DAWN PUTNAM SESSION

Putnam 1983 — Papers A and B

Putnam 120 Archive

Contents

Putnam 1983 — Paper A **5**
1983 A1 5
1983 A2 7
1983 A3 10
1983 A4 12
1983 A5 15
1983 A6 18

Putnam 1983 — Paper B **23**
1983 B1 23
1983 B2 25
1983 B3 28
1983 B4 31
1983 B5 33
1983 B6 36

Putnam 1983 — Paper A

1983 A1

1983 A1. How many positive integers n are there such that n is an exact divisor of at least one of the numbers 10^{40}, 20^{30}?

Fields: number theory (divisor counting, exponent lattices) **Position:** A1 — the opener **Answer:** 2301. Write $n = 2^i 5^j$ and translate "$n \mid 10^{40}$ or $n \mid 20^{30}$" into a condition on the pair (i, j); the count is $\tau(10^{40}) + \tau(20^{30}) - \tau(\gcd(10^{40}, 20^{30})) = 41^2 + 61 \cdot 31 - 41 \cdot 31 = 1681 + 1891 - 1271 = 2301$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You factor the two numbers: $10^{40} = 2^{40}5^{40}$ and $20^{30} = (2^2 \cdot 5)^{30} = 2^{60}5^{30}$. Counting divisors of each separately is immediate: $\tau(10^{40}) = (40+1)(40+1) = 1681$ and $\tau(20^{30}) = (60+1)(30+1) = 1891$. You know you cannot just add these, because some n divide both numbers and would be counted twice. So you need $\tau(10^{40}) + \tau(20^{30}) - \#\{n : n \mid 10^{40} \text{ and } n \mid 20^{30}\}$.

Here is where the honest attempt can go wrong without ever feeling wrong. You reach for the other number naturally associated to two integers with the same prime factors, and think: "the numbers most naturally attached to a pair a, b are $\gcd(a, b)$ and $\operatorname{lcm}(a, b)$ — which one governs *common divisors*?" If you answer that question by pattern-matching on the word "common" (common divisor, common multiple both sound alike) rather than by definition, it is entirely possible to write down $\operatorname{lcm}(10^{40}, 20^{30}) = 2^{60}5^{40}$ and start counting divisors of *that*, which counts something else entirely (divisors that divide the lcm, a much larger set, not divisors common to both numbers). Even without that misstep, actually computing $\{n : n \mid 10^{40} \text{ and } n \mid 20^{30}\}$ from scratch — rather than recognizing it as a single already-known number's divisor set — invites setting up a full case analysis on (i, j) pairs by hand, checking $i \le 40 \wedge i \le 60$ and $j \le 40 \wedge j \le 30$ term by term. It arrives at the right place eventually, but it re-derives, one inequality at a time, a fact that a single definition already hands you.

The model chain. Every n dividing a number of the form $2^a 5^b$ is itself of the form $n = 2^i 5^j$ with $0 \le i \le a$, $0 \le j \le b$; divisibility by n is exactly the pair of inequalities on exponents. So $n \mid 10^{40} = 2^{40}5^{40}$ iff $0 \le i \le 40$, $0 \le j \le 40$, and $n \mid 20^{30} = 2^{60}5^{30}$ iff $0 \le i \le 60$, $0 \le j \le 30$.

A divisor common to both numbers must satisfy both pairs of inequalities at once: $i \le \min(40, 60) = 40$ and $j \le \min(40, 30) = 30$. But "$i \le 40$ and $j \le 30$" is exactly the condition for n to divide $2^{40}5^{30} = \gcd(10^{40}, 20^{30})$. So

$$\{n : n \mid 10^{40}\} \cap \{n : n \mid 20^{30}\} = \{n : n \mid \gcd(10^{40}, 20^{30})\},$$

a fact that holds for any two positive integers a, b, not just these: a number divides both a and b precisely when it divides $\gcd(a, b)$, because that is what "greatest common divisor" means. Hence

$$\#\{n : n \mid \gcd(10^{40}, 20^{30})\} = \tau(2^{40}5^{30}) = (40+1)(30+1) = 1271.$$

By inclusion–exclusion,

$$\#\{n : n \mid 10^{40} \text{ or } n \mid 20^{30}\} = 1681 + 1891 - 1271 = 2301.$$

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree up to "I need $\#\{n : n \mid 10^{40} \text{ and } n \mid 20^{30}\}$." They split at the next question: what single number's divisor set is that? The model answers it by definition (*gcd*); the natural attempt is where a definition can get silently swapped for its dual (*lcm*), or skipped in favour of casework that reconstructs the same fact more slowly.

> *The set of divisors common to a and b is not some new object I have to build by casework — it is, by the very definition of greatest common divisor, exactly the divisor set of one specific number,* $\gcd(a, b)$*. I should name that number before I count anything.*

> *Divisibility by* $n = 2^i 5^j$ *is two independent inequalities, one per prime. Two divisibility conditions joined by* and *become two inequalities joined by* and*, which is a rectangle intersected with a rectangle in the* (i, j) *grid — and the corner of that intersection,* $(\min(40, 60), \min(40, 30))$*, is precisely the exponent vector of* $\gcd(10^{40}, 20^{30})$*.*

Layer 3 THE WHY

Why is "common divisors of a, b = divisors of $\gcd(a, b)$" not a trick but a restatement of a definition? Because $\gcd(a, b)$ *is defined as* the largest d with $d \mid a$ and $d \mid b$; the full claim (every common divisor, not just the largest one, divides $\gcd(a, b)$) follows immediately once a, b are written in prime-exponent form, since "$d \mid a$ and $d \mid b$" is componentwise "exponent of d at $p \leq$ exponent of p in a" and the same for b, and the largest exponent satisfying both bounds at every prime simultaneously is the componentwise minimum — which is exactly how gcd is built. There is no step here that deserves to feel clever: divisibility, translated into the exponent vector, is nothing but componentwise $\leq$, and the operations *and*/*or* on divisibility conditions become min/max on exponents by the plainest possible logic. Once you see that translation, gcd and lcm stop being two facts to memorise and become the same fact read in two directions: min for *divides both* (a divisor condition), max for *is divided by both* (a multiple condition). The trap in the natural attempt is reaching for the wrong one of that pair under time pressure, precisely because both are phrased with the word "common."

The second structural fact worth isolating is why the answer needs inclusion–exclusion at all, rather than some other combination. $\{n : n \mid 10^{40}\}$ and $\{n : n \mid 20^{30}\}$ are each order ideals (down-sets) in the exponent lattice $(\mathbb{Z}^2_{\geq 0}, \leq)$ — rectangles with a corner at the origin. The problem asks for the size of their *union*, and for any two finite sets, $|A \cup B| = |A| + |B| - |A \cap B|$ is not a number-theoretic fact at all, it is set-counting; the number theory only enters in identifying $|A \cap B|$ cheaply, which is exactly the gcd observation above. Recognising "or" in the problem statement as literally set union, before any factoring happens, is what tells you inclusion–exclusion is the right tool and that you will need gcd, not lcm, for the correction term.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is two numbers presented as powers built from the same primes ($10 = 2 \cdot 5$, $20 = 2^2 \cdot 5$) with an "or" between two divisibility conditions. Write down immediately: "convert to exponent pairs (i, j); *or* means I will need $|A| + |B| - |A \cap B|$."

2. The method is inclusion–exclusion on the two divisor sets, with the intersection identified as the divisor set of gcd of the two numbers.

3. Hesitation occurs exactly at naming the intersection: is it divisors of the gcd or the lcm? Reinforce it by re-deriving from the definition each time under pressure: gcd answers "divides both," lcm answers "is divided by both." The question here is the former.

4. The trap is substituting lcm for gcd in the correction term (or, more subtly, computing common divisors by casework instead of recognising them as the divisor set of a single number), either of which silently changes the answer.

5. Whenever a problem describes divisibility conditions on numbers with shared prime factors, translate immediately to exponent vectors and recognise divisibility as componentwise $\leq$; this turns every gcd/lcm question into min/max before any arithmetic starts.

6. Reusable rule: for counting problems phrased as "divides a" *or*/*and* "divides b," pass to exponent vectors, replace *or* with set union (inclusion–exclusion, correction term = divisors of gcd) and *and* with set intersection (correction term = divisors of gcd again, since intersecting two down-sets is itself a down-set at the componentwise minimum) — never reach for lcm when the connective governs *divisors*, only when it governs *multiples*.

▷ **Cue:** two divisibility conditions on numbers sharing the same prime factors, joined by *or* $\longrightarrow$ **Move:** convert to exponent vectors; count via $\tau(a) + \tau(b) - \tau(\gcd(a, b))$

▷ **Cue:** "n divides a and n divides b" $\longrightarrow$ **Move:** this set of n is exactly the divisor set of $\gcd(a, b)$, by definition of gcd — never substitute lcm here

▷ **Cue:** a number written as a product of prime powers $\longrightarrow$ **Move:** translate every divisibility statement about it into a componentwise inequality on the exponent vector

1983 A2

1983 A2. The hands of an accurate clock have lengths 3 and 4. Find the distance between the tips of the hands when that distance is increasing most rapidly.

Fields: calculus / plane geometry (rates of change, law of cosines, projection of velocity) **Position:** A2 — the trapdoor problem **Answer:** $\sqrt{7}$. If O is the pivot, A the tip of the length-4 hand and B the tip of the length-3 hand, the rate of change of AB is greatest exactly when $AB \perp OB$, and then $AB = \sqrt{4^2 - 3^2} = \sqrt{7}$ by Pythagoras.

Layer 1 THE WHAT — surface analysis

The natural attempt. You put the pivot at the origin and let $\theta(t)$ be the angle between the two hands at time t; since both hands rotate at constant (different) angular speeds, θ is a linear function of t, so $d\theta/dt$ is some nonzero constant — call it ω; you do not yet need its actual value. By the law of cosines, if $x(t)$ is the distance between the tips,

$$x^2 = 3^2 + 4^2 - 2 \cdot 3 \cdot 4 \cos\theta = 25 - 24\cos\theta.$$

You want the value of θ maximizing $|dx/dt|$. Differentiating implicitly,

$$2x\,\frac{dx}{d\theta} = 24\sin\theta \implies \frac{dx}{d\theta} = \frac{12\sin\theta}{x}.$$

Since $d\theta/dt = \omega$ is constant, maximizing $|dx/dt| = |\omega|\,|dx/d\theta|$ is the same as maximizing $|dx/d\theta|$ over θ. So you differentiate again: writing $x^2 = 25 - 24\cos\theta$ and $12\cos\theta - x\,d^2x/d\theta^2 + (dx/d\theta)^2 = (dx/d\theta)^2 \cdot (\text{terms})$, or more directly, differentiating $g(\theta) = \sin\theta/\sqrt{25 - 24\cos\theta}$ and setting $g'(\theta) = 0$, you land on

$$\cos\theta\,(25 - 24\cos\theta) - 12\sin^2\theta = 0 \implies 12\cos^2\theta - 25\cos\theta + 12 = 0,$$

a quadratic in $\cos\theta$ with roots $\cos\theta = \frac{3}{4}$ and $\cos\theta = \frac{4}{3}$ (rejected, out of range). This is correct, and it finishes the problem: $x^2 = 25 - 24(\frac{3}{4}) = 7$. But look at what you did to get here: you built a formula for x in terms of θ, differentiated it once to get a rate, then differentiated *that* to find where the rate itself is extremal — a second-derivative optimization in an angle variable that only entered the problem because you chose to work with $\cos\theta$ explicitly. (The Putnam committee's own account of this problem calls it a deliberate "trapdoor": a competent student can parametrize both tips' actual positions on the clock face, in real time, with the true angular

speeds of an hour and minute hand, and grind through a full two-variable calculus setup before ever getting this far. That approach also finishes — eventually — but it spends far more time than the problem is worth.)

The model chain. Fix the pivot O and, since only the *included angle* between the hands matters for the distance AB, work in the frame where the length-4 hand OA is motionless and the length-3 hand OB rotates alone at the constant relative angular speed ω. Then B traces a circle of radius 3 about O, so its velocity vector $\mathbf{v} = dB/dt$ is tangent to that circle: $\mathbf{v} \perp OB$ always, with constant magnitude $|\mathbf{v}| = 3|\omega|$.

Since A is fixed, the rate of change of the distance $AB = |A - B|$ is the component of $-\mathbf{v}$ along the direction from B to A — equivalently, the scalar projection of $\mathbf{v}$ onto line AB:

$$\frac{d}{dt}(AB) = -\,\mathbf{v} \cdot \widehat{BA}.$$

A vector of fixed length $|\mathbf{v}|$ has its component along a given direction maximized in absolute value exactly when it is parallel to that direction. So $|d(AB)/dt|$ is largest exactly when $\mathbf{v} \parallel AB$. But $\mathbf{v} \perp OB$ always; a vector parallel to AB and perpendicular to OB exists as the actual velocity direction only at the instant $AB \perp OB$ too. So the maximizing configuration is: *the segment AB is perpendicular to OB.*

At that instant, triangle OAB has a right angle at B, with hypotenuse $OA = 4$ and leg $OB = 3$. By the Pythagorean theorem,

$$AB = \sqrt{OA^2 - OB^2} = \sqrt{16 - 9} = \sqrt{7}.$$

Such a position is attained as B sweeps its full circle (the tangent line from A to the circle of radius 3 touches it exactly when $OA > OB$, which holds since $4 > 3$), so this maximum is achieved, and the answer is $\sqrt{7}$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree up to the law-of-cosines relation $x^2 = 25 - 24\cos\theta$ — or, in the model's telling, they agree that only the angle between the hands matters at all. They split at the very next move: the natural attempt treats x as an explicit function of θ and asks calculus (a second derivative) to find its extremal rate; the model asks instead what a *rate of change of a distance* means geometrically, and never writes x as a function of θ at all.

> *The rate of change of a distance between a fixed point and a moving point is nothing but the moving point's velocity, projected onto the line joining the two points — I don't need to write down $x(\theta)$ and differentiate it if I can reason about that projection directly.*

> *The velocity of a point moving on a circle around O is always perpendicular to the radius to O, and it has constant magnitude. A fixed-magnitude vector's projection onto a line is biggest exactly when the vector is parallel to that line — so I should ask when the velocity (perpendicular to OB) is parallel to AB, which is the same as asking when AB itself is perpendicular to OB.*

> *Once I know the maximizing configuration is a right angle at B, I don't need trigonometry to find the length — Pythagoras on triangle OAB gives it immediately.*

Layer 3 THE WHY

Why is the projection argument the natural route and not a trick? Go back to what a rate of change of distance *is*. If $B(t)$ moves and A is fixed, $AB(t)^2 = (A - B(t)) \cdot (A - B(t))$, so by the product rule

$$2\,AB \cdot \frac{d(AB)}{dt} = -2(A - B) \cdot \frac{dB}{dt} = -2\,AB \cdot \widehat{BA} \cdot \mathbf{v},$$

which is exactly the projection formula used above, derived from nothing but the definition of distance as a dot product. There is no step here that deserves to be called clever: differentiating a squared distance *always* produces a projection of velocity, because a squared distance is a dot product and the product rule on a dot product is a projection. The natural attempt reaches the same fact but buries it inside a differentiation of an explicit trigonometric formula, at the cost of a second derivative.

The second structural fact is why the maximizing configuration is a right angle at all: a vector $\mathbf{v}$ of fixed length has $|\mathbf{v} \cdot \hat{u}| \leq |\mathbf{v}|$ for any unit vector $\hat{u}$, with equality iff $\mathbf{v} \parallel \hat{u}$ — this is just the Cauchy–Schwarz inequality (or, more elementarily, $\mathbf{v} \cdot \hat{u} = |\mathbf{v}| \cos\phi$, extremal at $\phi = 0, \pi$). Combined with the structural fact that uniform circular motion has velocity always perpendicular to the radius, "$\mathbf{v} \parallel AB$" becomes "$AB \perp OB$" by pure substitution, with no case left unaccounted for. That the answer comes out via Pythagoras rather than trigonometry is a consequence, not a coincidence: a right angle is what perpendicularity *is*.

Finally, why does the problem punish the natural attempt so specifically? Because the setup invites you to introduce the clock's actual angular speeds (the minute hand turning 12 times faster than the hour hand) and to build $\theta(t)$ from two real rotation rates before you have asked whether that rate matters at all. It never does: ω cancels out of every extremization here, because we are asked only for a location (a value of θ, or a configuration of the triangle) where the rate is extremal, not for its numerical value. Bringing in real clock speeds is the "zombie maths" this problem was written to punish — the natural attempt above avoids that particular trap already, but still pays for treating the rate as a formula to be twice-differentiated rather than a vector to be projected.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "two segments from a common point, each rotating at its own constant rate; find where the rate of change of the distance between the endpoints is extremal." Write down immediately: "only the relative angle matters; freeze one hand and let the other move."

2. The method suggested is treating the rate of change of a distance to a fixed point as the projection of the moving point's velocity onto the line joining the points, then using that circular-motion velocity is always perpendicular to the radius.

3. Hesitation occurs at "am I allowed to freeze one hand?" Reinforce it by noting the distance AB depends on the two hands only through the angle between them, so only the *relative* angular velocity is physically relevant; fixing one hand and letting the other carry the full relative rate changes nothing about $AB(t)$.

4. The trap is parametrizing both tips' positions with the clock's actual angular speeds and grinding through a full two-hand calculus setup — technically correct, but the real angular speed never affects where the extremum occurs, so all of that setup is wasted work.

5. Whenever a rate of change of a *distance to a fixed point* is in question, translate immediately to "projection of velocity onto the connecting line," before writing any explicit formula in the angle or in time.

6. Reusable rule: to extremize the rate of change of the distance from a fixed point to a point in uniform circular motion, look for where the connecting segment is perpendicular to the radius (so the constant-magnitude, radius-perpendicular velocity is parallel to the segment); the length at that configuration then usually falls out of Pythagoras, not trigonometric optimization.

▷ **Cue:** rate of change of the distance between a fixed point and a moving point $\longrightarrow$ **Move:** project the moving point's velocity onto the line joining the two points, via $d(AB)/dt = -\mathbf{v} \cdot \widehat{BA}$, rather than differentiating an explicit distance formula

▷ **Cue:** a point in uniform circular motion about a centre O $\longrightarrow$ **Move:** its velocity is always perpendicular to the radius O–point and has constant magnitude

▷ **Cue:** maximizing the projection of a fixed-magnitude vector onto a variable line ⟶ **Move:** the maximum occurs exactly when the vector is parallel to that line (Cauchy–Schwarz equality case)

▷ **Cue:** two rotating arms from a common pivot, only their relative position mattering ⟶ **Move:** freeze one arm; the relative angular speed is the only quantity that survives into the answer, so its actual numeric value can be left unevaluated

1983 A3

1983 A3. Let p be in the set $\{3, 5, 7, 11, \dots\}$ of odd primes and let

$$F(n) = 1 + 2n + 3n^2 + \cdots + (p-1)n^{p-2}.$$

Prove that if a and b are distinct integers in $\{0, 1, 2, \dots, p-1\}$ then $F(a)$ and $F(b)$ are not congruent modulo p, that is, $F(a) - F(b)$ is not exactly divisible by p.

Fields: number theory (finite fields, telescoping sums, Fermat's little theorem) **Position:** A3 — past the halfway point **Answer:** F is injective on $\{0, \dots, p-1\}$ modulo p because $F(n) \equiv (1-n)^{-1} \pmod p$ for $n \not\equiv 1$, and $F(1) \equiv 0$; the map $n \mapsto (1-n)^{-1}$ is injective on nonzero residues and never equals 0, so no two values of F on $\{0, \dots, p-1\}$ can coincide mod p.

Layer 1 THE WHAT — surface analysis

The natural attempt. You want to show $F(a) \not\equiv F(b) \pmod p$ for $a \neq b$, so the direct object to study is the difference

$$F(a) - F(b) = \sum_{k=1}^{p-1} k \left(a^{k-1} - b^{k-1}\right).$$

Each term is divisible by $a - b$ (since $a^{k-1} - b^{k-1} = (a-b)(a^{k-2} + a^{k-3}b + \cdots + b^{k-2})$), so $F(a) - F(b)$ is divisible by $a - b$ — but that is a statement about ordinary integers, not about divisibility by p, and $a - b$ itself is generally not a multiple of p (indeed $|a - b| < p$ for $a, b \in \{0, \dots, p-1\}$), so this observation, true as it is, does not touch the actual claim. Pressing on, you could try to write out

$$\frac{F(a) - F(b)}{a - b} = \sum_{k=1}^{p-1} k \sum_{j=0}^{k-2} a^j b^{k-2-j}$$

and attempt to show this double sum is not divisible by p directly — a genuine polynomial identity in two variables, with no obvious mod-p simplification available, since nothing here has yet produced a term of the shape n^p or n^{p-1} where Fermat's little theorem could enter. You are doing correct algebra, and it is going nowhere useful in the time you have.

The model chain. Look at the coefficients of $F(n) = 1 + 2n + 3n^2 + \cdots + (p-1)n^{p-2}$: they are the consecutive integers $1, 2, \dots, p-1$, multiplying consecutive powers of n. Multiply $F(n)$ by n and subtract, so that the coefficients — which differ by exactly 1 at each step — collapse:

$$nF(n) = n + 2n^2 + 3n^3 + \cdots + (p-1)n^{p-1},$$

$$(1-n)F(n) = F(n) - nF(n) = 1 + n + n^2 + \cdots + n^{p-2} - (p-1)n^{p-1}.$$

The first block is now a plain geometric sum. Multiply by $(1-n)$ a second time to remove it as well: using $(1-n)(1 + n + \cdots + n^{p-2}) = 1 - n^{p-1}$,

$$(1-n)^2 F(n) = \left(1 - n^{p-1}\right) - (1-n)(p-1)n^{p-1} = 1 - p\,n^{p-1} + (p-1)n^p.$$

Now reduce modulo p. The middle term vanishes: $p\,n^{p-1} \equiv 0$. For the last term, $(p-1)n^p \equiv (-1)n^p \equiv -n^p$, and by Fermat's little theorem $n^p \equiv n \pmod p$ for every integer n (including $n \equiv 0$). So

$$(1-n)^2 F(n) \equiv 1 - n \pmod p.$$

Case $n \not\equiv 1 \pmod p$**:** divide both sides by $1-n$ (invertible mod p, since p is prime and $1-n \not\equiv 0$):

$$(1-n)F(n) \equiv 1 \implies F(n) \equiv (1-n)^{-1} \pmod p.$$

Case $n \equiv 1 \pmod p$**:** the division above is illegal, so compute $F(1)$ directly: $F(1) = 1 + 2 + \cdots + (p-1) = \frac{p(p-1)}{2}$. Since p is odd, $\frac{p-1}{2}$ is an integer, so $F(1) = p \cdot \frac{p-1}{2} \equiv 0 \pmod p$.

Now take distinct $a, b \in \{0, \ldots, p-1\}$. If neither equals 1, then $F(a) \equiv (1-a)^{-1}$ and $F(b) \equiv (1-b)^{-1}$; since $a \not\equiv b$, also $1-a \not\equiv 1-b \pmod p$, and the inverse map on the nonzero residues mod p is injective (it's a bijection of the group $(\mathbb{Z}/p\mathbb{Z})^\times$), so $F(a) \not\equiv F(b)$. If one of them equals 1 (say $a = 1$), then $F(a) \equiv 0$ while $F(b) \equiv (1-b)^{-1} \not\equiv 0$ (an inverse of a nonzero element is never 0), so again $F(a) \not\equiv F(b)$. This covers every pair, completing the proof.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start from the same object, $F(n)$, and both correctly note that $F(a) - F(b)$ is divisible by $a - b$ over the integers. They split immediately after that observation: the natural attempt keeps working with the *difference* $F(a) - F(b)$, trying to control it directly; the model abandons the difference altogether and instead multiplies the single function $F(n)$ by $(1-n)$, not because the target expression asks for it, but because the coefficients $1, 2, \ldots, p-1$ are begging to be telescoped.

> *The coefficients of F are consecutive integers multiplying a geometric progression of powers of n — this is an arithmetico-geometric sum, and the standard move for those is to multiply by $(1-n)$ and subtract: it turns a linear coefficient sequence into a constant one, because consecutive integers differ by exactly 1.*

> *After one telescoping I'm left with a geometric sum $1 + n + \cdots + n^{p-2}$ — I already know how to collapse that: multiply by $(1-n)$ again. Doing it twice should leave me with nothing but powers of n, and once I see n^{p-1} or n^p appear, Fermat's little theorem becomes available mod p for the first time.*

> *Dividing by $(1-n)$ is only valid when $n \not\equiv 1 \pmod p$ — I have to isolate that single residue and check it by direct summation instead of hoping the general formula covers it.*

Layer 3 THE WHY

Why is multiplying by $(1-n)$ the natural move here, rather than a trick pulled from nowhere? Because $F(n)$ *is*, by inspection of its coefficients, the finite-difference analogue of differentiating $1 + x + x^2 + \cdots + x^{p-1}$ term by term ($\frac{d}{dx}x^k = kx^{k-1}$ produces exactly the coefficient pattern $1, 2, 3, \ldots$). Multiplying a sum by $(1 - (\text{common ratio}))$ and subtracting is the standard algebraic device — Abel summation, or the schoolroom trick for summing an arithmetico-geometric series — for eliminating a coefficient that increases by a constant step each term; it is not particular to this problem, it is what you always do when the coefficients are in arithmetic progression and the base terms are in geometric progression. Once that telescoping is applied twice, the entire content of F collapses onto n^{p-1} and n^p, and those exponents are not incidental: they are exactly the exponents Fermat's little theorem was built to control mod p. The problem's own choice to stop the sum at $(p-1)n^{p-2}$ — one short of degree $p-1$ — is

what makes the two rounds of multiplying by $(1-n)$ land exactly on n^{p-1} and n^p rather than some unhelpful degree; the structure is aimed at Fermat from the start.

Why does working with the difference $F(a)-F(b)$ fail to find this route on its own? Because factoring out $(a-b)$ from each term is a true but mod-p-blind manipulation: it is an identity in the integers, valid before you ever reduce modulo p, and reducing afterwards gives you no new leverage, since nothing about the resulting double sum has a natural degree-p or degree-$(p-1)$ term for Fermat to act on. The natural attempt is not wrong so much as premature: it commits to comparing two values of F before it has found the single-variable identity that makes each value of F individually easy to compute mod p.

Finally, the case split at $n \equiv 1$ is not bookkeeping tacked onto the end; it is forced by the algebra itself. Dividing by $(1-n)$ is only legitimate when $1-n$ is invertible mod p, i.e. when $n \not\equiv 1$; the problem's insistence that p be an *odd* prime is exactly what makes $F(1) = p(p-1)/2$ come out as an integer multiple of p (if p were 2, $(p-1)/2$ would not be an integer and the sum $1+2+\cdots+(p-1)$ would need separate treatment) — so the hypothesis "p odd" and the case split at $n=1$ are the same fact seen from two sides.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is a finite sum whose coefficients are consecutive integers multiplying consecutive powers of a variable: $1,2,3,\ldots$ times $n^0, n^1, n^2, \ldots$. Write down immediately: "arithmetico-geometric sum; multiply by $(1-n)$ and subtract."

2. The method suggested is telescoping (Abel summation) applied twice — once to kill the linear coefficients, once to collapse the resulting geometric sum — followed by reduction modulo p via Fermat's little theorem, since the telescoping is engineered to produce exactly the exponents $p-1$ and p.

3. Hesitation occurs at dividing by $(1-n)$: is that legitimate mod p? Reinforce it by checking invertibility explicitly (valid iff $n \not\equiv 1 \pmod{p}$) and handling $n \equiv 1$ by direct summation rather than assuming the general formula extends.

4. The trap is working directly with the difference $F(a)-F(b)$, factoring out $(a-b)$, and trying to control the resulting double sum by brute force — a true identity that never produces a term Fermat's little theorem can simplify.

5. Whenever a sum's coefficients form an arithmetic progression multiplying a geometric one, and the sum's top degree is one less than a prime you are working modulo, expect two rounds of $(1-x)$-telescoping to expose exactly the degree-$(p-1)$ and degree-p terms that Fermat's little theorem needs.

6. Reusable rule: to compare values of a function mod p across a full residue system, don't attack the difference of two values directly — first find a closed form (via telescoping, differencing, or a known identity) for a single value mod p, then use the injectivity of a simple, already-understood map (here, the inverse map on $(\mathbb{Z}/p\mathbb{Z})^\times$) to conclude distinctness, isolating any point where an algebraic step (here, division by $1-n$) breaks down as its own case.

▷ **Cue:** a finite sum with coefficients in arithmetic progression multiplying a geometric progression $\longrightarrow$ **Move:** multiply the whole sum by $(1-\text{ratio})$ and subtract, to telescope the coefficients down to a constant; repeat if a geometric sum remains

▷ **Cue:** an expression reduced to n^{p-1} or n^p while working modulo an odd prime p $\longrightarrow$ **Move:** apply Fermat's little theorem ($n^{p-1} \equiv 1$ for $n \not\equiv 0$, or $n^p \equiv n$ for all n) to collapse it

▷ **Cue:** needing to prove a function is injective on a full set of residues mod p $\longrightarrow$ **Move:** find a closed form for its value mod p in terms of an already-known injective map (e.g. the inverse map on nonzero residues), and separately verify the one point where the closed form's derivation breaks down

1983 A4

1983 A4. Let k be a positive integer and let $m = 6k-1$. Let

$$S(m) = \sum_{j=1}^{2k-1} (-1)^{j+1} \binom{m}{3j-1}.$$

For example with $k = 3$,

$$S(17) = \binom{17}{2} - \binom{17}{5} + \binom{17}{8} - \binom{17}{11} + \binom{17}{14}.$$

Prove that $S(m)$ is never zero. [As usual, $\binom{m}{r} = \dfrac{m!}{r!\,(m-r)!}$.]

Fields: combinatorics / algebra (binomial sums, roots-of-unity filter, congruences) **Position:** A4 — mid-paper; ten minutes if you filter with roots of $z^3 = -1$, a long recursion if you do not **Answer:** $S(m) = 1 + (-1)^{k+1}\, 3^{3k-1}$ exactly, so $S(m) \equiv 1 \pmod 3$ and $|S(m)| \geq 8$; in particular $S(m) \neq 0$. (Checks: $S(5) = 10$, $S(11) = -242$.)

Layer 1 THE WHAT — surface analysis

The natural attempt. You compute small cases, because the problem gave you one. $k = 1$: $m = 5$, $S(5) = \binom{5}{2} = 10$. $k = 2$: $m = 11$, $S(11) = \binom{11}{2} - \binom{11}{5} + \binom{11}{8} = 55 - 462 + 165 = -242$. Nonzero, but the numbers carry no visible pattern: 10 and -242. You look for parity: 10 is even, so "S is odd" is dead. You try Pascal's rule, $\binom{m}{r} = \binom{m-1}{r} + \binom{m-1}{r-1}$, to relate $S(m)$ to $S(m-6)$, and discover that Pascal's rule does not move you in steps of six: it mixes the residues $r \equiv 2$ with $r \equiv 1 \pmod 3$, and those with $r \equiv 0$. You now have three interlocking sums and a recurrence among them. That is the official solution's route, and it works — it produces $T(m) = 3\big[T(m-1) - T(m-2)\big]$ for the full alternating sum T, whence $T \equiv 0 \pmod 3$ — but it takes three auxiliary sequences, a difference operator, and an expansion of ∇^3, and at no point does it tell you what $S(m)$ actually is. On first contact most students lose the thread among T_0, T_1, T_2 long before the factor 3 appears.

The model chain. *Step 1: complete the sum.* The indices $3j-1$ run over $2, 5, 8, \ldots, 6k-4$. The next index in the progression is $6k-1 = m$ itself, with $j = 2k$ and sign $(-1)^{2k+1} = -1$, contributing $-\binom{m}{m} = -1$. Beyond that, $\binom{m}{r} = 0$. So, with

$$T(m) = \sum_{j \geq 0} (-1)^j \binom{m}{3j+2} = \binom{m}{2} - \binom{m}{5} + \binom{m}{8} - \cdots,$$

we have $T(m) = S(m) - 1$, that is, $S(m) = T(m) + 1$.

Step 2: filter. The sign $(-1)^j$ attached to $r = 3j+2$ is what ζ^r produces when $\zeta^3 = -1$: $\zeta^{3j+2} = (-1)^j \zeta^2$. Let ζ run over the three roots of $z^3 + 1 = 0$, namely -1, $\omega = e^{i\pi/3}$, $\bar\omega = e^{-i\pi/3}$. Their sum is 0 (no z^2 term), and the sum of their squares is $0^2 - 2 \cdot 0 = 0$ (no z term). Hence $\sum_\zeta \zeta^s = 0$ for $s = 1, 2$ and $= 3$ for $s = 0$. Writing $r = 3j + 2 + s$ with $s \in \{0, 1, 2\}$,

$$\frac{1}{3} \sum_{\zeta^3 = -1} \zeta^{-2} (1+\zeta)^m = \frac{1}{3} \sum_r \binom{m}{r} \sum_\zeta \zeta^{r-2} = \sum_j (-1)^j \binom{m}{3j+2} = T(m).$$

Step 3: evaluate. For $\zeta = -1$ the term is $(1-1)^m = 0$. For $\zeta = \omega$: $1 + \omega = \frac{3}{2} + \frac{\sqrt{3}}{2} i = \sqrt{3}\, e^{i\pi/6}$, so $\omega^{-2}(1+\omega)^m = 3^{m/2} e^{i(m-4)\pi/6}$. The $\bar\omega$ term is its conjugate. So

$$T(m) = \tfrac{2}{3}\, 3^{m/2} \cos \frac{(m-4)\pi}{6}.$$

With $m = 6k-1$: $\frac{(m-4)\pi}{6} = k\pi - \frac{5\pi}{6}$, and $\cos(k\pi - \frac{5\pi}{6}) = (-1)^k \cos\frac{5\pi}{6} = (-1)^{k+1}\frac{\sqrt{3}}{2}$. Therefore

$$T(m) = \tfrac{2}{3} \cdot 3^{(6k-1)/2} \cdot (-1)^{k+1}\tfrac{\sqrt{3}}{2} = (-1)^{k+1}3^{3k-1}.$$

Step 4: conclude. $S(m) = 1 + (-1)^{k+1}3^{3k-1}$. Since $3k-1 \geq 2$, $|3^{3k-1}| \geq 9 > 1$, so $S(m) \neq 0$. (Equivalently: $S(m) \equiv 1 \pmod 3$.) Check: $k=1$ gives $1+9=10$; $k=2$ gives $1-243=-242$. Both match the direct computation.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start the same way: small cases, and the observation that the sum picks out every third binomial coefficient. They split at the phrase "every third, with alternating sign". The natural attempt reaches for Pascal's rule, which acts on m. The model reaches for a generating function, which acts on r, and asks which z makes z^r reproduce the pattern of the signs.

> *This is $\sum \binom{m}{r} c_r$ with a weight c_r that depends on r mod 6. Any such sum is a combination of $(1+z)^m$ at the sixth roots of unity. The sign flips every time r goes up by 3, so I want $z^3 = -1$. Three roots, three values of $(1+z)^m$, and one of them is zero.*

> *The sum stops at $6k-4$, one step short of $m = 6k-1$. That missing step is $-\binom{m}{m} = -1$. So the problem's S is the complete filtered sum plus 1, and the "1" is the whole reason it cannot vanish.*

Layer 3 THE WHY

Why is the roots-of-unity filter the natural tool and not a trick? Go back to what the binomial theorem says: $(1+z)^m = \sum_r \binom{m}{r} z^r$ is a machine that attaches the weight z^r to the r-th coefficient. If you want some other weight c_r, you need a combination of these machines. When c_r depends only on r modulo N, the weights z^r with $z^N = 1$ (or, here, $z^3 = -1$, a sub-family of the sixth roots) span every such pattern, because summing z^s over all roots of $z^N = c$ kills every s not divisible by N. That last fact is nothing deeper than Vieta: the roots of $z^3 + 1$ sum to minus the z^2 coefficient, which is 0. So "every third coefficient" is a statement about roots of a cubic, and "with alternating signs" only chooses which cubic.

Why does the answer come out as a power of 3? Because $|1+\omega| = \sqrt{3}$ for $\omega = e^{i\pi/3}$: the point $1+\omega$ is the long diagonal of a rhombus of side 1 with angle $60°$, of length $2\cos 30° = \sqrt{3}$. Raise it to the m-th power and you get $3^{m/2}$ times a rotation. The rotation angle $m\pi/6$ lands on a multiple of $\pi/6$, so the cosine is one of $0, \pm\frac{1}{2}, \pm\frac{\sqrt{3}}{2}, \pm 1$, and the $\sqrt{3}$ combines with $3^{m/2}$ (m odd) to give an integer power of 3. The divisibility by 3 that the official solution extracts from a recurrence is visible here as the modulus $\sqrt{3}$ of $1+\omega$.

Where does the official solution's recurrence come from? It introduces T_0, T_1, T_2 (the three residues of r mod 3) and derives $T_2(m) = 3[T_2(m-1) - T_2(m-2)]$. That recurrence has characteristic equation $x^2 - 3x + 3 = 0$, whose roots are $\frac{3 \pm i\sqrt{3}}{2} = \sqrt{3}e^{\pm i\pi/6} = 1+\omega, 1+\bar{\omega}$. The recurrence *is* the filter, written without complex numbers. The factor 3 in front is $|1+\omega|^2$. So the two solutions are the same mathematics; the filter gives you the closed form, and with it an exact value rather than a congruence.

Why is the "+1" the crux? The designers chose to stop the sum at $j = 2k-1$ rather than $j = 2k$. The complete sum is $\pm 3^{3k-1}$, which is divisible by 3. Removing the final term -1 shifts it to $\pm 3^{3k-1} + 1 \equiv 1 \pmod 3$. Had they included the last term, the sum would be nonzero anyway, but the problem would have told you its size. Truncating one term early disguises the structure — and the first thing to do with a truncated sum is to ask what completes it.

Layer $+\alpha$ BACKTESTING

1. The cue is "$\binom{m}{3j-1}$ with $(-1)^{j+1}$": binomial coefficients sampled along an arithmetic progression, with a sign pattern periodic in the index. Write down $(1+z)^m = \sum \binom{m}{r} z^r$ and the equation $z^3 = -1$.

2. The roots-of-unity filter: average $\zeta^{-2}(1+\zeta)^m$ over the three roots of $z^3 = -1$, then evaluate in polar form.

3. Hesitation comes at the upper limit $2k-1$, which looks arbitrary and blocks the use of a complete sum. Reinforce it away by writing out the next term: it is $\binom{m}{m}$, and it is the last nonzero term anyway.

4. The trap is Pascal's rule. It is the reflex for binomial identities and it does lead somewhere, but it forces three coupled sequences and hides the closed form. A second trap: after finding 10 and -242, trying a parity argument, which fails at once.

5. Any sum of binomial coefficients over r in a residue class, with or without periodic signs, is a finite combination of $(1+\zeta)^m$ over suitable roots of unity. Recognise the pattern by the words "every third", "every fourth", or a sign that flips with a period.

6. Reusable rule: when a sum is "nearly complete", complete it and account for what you added. The complete object usually has a closed form; the correction is usually the entire answer to a nonvanishing or divisibility question.

▷ **Cue:** a sum $\sum \binom{m}{r}$ restricted to $r \equiv a \pmod N$, possibly with a periodic sign $\longrightarrow$ **Move:** average $\zeta^{-a}(1+\zeta)^m$ over the roots of $z^N = \pm 1$ chosen to reproduce the sign

▷ **Cue:** a finite sum stopping one term before its natural end $\longrightarrow$ **Move:** add the missing term, evaluate the complete sum, subtract

▷ **Cue:** $1 + e^{i\pi/3}$ or $1 + e^{2\pi i/3}$ appears $\longrightarrow$ **Move:** write it in polar form ($\sqrt{3}e^{i\pi/6}$, $e^{i\pi/3}$) before raising to a power

▷ **Cue:** prove a quantity is never zero $\longrightarrow$ **Move:** find a modulus in which it is a fixed nonzero residue, or find it exactly

▷ **Cue:** a linear recurrence with characteristic roots of modulus $\sqrt{3}$ or $\sqrt{2}$ $\longrightarrow$ **Move:** those roots are $1+\zeta$ for a root of unity ζ; a binomial filter is behind it

1983 A5

1983 A5. Prove or disprove that there exists a positive real number u such that $\lfloor u^n \rfloor - n$ is an even integer for all positive integers n.

Here $\lfloor x \rfloor$ denotes the greatest integer less than or equal to x.

Fields: number theory / analysis (floor functions, algebraic integers with small conjugate, nested intervals) **Position:** A5 — hard end of the morning; the construction is short once you know why u^n can be close to an integer, and hopeless by trial **Answer:** Such u exists. One explicit example: $u = \frac{3+\sqrt{17}}{2} \approx 3.5616$, for which $\lfloor u^n \rfloor = 3, 12, 45, 160, 573, 2040, \ldots$ and $\lfloor u^n \rfloor \equiv n \pmod 2$ for every $n \geq 1$. (The official solution instead proves existence by nested intervals.)

Layer 1 THE WHAT — surface analysis

The natural attempt. You try candidates. $u = 3$: $\lfloor 3 \rfloor - 1 = 2$, good; $\lfloor 9 \rfloor - 2 = 7$, bad. $u = \sqrt{10}$: $3 - 1 = 2$, $10 - 2 = 8$, $\lfloor 10\sqrt{10} \rfloor - 3 = 31 - 3 = 28$, $100 - 4 = 96$, $\lfloor 100\sqrt{10} \rfloor - 5 = 316 - 5 = 311$, bad. You notice that every fixed guess must satisfy infinitely many independent parity conditions, each a coin flip, and you begin to suspect the answer is "disprove". But you cannot find an obstruction: nothing about $\lfloor u^n \rfloor$ for large n is forced by the small cases. Then you swing to construction: choose u in $[3, 4)$ so that $n = 1$ works; then shrink to the part of $[3, 4)$ where $\lfloor u^2 \rfloor$ is even; then shrink again for $n = 3$. This is the right instinct — it is the official solution — but on first contact it is unclear whether the surviving set stays nonempty. You need

a quantitative statement: at each stage the admissible set, pushed forward to u^{n+1}, is an interval of length at least 2, so it contains a full unit interval $[a, a+1)$ with a of either parity. Without that estimate the nesting argument is a hope, not a proof.

The model chain. *Idea.* Choose u so that u^n is always within distance 1 of a known integer, with a known sign for the error.

Choice. Let u, u' be the roots of $x^2 - 3x - 2 = 0$:

$$u = \frac{3+\sqrt{17}}{2} \approx 3.5616, \qquad u' = \frac{3-\sqrt{17}}{2} \approx -0.5616.$$

So $-1 < u' < 0$. (Check: the quadratic $p(x) = x^2 - 3x - 2$ has $p(-1) = 2 > 0$ and $p(0) = -2 < 0$.)

Integer traces. Let $s_n = u^n + u'^n$. Then $s_0 = 2$, $s_1 = u + u' = 3$, and since $u^2 = 3u + 2$ we have $u^n = 3u^{n-1} + 2u^{n-2}$, likewise for u'; adding,

$$s_n = 3s_{n-1} + 2s_{n-2} \qquad (n \geq 2).$$

By induction every s_n is an integer. Modulo 2 the recurrence reads $s_n \equiv s_{n-1}$, so from $s_1 = 3$ every s_n with $n \geq 1$ is odd.

The floor. $u^n = s_n - u'^n$, and $0 < |u'^n| < 1$.

- n odd: $u'^n \in (-1, 0)$, so $u^n = s_n + |u'^n| \in (s_n, s_n + 1)$ and $\lfloor u^n \rfloor = s_n$, which is odd. Then $\lfloor u^n \rfloor - n$ is odd minus odd, even.
- n even: $u'^n \in (0, 1)$, so $u^n \in (s_n - 1, s_n)$ and $\lfloor u^n \rfloor = s_n - 1$, which is even. Then $\lfloor u^n \rfloor - n$ is even minus even, even.

So $\lfloor u^n \rfloor - n$ is even for every $n \geq 1$. The statement is true. (Values: $s_n = 3, 13, 45, 161, 573, 2041$, floors $3, 12, 45, 160, 573, 2040$.)

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains begin by testing candidate values of u and seeing parity fail somewhere. They split at the question "what kind of u would let me *know* $\lfloor u^n \rfloor$ for every n at once?" The natural attempt does not ask it, and so either gives up or builds u by infinite approximation. The model asks it and answers: a u whose powers stay within 1 of integers — a root of a monic integer quadratic whose other root is small.

> *I cannot control $\lfloor u^n \rfloor$ for all n unless I know u^n to within 1. The only numbers whose powers I know that precisely are ones like $1 + \sqrt{2}$, where $u^n + u'^n$ is an integer and u'^n is tiny. So I want a quadratic $x^2 - ax - b$ with one root big and the other in $(-1, 1)$, and I get to choose a and b to fix the parities.*

> *Now the parity design. If u' is negative, the error $-u'^n$ alternates in sign, so the floor is s_n for odd n and $s_n - 1$ for even n. That flips parity every other step — exactly the n in "$\lfloor u^n \rfloor - n$". So I need all s_n odd: a odd, b even.*

Layer 3 THE WHY

Why is "a root with small conjugate" the natural object, not a piece of folklore? Start from what the floor function needs. $\lfloor u^n \rfloor$ is determined once you know an integer N and the fact $N \leq u^n < N + 1$. For a general real u you know neither; u^n is a real number whose fractional part wanders unpredictably. To know $\lfloor u^n \rfloor$ for *every* n with one argument, you need a formula that writes u^n as (integer) + (small, controlled error). The binomial theorem gives exactly this

for quadratic irrationals: $(a+\sqrt{d})^n + (a-\sqrt{d})^n$ has all odd powers of $\sqrt{d}$ cancelling, so it is an integer. If the conjugate $a-\sqrt{d}$ has modulus less than 1, its powers are the error term, and they are in $(-1,1)$. That is not a trick; it is the only mechanism in the A-level toolkit that makes u^n computable to within 1 for all n simultaneously. (Numbers with this property — an algebraic integer greater than 1 whose other conjugates lie inside the unit circle — are called Pisot numbers. You do not need the name.)

Why negative u'? Look at the target: $\lfloor u^n \rfloor \equiv n \pmod 2$, a parity that alternates with n. You could try to make the traces s_n themselves alternate in parity, but that constrains the recurrence mod 2 at every step. A negative u' supplies it for free: the correction "-1 or 0" alternates with n. Then the traces only need to be constant mod 2, and constant odd is the easiest pattern to arrange (a odd, b even). A positive u' would give floor $s_n - 1$ always, so you would need $s_n \equiv n+1$, but $s_2 = a^2 + 2b \equiv a \equiv s_1 \pmod 2$ — the parities of s_1 and s_2 agree, when they must differ. So positive u' cannot work with a quadratic; the sign of the conjugate is forced by the problem's phrasing.

Why does the official nested-interval proof work, and what does it teach? The official existence argument rests on one inequality: if $I_n = [a_n^{1/n}, (a_n+1)^{1/n})$ is the set of u with $\lfloor u^n \rfloor = a_n$, then the set of values u^{n+1} for $u \in I_n$ is an interval of length

$$(a_n+1)^{(n+1)/n} - a_n^{(n+1)/n} \;\geq\; a_n^{1/n} \;\geq 3$$

(the first inequality by the mean value theorem, the second because $a_n \geq 3^n$ is maintained inductively). An interval of length 3 contains a full $[a, a+1)$ for integers a of both parities, so the next condition can always be met, and the intervals I_n nest. Their intersection is nonempty (the left endpoints increase and are bounded), and any u in it works. The official proof shows that a "random" parity requirement can be met by some u — in fact any prescribed parity sequence can — which is a stronger statement than the problem asks. The explicit u is cheaper and tells you more about why: the floor of u^n is governed by an integer recurrence.

Caveat. The explicit example $u = (3+\sqrt{17})/2$ is this archive's construction, not the official one. It was proved above and also checked numerically (200-digit arithmetic) for $1 \leq n \leq 199$. The official nested-interval proof, which the OCR renders legibly, was checked step by step and is correct in substance; one detail it leaves implicit is the endpoint case (the intervals are half-open, so the limit u could in principle sit on a common right endpoint). Choosing each a_{n+1} so that $[a_{n+1}, a_{n+1}+1]$ lies strictly inside the image interval — possible because that interval has length at least 3 — makes the right endpoints strictly decrease and closes the gap.

Layer $+\alpha$ BACKTESTING

1. The cue is $\lfloor u^n \rfloor$ with a condition for *all* n. Write down "I need u^n = integer + error with $|\text{error}| < 1$ for every n", and immediately next to it $u^n + u'^n \in \mathbb{Z}$ for conjugate quadratic irrationals.

2. Construct u as the large root of $x^2 - ax - b$ with integers a, b, the other root in $(-1,1)$; use the recurrence $s_n = as_{n-1} + bs_{n-2}$ for the traces and read off parities mod 2.

3. Hesitation comes at "prove or disprove". Fixed guesses fail, and you drift towards disproof. Reinforce it away: one failed guess disproves nothing, and a statement about all n with only a parity condition is usually satisfiable by a structured u or by nested intervals. Try to construct before trying to refute.

4. Two traps. Choosing u' positive (for instance $u = 2+\sqrt{3}$): then $\lfloor u^n \rfloor = s_n - 1$ always, and the parity of s_n cannot alternate as required. And $u = 1+\sqrt{2}$: negative conjugate, but all s_n are even, so the parity comes out exactly wrong ($\lfloor u^n \rfloor - n$ is always odd). Get both the sign of u' and the parity of s_n right.

5. Floor or fractional part of u^n, $\lfloor (2+\sqrt{3})^n \rfloor$, "show $\{(\text{something})^n\} \to 1$", "last digit of $\lfloor (\ldots)^n \rfloor$": all are the conjugate trick. Any time a question wants exact integer information about a power of an irrational, add the conjugate power.

6. Reusable rule: to control a quantity for all n simultaneously, find a structure in which it is given exactly by a recurrence. If none exists, fall back to a nested-interval construction, and the key step there is an estimate that each new constraint cuts out an interval long enough to contain what you need.

▷ **Cue:** $\lfloor u^n \rfloor$ or $\{u^n\}$ required for every n $\longrightarrow$ **Move:** take u a quadratic irrational with conjugate $u' \in (-1,1)$; then $u^n = s_n - u'^n$ with s_n an integer recurrence

▷ **Cue:** a parity condition that alternates with n $\longrightarrow$ **Move:** make the error $-u'^n$ alternate in sign by choosing $u' < 0$

▷ **Cue:** the traces $s_n = u^n + u'^n$ of $x^2 - ax - b$ $\longrightarrow$ **Move:** $s_0 = 2$, $s_1 = a$, $s_n = as_{n-1} + bs_{n-2}$; reduce mod 2 to read parities

▷ **Cue:** "prove or disprove there exists u" with one condition per n $\longrightarrow$ **Move:** try nested intervals: show each condition leaves an interval, and the image under $u \mapsto u^{n+1}$ is long enough

▷ **Cue:** a monic integer quadratic, need a root in $(-1,0)$ $\longrightarrow$ **Move:** check the signs of $p(-1)$ and $p(0)$

1983 A6

1983 A6. Let $\exp(t)$ denote e^t and

$$F(x) = \frac{x^4}{\exp(x^3)} \int_0^x \int_0^{x-u} \exp(u^3 + v^3)\, dv\, du.$$

Find $\lim_{x\to\infty} F(x)$ or prove that it does not exist.

Caveat. The OCR of the source volume drops the displayed formula for $F(x)$. The statement above is the standard wording of 1983 A6, and it is consistent with the official solution as printed (the substitution $s = u - v$, $t = u + v$ over the triangle $u, v \geq 0$, $u + v \leq x$, and the denominator $E(x) = 2x^{-4}\exp(x^3)$). The limit 2/9 below was also confirmed numerically: $F(3) \approx 0.2409$, $F(4) \approx 0.2295$, $F(6) \approx 0.2243$, approaching 0.2222.

Fields: analysis (double integrals, L'Hôpital's rule, Laplace-type asymptotics) **Position:** A6 — last of the morning; routine machinery, but only if you stop differentiating at the right moment **Answer:** The limit exists and equals $\frac{2}{9}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You try to do the inner integral, $\int_0^{x-u} e^{v^3}\, dv$, and find that e^{v^3} has no elementary antiderivative. Swapping the order does not help; the integrand is symmetric in u and v. So you reach for L'Hôpital, since $F = G/E$ with $G(x)$ the double integral and $E(x) = x^{-4}e^{x^3}$, both tending to infinity. You need $G'(x)$, and differentiating a double integral whose region depends on x is unfamiliar. Suppose you get it right: $G'(x) = \int_0^x e^{u^3+(x-u)^3}\, du$. This is still an integral without a closed form, so you apply L'Hôpital again, which means differentiating G' under the integral sign; you get e^{x^3} from the endpoint plus $\int_0^x 3(x-u)^2 e^{u^3+(x-u)^3}\, du$, which is worse than what you started with. The ratio is now a sum of an explicit term and another intractable integral, and each further differentiation brings another. You are pushing the same unknown integral around.

The model chain. *Step 1: slice along the hypotenuse.* Write $G(x) = \int_0^x \int_0^{x-u} e^{u^3+v^3}\,dv\,du$, the integral over the triangle $T_x = \{u, v \geq 0,\ u+v \leq x\}$. Change variables to (u, s) with $s = u+v$ (Jacobian 1); T_x becomes $0 \leq u \leq s \leq x$. So

$$G(x) = \int_0^x K(s)\,ds, \qquad K(s) = \int_0^s e^{u^3+(s-u)^3}\,du,$$

and by the fundamental theorem of calculus $G'(x) = K(x)$.

Step 2: simplify the exponent. $u^3 + (x-u)^3 = x^3 - 3x^2u + 3xu^2 = x^3 - 3xu(x-u)$. Hence

$$G'(x) = e^{x^3}\int_0^x e^{-3xu(x-u)}\,du = 2e^{x^3}J(x), \qquad J(x) = \int_0^{x/2} e^{-3xu(x-u)}\,du,$$

using the symmetry $u \leftrightarrow x-u$.

Step 3: one application of L'Hôpital. $E(x) = x^{-4}e^{x^3} \to \infty$ and $E'(x) = e^{x^3}(3x^{-2} - 4x^{-5}) > 0$ for $x > (4/3)^{1/3}$. So if G'/E' has a limit, $F = G/E$ has the same limit. Now

$$\frac{G'(x)}{E'(x)} = \frac{2J(x)}{3x^{-2} - 4x^{-5}} = \frac{2\,x^2J(x)}{3 - 4x^{-3}}.$$

Step 4: estimate J. Put $u = t/x^2$. Then $3xu(x-u) = 3t - 3t^2/x^3$ and

$$x^2J(x) = \int_0^{x^3/2} \exp\Big(-3t + \frac{3t^2}{x^3}\Big)\,dt.$$

Lower bound: the integrand is at least e^{-3t}, so $x^2J \geq \frac{1}{3}(1 - e^{-3x^3/2})$. *Upper bound* (for $x \geq \sqrt{2}$): on $0 \leq t \leq x$, $3t^2/x^3 \leq 3/x$, so that part is at most $e^{3/x}\int_0^\infty e^{-3t}dt = \frac{1}{3}e^{3/x}$. On $x \leq t \leq x^3/2$, $t/x^3 \leq \frac{1}{2}$, so the exponent $-3t(1 - t/x^3) \leq -\frac{3}{2}t$, and that part is at most $\int_x^\infty e^{-3t/2}dt = \frac{2}{3}e^{-3x/2}$. Both bounds tend to $\frac{1}{3}$, so $x^2J(x) \to \frac{1}{3}$.

Step 5: conclude. $\dfrac{G'}{E'} \to \dfrac{2\cdot\frac{1}{3}}{3} = \dfrac{2}{9}$, hence $\lim_{x\to\infty} F(x) = \dfrac{2}{9}$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains apply L'Hôpital once and reach $G'(x) = \int_0^x e^{u^3+(x-u)^3}\,du$. They split there. The natural attempt differentiates again, because L'Hôpital is the tool in hand. The model stops differentiating and *estimates* G' directly, after simplifying the exponent to $x^3 - 3xu(x-u)$.

> *I will never get a closed form for these integrals, and I do not need one: I need their size as $x \to \infty$. An integral of $e^{(huge)\cdot(something)}$ is dominated by where the something is largest. So where is $u^3 + (x-u)^3$ largest on $[0, x]$, and how fast does it fall off?*

> *It is largest at the ends, $u = 0$ and $u = x$, where it equals x^3, and it falls off like $3x^2u$. So near each end the integral looks like $e^{x^3}\int_0^\infty e^{-3x^2u}du = e^{x^3}/(3x^2)$. Two ends: $2e^{x^3}/(3x^2)$. That is the whole answer; the rest is bounding the error.*

Layer 3 THE WHY

Why does the problem have a limit at all, and why does x^4 appear? Look at what the double integral measures. The integrand $e^{u^3+v^3}$ is astronomically larger near the points of the triangle where $u^3 + v^3$ is largest. Since $u^3 + v^3$ is convex on the triangle, its maximum is at a vertex: at $(x, 0)$ and $(0, x)$ it equals x^3, and at $(0, 0)$ it is 0. Near $(x, 0)$, write $u = x - a$ with a small; the

constraint $u+v \le x$ becomes $0 \le v \le a$, and $u^3+v^3 \approx x^3 - 3x^2a$. So the integral near that corner is about

$$e^{x^3}\int_0^\infty\int_0^a e^{-3x^2a}\,dv\,da = e^{x^3}\int_0^\infty a\,e^{-3x^2a}\,da = \frac{e^{x^3}}{9x^4}.$$

The corner region that matters is a tiny triangle with legs of order $1/x^2$, so its area is of order $1/x^4$. That is where the x^4 in the problem comes from: the problem-setter has normalised by the exact size of the dominant contribution. Two corners give $2/(9x^4)$, and $F \to 2/9$. The rigorous proof is this heuristic with bounds.

Why is the "9" split as 3×3, and why does one L'Hôpital produce it cleanly? One factor 3 is the rate $3x^2$ at which the exponent falls off moving along the hypotenuse; it appears in $J \approx 1/(3x^2)$. The other factor 3 comes from differentiating e^{x^3}, which is $3x^2e^{x^3}$; it appears in E'. The first L'Hôpital converts the *area* integral into a *line* integral along the hypotenuse — that is exactly what differentiating in x does, since increasing x adds a thin strip along $u+v=x$. The second application, in the official solution, is used on the remaining one-dimensional Gaussian-type integral after rotating coordinates; it works, but only because the official solution first rewrote G' with the substitution $s = w/\sqrt{x}$ into $\exp(3x^3/4)\int \exp(3w^2/4)\,dw$ form. Without that preparation, a second differentiation drags in an integral as hard as the first, which is where the natural attempt drowns.

Why slice along $u+v=s$? Because the region is a triangle whose moving side is $u+v=x$. The official solution uses $s=u-v$, $t=u+v$; I use (u,s) with $s=u+v$. Both are the same decision: choose coordinates in which the moving boundary is a coordinate line. Then the derivative of the region integral is an integral over that line, by the fundamental theorem of calculus, and no Leibniz rule for variable regions is needed.

Why can I trust the bounds without dominated convergence? The exponent $-3t+3t^2/x^3$ is at most $-3t+3/x$ where $t \le x$, and at most $-\frac{3}{2}t$ where $t \ge x$. Each piece is an exponential integral you can do by hand. That is the standard way to turn a Laplace heuristic into a proof at this level: split at a point that is large compared with the scale of the peak (1 in the variable t) but small compared with where the approximation breaks (x^3).

Layer $+\alpha$ BACKTESTING

1. The cue is $\exp(x^3)$ in the denominator facing an integral of $\exp(u^3+v^3)$: an exponential of a large quantity, normalised by its maximum. Write down "where is u^3+v^3 maximal on the region, and what is it there?" — answer x^3, at two corners.

2. Laplace's method, made rigorous with elementary bounds, preceded by one application of L'Hôpital (or the fundamental theorem of calculus) to reduce the double integral to a line integral.

3. Hesitation comes immediately after the first L'Hôpital, when G' is still an integral. Reinforce it away: an integral you cannot evaluate is still an integral you can bound. Ask for its size, not its value.

4. The trap is repeated L'Hôpital. Each differentiation of an integral with a variable integrand produces an endpoint term plus a new integral, and the new integral is no easier. A second trap: assuming the maximum of u^3+v^3 is in the interior or on the diagonal $u=v$; there it is only $x^3/4$, and the diagonal contributes nothing in the limit.

5. Any limit of the form $\frac{\int e^{\lambda\phi}}{e^{\lambda \max\phi}} \times (\text{power of } \lambda)$: find where ϕ is maximal, linearise (at a boundary) or take the quadratic term (at an interior maximum), and integrate the resulting exponential. The power of λ in the problem tells you which case you are in.

6. Reusable rule: when an integral has no closed form, stop trying to evaluate it and estimate it. The mass of $e^{\text{large}\times\phi}$ sits where ϕ is largest; the answer depends only on the local shape of ϕ there.

▷ **Cue:** an integral of $e^{\lambda\phi}$ with $\lambda\to\infty$, divided by $e^{\lambda\max\phi}$ $\longrightarrow$ **Move:** locate the maximiser of ϕ, expand ϕ there, integrate the local exponential (Laplace's method)

- ▷ **Cue:** a double integral over a region whose boundary moves with x $\longrightarrow$ **Move:** choose coordinates in which the moving edge is a coordinate line; then d/dx is a line integral by the FTC
- ▷ **Cue:** after one L'Hôpital, the derivative is still an unevaluable integral $\longrightarrow$ **Move:** stop differentiating; bound the integral from above and below
- ▷ **Cue:** $u^3 + (x-u)^3$ or similar symmetric expression $\longrightarrow$ **Move:** expand: $x^3 - 3xu(x-u)$; the variable part is a product vanishing at both ends
- ▷ **Cue:** an x^k normalisation in a limit problem $\longrightarrow$ **Move:** it measures the area or length of the region where the integrand is near its maximum; use it to check your local expansion
- ▷ **Cue:** need to make a Laplace heuristic rigorous $\longrightarrow$ **Move:** split the integral at a point far beyond the peak's width but well inside where the approximation is valid; bound each piece by an explicit exponential

Putnam 1983 — Paper B

1983 B1

1983 B1. Let v be a vertex (corner) of a cube C with edges of length 4. Let S be the largest sphere that can be inscribed in C. Let R be the region consisting of all points p between S and C such that p is closer to v than to any other vertex of the cube. Find the volume of R.

Fields: solid geometry (spheres inscribed in cubes, symmetry decomposition) **Position:** B1 — the opener of Paper B **Answer:** $\mathrm{vol}(R) = 8 - \dfrac{4\pi}{3}$. The region closer to v than to every other vertex is exactly the corner sub-cube of side 2, and the largest inscribed sphere meets that sub-cube in precisely one eighth of the ball, because the sphere's centre is the common vertex of the cube's eight symmetric octants.

Layer 1 THE WHAT — surface analysis

The natural attempt. Place v at the origin and let $C = [0,4]^3$, so the sphere S has centre $(2,2,2)$ and radius 2 (diameter 4, touching all six faces). "Closer to v than to any other vertex" is seven conditions, one per other vertex, each a half-space bounded by a perpendicular bisector plane. You write all seven out. Against $(4,0,0)$: $x^2 < (x-4)^2$, i.e. $x < 2$. Against $(0,4,0)$: $y < 2$. Against $(0,0,4)$: $z < 2$. Against a face-diagonal vertex such as $(4,4,0)$: $x^2 + y^2 < (x-4)^2 + (y-4)^2$, which simplifies to $x + y < 4$. Against $(4,0,4)$: $x + z < 4$. Against $(0,4,4)$: $y + z < 4$. Against the far vertex $(4,4,4)$: $x + y + z < 6$.

Seven inequalities is more than you want to carry into a volume computation, so before integrating you check whether some are redundant: given $x < 2$ and $y < 2$, certainly $x + y < 4$, so the face-diagonal conditions add nothing; given all three of $x, y, z < 2$, certainly $x + y + z < 6$. So the region reduces to the corner sub-cube $C' = [0,2]^3$. Good — but now you must find the volume of C' minus the part of the ball $\|(x,y,z) - (2,2,2)\| \le 2$ that lies in C', and the natural next step is to set up a triple integral: shift to $u = x-2, v = y-2, w = z-2$, so C' becomes $[-2,0]^3$ and you need

$$\mathrm{vol}\big(\{u,v,w \in [-2,0]\} \cap \{u^2 + v^2 + w^2 \ge 4\}\big).$$

You can grind this in spherical coordinates restricted to one octant, tracking that $u, v, w \le 0$ forces the polar and azimuthal angles into a specific quadrant range, and it does terminate — but you are one sign error in the angle bounds away from a wrong answer, and you have done a three-dimensional integral to answer a question that turns out to need none.

The model chain. Everything up to the reduction $R \cap (\text{near-}v \text{ octant}) = C' = [0,2]^3$ is unavoidable and is done exactly as above. The shortcut is in what comes next. The three planes $x = 2, y = 2, z = 2$ that carve C' out of C are also planes of symmetry of the sphere S, because S is centred exactly at $(2,2,2)$, the point where all three planes meet. Reflection in any of these three planes is an isometry that fixes the centre of S, hence maps S to itself, and the three reflections together generate a group of order 8 (independently flip the sign of $u = x-2$, $v = y-2$, $w = z-2$) which partitions space around $(2,2,2)$ into exactly eight congruent octant

cubes, one of which is C'. Since S is invariant under every element of this group and the eight octants are permuted simply transitively by it, S meets each octant in an equal volume: $\frac{1}{8}\,\mathrm{vol}(S)$. No integral is required. Hence

$$\mathrm{vol}(R) = \mathrm{vol}(C') - \tfrac{1}{8}\,\mathrm{vol}(S) = 2^3 - \frac{1}{8}\cdot\frac{4}{3}\pi\cdot 2^3 = 8 - \frac{4\pi}{3}.$$

Equivalently, summing over all eight congruent copies of R gives $8\,\mathrm{vol}(R) = \mathrm{vol}(C) - \mathrm{vol}(S) = 64 - \frac{32\pi}{3}$, the same answer.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths agree up to the identification of the near-v region as the sub-cube $C' = [0,2]^3$. They split at the next question: "how do I find the volume of $S \cap C'$?" The natural attempt reaches for an integral. The model asks first whether the group of symmetries defining C' already fixes the centre of S — and it does, which ends the problem without integration.

> *The three bisector planes that cut out C' are exactly the coordinate planes through the centre of the sphere. A sphere is invariant under every reflection through its centre. So reflecting C' across these planes maps S to itself while permuting the eight octants — meaning S must distribute its volume equally among them.*

> *I don't need to know the shape of $S \cap C'$ at all, only that eight isometric copies of it tile S exactly. A fraction that tiles a known whole into equal congruent pieces is $1/(\text{number of pieces})$, full stop.*

Layer 3 THE WHY

Why is it obvious, once stated, that the octant planes are symmetry planes of S? Because S is the largest sphere that fits in a cube of edge 4: it must be tangent to all six faces, which forces its centre to be equidistant from every pair of opposite faces, i.e. at the cube's own centre. The cube's centre is also the unique point fixed by the three reflections that generate the octant decomposition (each reflection swaps a pair of opposite faces and fixes their common perpendicular bisector plane, and all three bisector planes meet only at the centre). So the sphere's centre being forced to the cube's centre is not a coincidence to be checked case by case — it is a consequence of "largest inscribed sphere" meaning "tangent to all six faces", and once the centres coincide, invariance under the octant group is immediate from the definition of a sphere as a set of points at fixed distance from its centre: distance is unchanged by any isometry fixing that centre.

This is also why the seven vertex-inequalities collapse to three. "Closer to v than to w" is the half-space on v's side of the perpendicular bisector of vw. For the three vertices adjacent to v along an edge, that bisector is a coordinate mid-plane ($x = 2$, $y = 2$, $z = 2$). For a vertex reached along a face diagonal, such as $(4,4,0)$, the bisector condition $x + y < 4$ is the sum of the two edge conditions and is therefore implied by them, never independently binding; the same holds for the space-diagonal vertex with $x + y + z < 6$. The redundancy is not an accident to verify numerically each time — it follows because $x < 2, y < 2 \Rightarrow x + y < 4$ is just monotonicity of addition, so any inequality obtained by summing coordinates of already-satisfied strict inequalities is automatically satisfied. Recognising this in advance is what keeps the case list at three planes rather than seven.

Finally, why does the answer come out as a whole number minus a clean multiple of π rather than something uglier? Because the problem is engineered so that the polyhedral symmetry used to define R (bisecting the cube by vertex) is exactly the symmetry the inscribed sphere already possesses by construction. Whenever a region's defining symmetry group coincides with a symmetry group of the sphere sitting inside it, the sphere's volume must

divide evenly among the pieces the group creates: that is what "group acts simply transitively on interchangeable congruent regions, and the sphere is a level set of a group-invariant function" delivers for free, without any computation of the intersection's actual shape.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "largest sphere inscribed in a cube of edge 4." Write immediately: diameter $= 4$, radius $= 2$, centre $=$ centre of the cube. Do not re-derive this from tangency conditions each time; it is a fixed fact about inscribed spheres in cubes.

2. That cue, combined with "region closer to one vertex than to all others," suggests decomposing the cube into the eight congruent sub-cubes cut out by the three mid-planes, and asking how the sphere (already known to be centred at the common vertex of those sub-cubes) sits relative to that decomposition.

3. Hesitation occurs exactly at "how do I compute the volume of the ball inside one sub-cube?" Reinforce it away by asking first whether the sub-cube's bounding planes are symmetry planes of the ball — check this before reaching for an integral, because if they are, the intersection's fraction is forced without knowing its shape.

4. Two traps. First, treating all seven "closer to v" inequalities as independent and needing separate treatment, rather than checking for redundancy among coordinate-sum conditions. Second, setting up a triple integral for $S \cap C'$ out of habit, when the volume follows from a one-line symmetry count.

5. Whenever a sphere's centre coincides with the fixed point of a reflection or rotation group used to cut up the surrounding space into congruent pieces, expect the sphere's volume to split evenly among those pieces — check the fixed point first, before setting up any integral.

6. Reusable rule: to find the volume of (ball) $\cap$ (region bounded by planes or symmetries through the ball's centre), do not integrate. Identify the symmetry group generated by those bounding planes, confirm it fixes the ball's centre, count how many congruent regions it partitions space into near that centre, and divide the ball's volume by that count.

▷ **Cue:** a region cut from a solid by planes that all pass through the centre of an inscribed or circumscribed sphere $\longrightarrow$ **Move:** use the reflection/rotation symmetry fixing that centre to see the sphere's volume as an equal share of the pieces the planes create — skip integration entirely

▷ **Cue:** largest sphere inscribed in a cube of edge a $\longrightarrow$ **Move:** diameter $= a$; the sphere's centre is the cube's centre, automatically the fixed point of the cube's full symmetry group

▷ **Cue:** "closer to point A than to points $B_1, \ldots, B_k$" with the B_i symmetric about A $\longrightarrow$ **Move:** write only the perpendicular-bisector half-spaces for the "nearest neighbours" of A; check whether bisector conditions for farther, sum-type points are implied (often via summing coordinates) before treating them as separate cases

▷ **Cue:** a volume that reduces, after simplification, to (polyhedron) minus (a fixed fraction of a ball) $\longrightarrow$ **Move:** look for a symmetry group shared by the polyhedron's cutting planes and the ball before attempting any coordinate integral

1983 B2

1983 B2. For positive integers n, let $C(n)$ be the number of representations of n as a sum of nonincreasing powers of 2, where no power can be used more than three times. For example, $C(8) = 5$ since the representations for 8 are: 8, $4+4$, $4+2+2$, $4+2+1+1$, and $2+2+2+1+1$. Prove or disprove that there is a polynomial $P(x)$ such that $C(n) = [P(n)]$ for all positive integers n; here $[u]$ denotes the greatest integer less than or equal to u.

Fields: combinatorics (restricted-digit representations, two-term recursion) **Position:** B2 **Answer:** Such a polynomial exists: $P(x) = \frac{x}{2} + 1$, giving $C(n) = \left[\frac{n}{2} + 1\right] = \left\lfloor \frac{n}{2} \right\rfloor + 1$ for every positive integer n

(and for $n = 0$, by the same formula, if one extends the count to the empty sum).

Layer 1 THE WHAT — surface analysis

The natural attempt. You compute the first several values directly by listing representations: $C(1) = 1$ (1), $C(2) = 2$ (2; 1+1), $C(3) = 2$ (2+1; 1+1+1), $C(4) = 3$ (4; 2+2; 2+1+1), $C(5) = 3$ (4+1; 2+2+1; 2+1+1+1), $C(6) = 4$, $C(7) = 4$, and the given $C(8) = 5$. The values $1, 2, 2, 3, 3, 4, 4, 5$ jump out as $\lfloor n/2 \rfloor + 1$, and you check it against $C(8) = 5 = \lfloor 8/2 \rfloor + 1$. It matches every computed value, so you write down $P(x) = x/2 + 1$ and move on.

This is dangerous, not wrong: you have found the right answer by pattern-matching eight data points, but you have proved nothing. A Putnam grader wants to know *why* $C(n)$ must equal $\lfloor n/2 \rfloor + 1$ for every n, not merely that it does for $n \leq 8$; sequences that agree on their first eight terms and then diverge are the entire reason induction exists. The other natural reflex is to reach for the generating function directly: since each power 2^k contributes a factor $1 + x^{2^k} + x^{2 \cdot 2^k} + x^{3 \cdot 2^k}$ to the count, $\sum_n C(n) x^n = \prod_{k \geq 0} \left(1 + x^{2^k} + x^{2 \cdot 2^k} + x^{3 \cdot 2^k}\right)$. This is correct, but manipulating an infinite product of this shape into a closed form for its coefficients is a much larger undertaking than the problem needs, and it is easy to spend the exam period simplifying $\frac{1 - x^{4 \cdot 2^k}}{1 - x^{2^k}}$ factors without ever landing on $\lfloor n/2 \rfloor + 1$.

The model chain. Fix n and ask only about the coefficient of $2^0 = 1$ in a representation of n: call it $e_0 \in \{0, 1, 2, 3\}$. All the other terms in the sum are even, so $n - e_0$ must be even, meaning e_0 has the same parity as n — leaving exactly two allowed values of e_0 (they differ by exactly 2, since the cap is 3). Once e_0 is fixed, the remaining terms form a representation of $(n - e_0)/2$ using powers $1, 2, 4, \ldots$ again (divide every remaining power of 2 by 2), each still capped at three uses: exactly a $C(\cdot)$-count of a smaller number.

If $n = 2m$ is even, $e_0 \in \{0, 2\}$, giving $C(2m) = C(m) + C(m - 1)$. If $n = 2m + 1$ is odd, $e_0 \in \{1, 3\}$, giving $C(2m + 1) = C(m) + C(m - 1) = C(2m)$.

So $C(2m) = C(2m + 1) = C(m) + C(m - 1)$, with base values $C(0) = 1$ (empty sum) and $C(1) = 1$. Now prove $C(n) = \lfloor n/2 \rfloor + 1$ by strong induction. It holds for $n = 0, 1$. For $n = 2m \geq 2$, by hypothesis $C(m) = \lfloor m/2 \rfloor + 1$ and $C(m - 1) = \lfloor (m - 1)/2 \rfloor + 1$; summing these two values, one for m and one for $m - 1$, always gives $m + 1$ regardless of the parity of m (check both cases directly), so $C(2m) = m + 1 = \lfloor 2m/2 \rfloor + 1$. For $n = 2m + 1$, $C(2m + 1) = C(2m) = m + 1 = \lfloor (2m + 1)/2 \rfloor + 1$ since $\lfloor (2m + 1)/2 \rfloor = m$. Both cases close the induction, so $C(n) = \lfloor n/2 \rfloor + 1 = \lfloor n/2 + 1 \rfloor$ for every $n \geq 0$, and $P(x) = x/2 + 1$ answers the problem.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths start by tabulating $C(1), \ldots, C(8)$ and noticing the same numbers. They split at what to do with that pattern. The natural attempt treats the pattern as the destination and stops once it matches the eight given values. The model treats the pattern as a hypothesis and asks: *what forces this recursively?* — the bifurcating line is "how many 1's can I use?", asked about the representation itself rather than about the sequence of totals.

> *Every term in the sum except the count of 1's is even, so the number of 1's used must match n's parity exactly; with a cap of three, that leaves only two legal counts, and they differ by two.*

> *Once the count of 1's is fixed, everything left over is even, so dividing by two turns it into a representation of a strictly smaller number by exactly the same rules — the problem is self-similar under $n \mapsto n/2$, and self-similarity is what a recursion is.*

Layer 3 THE WHY

Why does dividing the leftover by two send a valid representation of $n - e_0$ to a valid representation of $(n - e_0)/2$, rather than to something in a different family? Because the

allowed part-sizes are exactly the powers of 2: $1, 2, 4, 8, \ldots$. Halving every remaining term $2, 4, 8, \ldots$ gives $1, 2, 4, \ldots$ back — the same set of allowed sizes, each still used at most three times, because halving is a bijection on $\{2^k : k \geq 1\}$ landing on $\{2^k : k \geq 0\}$ that preserves multiplicities. If the allowed sizes had been, say, $1, 3, 9, 27, \ldots$ (powers of 3), the same trick would work with $e_0 \in \{0, 1, 2, 3\}$ forced by n mod 3 up to the usage cap, and if the sizes were not a geometric progression at all, no such uniform "peel the smallest, divide the rest" step would exist. The geometric ratio 2 between consecutive allowed sizes is exactly what makes C satisfy a two-term recursion in n rather than needing the full machinery of the generating function.

That also explains why the closed form is linear (up to a floor): the recursion $C(2m) = C(m) + C(m-1)$, together with $C(2m+1) = C(2m)$, forces $C(n) - C(n-1)$ to alternate between 0 and 1 as n increases by one (verify directly from the values above), and a sequence whose successive differences alternate between 0 and 1 is precisely $\lfloor n/2 \rfloor$ plus a constant. The official solution's remark that "an easy induction now shows $C(n) = [1 + n/2]$" is the one place where the archive should supply the missing step: the induction is easy only once you have proved $C(m) + C(m-1) = m + 1$ for both parities of m, which is itself a one-line check once the closed forms for $C(m)$ and $C(m-1)$ are assumed by strong induction — the archive states that check explicitly above rather than waving at "easy".

Finally, why is guessing from eight data points genuinely risky here, even though it happened to work? Because the recursion $C(2m) = C(2m+1)$ means the sequence is constant on consecutive pairs $(2m, 2m+1)$ — a coarser structure than $\lfloor n/2 \rfloor + 1$ alone reveals from raw values, and a problem-setter could easily have built a sequence that agrees with $\lfloor n/2 \rfloor + 1$ up to $n = 8$ (the given example) and departs afterward. The only way to rule that out is the recursion; the pattern-match is evidence, not proof.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "powers of 2, each used at most three times." Write immediately: the count of the smallest part (1) is pinned down by n's residue, up to the usage cap, to exactly two candidate values differing by 2.

2. That cue suggests a two-term recursion in n obtained by peeling off the 1's and halving what remains, not a generating-function expansion.

3. Hesitation occurs at "does dividing by 2 really give back a representation of the same type?" Reinforce it by checking explicitly that halving the multiset $\{2, 4, 8, \ldots\}$ (each used ≤ 3 times) is a bijection onto $\{1, 2, 4, \ldots\}$ (each used ≤ 3 times) — the geometric ratio 2 is what makes this true.

4. The trap is stopping at a pattern that fits every given data point (here, all the way to the example $C(8) = 5$) without deriving the recursion that would make the pattern provably permanent.

5. Whenever a counting problem caps how many times each part of a geometric sequence of sizes (ratio b) may be used, expect the count of the smallest part to be forced by n mod b up to the cap, giving a b-term recursion in n after dividing the remainder by b.

6. Reusable rule: for representations using parts from a geometric progression with a per-part usage cap, do not search for a generating-function closed form first. Peel the smallest part (its multiplicity is forced, up to the cap, by n modulo the ratio), divide the remainder by the ratio, and solve the resulting short recursion by induction.

▷ **Cue:** sum of powers of a fixed base b, each power used at most k times $\longrightarrow$ **Move:** the multiplicity of the smallest part is forced by n mod b up to the cap k; peel it off and recurse on the quotient by b

▷ **Cue:** a closed form guessed from a handful of computed terms in a Putnam counting problem $\longrightarrow$ **Move:** treat the guess only as a hypothesis; do not submit it until an explicit recursion forces it for all n, not just the values checked

▷ **Cue:** counting problem with a per-part usage cap that tempts a generating-function attack
⟶ **Move:** use the generating function only as a sanity check on small cases; solve via the digit-peeling recursion for the actual closed form

1983 B3

1983 B3. Assume that the differential equation $y''' + p(x)y'' + q(x)y' + r(x)y = 0$ has solutions $y_1(x)$, $y_2(x)$, and $y_3(x)$ on the whole real line such that

$$y_1^2(x) + y_2^2(x) + y_3^2(x) = 1$$

for all real x. Let

$$f(x) = y_1'(x)^2 + y_2'(x)^2 + y_3'(x)^2.$$

Find constants A and B such that $f(x)$ is a solution to the differential equation $y' + Ap(x)y = Br(x)$.

Caveat. The scanned source has an OCR gap exactly where $f(x)$ is defined — the line survives only as "Let … Find constants A and B…". The definition given above, $f(x) = y_1'^2 + y_2'^2 + y_3'^2$, is not a guess: it is the unique choice for which the archive's own from-scratch derivation below (done before consulting the official solution) closes into a first-order linear equation for f at all, and it reproduces the official solution's arithmetic ($A = B = 2/3$) term for term, including the vanishing of q. Both facts together make the reconstruction certain, but the statement above is reconstructed, not transcribed, and should be flagged as such.

Fields: ordinary differential equations (differentiating a symmetric constraint), algebraic identity-chasing
Position: B3 **Answer:** $A = B = \frac{2}{3}$, i.e. f satisfies $f' + \frac{2}{3}p(x)f = \frac{2}{3}r(x)$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You see a third-order linear ODE and a quantity f built from first derivatives of its solutions, and your first instinct is to work forward from the ODE itself: multiply the equation for y_i by something and sum over i, hoping to land directly on an equation involving f. The most natural such move is to multiply $y_i''' + py_i'' + qy_i' + ry_i = 0$ by y_i' and sum over $i = 1, 2, 3$:

$$\sum_i y_i'y_i''' + p\sum_i y_i'y_i'' + q\sum_i (y_i')^2 + r\sum_i y_iy_i' = 0.$$

The third term is exactly qf, which looks promising — you seem to be closing in on an equation for f. But look at the other two sums: $\sum_i y_i'y_i'' = \frac{1}{2}f'$ is fine, but $\sum_i y_i'y_i'''$ is a genuinely new quantity, not expressible in terms of f, f' alone without more information, and $\sum_i y_iy_i'$ is likewise unknown to you at this point unless you have separately used the constraint $\sum y_i^2 = 1$. You are stuck with an equation

$$\sum_i y_i'y_i''' + \tfrac{p}{2}f' + qf + r\sum_i y_iy_i' = 0$$

that mixes f, f' with two quantities you cannot eliminate using the ODE alone, because the ODE never relates y_i to y_j for $i \neq j$ or ties $\sum y_iy_i'$ to anything — it is a statement about each y_i in isolation. You can multiply by y_i instead of y_i', or try other combinations, but every version of "manipulate the ODE and sum" runs into the same wall: the ODE alone carries no information that couples the three solutions together, yet f is a coupled, symmetric quantity. Something else in the problem must be supplying that coupling, and it is not the differential equation.

The model chain. The one piece of data that ties y_1, y_2, y_3 together is the constraint $\sum_i y_i^2 = 1$. Differentiate it, not the ODE, and see how far it goes before the ODE is needed at all.

$$\sum_i y_i^2 = 1 \implies \sum_i y_i y_i' = 0. \tag{1}$$

Differentiate (1):

$$\sum_i (y_i')^2 + \sum_i y_i y_i'' = 0 \implies \sum_i y_i y_i'' = -f. \tag{2}$$

Differentiate (2), using $\frac{d}{dx}\sum_i (y_i')^2 = f'$:

$$f' + \sum_i y_i' y_i'' + \sum_i y_i y_i''' = -f' \cdot 0 \text{ (rewrite carefully).}$$

More precisely: $\frac{d}{dx}\sum_i y_i y_i'' = \sum_i y_i' y_i'' + \sum_i y_i y_i'''$, and $\sum_i y_i' y_i'' = \frac{1}{2} f'$, so differentiating (2) gives

$$\tfrac{1}{2} f' + \sum_i y_i y_i''' = -f' \implies \sum_i y_i y_i''' = -\tfrac{3}{2} f'. \tag{3}$$

Only now does the ODE enter, exactly once, at the top order. Multiply $y_i''' + p y_i'' + q y_i' + r y_i = 0$ by y_i and sum:

$$\sum_i y_i y_i''' + p \sum_i y_i y_i'' + q \sum_i y_i y_i' + r \sum_i y_i^2 = 0.$$

Substitute (1), (2), (3): $-\frac{3}{2} f' + p(-f) + q(0) + r(1) = 0$, i.e.

$$-\tfrac{3}{2} f' - pf + r = 0 \implies f' + \tfrac{2}{3} pf = \tfrac{2}{3} r.$$

So $A = B = \frac{2}{3}$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths agree that eventually you must multiply the ODE by y_i (or y_i') and sum, since that is the only way to invoke p, q, r at all. They split at what you differentiate *before* that step. The natural attempt multiplies the ODE straight away and only afterwards discovers it needs $\sum y_i y_i'$ and other cross-terms it has no source for. The model differentiates the constraint $\sum y_i^2 = 1$ three times *first*, manufacturing exactly the cross-terms $\sum y_i y_i' = 0$, $\sum y_i y_i'' = -f$ that the ODE step will need, and only then invokes the ODE once, at the top order.

> *f is built symmetrically from all three solutions, but the ODE says nothing about any y_i in relation to y_j — it is three copies of the same uncoupled statement. Any equation that couples them has to come from the one place they actually are coupled: the identity $\sum y_i^2 = 1$. So I should be differentiating that, not the ODE, until I have everything I need.*

> *f is a sum of squares of first derivatives, so it is an "order-2" object built from "order-1" data; the ODE is order 3. I should differentiate the constraint exactly enough times to reach an order-3 cross-term $\sum y_i y_i'''$, because that is the only place the ODE can be substituted in.*

Layer 3 THE WHY

Why must the coupling come from the constraint and not the ODE? Because y_1, y_2, y_3 each individually satisfy the identical linear ODE $y''' + py'' + qy' + ry = 0$; nothing in that equation refers to any function other than y_i itself and the fixed coefficients p, q, r. Three solutions of the same uncoupled equation carry no information relating one to another beyond what any single

solution already has. The quantity $f = \sum (y_i')^2$, by contrast, cannot even be evaluated without knowing all three functions simultaneously — it is intrinsically a joint quantity. A joint quantity can only obey a nontrivial law if it is fed by a joint hypothesis, and the only joint hypothesis in the problem is $\sum y_i^2 = 1$. Once this is seen, the order of operations is forced before any algebra: differentiate the constraint until it is useful, invoke the ODE once.

Why exactly three times, and why does the ODE enter only at the top? Track the "derivative order" carried by each object: $\sum y_i^2$ has order $0+0$, $\sum y_i y_i'$ has order $0+1$, $\sum y_i y_i''$ has order $0+2$, $\sum y_i y_i'''$ has order $0+3$. Differentiating the constraint once raises this total by exactly one each time (product rule: $\frac{d}{dx}(y_i y_i^{(k)}) = y_i' y_i^{(k)} + y_i y_i^{(k+1)}$, and the first term is itself expressible via f or its derivative once $k \geq 1$). The ODE lets you replace y_i''' by lower-order terms only when it is multiplied against something and summed, and the only combination that produces a quantity you can control is $\sum y_i y_i'''$ — so you need to reach order 3 in the constraint chain, no more and no less, before substituting.

Why does q vanish from the final relation, leaving $A = B$ rather than three independent constants? Because q is multiplied, in the summed ODE, by $\sum_i y_i y_i'$ — and that quantity is identically 0 by (1), the very first differentiation of the constraint. This is not a coincidence to double-check numerically; it is forced the instant you notice that q's coefficient in the coupled sum is exactly the derivative of the constraint itself, which must vanish because the constraint is a constant. The archive's point here is that the official solution's arithmetic "$q \cdot 0$" is not a lucky cancellation but a structural fact about which term of the ODE lines up with which derivative of the normalisation identity: r pairs with $\sum y_i^2 = 1$ itself (order 0), q pairs with its first derivative (order 1, always 0), p pairs with its second derivative (order 2, equal to $-f$), and the leading term pairs with the third derivative (order 3, equal to $-\frac{3}{2}f'$). Reading off which coefficient multiplies which known quantity is what produces $A = B = \frac{2}{3}$ without needing to solve anything further.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is three functions individually solving an *identical, uncoupled* ODE, tied together only by a symmetric algebraic identity ($\sum y_i^2 = 1$). Write immediately: any equation coupling them must come from differentiating that identity, not from the ODE.

2. That cue suggests the method: differentiate the constraint repeatedly, tracking $\sum y_i y_i^{(k)}$ at each order, and invoke the ODE exactly once, at the highest order reached.

3. Hesitation occurs at deciding how many times to differentiate before substituting the ODE. Reinforce it by counting derivative orders: f itself sits at order 2 (sum of squares of first derivatives), the ODE is order 3, so the constraint chain must be pushed to its order-3 term $\sum y_i y_i'''$ before the ODE has anything to substitute into.

4. The trap is substituting the ODE too early — multiplying it by y_i or y_i' before the constraint's derivatives have supplied values for $\sum y_i y_i'$, $\sum y_i y_i''$ — which leaves an equation mixing individual y_i, y_i', y_i'' that never resolves back into f and f' alone.

5. Whenever several objects individually solve identical decoupled equations but are tied by one symmetric algebraic constraint, expect that constraint (differentiated enough times to match the equation's order) to be the sole source of any cross-object identity.

6. Reusable rule: to find an ODE satisfied by a symmetric combination S of derivatives of solutions coupled only by an algebraic identity, differentiate the identity until its derivative order matches the order needed to invoke the original ODE, substitute the ODE exactly once at that top order, and read the coefficients off by matching which term of the ODE lines up with which derivative of the identity.

▷ **Cue:** several solutions of an otherwise-uncoupled ODE tied together by a symmetric algebraic identity (e.g., $\sum y_i^2 = 1$) $\longrightarrow$ **Move:** differentiate the identity repeatedly, not the ODE, to generate the cross-terms; substitute the ODE only once, at the highest derivative order reached

▷ **Cue:** a term like $\sum y_i y_i'$ appearing after one differentiation of $\sum y_i^2 = \text{const}$ $\longrightarrow$ **Move:** recognise it as the derivative of a constant, hence identically zero, and track which coefficient of the original ODE it will silently kill

▷ **Cue:** need an ODE for $f = \sum(\text{derivative of } y_i)^2$ built from several solutions of an nth-order linear ODE under a symmetric constraint $\longrightarrow$ **Move:** differentiate the constraint exactly (order of the original ODE) times before substituting; match coefficients by derivative order

1983 B4

1983 B4. Let $f(n) = n + [\sqrt{n}\,]$ where $[x]$ is the largest integer less than or equal to x. Prove that, for every positive integer m, the sequence

$$m,\ f(m),\ f(f(m)),\ f(f(f(m))),\ \ldots$$

contains at least one square of an integer.

Fields: number theory; floor functions; iteration and descent **Position:** B4 — middle of the paper; a statement about an infinite orbit that becomes a finite descent once each integer is written relative to the square below it **Answer:** Always true. Write $m = k^2 + j$ with $0 \le j \le 2k$. If $k < j \le 2k$, one step of f gives $(k+1)^2 + (j-k-1)$ with $0 \le j-k-1 \le k-1$. If $1 \le j \le k$, two steps give $(k+1)^2 + (j-1)$. So the excess over the nearest square below falls by 1 every two steps and reaches 0 within $2j$ steps.

Layer 1 THE WHAT — surface analysis

The natural attempt. The strong student computes, because a statement about "the sequence contains a square" invites examples.

$$2 \to 3 \to 4, \qquad 5 \to 7 \to 9, \qquad 6 \to 8 \to 10 \to 13 \to 16,$$
$$10 \to 13 \to 16, \qquad 11 \to 14 \to 17 \to 21 \to 25, \qquad 12 \to 15 \to 18 \to 22 \to 26 \to 31 \to 36.$$

Every orbit lands on a square, and the number of steps varies irregularly: 2, 2, 4, 2, 4, 6. The student now looks for a global handle. Three reasonable tries, each taken far enough to see the wall.

A closed form. f adds $[\sqrt{n}\,]$, which changes as n grows, so $f^{(r)}(m)$ has no formula in r one can write down; the increments $3, 3, 4, 4, 5, 5, \ldots$ in the 12-orbit look regular, but the student cannot yet say why they come in pairs.

Contradiction by growth. Suppose the orbit misses every square. It is strictly increasing and unbounded, so it passes every square s^2; it must jump over each one, i.e. some term $n < s^2$ has $f(n) > s^2$. That is certainly possible for a single n ($f(7) = 9$ lands, but $f(8) = 10$ jumps 9), so growth alone gives no contradiction. The student needs to know which n jump and which land — which is the whole problem.

Residues. Squares occupy only the residues $0, 1, 4 \pmod 8$; perhaps the orbit is forced through one of them. But the increment $[\sqrt{n}\,]$ is not periodic in n, so no modulus is preserved. Dead end.

The honest diagnosis: all three tries treat n as a single number, while f does not. f behaves one way on the block $[k^2, (k+1)^2)$ and another on the next block.

The model chain. Coordinates. Every positive integer n lies in exactly one block $k^2 \le n < (k+1)^2 = k^2 + 2k + 1$ with $k \ge 1$, so it has a unique representation

$$n = k^2 + j, \qquad 0 \le j \le 2k, \qquad k = [\sqrt{n}\,].$$

Call j the *excess* of n. Then n is a square iff its excess is 0, and $f(n) = k^2 + j + k$.

Upper half of a block: $k < j \le 2k$. Then

$$f(n) = k^2 + k + j = (k+1)^2 + (j-k-1), \qquad 0 \le j-k-1 \le k-1 < 2(k+1),$$

so $f(n)$ lies in block $k+1$ with excess $j-k-1 \le k-1$, i.e. in the lower half of that block (excess at most $k+1$). If $j = k+1$ it is the square $(k+1)^2$.

Lower half of a block: $1 \le j \le k$. Then $f(n) = k^2 + k + j \le k^2 + 2k < (k+1)^2$, so $f(n)$ is still in block k and $[\sqrt{f(n)}\,] = k$. A second step gives

$$f(f(n)) = k^2 + 2k + j = (k+1)^2 + (j-1), \qquad 0 \le j-1 \le k-1.$$

So $f^{(2)}(n)$ is either the square $(k+1)^2$ (when $j = 1$) or again in the lower half of its block (excess $j-1 \le (k+1)$), with excess exactly one smaller.

Descent. Start from $m = k^2 + j$. If $j = 0$ we are done. If $j > k$, one step moves to a lower-half number (or a square). From a lower-half number of excess $e \ge 1$, two steps produce a square or a lower-half number of excess $e-1$. The excess is a non-negative integer that strictly decreases every two steps, so it reaches 0: some iterate $f^{(r)}(m)$, with $r \le 2j$ (one step from the upper half, then $2(j-k-1)$ more; or $2j$ steps from the lower half), is a perfect square. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains compute the same examples. They split at the sentence "$f(n) = n + [\sqrt{n}\,]$ and $[\sqrt{n}\,]$ changes as n grows." The natural attempt treats the floor as an obstacle to a formula and looks for a global handle on the whole orbit. The model chain treats the floor as a *coordinate*: it names the block $k = [\sqrt{n}\,]$ and the position j inside it, and on each block f is a translation by the constant k.

> *The floor $[\sqrt{n}\,]$ is constant between consecutive squares. So I will stop writing n and start writing $n = k^2 + j$; inside one block f is nothing but "add k", and the only question is whether adding k leaves the block.*

> *The block has length $2k+1$ and each step moves k. So from the bottom half I stay in the block for one step, and from the top half I cross into the next block. Two steps cost $2k$, and the next square is $2k+1$ away — I lose exactly one unit of excess per double step. That is a descent.*

- ▷ **Cue:** a map built from a floor $[\sqrt{n}\,]$, $[n/k]$, $[\log_2 n]$, … ⟶ **Move:** write n as (block index, offset in block); on each block the map is affine, so track the offset
- ▷ **Cue:** "the orbit contains at least one X" for an increasing orbit ⟶ **Move:** find an integer quantity measured relative to the X's that strictly decreases and is 0 exactly at an X
- ▷ **Cue:** step size about k inside gaps of size $2k+1$ ⟶ **Move:** two steps fall exactly one short of the gap: look at what happens every *two* steps, not every one

Layer 3 THE WHY

Why is "write $n = k^2 + j$" not a trick? Go to the definition. $[\sqrt{n}\,] = k$ means precisely $k^2 \le n < (k+1)^2$. So the floor function partitions the positive integers into blocks between consecutive squares, and the value of the floor is the name of the block. The instant a floor appears in a map, the map is piecewise and the pieces are those blocks; recording which piece you are on is not an extra idea, it is reading the definition. The remaining coordinate $j = n - k^2$ is forced too: the problem asks about squares, and a square is exactly an integer at offset 0 in its block. The problem's target has an address in these coordinates (excess 0), and nowhere else does it have one so simply.

Why look at pairs of steps? Count. The gap from k^2 to $(k+1)^2$ is $2k+1$, and f moves by k while inside it. The number $2k+1$ is $k+k+1$: two steps of size k fall one short of a full

block. So any argument must be about two steps at a time, and "one short" is exactly the unit of descent. This is also why the increments in the natural attempt's 12-orbit came in pairs, $3, 3, 4, 4, 5, 5$: each block is crossed in two strides. The pattern the student saw and could not explain is the proof.

Why does the lower half matter? Because the second step must still use the same stride k. That needs $f(n) < (k+1)^2$, i.e. $j \le k$. The upper half fails this, and that is why it needs the separate one-step reduction: it spills into the next block, and a quick computation shows it lands in the lower half there. So the case split is not a nuisance; it is the statement "the stride stays constant for two steps only if you start low enough".

The official solution sets up exactly these classes (A for $0 \le j \le k$, B for $k < j \le 2k$) without saying why. Here is where they come from: A is the set of n from which two steps stay on one stride, and B is everything else, which one step sends into A.

Layer $+\alpha$ BACKTESTING

1. The first cue is the floor of $\sqrt{n}$. On seeing it, write $n = k^2 + j$, $0 \le j \le 2k$, and record $f(n) = k^2 + j + k$ and "square $\iff$ $j = 0$".
2. Piecewise analysis on the blocks of the floor, followed by a descent (strictly decreasing non-negative integer) on the offset.
3. Hesitation comes after the first step, when $f(n)$ may or may not have left the block and the student cannot see a monotone quantity. It is reinforced away by computing two steps from the lower half and noticing the -1 in $(k+1)^2 + (j-1)$.
4. The trap is trying to prove it by growth or by contradiction ("the orbit jumps every square"), which is consistent with everything known until you understand which integers jump. A secondary trap is forgetting to check that the upper half lands in the *lower* half of the next block; without that the descent has a gap.
5. Whenever a map is defined through a floor, rewrite the input in (piece, offset) coordinates before doing anything else; and when a stride is roughly half the length of a piece, look at double steps.
6. Reusable decision rule: *to show an orbit hits a target set, measure the distance to the target in the coordinates the map is piecewise-affine in, and find the number of steps after which that distance drops by a fixed integer.*

1983 B5

1983 B5. Let $\|u\|$ denote the distance from the real number u to the nearest integer. (For example, $\|2.8\| = .2 = \|3.2\|$.) For positive integers n, let

$$a_n = \frac{1}{n} \int_1^n \left\| \frac{n}{x} \right\| dx.$$

Determine $\lim_{n\to\infty} a_n$. You may assume the identity

$$\frac{2}{1} \cdot \frac{2}{3} \cdot \frac{4}{3} \cdot \frac{4}{5} \cdot \frac{6}{5} \cdot \frac{6}{7} \cdot \frac{8}{7} \cdot \frac{8}{9} \cdots = \frac{\pi}{2}.$$

Fields: real analysis; integration of piecewise-linear functions; infinite products **Position:** B5 — late in the paper; a computation whose difficulty is entirely in choosing the variable, after which it is a sum of logarithms telescoping into Wallis's product **Answer:** $\lim_{n\to\infty} a_n = \ln \frac{4}{\pi} \approx 0.2416$. Substituting $u = n/x$ gives $a_n = \int_1^n \|u\|\, u^{-2}\, du$; the integral over $[k, k+1]$ is $\ln \frac{(2k+1)^2}{2k(2k+2)}$, and the product of these ratios is 2 divided by a partial Wallis product.

Layer 1 THE WHAT — surface analysis

The natural attempt. The student sees an integral of a distance-to-nearest-integer and does what one does with $\|\cdot\|$: find where its argument is an integer or a half-integer and integrate piece by piece, in the variable x given.

n/x is the integer k at $x = n/k$ and the half-integer $k+\frac12$ at $x = \frac{2n}{2k+1}$. For $x \in [1,n]$, n/x runs over $[1,n]$, so the break points are

$$1 = \frac{n}{n} < \cdots < \frac{n}{k+1} < \frac{2n}{2k+1} < \frac{n}{k} < \cdots < \frac{n}{1} = n.$$

On $\left[\frac{n}{k+1}, \frac{2n}{2k+1}\right]$, $n/x \in [k+\frac12, k+1]$ and $\|n/x\| = k+1-\frac{n}{x}$; on $\left[\frac{2n}{2k+1}, \frac{n}{k}\right]$, $\|n/x\| = \frac{n}{x} - k$. Each piece integrates to a constant times n plus n times a logarithm, e.g.

$$\int_{2n/(2k+1)}^{n/k} \Big(\frac{n}{x} - k\Big)\,dx = n \ln\frac{2k+1}{2k} - k\Big(\frac{n}{k} - \frac{2n}{2k+1}\Big) = n\Big(\ln\frac{2k+1}{2k} - \frac{1}{2k+1}\Big).$$

Dividing by n and doing the same for the other half-piece gives the terms of a sum over k. This is exactly the official solution's route, and it works; but every line carries n in two places, the division by n has to be remembered at the end, and the student has to spot that the n's cancel term by term.

There is a second natural attempt, more dangerous because it is fast: "$\|\cdot\|$ averages $\frac14$ over a period, and n/x oscillates rapidly, so $a_n \approx \frac1n \int_1^n \frac14\,dx \to \frac14$." It is wrong. For $x \in [n/2, n]$ — half of the range — n/x only runs through $[1,2]$, one period, not many; there is no averaging there. The true answer, $\ln(4/\pi) \approx 0.2416$, is close to 0.25, which makes the error hard to catch by estimation.

The model chain. *Change variable.* The integrand depends on x only through n/x, so put $u = n/x$, $x = n/u$, $dx = -n\,u^{-2}\,du$; $x = 1 \mapsto u = n$, $x = n \mapsto u = 1$. Then

$$a_n = \frac1n \int_n^1 \|u\| \Big(-\frac{n}{u^2}\Big)\,du = \int_1^n \frac{\|u\|}{u^2}\,du.$$

The parameter n has left the integrand and survives only as the upper limit.

One period. For an integer $k \ge 1$, $\|u\| = u - k$ on $[k, k+\frac12]$ and $\|u\| = k+1-u$ on $[k+\frac12, k+1]$. So

$$\int_k^{k+1/2} \Big(\frac1u - \frac{k}{u^2}\Big)\,du = \ln\frac{2k+1}{2k} + k\Big(\frac{2}{2k+1} - \frac1k\Big) = \ln\frac{2k+1}{2k} - \frac{1}{2k+1},$$

$$\int_{k+1/2}^{k+1} \Big(\frac{k+1}{u^2} - \frac1u\Big)\,du = (k+1)\Big(\frac{2}{2k+1} - \frac{1}{k+1}\Big) - \ln\frac{2k+2}{2k+1} = \frac{1}{2k+1} - \ln\frac{2k+2}{2k+1}.$$

The rational parts cancel and

$$I_k := \int_k^{k+1} \frac{\|u\|}{u^2}\,du = \ln\frac{(2k+1)^2}{(2k)(2k+2)}.$$

Sum. Since n is an integer, $[1,n]$ is the union of $[k,k+1]$, $1 \le k \le n-1$, and

$$a_n = \sum_{k=1}^{n-1} I_k = \ln P_{n-1}, \qquad P_N = \prod_{k=1}^{N} \frac{2k+1}{2k}\cdot\frac{2k+1}{2k+2} = \frac32\cdot\frac34\cdot\frac54\cdot\frac56\cdots\frac{2N+1}{2N}\cdot\frac{2N+1}{2N+2}.$$

Wallis. Let W_M be the product of the first M factors of $\frac21\cdot\frac23\cdot\frac43\cdot\frac45\cdots$; by hypothesis $W_M \to \pi/2$. The factors of P_N are exactly the reciprocals of the 2nd through $(2N+1)$th Wallis factors, and the 1st is $\frac21$, so

$$P_N = \frac{2}{W_{2N+1}} \longrightarrow \frac{2}{\pi/2} = \frac{4}{\pi}.$$

Since ln is continuous on $(0,\infty)$,

$$\lim_{n\to\infty} a_n = \lim_{N\to\infty} \ln P_N = \ln \frac{4}{\pi}. \qquad \square$$

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains begin by locating where $\|\cdot\|$ changes formula. They split on the very first line: the natural attempt locates those points in the x-variable ($x = n/k$, $x = 2n/(2k+1)$) and keeps n inside every piece; the model first writes $u = n/x$, and the break points become the fixed integers and half-integers k, $k+\frac{1}{2}$, independent of n.

> *The integrand only sees n/x. So n/x is the real variable, and I should integrate in it. If I substitute $u = n/x$ the $\frac{1}{n}$ in front cancels the n in dx — which is a strong hint that the problem was built this way — and the limit $n \to \infty$ becomes nothing more than the upper limit of an integral going to infinity.*

> *They give me Wallis's product, so the answer is a logarithm of a product of ratios like $\frac{2k}{2k\pm1}$. An integral of $\|u\|/u^2$ over a period gives* ln *of ratios of the break points k, $k+\frac{1}{2}$, $k+1$ — that is $2k : 2k+1 : 2k+2$. The rational parts must cancel.*

- ▷ **Cue:** an integrand depending on x only through n/x (or x/n, nx) $\longrightarrow$ **Move:** substitute $u = n/x$; the scaling parameter moves from the integrand to a limit of integration
- ▷ **Cue:** "determine $\lim a_n$" where a_n is an integral whose range grows with n $\longrightarrow$ **Move:** rewrite as $\int_1^n g(u)\,du$ with g independent of n; the limit is the improper integral $\int_1^\infty g$
- ▷ **Cue:** a problem hands you an infinite product such as Wallis's $\longrightarrow$ **Move:** the answer is its logarithm or its value at a partial product; aim your computation at producing a telescoping product of the same ratios
- ▷ **Cue:** an integral of $\|u\|$ or $\{u\}$ against a weight $\longrightarrow$ **Move:** split at integers and half-integers and compute one period in closed form

Layer 3 THE WHY

Why substitute $u = n/x$? Because of what the integrand *is*. $\|n/x\|$ is a function of one quantity, n/x; the variable x is only a way of parametrising it. When a function depends on the variable of integration through a single expression, that expression is the natural variable — this is the definition of what substitution is for. The problem then confirms the choice in two ways. First, the Jacobian: $dx = -n\,du/u^2$, and the n exactly cancels the $\frac{1}{n}$ in front. A factor that cancels on substitution is a design signature: the setter wrote $\frac{1}{n}\int dx$ because $\int du/u^2$ was the intended object. Second, the limits: $x \in [1,n]$ becomes $u \in [1,n]$, so the only place n remains is the range. A sequence of integrals over growing ranges of a fixed integrand converges to the improper integral when the integrand is non-negative and integrable, and here $0 \le \|u\|/u^2 \le \frac{1}{2u^2}$.

Why does the averaging heuristic fail, precisely? In the u-variable the answer is $\int_1^\infty \|u\|u^{-2}\,du$, and the weight u^{-2} is heavy near $u = 1$: the interval $[1,2]$ carries half the total weight $\int_1^\infty u^{-2}du = 1$. On $[1,2]$ the function $\|u\|$ is a single tent, not a rapidly oscillating function, so replacing it by its average $\frac{1}{4}$ is not justified. Replacing it anyway gives $\frac{1}{4}$; the correct value is smaller because u^{-2} is convex: on each period it puts more weight near the two endpoints, where $\|u\| = 0$, than near the midpoint, where $\|u\| = \frac{1}{2}$. (Pair u with $2k+1-u$: the paired weight grows with distance from $k+\frac{1}{2}$ while $\|u\| - \frac{1}{4}$ falls, so $\int_k^{k+1}(\|u\| - \frac{1}{4})u^{-2}\,du < 0$.)

Why must the answer be a logarithm of Wallis's product? Integrating a piecewise-linear function against u^{-2} produces $\int u^{-1}$ (a logarithm) and $\int u^{-2}$ (a rational number). The break points are $k, k+\frac{1}{2}, k+1$, i.e. in the ratio $2k : 2k+1 : 2k+2$, so the logarithms are of $\frac{2k+1}{2k}$ and $\frac{2k+2}{2k+1}$

— exactly the factors of Wallis's product. The rational parts $\mp\frac{1}{2k+1}$ must cancel, and integration by parts says why: $\int_k^{k+1} \|u\|u^{-2}\,du = \left[-\|u\|/u\right]_k^{k+1} + \int_k^{k+1} \|u\|' u^{-1}\,du$, the boundary term vanishes because $\|k\| = \|k+1\| = 0$, and $\|u\|' = \pm 1$ leaves only $\int \pm u^{-1}\,du$ — pure logarithms. (This one line is a cheaper derivation of I_k than the two separate integrals in the model chain.) So once the substitution is made, the hint is not a surprise; it is the only product the computation could produce.

The official solution integrates in x and writes the sum without saying why the n's vanish. Here is why: the integral is scale-invariant in disguise, and the substitution $u = n/x$ removes the disguise.

Layer $+\alpha$ BACKTESTING

1. The first cue is $\|n/x\|$ under $\frac{1}{n}\int \cdots dx$. On seeing it write $u = n/x$ and the transformed integral $a_n = \int_1^n \|u\|u^{-2}\,du$.
2. Substitution to strip the parameter from the integrand, then period-by-period integration of a piecewise-linear function, then a telescoping product matched to the supplied Wallis identity.
3. Hesitation comes at the product $\prod \frac{(2k+1)^2}{2k(2k+2)}$, which does not look like the given identity. It is reinforced away by writing both products out factor by factor, $\frac{3}{2}\cdot\frac{3}{4}\cdot\frac{5}{4}\cdots$ against $\frac{2}{1}\cdot\frac{2}{3}\cdot\frac{4}{3}\cdots$, and seeing reciprocals shifted by one.
4. Two traps. The averaging heuristic gives $\frac{1}{4}$, wrong but numerically close. And in matching Wallis the stray first factor $\frac{2}{1}$ is easy to drop, which would give $\ln(2/\pi) < 0$ — a negative limit of non-negative a_n, which should be caught immediately.
5. Whenever an integrand is a function of a ratio of the variable and a parameter, change to the ratio first; whenever a problem supplies a named constant as an infinite product, plan to arrive at a product of the same factors.
6. Reusable decision rule: *move every parameter out of the integrand before taking a limit; if it can be pushed entirely into the limits of integration, the limit of the sequence is an improper integral of a fixed function.*

1983 B6

1983 B6. Let k be a positive integer, let $m = 2^k + 1$, and let $r \neq 1$ be a complex root of $z^m - 1 = 0$. Prove that there exist polynomials $P(z)$ and $Q(z)$ with integer coefficients such that

$$(P(r))^2 + (Q(r))^2 = -1.$$

Fields: algebra; roots of unity; sums of two squares **Position:** B6 — last problem of the paper; hard to start, short once the multiplicative structure of sums of two squares is brought in **Answer:** Always possible, with P, Q independent of which r is chosen. From $1 + r + \cdots + r^{m-1} = 0$ and $m - 1 = 2^k$, $-1 = (r + r^2)\prod_{i=1}^{k-1}(1 + r^{2^i})$. Here $r + r^2 = \left(r^{2^{k-1}+1}\right)^2 + r^2$ (as $r^{m+1} = r$) and $1 + r^{2^i} = 1^2 + \left(r^{2^{i-1}}\right)^2$; multiply using $(a^2 + b^2)(c^2 + d^2) = (ac - bd)^2 + (ad + bc)^2$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The student starts with the smallest case, $k = 1$, $m = 3$. Then r is a primitive cube root of unity, so $r^2 + r + 1 = 0$ and every polynomial in r reduces to $a + br$. Try $P = a + br$, $Q = c + dr$:

$$P^2 + Q^2 = (a^2 + c^2) + 2(ab + cd)r + (b^2 + d^2)r^2 = (a^2 + c^2 - b^2 - d^2) + (2ab + 2cd - b^2 - d^2)\,r.$$

Since $1, r$ are linearly independent over $\mathbb{Q}$, the target -1 requires

$$a^2 + c^2 - b^2 - d^2 = -1, \qquad 2ab + 2cd = b^2 + d^2.$$

A short search finds $(a,b,c,d) = (0,1,1,1)$: $P = r$, $Q = 1+r$, and indeed $r^2 + (1+r)^2 = 2(r^2+r+1) - 1 = -1$. Case $k = 1$ done.

Now $k = 2$, $m = 5$. If r is a primitive fifth root, $1, r, r^2, r^3$ are independent and P, Q have four unknown coefficients each: eight unknowns, four simultaneous quadratic Diophantine equations. The search that worked for $m = 3$ is now a wall, and $k = 3$ ($m = 9$) is worse — and there r need not even be primitive (it may be a cube root of unity), so the reduction rule changes with r.

A second attempt: "$-1 = i^2$, so find i as a polynomial in r and take $Q = 0$." For $m = 3$ this is impossible: $\mathbb{Q}(r) = \mathbb{Q}(\sqrt{-3})$ does not contain i. So a genuinely two-square representation is needed.

The honest diagnosis: case-by-case coefficient matching scales with m, and m grows exponentially in k. The proof must use structure that is uniform in k.

The model chain. Throughout, "x is a sum of two squares" means $x = P(r)^2 + Q(r)^2$ with $P, Q \in \mathbb{Z}[z]$. Let S be the set of such x.

Fact 1: S is closed under multiplication. For any complex numbers,

$$(a^2+b^2)(c^2+d^2) = (ac-bd)^2 + (ad+bc)^2,$$

and if a, b, c, d are integer polynomials in r, so are $ac - bd$ and $ad + bc$.

Fact 2: an expression for -1. Since $r^m = 1$ and $r \neq 1$,

$$0 = \frac{r^m - 1}{r-1} = 1 + r + r^2 + \cdots + r^{m-1}, \qquad \text{so} \qquad -1 = r\big(1 + r + \cdots + r^{m-2}\big).$$

Fact 3: factor by doubling. As $m - 2 = 2^k - 1$,

$$1 + r + \cdots + r^{2^k-1} = (1+r)(1+r^2)(1+r^4)\cdots\big(1+r^{2^{k-1}}\big),$$

because expanding the right side gives every r^j, $0 \le j \le 2^k - 1$, exactly once — one term for each binary expansion of j. Hence

$$-1 = (r+r^2)\prod_{i=1}^{k-1}\big(1+r^{2^i}\big) \qquad \text{(empty product} = 1 \text{ when } k = 1\text{)}.$$

Fact 4: each factor lies in S. For $1 \le i \le k-1$, $1 + r^{2^i} = 1^2 + \big(r^{2^{i-1}}\big)^2$. For the first factor, m is odd, $m + 1 = 2(2^{k-1}+1)$, and $r^{m+1} = r$, so

$$r + r^2 = \big(r^{2^{k-1}+1}\big)^2 + (r)^2.$$

Conclusion. By Facts 1 and 4, applied $k-1$ times, the product in Fact 3 is $P(r)^2 + Q(r)^2$ with $P, Q \in \mathbb{Z}[z]$, and it equals -1. The polynomials are built from the exponents alone, so the same P, Q work for every root $r \neq 1$. □

Check. $k = 1$: $P = z^2$, $Q = z$, and $r^4 + r^2 = r + r^2 = -1$. $k = 2$: $(r^3)^2 + r^2$ times $1^2 + r^2$ gives $P = r^3 - r^2$, $Q = r^4 + r$, and $P^2 + Q^2 = r^6 + r^4 + r^8 + r^2 = r + r^2 + r^3 + r^4 = -1$ using $r^5 = 1$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains write down the goal $P^2 + Q^2 = -1$. The natural attempt treats it as an equation to be solved for the coefficients of P and Q. The model treats "is a sum of two squares" as a *property* closed under multiplication, and asks instead for a factorisation of -1 into pieces that visibly have the property. The split is at the first line after the goal: "let $P = a + br$" versus "sums of two squares multiply".

> *I am not going to find eight coefficients. Sums of two squares are closed under multiplication, $(a^2+b^2)(c^2+d^2) = (ac-bd)^2+(ad+bc)^2$. So I only need -1 as a product of things that are each visibly a sum of two squares.*

> *What do I know that equals -1? Only $1+r+\cdots+r^{m-1}=0$. And $m-1=2^k$ is a power of two — that is the doubling factorisation $1+x+\cdots+x^{2^k-1}=\prod(1+x^{2^i})$, and $1+x^{2j}$ is $1^2+(x^j)^2$.*

> *The one factor left over is $r(1+r) = r+r^2$. r^2 is a square; is r? m is odd, so $r = r^{m+1} = (r^{(m+1)/2})^2$. Every root of unity of odd order is a square.*

- ▷ **Cue:** "prove x is a sum of two squares" in any commutative ring $\longrightarrow$ **Move:** sums of two squares are closed under multiplication; factor x into pieces of the form s^2+t^2, $1+s^2$, or a square
- ▷ **Cue:** $1+x+x^2+\cdots+x^{2^k-1}$, or an exponent 2^k in the hypothesis $\longrightarrow$ **Move:** factor as $(1+x)(1+x^2)(1+x^4)\cdots(1+x^{2^{k-1}})$ — binary expansion
- ▷ **Cue:** r a root of unity of odd order m $\longrightarrow$ **Move:** r is a square: $r=(r^{(m+1)/2})^2$
- ▷ **Cue:** $r^m=1$, $r\neq 1$, and a need for an identity in r $\longrightarrow$ **Move:** divide out $r-1$: $1+r+\cdots+r^{m-1}=0$ is the only relation guaranteed for every such r

Layer 3 THE WHY

Why is multiplicativity the natural first thought rather than a trick? Look at what the goal says. P^2+Q^2 is the *norm* form: $P^2+Q^2 = (P+iQ)(P-iQ)$, and a product of two norms is the norm of the product, $(a+ib)(c+id) = (ac-bd)+i(ad+bc)$, whose norm is $(ac-bd)^2+(ad+bc)^2$. That is the whole content of the two-square identity; it is the statement $|zw|=|z||w|$ squared, with i adjoined formally so that it works for complex a,b,c,d too. So "sum of two squares" is not a single equation to solve — it is membership in a set closed under multiplication, containing all squares and everything of the form $1+s^2$. Any membership question for such a set is answered by factoring. The natural attempt could not see this because it expanded P^2+Q^2 into coefficients immediately, destroying the product structure.

Why is the power of two not an arbitrary feature? It is the exact hypothesis for the doubling factorisation. $\prod_{i=0}^{k-1}(1+x^{2^i}) = \sum_{j=0}^{2^k-1} x^j$ is the statement that every integer $0\le j<2^k$ has exactly one binary expansion, and $1+x^{2^i}$ for $i\ge 1$ is $1+(\text{square})$, which is a sum of two squares with nothing to do. The $+1$ in $m=2^k+1$ is exactly what makes the sum of the nonzero powers, $r+\cdots+r^{m-1}$, equal to r times a sum of length 2^k. So both parts of $m=2^k+1$ are used: 2^k for the factorisation, and oddness of m for the leftover factor.

Why must the last factor be $r+r^2$ and not r alone? Because the doubling product starts with $1+r$, which is $1+(\text{non-square})$, and cannot be handled like the others. Absorbing the stray r into it produces $r+r^2$, and now both terms are squares because r itself is a square. That r is a square is again forced by the hypothesis: in the cyclic group of m-th roots of unity with m odd, squaring is a bijection (its inverse is raising to the power $(m+1)/2$), so everything is a square.

The official solution writes $-1=(r+r^2)(1+r^2)(1+r^4)\cdots$ in one line. Here is where it comes from: the goal is a norm, norms multiply, and the only available relation for -1 is the geometric sum, whose length is a power of two.

Layer $+\alpha$ BACKTESTING

1. The first cue is the target P^2+Q^2, a sum of two squares. On seeing it write the two-square identity $(a^2+b^2)(c^2+d^2) = (ac-bd)^2+(ad+bc)^2$ in the margin. The second cue is $m=2^k+1$; write $0=1+r+\cdots+r^{2^k}$.

2. Multiplicative closure: factor -1 into factors each of which is a sum of two squares, then multiply.

3. Hesitation comes at the leftover factor $1 + r$ (or r), which is not of the form $1 + (\text{square})$. It is reinforced away by asking "is r a square?" and remembering that in a group of odd order every element is a square.

4. The trap is coefficient-matching: it succeeds for $m = 3$ and so feels like the method, then explodes at $m = 5$. A second trap is seeking $\sqrt{-1}$ in $\mathbb{Z}[r]$; it does not exist in general. A third is assuming r primitive; the proof above uses only $r^m = 1$, $r \neq 1$.

5. Whenever a problem asks for a representation by a form that is multiplicative (sums of two squares, sums of four squares, norms $a^2 - db^2$), stop solving for coefficients and start factoring the target.

6. Reusable decision rule: *if the set of allowed answers is closed under an operation, decompose the target with respect to that operation into pieces that are visibly allowed; exponents of the form* 2^k *point to the doubling factorisation.*