(0))$ has length 2 — this is a point on the SHM energy circle $y^2 + (y')^2 = 4$. If f actually solved $f'' + f = 0$ it would be $2\cos(x - \theta)$, which hits 2 somewhere; but $|f| \leq 1$ forbids that. So $f'' + f = 0$ cannot hold identically, and the honest object to track is the energy $G = f^2 + f'^2$ itself, not a term-by-term comparison of f against the model solution.*

> *$G'(x) = 2f'(x)\big(f(x) + f''(x)\big)$ — this factors the exact quantity I need to vanish, $f + f''$, as one of two factors. So I do not need G' to vanish for its own sake; I need a point where $G' = 0$ and I can rule out the other factor, $f' = 0$, being the reason.*

> *I don't need to control f everywhere to compare G against $G(0)$ — I only need f' to be small at two isolated points on either side of 0, and the Mean Value Theorem hands me exactly that, for free, from the single hypothesis $|f| \leq 1$: the secant slope over any interval of length 2 is at most 1 in size.*

> *The maximum of G on $[d, c]$ cannot sit at d or c, because I already know a point in the interior, namely 0, where G is strictly larger than both endpoint values — so the maximiser is interior, G' vanishes there, and the factorisation from the first thought does the rest: $f' = 0$ there would force $G \leq 1$, contradicting that this is the maximum and it is at least $G(0) = 4$.*

Layer 3 THE WHY

Why build G at all, rather than work with f and f'' directly? Because the problem does not hand you $f(0)$ and $f'(0)$ separately — it hands you their sum of squares. A hypothesis stated as an invariant combination of a function and its derivative is a direct instruction about which auxiliary function to build: the combination itself, promoted from a number to a function of x. This is not a trick pulled from nowhere; it is the definition of the phase-space energy for the equation $y'' + y = 0$, and the problem's hypothesis $(f(0))^2 + (f'(0))^2 = 4$ is stated in exactly the language that makes that energy the natural object. Once G exists as a function, differentiating it is close to forced (it is the only thing you can do with it that connects it back to f''), and the factorisation $G' = 2f'(f + f'')$ is not a coincidence: it is the work–energy identity in disguise, $\frac{d}{dx}(\text{speed}^2 + \text{position}^2) = 2 \cdot \text{speed} \cdot \text{net acceleration relative to the restoring term}$. Recognising the mechanical analogy is what tells you to compute G' rather than, say, $f'' - f$ or some other combination.

Why does an interior extremum settle the question, rather than merely relate G' to zero? Because Fermat's theorem is the only tool available that converts "G is largest here" into "G' vanishes here" without any assumption on what kind of function G is — it needs nothing but differentiability at an interior point. The entire argument is designed to manufacture such a point cheaply: $G(0) = 4$ is large, and the Mean Value Theorem, using only $|f| \leq 1$, shows G must be smaller a bounded distance away on both sides. A maximum of a continuous function on a closed interval always exists; the only work is making sure it cannot hide at an endpoint,

and that is done by a single strict inequality, $G(0) > G(d), G(c)$, which the numbers $4 > 2$ deliver with room to spare.

Why does $f'(x_0) = 0$ have to fail at the maximiser? Because the hypothesis $|f(x)| \leq 1$ is being used a second time, in a different guise. The first use (Step 2) bounds a *secant slope* of f, through the Mean Value Theorem, to control G at two flanking points. The second use (Step 4) bounds a *value* of f, directly, to control G at the point that matters. The same single hypothesis is doing two structurally different jobs, and noticing that a hypothesis can be reused in a different form is itself a transferable habit: before discarding a condition as "already used," ask whether it bounds something else — a slope as well as a value, a sum as well as a product — that the rest of the argument still needs.

One more thing is worth verifying independently, because it sharpens the problem and explains why 4 was chosen generously rather than tightly. The argument above used a fixed interval of length 2 on each side, giving $|f'| \leq 1$ there and hence $G \leq 2$ at the flanking points. But the interval length was never forced to be 2: on $[0, L]$, the Mean Value Theorem gives $|f'(c)| \leq 2/L$, so $G(c) \leq 1 + 4/L^2$, and this bound tends to 1 as $L \to \infty$. So for *any* hypothesis $G(0) > 1$, a sufficiently large L makes $1 + 4/L^2 < G(0)$, and the same four-step argument goes through verbatim with $[d, c]$ replaced by a longer interval. The theorem therefore holds under the weaker hypothesis $G(0) > 1$, and this is sharp: $f \equiv 1$ satisfies $|f| \leq 1$ and $G(0) = 1 + 0 = 1$, yet $f(x) + f''(x) = 1 + 0 = 1 \neq 0$ for every x. The constant 4 in the original statement is not load-bearing; it was chosen only to make the fixed interval $[-2, 2]$ work without any optimisation over L, keeping the exam solution self-contained.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is $(f(0))^2 + (f'(0))^2 = 4$: a sum of squares of a function and its derivative, evaluated at a single point. Write down immediately: "this is $G(0)$ for $G = f^2 + f'^2$; this expression is worth promoting to a function of x."

2. That cue suggests differentiating G and seeing what its derivative is built from, in the hope that the factorisation lands on the quantity the problem asks about. It does: $G' = 2f'(f + f'')$.

3. Hesitation occurs at deciding whether an interior critical point of G actually proves the claim, since $G' = 0$ could come from $f' = 0$ instead of $f + f'' = 0$. It is reinforced away by checking what $f' = 0$ would force: $G(x_0) = f(x_0)^2 \leq 1$ at the maximiser, contradicting that the maximiser's value is at least $G(0)$.

4. The trap is trying to compare f to the exact SHM solution $2\cos(x - \theta)$ that shares its initial energy, which demands a comparison theorem for differential inequalities and rarely closes cleanly on a timed paper. A second, quieter trap is fixing the comparison interval at $[-2, 2]$ as though that length were forced by the constant 4, rather than seeing it as one convenient choice among many.

5. Whenever a hypothesis is stated as an invariant combination of a function and its derivative at a point, recognise it as the value of a phase-space "energy" function and build that function on the whole domain, rather than trying to reconstruct the function itself from the invariant.

6. If a hypothesis reads $y(a)^2 + y'(a)^2 = k$ and you must produce a point where $y + y''$ (or any similarly-factoring quantity) vanishes, form $G = y^2 + (y')^2$, compute $G' = 2y'(y + y'')$, use any independent bound on y (here $|y| \leq 1$) via the Mean Value Theorem to make G smaller a bounded distance from a on both sides, then take an interior extremum of G between those two points: the extremum forces $G' = 0$, and the same independent bound on y eliminates the unwanted factor $y' = 0$, leaving the wanted factor.

▷ **Cue:** hypothesis of the form $y(a)^2 + y'(a)^2 = k$, a sum of squares of a function and its derivative at a point $\longrightarrow$ **Move:** treat k as the value at a of the auxiliary function $G = y^2 + (y')^2$ defined on the whole domain, and study G's critical points

▷ **Cue:** $(f(0), f'(0))$ with $f(0)^2 + f'(0)^2$ equal to a value larger than $\sup |f|^2$ $\longrightarrow$ **Move:** f cannot solve $y'' + y = 0$ everywhere on this energy level without leaving the bound on $|f|$; the restoring identity must fail somewhere, and the failure is what is being asked for

- ▷ **Cue:** need to bound f' at a specific point using only a bound on $|f|$ $\longrightarrow$ **Move:** apply the Mean Value Theorem on an interval of chosen length L starting or ending at that point: $|f'| \leq 2 \sup |f| / L$, and L can be enlarged to make the bound as close to $\sup |f| \cdot 0$ as needed
- ▷ **Cue:** a differentiable function is known to be strictly larger at one interior point than at both ends of a closed interval $\longrightarrow$ **Move:** its maximum on that interval is attained at an interior point, and Fermat's theorem gives a zero derivative there
- ▷ **Cue:** a derivative factors as (something)$\times$(the quantity you must show vanishes), at a forced critical point $\longrightarrow$ **Move:** rule out the "something" factor using an independent bound, leaving the wanted factor as the only surviving alternative

Putnam 1976 — Paper B

1976 B1

1976 B1. Evaluate

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \left(\left[\frac{2n}{k} \right] - 2 \left[\frac{n}{k} \right] \right)$$

and express your answer in the form $\log a - b$, with a and b positive integers. Here $[x]$ is defined to be the integer such that $[x] \le x < [x] + 1$, and $\log x$ is the logarithm of x to base e.

Fields: Analysis (floor functions, Riemann sums) **Position:** B1 — opening problem of Paper B **Answer:** $\log 4 - 1$

Caveat. The book's OCR renders the summand as "`([ 2n]- 2 (!!.]) ·- n ,_, k k`", with the fraction bars and index dropped. The reconstruction above is fixed by three independent checks: (i) it is the only reading under which the stated answer form "$\log a - b$" is even well-posed, (ii) the book's own solution opens "Let $f(x) = [2/x] - 2[1/x]$. Then the desired limit L equals $\int_0^1 f(x)\,dx$", which only parses if the summand is $f(k/n)$ for f exactly as reconstructed, and (iii) I solved the resulting limit independently below and it produces $\log 4 - 1$, matching the book's stated $a = 4, b = 1$ before I read the details of its argument.

Layer 1 THE WHAT — surface analysis

The natural attempt. A strong student sees $[2n/k] - 2[n/k]$ and reads it as a discrete object: a sum over $k = 1, \ldots, n$ of an integer-valued quantity, divided by n. The instinct is to attack it arithmetically — write $n/k = q + r/k$ with $0 \le r < k$, expand $[2n/k]$ in terms of q and r, and try to get a clean closed form for the sum as a function of n, perhaps via a number-theoretic identity (sums of $[n/k]$ count divisors, after all, and this is a familiar shape from counting lattice points under a hyperbola). This is not a bad instinct in general — $\sum [n/k]$ genuinely does have that divisor-counting meaning — but here it fights the problem. $[2n/k] - 2[n/k]$ does not telescope or factor nicely in k for fixed n; the natural attempt spends real time trying to find a closed form for the inner sum before dividing by n and taking the limit, and that closed form does not exist in an elementary way. The attempt stalls, or produces an estimate that is correct in leading order but cannot pin down the additive constant that the answer's "$-b$" demands.

The model chain. Fix $x = k/n \in (0, 1]$ and notice that $[2n/k] - 2[n/k] = [2/x] - 2[1/x]$: the expression is a function of the ratio k/n alone, not of k and n separately. Define $f(x) = [2/x] - 2[1/x]$ for $x \in (0, 1]$. Then

$$\frac{1}{n} \sum_{k=1}^{n} \left([2n/k] - 2[n/k] \right) = \frac{1}{n} \sum_{k=1}^{n} f(k/n),$$

which is a right-endpoint Riemann sum for $\int_0^1 f(x)\,dx$ over the partition into n equal subintervals. Since f is bounded (it takes only the values 0 and 1, shown next) and has finitely many discontinuities on any compact subinterval away from 0, the Riemann sum converges to the integral provided the accumulation near $x = 0$ is controlled, which it is because f is bounded. So $L = \int_0^1 f(x)\,dx$.

Now evaluate f. Write $1/x = t \geq 1$ and $t = m + \theta$ with $m = [t]$ and $\theta \in [0,1)$. Then $2t = 2m + 2\theta$, so $[2t] = 2m + [2\theta]$, and $[2\theta] = 0$ if $\theta < 1/2$, $= 1$ if $\theta \geq 1/2$. Hence $f(x) = [2t] - 2[t] = [2\theta] \in \{0,1\}$, and $f(x) = 1$ exactly when $\theta = \{1/x\} \geq 1/2$.

For each integer $n \geq 1$, on the interval $x \in (1/(n+1), 1/n]$ we have $1/x \in [n, n+1)$, so $\theta = 1/x - n$, and $f(x) = 1$ iff $1/x \geq n + \frac{1}{2}$ iff $x \leq \dfrac{2}{2n+1}$. Since $\dfrac{1}{n+1} < \dfrac{2}{2n+1} < \dfrac{1}{n}$, the interval splits: $f = 1$ on $\left(\dfrac{1}{n+1}, \dfrac{2}{2n+1}\right]$, of length $\dfrac{1}{(n+1)(2n+1)}$, and $f = 0$ on the rest. So

$$L = \int_0^1 f(x)\,dx = \sum_{n=1}^{\infty} \frac{1}{(n+1)(2n+1)}.$$

Partial fractions give $\dfrac{1}{(n+1)(2n+1)} = \dfrac{2}{2n+1} - \dfrac{1}{n+1}$. Summing to N and writing everything in harmonic numbers $H_N = \sum_{j=1}^{N} 1/j$: the odd-reciprocal sum $\sum_{n=1}^{N} \frac{1}{2n+1} = H_{2N+1} - \frac{1}{2}H_N - 1$ (all reciprocals up to $2N+1$, minus the even ones, minus the $n = 0$ term $1/1$), so

$$\sum_{n=1}^{N} \left(\frac{2}{2n+1} - \frac{1}{n+1}\right) = 2H_{2N+1} - H_N - 2 - (H_{N+1} - 1) = 2H_{2N+1} - H_N - H_{N+1} - 1.$$

As $N \to \infty$, $H_{2N+1} = \ln(2N) + \gamma + o(1)$ and $H_N, H_{N+1} = \ln N + \gamma + o(1)$, so $2H_{2N+1} - H_N - H_{N+1} \to 2\ln 2 + 2\gamma - 2\ln N - 2\gamma = 2\ln 2 = \ln 4$ (the $\ln N$ terms cancel exactly because the coefficients $2 - 1 - 1 = 0$). Hence $L = \ln 4 - 1$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree that the object is a sum of a floor-function expression over k. They split at the very first move: whether to treat n as fixed and hunt for a closed form in k, or to recognise that $[2n/k] - 2[n/k]$ depends only on the ratio $x = k/n$, which reframes the whole sum as a Riemann sum in disguise.

> *This is $[2n/k] - 2[n/k]$, and if I write $x = k/n$ this is exactly $[2/x] - 2[1/x]$ — the n has vanished from the formula, which means dividing by n and summing n terms indexed by $k/n \in (0,1]$ is not an accident of the problem's phrasing, it is a Riemann sum wearing floor-function clothing, and I should stop trying to sum in k and start trying to integrate in x.*

Layer 3 THE WHY

The phrase "express your answer in the form $\log a - b$" is itself a structural hint that the natural attempt misses and the model chain is built to exploit. A closed-form arithmetic identity for $\sum_k([2n/k] - 2[n/k])$, if one existed, would most likely be a rational function of n or a sum of divisor-counting terms — objects that do not naturally produce a logarithm. A logarithm appearing in the limit is the fingerprint of a harmonic sum, and harmonic sums are what you get from integrating $1/x$-type functions or from Riemann sums of functions with a $1/x$ flavour. So the answer's own shape is evidence, available before any computation, that the k-sum should be reorganised as an integral in the variable $x = k/n \in (0,1]$ rather than manipulated as a function of the integer k.

The deeper reason the reframing is legitimate, not merely convenient, is that $[2n/k] - 2[n/k]$ is *scale-covariant*: replacing (n,k) by (cn,ck) for a positive integer c leaves the expression

unchanged, because $[2n/k] = [2(cn)/(ck)]$. A quantity built only from the ratio k/n is, by definition, a function sampled along a partition of $[0,1]$, and $\frac{1}{n}\sum$ of n samples along an equal partition of $[0,1]$ is the definition of a right Riemann sum. Once $f(x) = [2/x] - 2[1/x]$ is written down, its structure — an integer part of $1/x$ split by the fractional-part threshold $1/2$ — is forced by the definition of the floor function itself: $[2t] - 2[t]$ depends only on whether $\{t\}$ is above or below $1/2$, because that is precisely what doubling does to a fractional part. Nothing here is a trick; it is floor arithmetic read at the level of its definition.

- ▷ **Cue:** a floor-function expression evaluated at k turns out, on inspection, to depend only on k/n ⟶ **Move:** stop summing in k; substitute $x = k/n$ and recognise $\frac{1}{n}\sum f(k/n)$ as a Riemann sum for $\int_0^1 f(x)\,dx$
- ▷ **Cue:** $[ct] - c[t]$ for integer c, as a function of the fractional part $\theta = \{t\}$ ⟶ **Move:** it equals $[c\theta]$, so it only depends on which of the c equal subintervals of $[0,1)$ contains θ

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that the summand, written as $[2n/k] - 2[n/k]$, is unchanged under $(n,k) \mapsto (cn,ck)$. On seeing that, the move to write down is: "this depends only on k/n."
2. That cue suggests treating the sum as a Riemann sum over the partition $x = 1/n, 2/n, \ldots, 1$ and passing to $\int_0^1 f(x)\,dx$ with $f(x) = [2/x] - 2[1/x]$.
3. Hesitation occurs right after defining f: it is not obvious f only takes the values 0 and 1, and a student can waste time trying to evaluate f at several sample points numerically instead of reasoning about $[2t] - 2[t]$ structurally. What reinforces it away is remembering that $[2t]$, written as $2[t] + [2\{t\}]$, isolates exactly the fractional part's effect, and $[2\theta]$ for $\theta \in [0,1)$ can only be 0 or 1 — there is no other case to check.
4. The trap is committing to an arithmetic (number-theoretic) attack on the sum before checking whether the summand actually depends on n and k jointly or only on their ratio. Divisor-sum intuition from $\sum[n/k]$ is a false friend here: it is the right intuition for a different, additive-in-n problem, and applying it wastes the time that should go to noticing the scale-covariance.
5. To recognise this faster next time: whenever a floor-function summand contains n and k only through the combinations n/k (or cn/k for fixed constants c), write $x = k/n$ immediately, before attempting any other manipulation, and check whether the summand becomes a function of x alone.
6. The reusable decision rule: a sum of the form $\frac{1}{n}\sum_{k=1}^{n} g(n,k)$ is a Riemann sum candidate whenever $g(n,k) = h(k/n)$ for some fixed function h independent of n; the presence of a logarithm in a proposed closed-form answer to such a limit is itself evidence that the correct route is integration of a $1/x$-type function rather than direct summation.

1976 B2

1976 B2. Suppose that G is a group generated by elements A and B, that is, every element of G can be written as a finite "word" $A^{n_1}B^{n_2}A^{n_3}\cdots B^{n_s}$, where $n_1, \ldots, n_s$ are any integers, and $A^0 = B^0 = I$ as usual. Also, suppose that $A^4 = B^7 = ABA^{-1}B = I$, $A^2 \neq I$, and $B \neq I$.

(a) How many elements of G are of the form C^2 with C in G?

(b) Write each such square as a word in A and B.

Fields: Abstract algebra (finite group theory) **Position:** B2 **Answer:** (a) 8 elements; (b) I, A^2, B, B^2, B^3, B^4, B^5, B^6

Layer 1 THE WHAT — surface analysis

The natural attempt. The relation $ABA^{-1}B = I$ is unfamiliar in that exact form, and the natural first move is to massage it into something recognisable: multiply on the right by B^{-1} to get

$ABA^{-1} = B^{-1}$, i.e. conjugation by A inverts B. A student who has seen dihedral groups recognises this shape and reaches for the dihedral group of order $2n$ as a model, since there $rsr^{-1} = s^{-1}$ for a rotation r and reflection s. But here A has order 4, not 2, and B has order 7 (forced: $B^7 = I$, $B \neq I$, and 7 is prime, so the order of B divides 7 and is not 1, hence is 7). This is not literally a dihedral group – it is a semidirect product $\mathbb{Z}_7 \rtimes \mathbb{Z}_4$ with A acting on B by inversion, an action of order 2 mapped from a group of order 4, so A^2 acts trivially and lies in the centre. A student who does not stop to build the abstract structure and instead tries to compute $(A^iB^j)^2$ by brute matrix-free symbol pushing can get the exponent bookkeeping wrong, in particular losing track of which power of A actually appears after commuting past the B's, and can undercount or overcount the squares. The computation is short but exact bookkeeping is the entire game.

The model chain. From $ABA^{-1} = B^{-1}$, conjugation by A sends $B \mapsto B^{-1}$, so by induction $A^jBA^{-j} = B^{(-1)^j}$ for every integer j, i.e.

$$A^jB^i = B^{i(-1)^j}A^j \qquad \text{for all integers } i, j.$$

This single identity lets every word $A^{n_1}B^{n_2}A^{n_3}\cdots$ be rewritten with all the B's moved to the left of all the A's, so $G = \{B^iA^j : 0 \le i \le 6,\ 0 \le j \le 3\}$, a set of at most 28 elements that is closed under multiplication by the identity above (product $(B^iA^j)(B^kA^l) = B^{i+k(-1)^j}A^{j+l}$), hence is exactly G: $|G| = 28$, with every element uniquely B^iA^j.

Now compute $(B^iA^j)^2 = B^iA^jB^iA^j = B^i\big(A^jB^i\big)A^j = B^i \cdot B^{i(-1)^j} \cdot A^{2j} = B^{i(1+(-1)^j)}A^{2j}$.

Case j even ($j = 0, 2$): $(-1)^j = 1$, so the square is $B^{2i}A^{2j}$. If $j = 0$, $A^{2j} = I$; if $j = 2$, $A^{2j} = A^4 = I$. Either way the square is B^{2i}. As i ranges over $0, \ldots, 6$, since $\gcd(2, 7) = 1$, $2i \bmod 7$ ranges over all of $0, \ldots, 6$: this produces exactly the 7 elements $I, B, B^2, \ldots, B^6$.

Case j odd ($j = 1, 3$): $(-1)^j = -1$, so the B-exponent is $i(1-1) = 0$ regardless of i, and the square is A^{2j}. If $j = 1$, this is A^2; if $j = 3$, $A^{2j} = A^6 = A^4 \cdot A^2 = A^2$. So *every* element with j odd squares to the single element A^2, no matter what i is.

Collecting: the set of squares is $\{I, B, B^2, B^3, B^4, B^5, B^6\} \cup \{A^2\}$, eight elements, and none of these eight coincide (the seven powers of B are distinct since B has order 7, and $A^2 \neq I$ by hypothesis and $A^2 \notin \langle B \rangle$ since $\langle B \rangle$ has order 7 and A^2 has order dividing 2, forcing $A^2 = I$ if it were in $\langle B \rangle$, contradiction). So (a) is 8, and (b) is exactly the list above.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains reach $ABA^{-1} = B^{-1}$ immediately – that step is forced by the given relation and is not where the split happens. The split is at what to do with that identity: the natural attempt treats it as a fact to invoke case-by-case while manipulating $(A^iB^j)^2$ directly, re-deriving the commutation each time it is needed; the model chain instead converts it once into the general rule $A^jB^i = B^{i(-1)^j}A^j$ and uses that single rule mechanically for every element of G, which is what makes the case split on the parity of j visible and exhaustive rather than something to be rediscovered per instance.

> *I don't want to recompute this conjugation identity separately for every value of i and j I might need – I have one relation, $ABA^{-1} = B^{-1}$, and it says conjugation by A is an automorphism of $\langle B \rangle$ of order dividing 2; if I write down the general formula $A^jB^i = B^{i(-1)^j}A^j$ once, every square I ever need to compute becomes a two-line substitution instead of a fresh derivation.*

Layer 3 THE WHY

The relation $ABA^{-1}B = I$ is not an arbitrary equation to be solved for one unknown; read structurally, it says that the map $\beta \mapsto A\beta A^{-1}$ is an automorphism of the cyclic group $\langle B \rangle$, because conjugation by any fixed group element is always an automorphism of any subgroup it normalises, and here it sends the generator B to B^{-1}, which is itself a generator of the same

cyclic group. An automorphism of a cyclic group of prime order 7 is completely determined by where it sends the generator, so knowing $ABA^{-1} = B^{-1}$ pins down $ABA^{-1} = B^{-1}$ for that one generator *and thereby for every power of B simultaneously*: $A(B^i)A^{-1} = (ABA^{-1})^i = B^{-i}$. That is the axiom-level reason the single relation upgrades, for free, to the general commutation rule used in the model chain – it is not a separate fact that happens to also hold, it is what "conjugation is a homomorphism" forces.

The case split on the parity of j is, in turn, forced by A having order 4 while the automorphism $\beta \mapsto \beta^{-1}$ has order exactly 2: A generates a cyclic group of order 4 acting on $\langle B \rangle$ through a quotient of order 2 (the kernel of the action is $\langle A^2 \rangle$). So A^2, and only A^2, acts trivially on B and hence commutes with everything – which is precisely why A^2, and not A itself or A^3, is the extra square that shows up outside $\langle B \rangle$. The mismatch between A's order (4) and the order of its action on B (2) is the entire source of the answer's asymmetry between the j-even and j-odd cases.

- ▷ **Cue:** a relation of the form $XYX^{-1}Y = I$ for a proposed generator Y of prime order $\longrightarrow$ **Move:** rewrite as $XYX^{-1} = Y^{-1}$ and treat it as "conjugation by X is the inversion automorphism of $\langle Y \rangle$", then upgrade immediately to $X^jY^i = Y^{i(-1)^j}X^j$ for all i, j
- ▷ **Cue:** an element X of order m acting on a subgroup through an automorphism of order $d < m$ $\longrightarrow$ **Move:** expect the kernel of that action, $\langle X^{m/d} \rangle$, to behave specially – here it is exactly the source of the extra square A^2 outside $\langle B \rangle$

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the relation $ABA^{-1}B = I$ itself; on seeing it, what should be written down is $ABA^{-1} = B^{-1}$, i.e. conjugation by A inverts B.
2. That cue suggests treating G as (a quotient of) a semidirect product $\langle B \rangle \rtimes \langle A \rangle$, and suggests the method of reducing every word to the normal form B^iA^j before computing anything.
3. Hesitation occurs when squaring B^iA^j: it is easy to move the middle A^j past the following B^i incorrectly, or to forget that A^{2j} must itself be reduced mod 4. What reinforces it away is deriving the general commutation identity $A^jB^i = B^{i(-1)^j}A^j$ once, in full generality, before touching any specific square, so that every subsequent computation is substitution, not fresh algebra.
4. The trap is assuming this is literally the dihedral group (because the relation looks dihedral) and importing dihedral-group facts wholesale, in particular assuming every non-identity element outside $\langle B \rangle$ has order 2 – false here, since A itself has order 4, only A^2 does.
5. To recognise this faster next time: whenever a relation reduces to "conjugation by a generator inverts another generator of prime order", immediately identify the kernel of that action as the subgroup generated by the appropriate power of the first generator, and expect that kernel element to be the one extra "non-obvious" element showing up in any subsequent classification (squares, centre, commutator subgroup, etc.).
6. The reusable decision rule: given a semidirect-product-flavoured relation $XYX^{-1} = Y^{-1}$ with Y of prime order p and X of order m, first find $d =$ order of the induced automorphism on $\langle Y \rangle$ (here $d = 2$ since inversion is order 2), then the kernel $\langle X^{m/d} \rangle$ is central and behaves like a separate, independent factor for any question about powers, centre, or conjugacy classes.

1976 B3

1976 B3. Suppose that we have n events $A_1, \ldots, A_n$, each of which has probability at least $1 - a$ of occurring, where $a < 1/4$. Further suppose that A_i and A_j are mutually independent if $|i - j| > 1$, although A_i and A_{i+1} may be dependent. Assume as known that the recurrence $u_{k+1} = u_k - au_{k-1}$, $u_0 = 1$, $u_1 = 1 - a$, defines positive real numbers u_k

for $k = 0, 1, \ldots$. Show that the probability of all of $A_1, \ldots, A_n$ occurring is at least u_n.

Fields: Probability, combinatorics **Position:** B3 **Answer:** As literally stated, the claim is false for $n \geq 5$; it is provable only under a strengthened independence hypothesis, and only $n \leq 4$ is verified rigorously here

Caveat. This is a documented erratum. The book's own solution opens: "The statement to be proved is false for $n \geq 5$ unless the hypothesis is strengthened to state that A_i is independent of the conjunction of $A_1, A_2, \ldots, A_{i-2}$ for $3 \leq i \leq n$," and supplies a counterexample for $n = 5$ credited to Professor David M. Bloom, built from a probability table with weights involving $h = 33/37$ and $k = 1/(64 + h)$. The OCR of that table is badly corrupted (columns and entries are illegible past recognising the general shape: a joint distribution over five events engineered so pairwise independence for $|i - j| > 1$ holds exactly while the conclusion fails), so I have not attempted to reconstruct or re-verify Bloom's specific numbers; I report only the qualitative fact, which the book states outright and which is corroborated by the difficulty I ran into below trying to prove the general case myself. Treat the numeric counterexample as attributed to the source, not independently re-derived. This part file therefore departs from the usual archive format: instead of a verified model chain for all n, it gives a verified argument for small n and an honest account of exactly where a natural attempt at the general case runs into a gap – which is, itself, the reusable lesson.

Layer 1 THE WHAT — surface analysis

The natural attempt. The recurrence $u_{k+1} = u_k - au_{k-1}$ strongly suggests induction on n, matching each new event A_n against a "loss" term of size roughly $a \cdot$ (something two steps back). Write $p_n = P(A_1 \wedge \cdots \wedge A_n)$. The base cases are clean: $p_1 = P(A_1) \geq 1 - a = u_1$, and by inclusion–exclusion, $p_2 = P(A_1 \wedge A_2) \geq P(A_1) + P(A_2) - 1 \geq 2(1 - a) - 1 = 1 - 2a$, which equals $u_2 = u_1 - au_0 = (1 - a) - a = 1 - 2a$ exactly – so $p_2 \geq u_2$, tight.

For the inductive step, the natural move is

$$p_n = P(A_1 \wedge \cdots \wedge A_{n-1}) - P(A_1 \wedge \cdots \wedge A_{n-1} \wedge \neg A_n) = p_{n-1} - P(A_1 \wedge \cdots \wedge A_{n-1} \wedge \neg A_n),$$

and then to bound the subtracted term above, using that A_n is independent of the earlier events far enough back:

$$P(A_1 \wedge \cdots \wedge A_{n-1} \wedge \neg A_n) \leq P(A_1 \wedge \cdots \wedge A_{n-2} \wedge \neg A_n) = P(A_1 \wedge \cdots \wedge A_{n-2}) \cdot P(\neg A_n) \leq P(A_1 \wedge \cdots \wedge A_{n-2}) \cdot a$$

the first inequality because dropping the conjunct A_{n-1} can only raise the probability, and the equality using independence of A_n from the earlier block. This is exactly the shape of the target recurrence – *if* $P(A_1 \wedge \cdots \wedge A_{n-2})$ could be replaced by u_{n-2} here, the induction would close immediately: $p_n \geq p_{n-1} - a \cdot u_{n-2} \geq u_{n-1} - au_{n-2} = u_n$. The attempt feels one line from finished.

Layer 2 THE HOW — the bifurcation

Bifurcating point. The chain reaches, cleanly and correctly,

$$p_n \geq p_{n-1} - a \cdot P(A_1 \wedge \cdots \wedge A_{n-2}).$$

Both a correct and an incorrect proof agree up to exactly this line. The split is what happens next: it is tempting to substitute the inductive hypothesis $P(A_1 \wedge \cdots \wedge A_{n-2}) \geq u_{n-2}$ directly into this inequality and declare victory. That substitution is invalid, and noticing why it is invalid is the entire content of this problem.

> *I have a* lower *bound on $P(A_1 \wedge \cdots \wedge A_{n-2})$ from the induction hypothesis, but the quantity I actually need to control here is being multiplied by a and then* subtracted *– so to get a lower bound on p_n I need an* upper *bound on $P(A_1 \wedge \cdots \wedge A_{n-2})$, and induction on a chain of lower bounds does not hand me an upper bound for free; substituting u_{n-2} in for $P(A_1 \wedge \cdots \wedge A_{n-2})$ on the wrong side of an inequality is not a shortcut, it is a non sequitur, and I have to stop and ask what actually bounds $P(A_1 \wedge \cdots \wedge A_{n-2})$ from above before I can finish this step.*

Layer 3 THE WHY

The trap is a mismatch of inequality direction that is easy to miss because both quantities involved – p_{n-2} and u_{n-2} – are "the same kind of thing" notationally (both indexed by $n-2$, both appearing in the recurrence), so the eye wants to swap them freely. But an induction hypothesis of the form $p_{n-2} \geq u_{n-2}$ only licenses substitution into an expression that is monotonically *increasing* in p_{n-2}; here p_{n-2} appears with a *negative* coefficient (it is the size of a term being subtracted off), so the induction hypothesis, used naively, would only give an *upper* bound on p_n, not the lower bound the problem demands. This is not a technicality to route around with a slightly cleverer estimate – it is the actual reason the published solution retracts the problem as stated: the book's own words, "the statement is false for $n \geq 5$ unless the hypothesis is strengthened," say plainly that no amount of massaging the given hypotheses repairs this gap in general, because Bloom's counterexample exhibits genuine events satisfying every stated hypothesis for which the conclusion fails.

What the stated hypothesis ("A_i, A_j independent whenever $|i-j| > 1$") buys you is only *pairwise* independence at distance ≥ 2; what the induction above actually needs, to legitimately bound $P(A_1 \wedge \cdots \wedge A_{n-2})$ or otherwise control the interaction between A_n and the whole preceding block, is a form of *joint* or *mutual* independence between A_n and the entire conjunction $A_1 \wedge \cdots \wedge A_{n-2}$ – a strictly stronger statement that pairwise independence at distance ≥ 2 does not imply (pairwise independence never implies mutual independence for three or more events in general). That gap between "independent of each earlier event individually" and "independent of the conjunction of all earlier-but-one events" is exactly the axiom-level fact the problem's own phrasing elides, and it is precisely what the book's strengthened hypothesis supplies by fiat rather than derives.

- ▷ **Cue:** an induction hypothesis of the form $p_k \geq u_k$ is about to be substituted into an expression where p_k appears with a negative coefficient (as a term being subtracted) $\longrightarrow$ **Move:** stop: a lower bound cannot be substituted there to produce a lower bound on the outer quantity; either find an independent upper bound on p_k, or restructure the induction to avoid needing one
- ▷ **Cue:** a hypothesis states pairwise independence among a family of events at some index distance $\longrightarrow$ **Move:** do not silently upgrade this to joint/mutual independence of one event from the conjunction of several others – that upgrade needs to be an explicit, separately stated hypothesis

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the shape of the recurrence $u_{k+1} = u_k - a u_{k-1}$ itself: on seeing it, what should be written down is "the inductive step will need to bound a subtracted loss term by a times something two steps back," which correctly predicts the algebra up to the point of failure.
2. That cue suggests strong induction on n, isolating $P(A_1 \wedge \cdots \wedge A_{n-1} \wedge \neg A_n)$ as the loss term and trying to bound it using the given independence-at-distance hypothesis.
3. Hesitation should occur exactly at the substitution step, the instant p_{n-2} is about to be replaced by u_{n-2} inside an expression where it carries a minus sign. What reinforces the hesitation away from a silent, wrong substitution is asking explicitly: "is p_{n-2} appearing here with a positive or negative coefficient in what I am trying to bound?" – a check that costs one line and catches the entire error.

4. The trap is the visual similarity between p_k and u_k, both indexed by the same k, inviting an automatic swap regardless of which direction of inequality is actually licensed at that point in the argument.

5. To recognise this faster next time: whenever an induction hypothesis is a one-directional bound (here, $p_k \geq u_k$, a lower bound only), explicitly check the sign of the coefficient the bounded quantity carries in the expression being substituted into, before substituting – a lower bound only transfers through positive coefficients, an upper bound only through negative ones.

6. The reusable decision rule: pairwise independence of a family of events ("any two at distance > 1 are independent") is strictly weaker than joint independence of one event from the conjunction of several others, and an inductive or recursive probability bound that needs to multiply probabilities across more than two events at once must say explicitly which of these two hypotheses it is using – silently treating pairwise independence as if it licensed factoring a joint probability is the single most common invisible gap in this class of argument.

1976 B4

1976 B4. For a point P on an ellipse, let d be the distance from the center of the ellipse to the line tangent to the ellipse at P. Prove that $(PF_1)(PF_2)d^2$ is constant as P varies on the ellipse, where PF_1 and PF_2 are the distances from P to the foci F_1 and F_2 of the ellipse.

Fields: coordinate geometry; conics; symmetric functions **Position:** B4 — middle of the paper; a computation that rewards knowing which quantities to compute and which to leave alone **Answer:** The constant is a^2b^2, where a and b are the semi-axes. With $r_i = PF_i$ and $P = (x, y)$, both r_1r_2 and a^2b^2/d^2 equal $a^2 + b^2 - x^2 - y^2$.

Layer 1 THE WHAT — surface analysis

The natural attempt. Put the ellipse in standard position, $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ with $a \geq b > 0$, foci $F_{1,2} = (\mp c, 0)$, $c^2 = a^2 - b^2$, and $P = (x, y)$ on it. The student writes down the three things the problem names:

$$PF_1 = \sqrt{(x+c)^2 + y^2}, \qquad PF_2 = \sqrt{(x-c)^2 + y^2}, \qquad d = \frac{1}{\sqrt{x^2/a^4 + y^2/b^4}},$$

the last from the tangent line $\frac{xX}{a^2} + \frac{yY}{b^2} = 1$ and the point-to-line distance formula. Then the product. Multiplying the two roots gives

$$PF_1 \cdot PF_2 = \sqrt{\big((x+c)^2 + y^2\big)\big((x-c)^2 + y^2\big)} = \sqrt{(x^2 + y^2 + c^2)^2 - 4c^2x^2},$$

a square root of a quartic. Now the student has to show that this, times d^2, is constant, i.e. that

$$(x^2 + y^2 + c^2)^2 - 4c^2x^2 \;=\; K^2\Big(\frac{x^2}{a^4} + \frac{y^2}{b^4}\Big)^2$$

for some constant K, using the ellipse equation. It can be done: substitute $y^2 = b^2 - \frac{b^2}{a^2}x^2$ everywhere, expand both sides as polynomials in x^2, and compare. But it is a page of algebra with a square root hanging over it, and there is no sign on the way that you are winning until the very last line. Under time pressure this is where students make a sign error in $c^2 = a^2 - b^2$ and conclude that the statement is false.

A second natural attempt parametrises: $P = (a\cos t, b\sin t)$. This is better, but if the student still computes PF_1 as $\sqrt{(a\cos t + c)^2 + b^2\sin^2 t}$ and does not simplify the radicand, the same quartic appears in trigonometric dress.

The model chain. Keep the set-up above; write $r_1 = PF_1$, $r_2 = PF_2$.

Step 1: the product r_1r_2 without square roots. By the definition of the ellipse, $r_1 + r_2 = 2a$. The squares r_1^2 and r_2^2 are polynomials in x, y. So compute the product from the sum and the squares:

$$r_1r_2 = \tfrac{1}{2}\left[(r_1+r_2)^2 - r_1^2 - r_2^2\right] = \tfrac{1}{2}\left[4a^2 - (x+c)^2 - y^2 - (x-c)^2 - y^2\right] = 2a^2 - x^2 - y^2 - c^2.$$

Since $c^2 = a^2 - b^2$,

$$r_1r_2 = a^2 + b^2 - x^2 - y^2. \tag{1}$$

Step 2: d^2. The tangent at P is $\dfrac{x}{a^2}X + \dfrac{y}{b^2}Y = 1$, so the distance from the origin to it is

$$d^2 = \frac{1}{x^2/a^4 + y^2/b^4} = \frac{a^4b^4}{b^4x^2 + a^4y^2}.$$

Now use the ellipse equation in the form $b^2x^2 + a^2y^2 = a^2b^2$, twice:

$$b^4x^2 + a^4y^2 = b^2(a^2b^2 - a^2y^2) + a^2(a^2b^2 - b^2x^2) = a^2b^2\,(a^2 + b^2 - x^2 - y^2).$$

Hence

$$d^2 = \frac{a^2b^2}{a^2 + b^2 - x^2 - y^2}. \tag{2}$$

Step 3. The denominator in (2) is positive, because by (1) it equals r_1r_2 and both foci lie strictly inside the ellipse, so $r_1, r_2 > 0$. Multiplying (1) and (2),

$$(PF_1)(PF_2)\,d^2 = a^2b^2,$$

which does not depend on P. □

Checks. At $P = (a, 0)$: $r_1r_2 = (a+c)(a-c) = b^2$ and $d = a$, product a^2b^2. At $P = (0, b)$: $r_1 = r_2 = a$ and $d = b$, product a^2b^2. For a circle ($a = b = r$), $r_1 = r_2 = d = r$ and the product is r^4.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains write $PF_1 = \sqrt{(x+c)^2 + y^2}$. They split at the next line. The natural attempt multiplies the two square roots. The model chain refuses to handle r_1 and r_2 separately at all: it takes the known sum $r_1 + r_2 = 2a$ and the polynomial squares r_1^2, r_2^2, and gets the product from $2r_1r_2 = (r_1+r_2)^2 - r_1^2 - r_2^2$. There is a second, smaller split in Step 2: the model uses the ellipse equation to turn the denominator $b^4x^2 + a^4y^2$ into a multiple of the same expression $a^2 + b^2 - x^2 - y^2$, instead of eliminating y.

> *I don't want PF_1 and PF_2 one at a time; each is a square root. I want their product. What do I already know about the pair? Their sum is $2a$ — that's the definition of the ellipse. And their squares are polynomials. Sum and squares give the product with no roots at all.*

> *The answer has to be constant, so $1/d^2$ must be a constant multiple of whatever I got for r_1r_2. I'll push $b^4x^2 + a^4y^2$ towards $a^2 + b^2 - x^2 - y^2$ using $b^2x^2 + a^2y^2 = a^2b^2$, and see if it lands.*

▷ **Cue:** a product of two distances, each a square root, and their sum is known ⟶ **Move:** write $2r_1r_2 = (r_1+r_2)^2 - r_1^2 - r_2^2$; the squares are polynomials, so the product is a polynomial

▷ **Cue:** "prove this expression is constant" and you have found one factor in closed form ⟶ **Move:** the other factor must be the reciprocal of the same expression up to a constant; aim the algebra at that expression instead of expanding blind

▷ **Cue:** any quantity at a point of an ellipse or hyperbola involving focal distances ⟶ **Move:** use $r_1 + r_2 = 2a$ (or $|r_1 - r_2| = 2a$), or the focal-radius form $r_{1,2} = a \pm ex$; never keep $\sqrt{(x \pm c)^2 + y^2}$

Layer 3 THE WHY

Why is "sum and squares" the natural move and not a trick? Go back to what an ellipse is. It is the set of points whose distances to two fixed points add up to $2a$. So the one exact fact you are guaranteed about r_1 and r_2 is a statement about their *sum*. The problem asks about their *product*. The link between sum and product of two numbers is the most basic identity in algebra, $(r_1 + r_2)^2 = r_1^2 + 2r_1r_2 + r_2^2$, and it needs the squares, which the distance formula hands you free of roots. So every ingredient of Step 1 is either the definition of the object or the definition of distance. The natural attempt fails because it uses the distance formula and never uses the definition of the ellipse.

There is an equivalent way to see the same thing. From $r_1^2 - r_2^2 = 4cx$ and $r_1 + r_2 = 2a$ you get $r_1 - r_2 = 2cx/a$, so

$$r_1 = a + \frac{c}{a}x, \qquad r_2 = a - \frac{c}{a}x.$$

These are the *focal radius* formulas: on an ellipse, distance to a focus is a *linear* function of x. Once you know that, $r_1r_2 = a^2 - \frac{c^2}{a^2}x^2$ is immediate, and you can check it agrees with (1). The square roots in the natural attempt were never really there; they cancel because P lies on the curve. The model chain lets them cancel before they are written.

Now Step 2. Why aim $b^4x^2 + a^4y^2$ at $a^2 + b^2 - x^2 - y^2$? Because the problem says the product is constant. That is information. If $r_1r_2 = Q(x, y)$ and $r_1r_2d^2$ is constant, then d^2 must be a constant over Q. So before touching the algebra you know the target. That is the general principle behind the second critical thought: a "prove it is constant" problem tells you the shape of every intermediate answer, and you should use that to steer.

Why a^2b^2? The official solution does exactly the computation above and gives the constant without comment. There is also a synthetic reason, which tells you what the constant *means*. Let p_1, p_2 be the distances from F_1, F_2 to the tangent. By the reflection property, PF_1 and PF_2 make the same angle θ with the tangent, so $p_i = r_i \sin\theta$. Both foci lie on the same side of the tangent, and O is the midpoint of F_1F_2, so $d = \frac{1}{2}(p_1 + p_2) = \frac{1}{2}(r_1 + r_2)\sin\theta = a\sin\theta$. The classical fact $p_1p_2 = b^2$ (the feet of these perpendiculars lie on the auxiliary circle $x^2 + y^2 = a^2$, and p_1p_2 is the power of F_1 with respect to that circle, $a^2 - c^2 = b^2$) then gives

$$r_1r_2\,d^2 = r_1r_2\,a^2\sin^2\theta = a^2\,(r_1\sin\theta)(r_2\sin\theta) = a^2p_1p_2 = a^2b^2.$$

So the problem is a disguised form of "the product of the focal distances to a tangent is b^2". The coordinate route is shorter to write and needs no remembered theorem; the synthetic route explains the constant. On the day, the coordinate route is the one to take.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the word "foci" next to a *product* $(PF_1)(PF_2)$. On seeing it, write $r_1 + r_2 = 2a$ and $r_1r_2 = \frac{1}{2}[(r_1 + r_2)^2 - r_1^2 - r_2^2]$.

2. Coordinates in standard position, with the focal-distance product and the tangent-distance both reduced to one expression using the curve's equation.

3. Hesitation comes after $d^2 = a^4b^4/(b^4x^2 + a^4y^2)$: the denominator does not look like anything. What removes it is the knowledge that the final answer is constant, so the denominator must be a constant times r_1r_2. Try to make it so.

4. The trap is the distance formula. It is the first thing you know about PF_1 and it leads to $\sqrt{\text{quartic}}$. The second trap is eliminating y by $y^2 = b^2(1 - x^2/a^2)$ everywhere, which works but hides the symmetric expression $a^2 + b^2 - x^2 - y^2$.

5. Whenever two quantities have a known sum (or difference) and you need their product, the product is one identity away. Look for the defining relation of the object; it often supplies the sum.

6. Reusable decision rule: *before computing quantities one at a time, ask what symmetric function of them the problem needs and what symmetric function the definition gives you; compute the needed one from the given one. In a "show it is constant" problem, use the constancy to predict the form of each factor.*

1976 B5

1976 B5. Evaluate

$$\sum_{k=0}^{n} (-1)^k \binom{n}{k} (x-k)^n.$$

Fields: combinatorics; finite differences; polynomials **Position:** B5 — late in the paper, but one of the shortest problems on it once the right object is named **Answer:** $n!$, for every x. The sum is the nth backward difference of x^n, and each difference lowers the degree by one and multiplies the leading coefficient by the degree.

Layer 1 THE WHAT — surface analysis

The natural attempt. The student sees a binomial coefficient and a power, and expands the power by the binomial theorem:

$$(x-k)^n = \sum_{j=0}^{n} \binom{n}{j} x^{n-j}(-k)^j.$$

Swapping the order of summation,

$$S(x) = \sum_{j=0}^{n} \binom{n}{j} (-1)^j x^{n-j}\, T_j, \qquad T_j = \sum_{k=0}^{n} (-1)^k \binom{n}{k} k^j.$$

So everything rests on the moment sums T_j. Small cases: $T_0 = (1-1)^n = 0$ for $n \geq 1$. $T_1 = \sum (-1)^k k \binom{n}{k} = -n(1-1)^{n-1} = 0$ for $n \geq 2$. For T_2 the student writes $k^2 = k(k-1) + k$ and uses $k(k-1)\binom{n}{k} = n(n-1)\binom{n-2}{k-2}$, getting 0 for $n \geq 3$. A pattern appears: $T_j = 0$ for $j < n$, and presumably $T_n = (-1)^n n!$. If both are true, only $j = n$ survives and

$$S(x) = \binom{n}{n} (-1)^n x^0\, (-1)^n n! = n!.$$

This is the right answer, and a strong student gets it. But to be a proof it needs the claim about T_j for general j, which means writing k^j in the basis $k(k-1)\cdots(k-i+1)$ (Stirling numbers of the second kind) or differentiating $(1-t)^n$ repeatedly with $t\,\frac{d}{dt}$. The claim about T_j is exactly as hard as the original problem, so the expansion has moved the difficulty rather than removing it.

The model chain. For a function f define the backward difference $(\nabla f)(x) = f(x) - f(x-1)$.

Lemma 1. For every $n \geq 0$ and every f,

$$(\nabla^n f)(x) = \sum_{k=0}^{n} (-1)^k \binom{n}{k} f(x-k).$$

Proof. Induction on n; $n = 0$ is $f(x) = f(x)$. If it holds for n, then

$$(\nabla^{n+1} f)(x) = \sum_{k=0}^{n} (-1)^k \binom{n}{k} f(x-k) - \sum_{k=0}^{n} (-1)^k \binom{n}{k} f(x-1-k).$$

Re-index the second sum with $k+1 \mapsto k$. The coefficient of $f(x-k)$ becomes $(-1)^k\left[\binom{n}{k} + \binom{n}{k-1}\right] = (-1)^k\binom{n+1}{k}$ by Pascal's rule (with $\binom{n}{-1} = \binom{n}{n+1} = 0$). □

Lemma 2. If p is a polynomial of degree $m \geq 1$ with leading coefficient α, then ∇p is a polynomial of degree $m-1$ with leading coefficient $m\alpha$. If p is constant, $\nabla p = 0$.

Proof. ∇ is linear, so it is enough to look at x^m and at lower powers. By the binomial theorem $x^m - (x-1)^m = mx^{m-1} - \binom{m}{2}x^{m-2} + \cdots$, which has degree $m-1$ and leading coefficient m; and ∇x^i has degree at most $i - 1 < m - 1$ for $i < m$. A constant has difference 0. □

Conclusion. Apply Lemma 2 n times to x^n: the degrees go $n, n-1, \ldots, 0$ and the leading coefficients go $1,\ n,\ n(n-1),\ \ldots,\ n!$. So $\nabla^n x^n$ is the constant $n!$. By Lemma 1,

$$\sum_{k=0}^{n} (-1)^k \binom{n}{k} (x-k)^n = (\nabla^n x^n) = n!.$$

(For $n = 0$ the sum is $(x)^0 = 1 = 0!$.) □

Checks. $n = 1$: $x - (x-1) = 1$. $n = 2$: $x^2 - 2(x-1)^2 + (x-2)^2 = 2$. Putting $x = n$ gives $\sum_k (-1)^k \binom{n}{k}(n-k)^n = n!$, the inclusion–exclusion count of surjections from an n-set onto an n-set, i.e. of permutations. That is an independent confirmation.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains look at the same summand. The natural attempt opens the *power* $(x-k)^n$. The model chain leaves the power closed and reads the *weights* $(-1)^k\binom{n}{k}$ attached to the shifted arguments $x-k$. That reading is the split: once you see "alternating binomial weights on $f(x), f(x-1), \ldots, f(x-n)$", the sum is $\nabla^n f$, and the rest is a statement about how ∇ acts on degree.

> *The summand is $f(x-k)$ with $f(t) = t^n$, and the weights are $(-1)^k\binom{n}{k}$. Those are the coefficients of $(1-E)^n$, where E is the shift $f(x) \mapsto f(x-1)$. This is an nth difference. I don't need to know anything about t^n except its degree and leading coefficient.*

> *The problem says "evaluate" and leaves x free. If the answer depended on x they would expect a polynomial in x as the answer, which would be ugly. So I expect the x to disappear — and n differences of a degree-n polynomial is exactly the thing that kills every power of x.*

- ▷ **Cue:** a sum $\sum_k (-1)^k\binom{n}{k} f(a \pm k)$ ⟶ **Move:** recognise $\Delta^n f$ or $\nabla^n f$; use that each difference lowers polynomial degree by one
- ▷ **Cue:** alternating binomial sum of k^j or of a polynomial in k of degree j ⟶ **Move:** it vanishes for $j < n$ and equals $(-1)^n n! \times$ leading coefficient for $j = n$; do not expand
- ▷ **Cue:** "evaluate" with a free variable that looks like it should matter ⟶ **Move:** guess the variable drops out, and look for the operator that kills it

Layer 3 THE WHY

Why is reading the weights, rather than expanding the power, the natural move? Start from the definition of a difference. $\nabla f(x) = f(x) - f(x-1)$ has weights $1, -1$. Apply it twice: $f(x) - 2f(x-1) + f(x-2)$, weights $1, -2, 1$. Three times: $1, -3, 3, -1$. The weights of ∇^n are the rows of Pascal's triangle with alternating signs, because applying ∇ once more is the Pascal step "add the entry to its left", with a sign. So $(-1)^k\binom{n}{k}$ next to a function evaluated at $x-k$ is

not a coincidence to be exploited. It is the fingerprint of the operator ∇^n, as surely as $\frac{f(x+h)-f(x)}{h}$ is the fingerprint of a derivative.

Why does ∇ lower the degree by exactly one? Because it is the discrete derivative. $x^m - (x-1)^m$ loses its top term, the same way $\frac{d}{dx}x^m$ does, and the next term is mx^{m-1}, the same as the derivative. So $\nabla^n x^n = n!$ is the discrete twin of $\frac{d^n}{dx^n}x^n = n!$. Once you have that analogy, the answer is predictable before any computation: n derivatives of x^n is $n!$, and n differences should be too.

The official solution is a single sentence: the sum is $n!$ "since it is an nth difference of a monic polynomial, x^n, of degree n". That sentence is correct, but it assumes two facts it never states: that alternating binomial sums of shifted values are iterated differences (Lemma 1), and that each difference lowers the degree by one and multiplies the leading coefficient by the degree (Lemma 2). Here is where the sentence comes from: the author saw the weights first. A student who expands $(x-k)^n$ never sees them, because the expansion scatters the binomial coefficients $\binom{n}{j}$ from the power across the sum and buries the $\binom{n}{k}$.

Why does the natural attempt stall? Because $T_j = \sum_k (-1)^k \binom{n}{k} k^j$ is itself, up to the sign $(-1)^n$, an nth difference: the forward difference $\Delta^n t^j$ evaluated at $t = 0$. The claim "$T_j = 0$ for $j < n$" is Lemma 2 in disguise. So the expansion does not avoid the difference operator; it makes you prove the same fact $n+1$ times, once for each j, without its name.

The reusable lesson is about which part of a summand to look at. In $\sum (-1)^k \binom{n}{k} (x-k)^n$ there are two binomial structures: the one in the weights and the one hidden inside the power. Expanding the power creates a double sum. Reading the weights collapses the single sum. Always read the explicit structure before creating new structure.

Layer $+\alpha$ BACKTESTING

1. The first cue is $(-1)^k \binom{n}{k}$ multiplying a function of $x-k$. On seeing it, write $\nabla^n f(x) = \sum_k (-1)^k \binom{n}{k} f(x-k)$ with $f(t) = t^n$.
2. Finite differences: iterate ∇ and track degree and leading coefficient.
3. Hesitation comes if the student does not trust that ∇^n has those weights. Reinforce it by writing $\nabla, \nabla^2, \nabla^3$ out once — Pascal's triangle with signs appears at once — or by the two-line induction of Lemma 1.
4. The trap is the binomial theorem on $(x-k)^n$. It looks like progress, produces the correct answer on a pattern, and leaves an unproved lemma ($T_j = 0$ for $j < n$) that is the whole problem.
5. Learn the fingerprints of the difference operator: alternating binomial weights, shifted arguments, and an answer that should not depend on the free variable. Also learn the checking identity $\sum_k (-1)^k \binom{n}{k} (n-k)^n = n!$ (surjections), which catches sign slips.
6. Reusable decision rule: *when a sum carries alternating binomial weights, recognise an nth difference before expanding anything; an nth difference kills polynomials of degree below n and sends degree n to n! times the leading coefficient.*

1976 B6

1976 B6. As usual, let $\sigma(N)$ denote the sum of all the (positive integral) divisors of N. (Included among these divisors are 1 and N itself.) For example, if p is a prime, then $\sigma(p) = p+1$. Motivated by the notion of a "perfect" number, a positive integer N is called "quasiperfect" if $\sigma(N) = 2N+1$. Prove that every quasiperfect number is the square of an odd integer.

Fields: number theory; divisor functions; quadratic residues **Position:** B6 — the last problem of the paper; two independent steps, the second of which needs one idea **Answer:** Parity of σ forces

$N = 2^a M^2$ with M odd. Reducing $\sigma(N) = 2N + 1$ modulo $2^{a+1} - 1$ gives $M^2 \equiv -1$; if $a \geq 1$ this modulus has a prime factor $p \equiv 3 \pmod 4$, for which -1 is not a square. So $a = 0$.

Layer 1 THE WHAT — surface analysis

The natural attempt. A strong student starts where the definition points: $\sigma(N) = 2N + 1$ is *odd*. They know, or quickly derive, that σ is multiplicative and that for an odd prime p,

$$\sigma(p^\beta) = 1 + p + \cdots + p^\beta$$

is a sum of $\beta + 1$ odd numbers, hence odd exactly when β is even; while $\sigma(2^a) = 2^{a+1} - 1$ is always odd. So every odd prime appears to an even power: $N = 2^a M^2$ with M odd. This is half the problem, and most competent solvers reach it.

Now they must show $a = 0$. The equation becomes

$$(2^{a+1} - 1)\,\sigma(M^2) = 2^{a+1} M^2 + 1.$$

The instinct is to work with powers of 2, since a is the exponent of 2. Modulo 2^{a+1} the right side is 1 and the left is $-\sigma(M^2)$, so $\sigma(M^2) \equiv -1 \pmod{2^{a+1}}$. That is a condition, not a contradiction: $\sigma(M^2)$ is odd and nothing stops it being $\equiv -1$. Modulo 4 or 8 with M odd ($M^2 \equiv 1 \pmod 8$) the same thing happens: every congruence obtained this way is satisfiable. The student tries bounding: $\sigma(M^2) > M^2$, so the left side exceeds $(2^{a+1} - 1)M^2$, which is less than the right side, so no contradiction from size either. The case $a = 1$ may fall by luck — modulo 3 the equation reads $0 \equiv 4M^2 + 1 \equiv M^2 + 1$, impossible because squares are 0 or 1 mod 3 — but the student often does not see why mod 3, and cannot generalise it.

The model chain. *Step 1: shape of N.* σ is multiplicative: if $N = \prod p_i^{\beta_i}$ then $\sigma(N) = \prod \sigma(p_i^{\beta_i})$, since each divisor of N is uniquely a product of divisors of the prime powers. Write $N = 2^a \prod_{i=1}^{k} p_i^{\beta_i}$ with the p_i distinct odd primes and $a \geq 0$. Since $\sigma(N) = 2N + 1$ is odd, each factor $\sigma(p_i^{\beta_i})$ is odd. But $\sigma(p_i^{\beta_i}) = 1 + p_i + \cdots + p_i^{\beta_i}$ is a sum of $\beta_i + 1$ odd numbers, which is odd if and only if β_i is even. Hence the odd part of N is a square:

$$N = 2^a M^2, \qquad M \text{ odd}.$$

Step 2: the equation. $\sigma(N) = \sigma(2^a)\sigma(M^2) = (2^{a+1} - 1)\sigma(M^2)$, and $2N + 1 = 2^{a+1} M^2 + 1$. So

$$(2^{a+1} - 1)\,\sigma(M^2) = 2^{a+1} M^2 + 1. \tag{$*$}$$

Step 3: reduce modulo $q = 2^{a+1} - 1$. Suppose $a \geq 1$. Then $q \geq 3$. Since $2^{a+1} \equiv 1 \pmod q$, $(*)$ gives

$$0 \equiv M^2 + 1 \pmod q.$$

Also $4 \mid 2^{a+1}$, so $q \equiv -1 \equiv 3 \pmod 4$.

Step 4: a prime $p \equiv 3 \pmod 4$ divides q. q is odd, so all its prime factors are odd, i.e. $\equiv 1$ or $3 \pmod 4$. A product of numbers $\equiv 1 \pmod 4$ is $\equiv 1 \pmod 4$. Since $q \equiv 3 \pmod 4$, some prime factor p of q satisfies $p \equiv 3 \pmod 4$. Then $M^2 \equiv -1 \pmod p$.

Step 5: contradiction. $p \nmid M$, since otherwise $M^2 + 1 \equiv 1 \pmod p$. By Fermat, $M^{p-1} \equiv 1 \pmod p$. But $(p - 1)/2$ is odd, so

$$M^{p-1} = (M^2)^{(p-1)/2} \equiv (-1)^{(p-1)/2} = -1 \pmod p,$$

and $1 \not\equiv -1$ because p is odd. Contradiction. Therefore $a = 0$ and $N = M^2$ with M odd. □

Remark. This proves a necessary condition. No quasiperfect number is known to exist; the problem does not claim that one does.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains reach $(*)$: $(2^{a+1} - 1)\sigma(M^2) = 2^{a+1} M^2 + 1$. They split on the choice of modulus. The natural attempt reduces modulo powers of 2, because a is "about 2". The model reduces modulo $2^{a+1} - 1$, the factor sitting in front of the unknown $\sigma(M^2)$. That modulus deletes the one term you know nothing about and turns 2^{a+1} into 1 on the other side.

> *I know nothing about $\sigma(M^2)$ — it's the one quantity I can't control. So work modulo the thing multiplying it, $2^{a+1}-1$. Then that side is 0, and on the other side 2^{a+1} becomes 1: I get $M^2+1\equiv 0$.*

> *$M^2\equiv -1$ modulo something that is 3 mod 4. Numbers that are 3 mod 4 must have a prime factor that is 3 mod 4, and -1 is never a square modulo such a prime. That's the contradiction.*

- ▷ **Cue:** an equation $A\cdot X=B$ with X an uncontrollable quantity (a divisor sum, an unknown integer) $\longrightarrow$ **Move:** reduce modulo A (or a prime factor of A) to delete X
- ▷ **Cue:** a modulus of the form 2^m-1, 10^m-1, b^m-1 $\longrightarrow$ **Move:** in that modulus $b^m\equiv 1$; use it to collapse powers of b on the other side
- ▷ **Cue:** a congruence $M^2\equiv -1 \pmod q$ with $q\equiv 3 \pmod 4$ $\longrightarrow$ **Move:** extract a prime factor $p\equiv 3 \pmod 4$ of q; -1 is a non-residue mod p, contradiction
- ▷ **Cue:** $\sigma(N)$ odd $\longrightarrow$ **Move:** N is a square or twice a square (odd primes to even powers)

Layer 3 THE WHY

Why is Step 1 natural? The definition hands you $\sigma(N)=2N+1$, and the cheapest information about an equation of integers is parity. $2N+1$ is odd whatever N is, so the equation immediately constrains $\sigma(N)$ and nothing else. Multiplicativity is the only structure σ has, so parity of $\sigma(N)$ becomes parity of each $\sigma(p^\beta)$, and a geometric series of odd terms is odd exactly when it has an odd number of terms. None of that is a trick; each step is the definition of the object being used.

Step 3 is the one the official solution states without motivation. It says "hence $2^{a+1}M^2+1=(2^{a+1}-1)\sigma(M^2)$ so that" $M^2+1\equiv 0 \pmod{2^{a+1}-1}$, and moves on. Here is where that comes from. In $(*)$ there are two kinds of symbol: a and M, which describe N, and $\sigma(M^2)$, which is a function of M we cannot compute. An equation with an uncontrollable term is best read modulo that term's coefficient, because then the term vanishes. The coefficient is $2^{a+1}-1$, and the bonus is automatic: the other side is built from 2^{a+1}, which is 1 modulo $2^{a+1}-1$. The equation was set up for this reduction. Working modulo powers of 2 fails for the opposite reason: modulo 2^{a+1}, $\sigma(M^2)$ survives with coefficient -1, and you have learnt only a congruence about the thing you cannot control.

Why 3 mod 4 and quadratic residues? The congruence $M^2\equiv -1$ asks whether -1 is a square. The one thing to know about that is Euler's criterion in its simplest case: modulo an odd prime p, -1 is a square only if $p\equiv 1 \pmod 4$, because $(-1)^{(p-1)/2}$ must equal 1. The modulus $2^{a+1}-1$ is 3 mod 4 as soon as $a\geq 1$, which is exactly the case we must exclude. So the hypothesis $a\geq 1$ produces a modulus of the wrong residue class, and that is the contradiction. Notice also the lucky $a=1$ case from the natural attempt: modulo 3 worked because $3=2^2-1$. It was never luck; it was Step 3 with $a=1$.

Why must $q\equiv 3 \pmod 4$ have a prime factor $\equiv 3 \pmod 4$? Because residues mod 4 multiply: $1\cdot 1=1$. If every prime factor were 1 mod 4, so would q be. This small closure argument appears constantly (the same argument proves there are infinitely many primes $\equiv 3 \pmod 4$) and is worth owning.

The difficulty rating is honest: this is a B6, but its difficulty lies in one decision — the modulus — and the rest is standard number theory. The decision is made by asking which term in the equation you cannot control.

Layer $+\alpha$ BACKTESTING

1. The first cue is $2N+1$: odd. Write "$\sigma(N)$ odd $\Rightarrow$ every odd prime of N to an even power $\Rightarrow N=2^aM^2$".

2. Multiplicativity plus parity, then a congruence argument to kill the power of 2.

3. Hesitation occurs at $(*)$, when congruences modulo $2, 4, 8, 2^{a+1}$ all come out consistent. What reinforces it away is the question "which quantity in this equation can I not compute?" — then reduce modulo its coefficient.

4. The trap is reducing modulo powers of 2 because a is the exponent of 2. It feels on-topic and gives only satisfiable congruences. A secondary trap is trying to prove the problem by finding quasiperfect numbers: there are none known.

5. Whenever an equation has the form (explicit factor)$\times$(unknown) $=$ (explicit expression), try the explicit factor as a modulus first. When a modulus is $b^m - 1$, remember that $b^m \equiv 1$ there.

6. Reusable decision rule: *choose the modulus that deletes the term you cannot control; then, if you land on $x^2 \equiv -1$, look for a prime factor $\equiv 3 \pmod 4$ of the modulus.*