DAWN PUTNAM SESSION

Putnam 2019 — Papers A and B

Putnam 120 Archive

Contents

Putnam 2019 — Paper A **5**
2019 A1 5
2019 A2 7
2019 A3 10
2019 A4 12
2019 A5 15
2019 A6 18

Putnam 2019 — Paper B **23**
2019 B1 23
2019 B2 25
2019 B3 28
2019 B4 30
2019 B5 33
2019 B6 37

Putnam 2019 — Paper A

2019 A1

2019 A1. Determine all possible values of the expression

$$A^3 + B^3 + C^3 - 3ABC$$

where A, B, and C are nonnegative integers.

Fields: algebra (factoring identities), elementary number theory (residues mod 9) **Position:** A1 – opener of the paper, purely algebraic, but the full answer needs both a bound and a construction **Answer:** All nonnegative integers except those congruent to 3 or 6 $\pmod 9$.

Layer 1 THE WHAT — surface analysis

The natural attempt. Three cubes and a triple product suggest plugging in small cases to get a feel for the range. $(0,0,0) \to 0$; $(1,0,0) \to 1$; $(1,1,0) \to 2$; $(1,1,1) \to 0$; $(2,1,0) \to 9$; $(2,1,1) \to 4$; $(2,2,1) \to 5$; $(3,1,0) \to 28$. The values $0, 1, 2, 4, 5, 7, 8, 9, 10, \ldots$ show up; 3 and 6 stubbornly do not. A reasonable student conjectures: every nonnegative integer occurs except those $\equiv 3$ or 6 $\pmod 9$, and reaches for the direct tool – residue casework. Since a cube mod 9 depends only on the base mod 3 (residues $0, 1, 2$ cube to $0, 1, 8$), it is possible, in principle, to fix A, B, C each mod 9 (or, smarter, mod 3, cutting the casework to 27 triples) and mechanically expand $A^3 + B^3 + C^3 - 3ABC$ mod 9 in every one of them. This is real, finishable work – and it will, eventually, confirm that 3 and 6 mod 9 never appear. But it is grinding through 27 near-identical expansions to extract one congruence fact, and worse, it only ever proves the *negative* half of the problem. "Determine all possible values" also demands showing that every permitted value genuinely is hit by some nonnegative triple, and nothing about checking residues mod 9 produces a triple that realises, say, $X = 1{,}000{,}000$. The casework route reaches the wall exactly there: a correct but partial answer, purchased at the cost of 27 expansions, with the harder half of the problem – construction – still completely open.

The model chain. Write $B = A + b$ and $C = A + c$ for integers b, c (any integers, since A itself is free to be taken large enough to keep $B, C \geq 0$). Then $A + B + C = 3A + b + c$, and using $A - B = -b$, $B - C = b - c$, $C - A = c$,

$$A^2 + B^2 + C^2 - AB - BC - CA = \tfrac{1}{2}\left[(A-B)^2 + (B-C)^2 + (C-A)^2\right] = \tfrac{1}{2}\left[b^2 + (b-c)^2 + c^2\right] = b^2 - bc + c^2.$$

So, using the classical factorisation $A^3 + B^3 + C^3 - 3ABC = (A + B + C)(A^2 + B^2 + C^2 - AB - BC - CA)$,

$$X = (3A + b + c)(b^2 - bc + c^2).$$

Nonnegativity is now free: $b^2 - bc + c^2 = (b - \frac{c}{2})^2 + \frac{3}{4}c^2 \geq 0$, and $3A + b + c = A + B + C \geq 0$, so $X \geq 0$ always.

Achievability. Choosing (b, c) fixes the second factor and lets A (hence the first factor) range over an arithmetic progression: $(b, c) = (0, 1)$ gives second factor 1, first factor $3A + 1$, hitting

every integer $\equiv 1 \pmod 3$; $(b,c) = (1,1)$ gives second factor 1, first factor $3A+2$, hitting every integer $\equiv 2 \pmod 3$; $(b,c) = (0,0)$ gives $X = 0$; $(b,c) = (1,2)$ gives second factor $1-2+4 = 3$, first factor $3A+3$, so $X = 9(A+1)$, hitting every positive multiple of 9. Together these realise every nonnegative integer not $\equiv 3, 6 \pmod 9$.

The bound. Suppose $X \equiv 0 \pmod 3$. Modulo 3, $b^2 - bc + c^2 \equiv b^2 + 2bc + c^2 = (b+c)^2$ (since $-3bc \equiv 0$), so

$$X = (3A + b + c)(b^2 - bc + c^2) \equiv (b+c)(b+c)^2 = (b+c)^3 \pmod 3\text{-detecting},$$

more precisely $X \equiv (b+c)\cdot(b+c)^2 \pmod 9$ once $3 \mid (b+c)$ or not is resolved: if $3 \nmid X$ we are done (not $\equiv 0$), and if $3 \mid X$ then since $X \equiv (3A+b+c)(b+c)^2 \pmod 3$ forces $3 \mid (b+c)$, both factors $3A+b+c$ and $b+c$ (hence $b^2 - bc + c^2 \equiv (b+c)^2 \equiv 0 \pmod 9$... isolating cleanly) are divisible by 3, so their product is divisible by 9. Hence every multiple of 3 that X attains is in fact a multiple of 9, ruling out $3, 6 \pmod 9$ for good.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains begin at the same place: compute small cases, notice $3, 6 \pmod 9$ never appear, want to prove it. The split happens at the very next decision: how to represent an *arbitrary* triple (A, B, C) for the purpose of computing X mod 9. The natural attempt represents it by three independent residues mod 9 (or mod 3) and checks all combinations. The model represents it by one free parameter A and two *differences* $b = B - A$, $c = C - A$, exploiting that X is a symmetric-in-shift expression whose behaviour should depend on the shape of the triple, not its absolute size.

> *The expression only involves sums and products of A, B, C symmetrically – if I shift all three by the same amount the expression changes in a controlled way, so instead of three independent unknowns I should track one "size" variable and the two gaps between the numbers, because the gaps are what a factorisation like $(A+B+C)(\dots)$ is actually built from.*

> *I already know (or can derive in one line) that $a^3 + b^3 + c^3 - 3abc$ factors as $(a+b+c)(a^2+b^2+c^2-ab-bc-ca)$ – so instead of computing X from scratch in 27 residue cases, I should substitute into that factorisation and let the substitution $B = A + b$, $C = A + c$ turn the second factor into a clean function of b, c alone, decoupled from A.*

Layer 3 THE WHY

The factorisation $A^3 + B^3 + C^3 - 3ABC = (A+B+C)(A^2+B^2+C^2-AB-BC-CA)$ is not an arbitrary trick to recall; it is what falls out of writing the second factor as $\frac{1}{2}[(A-B)^2 + (B-C)^2 + (C-A)^2]$, a sum of squares built entirely from *pairwise differences*. That is the structural fact the casework approach misses: the expression is symmetric and (after factoring) manifestly nonnegative because it is, up to the first factor, a sum of squares of differences – not a coincidence discovered by trying triples, but forced by the algebraic identity itself. Once the second factor is recognised as living in the differences of A, B, C, substituting $B = A + b$, $C = A + c$ is not a trick either: it is the natural coordinate system for an expression that only cares about how far apart the three numbers are, with A left to carry whatever residual freedom remains (and A's only job, once b, c are fixed, is to slide the first factor through an entire arithmetic progression, which is exactly why the achievability half of the problem falls out for free once the bound is established – the same substitution that proves the bound also parametrises every attainable value).

The mod-9 argument is likewise not a checked table but a one-line consequence of the same substitution: mod 3, $b^2 - bc + c^2 \equiv (b+c)^2$, so the two factors of X are, mod 3, $(3A+b+c)$ and $(b+c)^2$ – both controlled by the single quantity $b+c$. If $3 \mid b+c$, it divides both, forcing

$9 \mid X$; if not, $3 \nmid X$ at all. There is no case where X is divisible by exactly 3 and not 9, and this is visible in one congruence check rather than 27 separate expansions, precisely because the substitution reduced three unknowns to the one quantity ($b + c$ mod 3) that actually controls the answer.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the shape $a^3 + b^3 + c^3 - 3abc$ itself – on sight, this is a named identity (the norm form of $\mathbb{Z}[\omega]/(\omega^3 - 1)$-type factorisation), and the cue should trigger writing down the factorisation immediately rather than reaching for numeric trial.

2. That cue suggests representing the triple by one free size parameter and the pairwise gaps between the three numbers, since the second factor of the identity is built purely from those gaps.

3. Hesitation occurs right after the bound is proved: a student who stops at "$3 \mid X \Rightarrow 9 \mid X$" has answered only half the question. The reinforcement is to notice that the same substitution used for the bound already produces an arithmetic progression for each fixed (b, c), so achievability needs no separate construction – it was earned by the bound's own machinery.

4. The trap is treating "determine all possible values" as a bound-only question and stopping at the negative result, or alternatively doing 27-case residue casework that proves the bound but leaves achievability as an afterthought no one actually checks.

5. Recognisable faster next time by naming the family: any symmetric cubic expression in three variables that "feels like" a norm form should be tested against $(a + b + c)(a^2 + b^2 + c^2 - ab - bc - ca)$-type factorisations before any casework is attempted.

6. Reusable rule: when a problem asks to "determine all possible values" of a symmetric expression in several nonnegative-integer variables, look for a factorisation that separates a free "scale" parameter from the pairwise differences; the scale parameter usually supplies an entire arithmetic progression for free, turning the bound argument and the construction argument into the same piece of algebra instead of two separate tasks.

▷ **Cue:** a symmetric cubic expression of the shape $a^3 + b^3 + c^3 - 3abc$ $\longrightarrow$ **Move:** factor via $(a + b + c)(a^2 + b^2 + c^2 - ab - bc - ca) = (a + b + c) \cdot \frac{1}{2}\sum(a - b)^2$ before doing any numeric trial

▷ **Cue:** a symmetric expression in A, B, C where you need both a bound and a construction of every achievable value $\longrightarrow$ **Move:** substitute $B = A + b$, $C = A + c$ so one free "scale" variable A and the two gaps b, c separate cleanly, letting A sweep out an arithmetic progression for each fixed gap pair

2019 A2

2019 A2. In the triangle $\triangle ABC$, let G be the centroid, and let I be the center of the inscribed circle. Let α and β be the angles at the vertices A and B, respectively. Suppose that the segment IG is parallel to AB and that $\beta = 2\tan^{-1}(1/3)$. Find α.

Fields: triangle geometry (incenter/centroid distances, half-angle tangent identities) **Position:** A2 – second problem, synthetic geometry with a short algebraic finish **Answer:** $\alpha = \pi/2$: the triangle is right-angled at A.

Layer 1 THE WHAT — surface analysis

The natural attempt. G and I both have known coordinate formulas – $G = \frac{A+B+C}{3}$ and $I = \frac{aA+bB+cC}{a+b+c}$ – so the natural first move is to set up coordinates, $A = (0, 0)$, $B = (c, 0)$, and parametrise C using the given angle β at B: $C = (c - t\cos\beta,\ t\sin\beta)$ for some $t > 0$ related to side length a. Plugging into the centroid and incenter formulas gives explicit (messy) expressions for G_y and I_y in terms of a, b, c, β; the condition $IG \parallel AB$ becomes the single equation $G_y = I_y$.

This does turn the geometric hypothesis into an honest equation, and it is not wrong – but it produces one equation in three unknowns a, b, c (with β fixed and α still to be extracted via the Law of Sines or Law of Cosines), and the algebra to isolate α from $G_y = I_y$ combined with $\beta = 2\tan^{-1}(1/3)$ is a genuine slog: clearing denominators from the incenter weighted-average formula, substituting Law-of-Sines relations $a = 2R\sin\alpha$ etc., and simplifying a two-variable trig equation down to a single relation on α. It can be pushed through, but there is no natural stopping point that signals "you are nearly done" – the student is grinding coordinates without a structural reason to expect a clean value like $\pi/2$ to fall out.

The model chain. $IG \parallel AB$ says the entire line through I and G is parallel to AB; every point on a line parallel to AB sits at the same perpendicular distance from AB, so in particular I and G are equidistant from line AB – no coordinates needed to see this, it is immediate from what "parallel" means.

Now name both distances using facts already on hand. The incenter is, by definition, equidistant from all three sides, at distance r (the inradius); so the distance from I to AB is r. The centroid lies on the median from C to the midpoint M of AB, dividing CM in ratio $2:1$ from the vertex; since M itself lies on line AB (distance 0), and distance to AB varies linearly along the segment CM, the distance from G to AB is $\frac{1}{3}$ of the distance from C to AB, i.e. $\frac{1}{3}h_C$ where h_C is the altitude from C.

Equating: $r = \frac{1}{3}h_C$, i.e. $h_C = 3r$.

Convert to side lengths. Writing K for the area, s for the semiperimeter, $K = \frac{1}{2}c\,h_C = \frac{1}{2}c(3r) = \frac{3}{2}cr$, and also $K = rs$. So $rs = \frac{3}{2}cr \Rightarrow s = \frac{3}{2}c \Rightarrow a + b + c = 3c \Rightarrow a + b = 2c$.

This single relation, together with the half-angle tangent formula $\tan(\theta/2) = \dfrac{r}{s - (\text{opposite side})}$ (a standard consequence of the incircle's tangent lengths), finishes the problem. Since $a + b = 2c$: $s = \frac{3}{4}(a + b)$ and $s - c = \frac{1}{4}(a + b)$, so $\frac{s-c}{s} = \frac{1}{3}$. By Heron, $r^2 s = (s-a)(s-b)(s-c)$, so

$$(s-a)(s-b) = \frac{r^2 s}{s-c} = r^2 \cdot \frac{s}{s-c} = 3r^2.$$

But $(s-a)(s-b) = \dfrac{r}{\tan(\alpha/2)} \cdot \dfrac{r}{\tan(\beta/2)}$, so $\tan(\alpha/2)\tan(\beta/2) = \frac{1}{3}$. Given $\tan(\beta/2) = \frac{1}{3}$, this forces $\tan(\alpha/2) = 1$, so $\alpha/2 = \pi/4$ and $\alpha = \pi/2$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start from the same hypothesis, "IG is parallel to AB." The split is at the very next step: how is that hypothesis translated into an equation? The natural attempt translates it as a coordinate condition – $G_y = I_y$ in some chosen frame – which is correct but pulls in the full weighted-average formula for I and forces the rest of the solution to fight through that formula's denominators. The model translates the same hypothesis as a pure distance statement – "I and G are the same distance from line AB" – and then reaches directly for the two distances the solver already has closed forms for: the inradius r for I, and $\frac{1}{3}h_C$ for G. No coordinate frame is ever chosen.

> *"$IG \parallel AB$" is not first and foremost a slope condition to set up in coordinates – it's a statement that two points are equidistant from a fixed line, and I already have a name for the distance from each of these two specific points to a side: the inradius for the incenter, and one-third of the opposite altitude for the centroid. I should equate those two named quantities instead of computing either point's coordinates from scratch.*

> *Once I have the side relation $a + b = 2c$, I don't need to solve for the actual triangle – the target is a single angle α, and $\tan(\theta/2) = r/(s - \text{opposite side})$ already converts angles into ratios involving $s - a$, $s - b$, so I should look for $(s-a)(s-b)$ specifically, because Heron's formula hands me that*

> *product directly once I know* $(s-c)/s$.

Layer 3 THE WHY

The first critical thought is forced by the actual definition of "parallel": two lines are parallel exactly when every point of one is at constant perpendicular distance from the other, so a hypothesis phrased as "$IG \parallel AB$" is, definitionally, a statement about two distances-to-a-line being equal – extracting coordinates is optional machinery bolted on afterward, not required by the hypothesis itself. Recognising this converts an apparently two-dimensional condition (a parallel segment) into a one-dimensional one (two numbers are equal), which is exactly the reduction that makes the rest of the problem algebraic rather than geometric.

That the distance from G to AB is $\frac{1}{3}h_C$ is likewise not a formula to memorise but a direct consequence of the centroid's definition as the point dividing each median $2:1$ from the vertex: since M (the foot of the median on AB) has distance 0 from AB, and distance to a fixed line is an affine (linear) function of position along any straight segment, the distance from G scales exactly with its position fraction along CM, which is $\frac{1}{3}$ from M.

The half-angle identity $\tan(\theta/2) = r/(s - \text{opposite side})$ is where the archive should not simply invoke a memorised formula. It comes from the tangent-length property of the incircle: if the incircle touches AB at E, then $AE = s - a$ and $BE = s - b$ (a standard fact: the two tangent lengths from a vertex are equal, and the three tangent lengths sum in pairs to the three sides, which solves to exactly s minus the opposite side). Since $IE \perp AB$ and $\angle IBE = \beta/2$ (the incenter lies on the angle bisector from B), the right triangle IEB gives $\tan(\beta/2) = IE/BE = r/(s-b)$ directly from the definition of tangent in a right triangle. Once this is in hand, the entire rest of the problem is just recognising that Heron's formula, $r^2 s = (s-a)(s-b)(s-c)$, hands you $(s-a)(s-b)$ as soon as you know the ratio $(s-c)/s$ – which the side relation $a + b = 2c$ supplies in one line.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the phrase "IG parallel to AB" itself: on seeing a parallelism hypothesis between two named points and a side, write down immediately what parallel actually means (equal perpendicular distance to that line), rather than reaching for coordinates as the default translation.

2. That cue suggests naming the distance from each named point to the line using whatever closed-form each point already has: inradius for the incenter, one-third the opposite altitude for the centroid.

3. Hesitation occurs right after deriving $a + b = 2c$: it isn't obvious this alone determines α, since it is one equation in three unknowns. The reinforcement is to notice the target is not "solve the triangle" but "find one angle," and $\tan(\theta/2) = r/(s - \text{side})$ converts exactly that target into a ratio the Heron identity can supply without ever pinning down a, b, c individually.

4. The trap is treating the parallelism condition as inherently coordinate-based (it feels like a slope condition) and setting up a full coordinate system before asking what "parallel" structurally means for two named points with known distance formulas.

5. Recognisable faster next time by naming the mechanism: "a segment between two points is parallel to a line iff those two points are equidistant from that line" – any problem stating a parallelism condition between two triangle centers and a side should trigger a search for closed-form distance-to-that-side formulas for each center, before coordinates.

6. Reusable rule: when a hypothesis is "segment PQ is parallel to line ℓ" for two named points P, Q with known distance-to-ℓ formulas, translate it directly into "distance(P, ℓ) = distance(Q, ℓ)" rather than into a coordinate slope condition – this halves the dimension of the problem before any algebra starts.

▷ **Cue:** a segment between two named points is stated to be parallel to a line $\longrightarrow$ **Move:** equate the two points' perpendicular distances to that line directly, using known distance formulas for each point, instead of setting up coordinates

▷ **Cue:** target is a single angle of a triangle, and Heron's formula or the incircle is in play
⟶ **Move:** use $\tan(\theta/2) = r/(s - \text{opposite side})$ to convert the angle target into a ratio of semiperimeter-minus-side quantities

2019 A3

2019 A3. Given real numbers $b_0, b_1, \ldots, b_{2019}$ with $b_{2019} \neq 0$, let $z_1, z_2, \ldots, z_{2019}$ be the roots in the complex plane of the polynomial

$$P(z) = \sum_{k=0}^{2019} b_k z^k.$$

Let $\mu = (|z_1| + \cdots + |z_{2019}|)/2019$ be the average of the distances from $z_1, z_2, \ldots, z_{2019}$ to the origin. Determine the largest constant M such that $\mu \geq M$ for all choices of $b_0, b_1, \ldots, b_{2019}$ that satisfy

$$1 \leq b_0 < b_1 < b_2 < \cdots < b_{2019} \leq 2019.$$

Fields: complex analysis / algebra (Vieta's formulas, AM-GM) **Position:** A3 – third problem, an optimisation over a 2020-parameter family that collapses to a two-line bound **Answer:** $M = 2019^{-1/2019}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The hypothesis piles on information – 2020 real numbers, strictly increasing, confined to $[1, 2019]$ – and it is natural to suspect that all of that information is load-bearing and must be used to pin down where the roots can sit. A student who has seen theorems bounding root locations from coefficient monotonicity (an Eneström–Kakeya-type result: for a polynomial with positive, monotonically increasing coefficients, every root lies in the annulus $b_0/b_{2019} \leq |z| \leq 1$, roughly) reaches for exactly this kind of individual-root bound, attempting either to recall or to reprove an inequality controlling each $|z_i|$ on its own, then to sum 2019 such per-root bounds and divide by 2019 to bound the average μ. This is not a dead end – such theorems exist and are true – but it is heavy machinery: proving a per-root annulus bound from scratch (typically via a clever comparison like $(z-1)P(z)$ having controlled coefficients, or a Rouché-type contour argument) is itself a multi-step proof, and even after establishing it the student still has to check that the resulting bound on μ, obtained by averaging 2019 possibly-loose individual bounds, is actually tight – i.e. that some configuration attains it. The route works in outline but costs a nontrivial theorem's worth of proof just to reach the same numeric answer a two-line argument produces.

The model chain. The target μ is an *arithmetic* mean of the $|z_i|$, so the AM-GM inequality is the first thing to try, and it needs only a *geometric* mean:

$$\mu = \frac{|z_1| + \cdots + |z_{2019}|}{2019} \geq (|z_1| \cdots |z_{2019}|)^{1/2019} = \left| \prod_{i=1}^{2019} z_i \right|^{1/2019}.$$

The product of all roots is exactly what Vieta's formula for the constant term gives: writing $P(z) = b_{2019} \prod_{i=1}^{2019} (z - z_i)$, the constant term is $P(0) = b_{2019} \prod_i (-z_i) = (-1)^{2019} b_{2019} \prod_i z_i = -b_{2019} \prod_i z_i$ (since 2019 is odd), and $P(0) = b_0$. So $\prod_i z_i = -b_0/b_{2019}$, hence

$$\left| \prod_i z_i \right| = \frac{b_0}{b_{2019}} \qquad \text{(both positive, since } 1 \leq b_0 < b_{2019} \leq 2019\text{)}.$$

Therefore $\mu \geq (b_0/b_{2019})^{1/2019}$ for *every* admissible choice of coefficients, and since $b_0 \geq 1$ and $b_{2019} \leq 2019$,

$$\mu \geq \left(\frac{1}{2019}\right)^{1/2019} = 2019^{-1/2019}.$$

This uses only the two extreme coefficients; the strict increase of $b_1, \ldots, b_{2018}$ never enters the bound at all.

For the matching construction, take $b_k = 2019^{k/2019}$ for $k = 0, \ldots, 2019$; this is strictly increasing, with $b_0 = 1$ and $b_{2019} = 2019$, so it is admissible. Then $P\left(z/2019^{1/2019}\right) = \sum_{k=0}^{2019} z^k = \dfrac{z^{2020} - 1}{z - 1}$, whose roots are the 2020-th roots of unity other than 1 – all of modulus exactly 1. Undoing the rescaling, every root of P has modulus exactly $2019^{-1/2019}$, so $\mu = 2019^{-1/2019}$ exactly: the bound is attained, so it is the best possible constant.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start from the same reading of the problem: a lower bound on μ, controlled somehow by the coefficient conditions. The split is in what feature of μ is targeted. The natural attempt targets each $|z_i|$ individually – since bounding a sum is often easiest by bounding each term – and this pulls in machinery for locating roots one at a time. The model targets μ's identity as an *arithmetic mean* and asks what inequality converts an arithmetic mean into something Vieta's formulas can compute directly: the geometric mean, controlled by the product of *all* the roots at once, which needs no information about any individual root's location.

> *I don't need to know where any single root is – the quantity I actually have to bound is an average, and averages of nonnegative numbers are bounded below by their geometric mean via AM-GM, and the geometric mean of all the roots together is exactly a product, which is exactly what the constant-term-to-leading-coefficient ratio in Vieta's formula hands me for free, with zero work spent on locating any individual z_i.*

> *Once the bound only depends on b_0 and b_{2019}, the rest of the hypothesis – the strict increase of $b_1, \ldots, b_{2018}$ – is not there to be used in the inequality; it is there to make the equality case honest, i.e. to guarantee that a strictly increasing sequence with these two endpoints actually exists, and I should go looking for a construction achieving $b_0/b_{2019} = 1/2019$ exactly rather than expect the middle coefficients to sharpen the bound.*

Layer 3 THE WHY

AM-GM is the natural bridge here because μ is defined, by the problem's own notation, as an arithmetic mean – the definitional form of the quantity being bounded already announces which classical inequality applies. AM-GM's geometric-mean side is a product, and a product of *all* the roots of a polynomial is not something that needs root-by-root analysis to access: it is read off directly from the two coefficients that survive when a polynomial is written in factored form, $P(z) = b_{2019} \prod (z - z_i)$, evaluated at $z = 0$. This is why the per-root machinery (Eneström–Kakeya or similar) is more than the problem asks for: it answers a harder question – where is *each* root? – when the problem only asks about a symmetric function of all the roots together, and Vieta's formulas exist precisely to extract symmetric functions of roots without locating any of them individually. The strict-increase hypothesis on $b_1, \ldots, b_{2018}$ is, from this vantage point, a red herring for the bound itself (it plays no role in the inequality) and is only there to be checked for consistency in the equality case: the construction $b_k = 2019^{k/2019}$ needs to be strictly increasing to be admissible at all, and the geometric sequence is the natural candidate to check first, because it is exactly the sequence whose rescaled polynomial becomes the all-ones polynomial $\sum z^k$, whose roots (the roots of unity, minus 1) are the textbook example of a polynomial with every root on a fixed circle.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the definition of μ as an *arithmetic mean* of $|z_i|$'s – on seeing "average

of the $|z_i|$", write down AM-GM immediately as a candidate tool, since the problem has announced which side of the inequality it wants bounded.

2. That cue suggests replacing the arithmetic mean with the geometric mean, i.e. with $|\prod z_i|^{1/2019}$, and then asking what tool computes a product of *all* roots at once – Vieta's formula for the constant term over the leading coefficient.

3. Hesitation occurs at seeing the strict chain $b_0 < b_1 < \cdots < b_{2019}$ and assuming all 2020 inequalities must be used; the reinforcement is to check, after writing the AM-GM bound, exactly which coefficients actually appear in it (only b_0, b_{2019}) and to stop looking for a use for the rest.

4. The trap is over-trusting the amount of hypothesis given: a long chain of strict inequalities on 2020 numbers reads like it should all be used, and reaching for a heavy per-root location theorem is the natural (but unnecessary) response to that impression.

5. Recognisable faster next time by naming the pattern: whenever a problem asks for a bound on an arithmetic mean of root moduli (or any symmetric quantity built from roots), check Vieta's formulas for a symmetric-function shortcut before reaching for tools that locate individual roots.

6. Reusable rule: to bound the arithmetic mean of the moduli of all roots of a polynomial, apply AM-GM to convert it to a geometric mean, then evaluate that geometric mean via $|\prod z_i| = |b_0/b_n|$ from Vieta's formula – this needs only the two extreme coefficients, so check first whether a problem's apparently rich hypothesis (a long monotone chain, say) is only load-bearing for the equality case, not for the bound.

▷ **Cue:** a bound is requested on the arithmetic mean of the moduli of all roots of a polynomial $\longrightarrow$ **Move:** apply AM-GM to reduce to the geometric mean, then evaluate the product of all roots via Vieta's constant-term-over-leading-coefficient formula, avoiding any individual root-location argument

▷ **Cue:** a hypothesis lists many strict inequalities on the coefficients, but the derived bound only involves the two extremes $\longrightarrow$ **Move:** treat the remaining inequalities as constraints to satisfy in the matching construction, not as levers to sharpen the bound itself

2019 A4

2019 A4. Let f be a continuous real-valued function on $\mathbb{R}^3$. Suppose that for every sphere S of radius 1, the integral of $f(x, y, z)$ over the surface of S equals 0. Must $f(x, y, z)$ be identically 0?

Fields: analysis: surface integrals, spherical means; Fourier analysis (the sphere as a convolution kernel) **Position:** A4 — mid-paper. A yes/no question whose answer is "no", which means the whole problem is the construction of one counterexample and one surface integral. **Answer:** No. For example $f(x, y, z) = \sin(\pi z)$ integrates to 0 over every unit sphere.

Layer 1 THE WHAT — surface analysis

The natural attempt. The phrasing leans towards "yes". You have a family of averages that all vanish, and the family is huge: one condition for every centre in $\mathbb{R}^3$. The reflex is to shrink the sphere and use continuity — if the average over a tiny sphere around p is 0 then $f(p) = 0$. But the radius is fixed at 1. You cannot shrink anything. So you try to extract a pointwise statement from radius-1 averages alone.

Attempt one: mean-value property. Harmonic functions are exactly the functions whose sphere averages equal the centre value, so perhaps the hypothesis says f is harmonic with $f \equiv 0$ in some averaged sense, and Liouville finishes it. This is the wall: the mean-value characterisation of harmonic functions needs the identity for *all* small radii (or all radii), and

nothing here links f at the centre to the sphere average at all — the centre value never appears in the hypothesis.

Attempt two: differentiate in the centre. Write $A(c) = \iint_{|x-c|=1} f\,dS$; then $A \equiv 0$ so every derivative of A in c vanishes. Each derivative is again a sphere integral, of ∂f this time, and you are in the same position one level down. After twenty minutes the conditions have multiplied but none of them pins f at a point. That is the honest signal that the answer might be "no", and a student who does not pause here keeps trying to prove a false statement.

The model chain. *Step 1: look for a counterexample of the cheapest possible shape.* A function of one variable, $f(x,y,z) = g(z)$ with g continuous, reduces a surface integral in $\mathbb{R}^3$ to a single-variable integral. Try it.

Step 2: compute the sphere integral of $g(z)$. Let S be the unit sphere with centre (x_0, y_0, z_0) and parametrise it by

$$S(\phi,\theta) = (x_0,y_0,z_0) + (\sin\phi\cos\theta,\ \sin\phi\sin\theta,\ \cos\phi), \qquad \phi \in [0,\pi],\ \theta \in [0,2\pi].$$

The area element is $\left\|\partial_\phi S \times \partial_\theta S\right\| d\phi\,d\theta = \sin\phi\,d\phi\,d\theta$. Hence

$$\iint_S g(z)\,dS = \int_0^{\pi}\int_0^{2\pi} g(z_0+\cos\phi)\sin\phi\,d\theta\,d\phi = 2\pi\int_{-1}^{1} g(z_0+t)\,dt,$$

using $t = \cos\phi$, $dt = -\sin\phi\,d\phi$. This is Archimedes' hat-box theorem: the area of the slice of a unit sphere between heights t and $t+dt$ is $2\pi\,dt$, independent of t.

Step 3: choose g. We need $\int_{z_0-1}^{z_0+1} g(t)\,dt = 0$ for *every* z_0: every window of length 2 must have integral zero. Any continuous g with period 2 and $\int_0^2 g = 0$ does this, since a window of length one full period has the same integral wherever it starts. Take $g(t) = \sin(\pi t)$:

$$\int_{z_0-1}^{z_0+1} \sin(\pi t)\,dt = \frac{1}{\pi}\big[\cos(\pi(z_0-1)) - \cos(\pi(z_0+1))\big] = 0,$$

because cos has period 2π.

Step 4: conclude. $f(x,y,z) = \sin(\pi z)$ is continuous, not identically zero, and integrates to 0 over every sphere of radius 1. The answer is no.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree that the hypothesis is "all radius-1 sphere integrals vanish". At (†) — the decision about which direction to argue — they split. The natural attempt commits to proving "yes" because a question phrased "Must f be identically 0?" sounds like a uniqueness theorem. The model tests "no" first with the simplest functions available, one-variable functions, and the test succeeds in five lines.

> *The radius is fixed. A fixed-radius average is a blurring, and a blur of fixed width can kill a wave whose wavelength is matched to it. Before I try to prove anything, let me feed it a function of z alone and see what the sphere integral turns into.*

> *The sphere integral of $g(z)$ is $2\pi\int_{z_0-1}^{z_0+1} g$. So I need every window of length 2 to integrate to zero. That is exactly "period 2, mean zero". $\sin(\pi z)$.*

Layer 3 THE WHY

Why is "no" the natural expectation, once you look at the problem the right way? Because of what the hypothesis actually is. Define the spherical mean operator

$$(Tf)(c) = \iint_{|x-c|=1} f(x)\,dS(x).$$

It is linear and commutes with translation: shifting f shifts Tf. Any linear, translation-invariant operator acts on plane waves by multiplication. Feed it $e_\xi(x) = e^{i\xi\cdot x}$ and you get $Te_\xi = m(\xi)\, e_\xi$ for some number $m(\xi)$. So the question "is T injective?" is the question "does m ever vanish?". By rotational symmetry m depends only on $k = |\xi|$, so orient ξ along the z-axis and use Step 2 with $g(t) = e^{ikt}$:

$$m(\xi) = 2\pi \int_{-1}^{1} e^{ikt}\, dt = \frac{4\pi \sin k}{k}.$$

This vanishes at $k = \pi, 2\pi, 3\pi, \ldots$. So there is a whole sphere's worth of frequencies $|\xi| = n\pi$ that T annihilates, and the real part of $e^{i\pi z}$ is $\cos(\pi z)$, the imaginary part $\sin(\pi z)$ — the counterexample, found by computation rather than by inspiration.

This is the same mechanism as a moving average. The average of h over every window $[t-1, t+1]$ vanishes for $h(t) = \sin(\pi t)$, because a window of width 2 contains exactly one full period. The sphere is a three-dimensional window, and Archimedes' theorem says that, seen from the z-axis, it is *exactly* the flat window $[z_0 - 1, z_0 + 1]$ with uniform weight 2π. That is why the one-variable test is not a trick: the sphere measure projected onto a line is uniform, so a sphere integral of a function of z is literally a moving average on the line.

Why did "shrink the sphere" feel right and fail? The uniqueness results you know — the mean value property, Lebesgue differentiation — all use averages at scales tending to zero. A single scale is a single filter, and a single filter has a kernel. The phrase "for every sphere S of radius 1" is the cue: *fixed scale* means *filter*, and filters are rarely injective. Had the problem said "every sphere of every radius" the answer would be yes (let the radius tend to 0).

Kedlaya's solution opens with "Let g be any continuous function with $g(t+2) = g(t)$ and $\int_0^2 g = 0$". Here is where that comes from: the period 2 is the diameter of the sphere, forced by the window $[z_0 - 1, z_0 + 1]$, and the mean-zero condition is what makes a full-period window integrate to zero.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "radius 1" — a single fixed scale — together with a "Must ...?" question. Write down: *fixed-radius averaging is a translation-invariant linear filter; test plane waves or one-variable functions before trying to prove injectivity.*

2. Reduce dimension: try $f = g(z)$, compute the sphere integral in spherical coordinates, and obtain the moving-average condition $\int_{z_0-1}^{z_0+1} g = 0$ for all z_0.

3. Hesitation occurs at the choice of direction: "Must f be 0?" reads like a theorem. What reinforces it away is the rule that on a yes/no question you spend the first five minutes on the cheapest possible counterexample, and here the cheapest one costs one integral.

4. The trap is the mean-value property. It feels like the relevant theorem, but it needs all small radii and it involves the value at the centre, neither of which the hypothesis supplies. A second, smaller trap: getting the area element wrong ($d\phi\, d\theta$ instead of $\sin\phi\, d\phi\, d\theta$), which destroys the uniform-weight projection that makes the calculation work.

5. Recognise faster via the slogan: *an average at one fixed scale is a convolution, and a convolution kills every wave whose frequency is a zero of the kernel's transform.* The same thought answers questions about moving averages of sequences, averages over unit discs or unit balls, and difference operators like $f(x+1) - f(x)$.

6. Decision rule: **when a question asks whether a linear, translation-invariant condition forces a function to vanish, test exponentials (or sines) first.** If some frequency is annihilated, the answer is no and that wave is the counterexample; if none is, you have the heuristic for why the answer is yes.

▷ **Cue:** a condition imposed on averages at one fixed radius or window width, asked whether it forces $f \equiv 0$ $\longrightarrow$ **Move:** suspect "no"; the averaging is a convolution, so test $e^{i\xi\cdot x}$ or sin of one coordinate and find the frequencies it kills

▷ **Cue:** integral of a function of z alone over a sphere $\longrightarrow$ **Move:** use Archimedes: the sphere's area projects uniformly onto the axis, so $\iint_S g(z)\,dS = 2\pi r \int_{z_0-r}^{z_0+r} g(t)\,dt$

▷ **Cue:** every window of fixed length L must have integral zero $\longrightarrow$ **Move:** take g with period L (or L/n) and mean zero, e.g. $\sin(2\pi t/L)$

▷ **Cue:** a yes/no "must" question in analysis $\longrightarrow$ **Move:** spend the first minutes on the cheapest counterexample family (constants, one-variable functions, exponentials) before attempting a proof

2019 A5

2019 A5. Let p be an odd prime number, and let $\mathbb{F}_p$ denote the field of integers modulo p. Let $\mathbb{F}_p[x]$ be the ring of polynomials over $\mathbb{F}_p$, and let $q(x) \in \mathbb{F}_p[x]$ be given by

$$q(x) = \sum_{k=1}^{p-1} a_k x^k,$$

where

$$a_k = k^{(p-1)/2} \mod p.$$

Find the greatest nonnegative integer n such that $(x-1)^n$ divides $q(x)$ in $\mathbb{F}_p[x]$.

Fields: algebra: polynomials over $\mathbb{F}_p$; number theory: power sums modulo p, Fermat's little theorem, primitive roots **Position:** A5 — hard but short once the right coordinate is chosen. The Legendre-symbol dressing is a decoy; the problem is about the exponent $(p-1)/2$, not about quadratic residues. **Answer:** $n = \dfrac{p-1}{2}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You recognise a_k at once: by Euler's criterion $k^{(p-1)/2} \equiv \left(\frac{k}{p}\right) = \pm 1$, so q is the Fekete polynomial $\sum \left(\frac{k}{p}\right) x^k$. That recognition pulls you towards quadratic-residue machinery — Gauss sums, substituting $x \mapsto x^g$ for a primitive root g, pairing k with $p-k$ — and none of it touches the question, which is about the behaviour of q at the single point $x = 1$.

So you try small cases. For $p = 5$: $a_1, a_2, a_3, a_4 = 1, 4, 4, 1$, so $q = x + 4x^2 + 4x^3 + x^4 = x(x+1)(x^2+3x+1)$ over $\mathbb{F}_5$, and $x^2 + 3x + 1 = (x-1)^2$ there. Multiplicity 2. For $p = 7$ the multiplicity comes out as 3. The guess $(p-1)/2$ is on the table.

To prove it you take derivatives at 1. $q(1) = \sum k^{(p-1)/2}$, $q'(1) = \sum k \cdot k^{(p-1)/2}$, $q''(1) = \sum k(k-1)k^{(p-1)/2}$, and in general

$$q^{(j)}(1) = \sum_{k=1}^{p-1} k(k-1)\cdots(k-j+1)\,k^{(p-1)/2}.$$

Now the falling factorial has to be expanded into powers of k — Stirling numbers — and each power sum evaluated. You also start worrying whether "derivatives vanish at 1" even detects multiplicity in characteristic p, since $(x-1)^p$ has all derivatives zero at 1. The calculation is correct in outline and heavy in detail, and the worry is real enough to cost time.

The model chain. *Step 1: expand around $x = 1$.* Put $x = 1 + t$. Then $(x-1)^n \mid q(x)$ if and only if t^n divides $q(1+t)$, that is, if and only if the first n coefficients of $q(1+t)$ vanish. By the binomial theorem,

$$q(1+t) = \sum_{k=1}^{p-1} a_k (1+t)^k = \sum_{j \geq 0} c_j\, t^j, \qquad c_j = \sum_{k=1}^{p-1} a_k \binom{k}{j}.$$

No derivatives, so no characteristic-p worry. Since $\deg q \le p-1$, only $j \le p-1$ matter.

Step 2: the only fact needed about $\binom{k}{j}$. For $0 \le j \le p-1$,

$$\binom{k}{j} = \frac{k(k-1)\cdots(k-j+1)}{j!}$$

is a polynomial in k of degree j with leading coefficient $1/j!$, and $j!$ is invertible mod p because $j < p$. For $j \ge 1$ it has zero constant term. So, working in $\mathbb{F}_p$ and writing $m_0 = (p-1)/2$,

$$c_j = \sum_{k=1}^{p-1} k^{m_0} \binom{k}{j} = \sum_i \beta_i S_i, \qquad S_i = \sum_{k=1}^{p-1} k^i,$$

where the β_i are the coefficients of the polynomial $k^{m_0}\binom{k}{j}$; its monomials have degrees i with $m_0 \le i \le m_0 + j$, and $\beta_{m_0+j} = 1/j!$.

Step 3: power sums. For $i \ge 1$,

$$S_i \equiv \begin{cases} -1, & (p-1) \mid i, \\ 0, & \text{otherwise.} \end{cases}$$

Proof: if $(p-1) \mid i$ each term is 1 by Fermat and there are $p-1$ of them. Otherwise take a primitive root g; then $g^i \neq 1$, and $k \mapsto gk$ permutes $\{1, \dots, p-1\}$ mod p, so $g^i S_i = S_i$, forcing $S_i = 0$.

Step 4: read off the answer. Every degree i occurring in c_j satisfies $1 \le m_0 \le i \le m_0 + j$. If $j < m_0$ then $i \le 2m_0 - 1 = p-2$, so no i is a multiple of $p-1$ and $c_j = 0$. If $j = m_0$, the top degree is $i = p-1$ with coefficient $1/m_0!$, and every other degree lies in $[m_0, p-2]$, so

$$c_{m_0} = \frac{1}{m_0!} S_{p-1} = -\frac{1}{m_0!} \neq 0.$$

Hence t^{m_0} divides $q(1+t)$ and t^{m_0+1} does not: $n = \dfrac{p-1}{2}$.

Check. $p = 5$: $c_2 = -1/2! = -3 = 2$ in $\mathbb{F}_5$; directly, $q(1+t)$ has t^2-coefficient $\sum a_k \binom{k}{2} = 0 + 4 + 12 + 6 = 22 \equiv 2$. It agrees. (I also verified $n = (p-1)/2$ by machine for all odd primes $p \le 31$.)

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree on reducing the question to the Taylor data of q at $x = 1$. At (†), the moment $k(k-1)\cdots(k-j+1)$ appears, they split. The natural attempt expands the falling factorial term by term. The model asks what it actually needs to know about that polynomial — only its degree range and its top coefficient — because the power-sum lemma kills every monomial of degree below $p-1$ regardless of its coefficient.

> *I am going to sum a polynomial in k over all nonzero residues. That sum only sees the monomials of degree divisible by $p-1$. So I do not care what the coefficients of $\binom{k}{j}$ are; I care about the range of degrees, and the top one.*

> *Derivatives in characteristic p make me nervous. So I will not differentiate. I will substitute $x = 1 + t$ and read the coefficients of t^j; divisibility by $(x-1)^n$ is then literally divisibility by t^n.*

Layer 3 THE WHY

Why is the power sum the right lens? Because of what the question measures. "$(x-1)^n$ divides q" is a statement about q near $x = 1$, and near $x = 1$ the value of x^k is governed

by how k enters polynomially: $(1+t)^k = \sum_j \binom{k}{j} t^j$. So every Taylor coefficient of q at 1 is $\sum_k a_k \cdot$ (polynomial in k) — a moment of the sequence a_k. With $a_k = k^{(p-1)/2}$, every moment is a combination of power sums $\sum k^i$. And power sums over the multiplicative group $\mathbb{F}_p^\times$ are the most rigid objects in elementary number theory: they vanish unless the exponent is a multiple of the group order $p-1$. That is the whole mechanism. The j-th coefficient climbs in degree to $(p-1)/2+j$, and the first time it reaches $p-1$ is $j=(p-1)/2$. The answer is the distance from the given exponent to the group order.

This also shows why the Legendre-symbol reading is a decoy. Replace $(p-1)/2$ by any exponent m with $1 \le m \le p-2$ and the same argument gives multiplicity $p-1-m$. Quadratic residues play no role; only the exponent does.

Kedlaya's solution introduces the operator $D = x\frac{d}{dx}$ and writes $q = D^{(p-1)/2}s$ with $s = 1+x+\cdots+x^{p-1}$. Here is where that comes from. $D(x^k) = kx^k$, so D^m multiplies the coefficient of x^k by k^m. A polynomial whose coefficients are k^m is therefore D^m of the polynomial whose coefficients are all 1 — and that polynomial is the geometric series, which over $\mathbb{F}_p$ is

$$s(x) = \frac{x^p-1}{x-1} = \frac{(x-1)^p}{x-1} = (x-1)^{p-1},$$

using $(x-1)^p = x^p - 1$ (the binomial coefficients $\binom{p}{i}$, $0 < i < p$, are divisible by p). Now each application of D to $(x-1)^N r(x)$ with $r(1) \neq 0$ gives $(x-1)^{N-1}x\big(Nr + (x-1)r'\big)$, whose cofactor at $x=1$ is $Nr(1)$; so D lowers the multiplicity at 1 by exactly one provided $N \not\equiv 0 \pmod p$. Starting from $N = p-1$ and applying D $(p-1)/2$ times passes through $N = p-1, p-2, \ldots, (p+1)/2$, none divisible by p, and lands on $(p-1)/2$. This is a second complete proof, and it is the same fact seen from the other side: "coefficients are k^m" is the fingerprint of $x\frac{d}{dx}$, exactly as "coefficients are 1" is the fingerprint of a geometric series.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "$(x-1)^n$ divides". Write down $x = 1+t$ and $q(1+t) = \sum_j \big(\sum_k a_k \binom{k}{j}\big) t^j$. The second cue is "$k^{(p-1)/2}$ summed over $k = 1, \ldots, p-1$": write down the power-sum lemma.

2. Moments plus power sums over $\mathbb{F}_p^\times$: each Taylor coefficient is a combination of $\sum k^i$, and only exponents divisible by $p-1$ survive. Alternatively, $x\frac{d}{dx}$ acting on the geometric series.

3. Hesitation occurs at the falling factorial. You feel you must expand it. What reinforces this away is asking "what does the next step use?" — the power-sum lemma uses only degrees, so only the degree range and the top coefficient of $\binom{k}{j}$ matter. A second hesitation is whether derivative tests work in characteristic p; substituting $x = 1+t$ removes the question.

4. The Legendre-symbol trap: recognising a_k as $\left(\frac{k}{p}\right)$ and reaching for Gauss sums or reciprocity. A quieter trap is dividing by $j!$ without noting that $j < p$ makes it invertible, and applying "multiplicity = number of vanishing derivatives" without the bound $n \le p-1$.

5. Recognise faster via the slogan: *divisibility by $(x-a)^n$ is a question about moments of the coefficient sequence, and moments of k^m over $\mathbb{F}_p^\times$ are power sums, which vanish below degree $p-1$.* The same move answers questions about $\sum_k f(k)$ for polynomial f over a finite field (Chevalley–Warning style arguments), and about vanishing orders of $\sum a_k x^k$ at roots of unity.

6. Decision rule: **before expanding an auxiliary polynomial, ask what property of it the next step consumes; compute only that.** Companion rule: when coefficients of a polynomial are k^m, think $\left(x\frac{d}{dx}\right)^m$ applied to the polynomial with coefficients 1.

▷ **Cue:** "$(x-a)^n$ divides $q(x)$", find the largest n $\longrightarrow$ **Move:** substitute $x = a+t$ and find the first nonzero coefficient of $q(a+t)$; this avoids derivative tests entirely

- ▷ **Cue:** a sum $\sum_{k=1}^{p-1} P(k)$ with P a polynomial, working mod p $\longrightarrow$ **Move:** only monomials k^i with $(p-1) \mid i$, $i \geq 1$, contribute, each contributing -1 times its coefficient; so track degrees, not coefficients
- ▷ **Cue:** polynomial coefficients equal to k^m $\longrightarrow$ **Move:** write the polynomial as $\left(x\frac{d}{dx}\right)^m$ of $\sum x^k$ and use the closed form of the geometric series
- ▷ **Cue:** geometric series $1 + x + \cdots + x^{p-1}$ over $\mathbb{F}_p$ $\longrightarrow$ **Move:** it equals $(x-1)^{p-1}$, by the Frobenius identity $(x-1)^p = x^p - 1$

2019 A6

2019 A6. Let g be a real-valued function that is continuous on the closed interval $[0,1]$ and twice differentiable on the open interval $(0,1)$. Suppose that for some real number $r > 1$,

$$\lim_{x \to 0^+} \frac{g(x)}{x^r} = 0.$$

Prove that either

$$\lim_{x \to 0^+} g'(x) = 0 \qquad \text{or} \qquad \limsup_{x \to 0^+} x^r |g''(x)| = \infty.$$

Fields: real analysis: Taylor's theorem with remainder, interpolation between a function and its second derivative (Landau–Kolmogorov type inequality) **Position:** A6 — the last problem of the morning, and it is hard in the specific way that analysis problems are hard: every ingredient is first-year calculus, and the entire difficulty is choosing one step length. **Answer:** Proof. If $|g''(x)| \leq Mx^{-r}$ near 0, apply Taylor on $[x, x+h]$ with $h = x^r\sqrt{\varepsilon(2x)}$, where $\varepsilon(y) = \sup_{0<z\leq y} |g(z)|/z^r$; this gives $|g'(x)| \leq (1 + 2^r + M/2)\sqrt{\varepsilon(2x)} \to 0$.

Layer 1 THE WHAT — surface analysis

The natural attempt. An "either/or" conclusion is a contrapositive in disguise, and most strong students see that: assume the second alternative fails, so $\limsup x^r|g''(x)| < \infty$, and prove $g'(x) \to 0$. Concretely there are $M, \delta_0 > 0$ with

$$|g''(x)| \leq Mx^{-r} \qquad (0 < x < \delta_0).$$

The only tool that links g, g' and g'' at once is Taylor's theorem. For $h > 0$ and some $\xi \in (x, x+h)$,

$$g(x+h) = g(x) + g'(x)h + \tfrac{1}{2}g''(\xi)h^2, \quad \text{so} \quad |g'(x)| \leq \frac{|g(x+h)| + |g(x)|}{h} + \frac{h}{2}|g''(\xi)|.$$

Now choose h. The natural first choice is the scale of the problem, $h = x$. The first term is $o\big((2x)^r\big)/x = o(x^{r-1}) \to 0$, since $r > 1$. The second is at most $\frac{x}{2}Mx^{-r} = \frac{M}{2}x^{1-r}$, which tends to *infinity*. Wall.

So shrink the step. Try $h = x^r$, the scale suggested by the x^r in both hypotheses. The first term becomes $o(x^r)/x^r = o(1)$, good. The second becomes $\frac{x^r}{2}Mx^{-r} = \frac{M}{2}$. Bounded — but not tending to zero. You have proved that g' is bounded near 0, which is not what was asked, and a student who stops here writes a partial answer. The two terms are fighting: a larger h helps the first and hurts the second.

The model chain. Assume $\limsup_{x\to 0^+} x^r|g''(x)| < \infty$; we show $g'(x) \to 0$. Fix $M, \delta_0 > 0$ with $|g''(x)| \leq Mx^{-r}$ on $(0, \delta_0)$.

Step 1: a monotone modulus for the first hypothesis. Define

$$\varepsilon(y) = \sup_{0<z\le y} \frac{|g(z)|}{z^r}.$$

Since $g(z)/z^r \to 0$, $\varepsilon(y)$ is finite for small y and $\varepsilon(y) \to 0$ as $y \to 0^+$; by construction ε is nondecreasing and $|g(z)| \le \varepsilon(y) z^r$ for $0 < z \le y$. (Passing to the supremum is what makes the bound usable at the *two* points x and $x+h$ with a single number.)

Step 2: the step length. Take x so small that $2x < \min(\delta_0, 1)$, $\varepsilon(2x) \le 1$ and $x^{r-1} \le 1$ (possible because $r > 1$). If $\varepsilon(2x) = 0$ then $g \equiv 0$ on $(0, 2x]$ and $g'(x) = 0$. Otherwise set

$$h = x^r \sqrt{\varepsilon(2x)} \;\le\; x^r \le x,$$

so $[x, x+h] \subset (0, 2x] \subset (0, \delta_0)$.

Step 3: Taylor. Since g is twice differentiable on $(0,1)$, g' is continuous on $[x, x+h]$ and Taylor's theorem with Lagrange remainder gives $\xi \in (x, x+h)$ with $g(x+h) = g(x) + g'(x)h + \frac{1}{2}g''(\xi)h^2$. Hence

$$|g'(x)| \le \frac{|g(x+h)| + |g(x)|}{h} + \frac{h}{2}|g''(\xi)|.$$

Step 4: estimate. Using $x + h \le 2x$ and Step 1, $|g(x+h)| + |g(x)| \le \varepsilon(2x)\big((2x)^r + x^r\big) = (1+2^r)\varepsilon(2x)x^r$. Using $\xi > x$, $|g''(\xi)| \le M\xi^{-r} \le Mx^{-r}$. Substituting h:

$$|g'(x)| \le \frac{(1+2^r)\varepsilon(2x)x^r}{x^r\sqrt{\varepsilon(2x)}} + \frac{M}{2}\cdot\frac{x^r\sqrt{\varepsilon(2x)}}{x^r} = \Big(1 + 2^r + \frac{M}{2}\Big)\sqrt{\varepsilon(2x)}.$$

Step 5. As $x \to 0^+$, $\varepsilon(2x) \to 0$, so $g'(x) \to 0$. This proves the contrapositive, hence the dichotomy.

Sharpness of the dichotomy. Neither alternative can be dropped. With $a = 2r - 1 > 0$ and $g(x) = x^{2r}\sin(x^{-a})$ (and $g(0) = 0$), $g(x)/x^r = x^r \sin(x^{-a}) \to 0$, while

$$g'(x) = 2rx^{2r-1}\sin(x^{-a}) - (2r-1)\cos(x^{-a})$$

has no limit; so here the second alternative must hold (and indeed $x^r g''(x)$ is of order x^{-r}). Kedlaya–Ng record Derksen's example in which *both* hold, so the "or" is inclusive.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree through the Taylor inequality $|g'(x)| \le \frac{|g(x+h)| + |g(x)|}{h} + \frac{h}{2}|g''(\xi)|$. At (†) — the choice of h — they split. The natural attempt picks h from the problem's visible scales, x or x^r, and gets ∞ or a constant. The model treats h as a free parameter to be *optimised*: two terms, one decreasing in h and one increasing, are balanced where they are equal.

> *I have a bound of the form $A/h + Bh$ and I get to choose h. That is minimised at $h = \sqrt{A/B}$, where it equals $2\sqrt{AB}$. Here A is about εx^r and B is about Mx^{-r}, so $\sqrt{AB}$ is about $\sqrt{\varepsilon M}$ — the x^r cancels. That is the whole proof.*

> *"Either P or Q": assume not Q, prove P. The hypothesis I get to use is a* bound *on g'', and the only machine that turns a bound on g'' plus smallness of g into smallness of g' is Taylor's theorem on a short interval.*

> *$g(x)/x^r \to 0$ gives me a different ε at each point. I need one number that controls g at both x and $x+h$, so I take the supremum up to $2x$ and make it monotone.*

Layer 3 THE WHY

Why is balancing the step length the natural move and not a trick? Because of what a first derivative *is* relative to the function and the second derivative. g' sits halfway between g and g''. On an interval of length h, a function of size A and second derivative of size B has a derivative bounded by roughly A/h (difference quotient) plus Bh (how far the slope can drift). The derivative is then pinned by $\sqrt{AB}$ — this is the Landau–Kolmogorov interpolation inequality, $\|g'\|^2 \le C\,\|g\|\,\|g''\|$. The problem has been built so that this geometric mean is small: near x, the size of g is $o(x^r)$ and the size of g'' is $O(x^{-r})$, and their product is $o(1)$. The matching exponents r and $-r$ in the two hypotheses are the cue. Read them together and the answer is on the page: $\sqrt{o(x^r)\cdot O(x^{-r})} = o(1)$.

That also explains both walls in the natural attempt. $h = x$ ignores the interpolation and sits far from the optimum. $h = x^r$ is the optimum for the *worst-case* size x^r of g, so it only yields $|g'| \lesssim \sqrt{x^r \cdot Mx^{-r}} = O(1)$. To get $o(1)$ you must feed in the *actual* size εx^r, and that shrinks the optimal step by $\sqrt{\varepsilon}$.

Where does $r > 1$ enter? Only in keeping the step inside the region where the hypotheses are comparable: we need $h \le x$ so that $x + h \le 2x$ (to bound $g(x+h)$ by $(2x)^r$) and $\xi \ge x$ (to bound $g''(\xi)$ by x^{-r}). With $h = x^r\sqrt{\varepsilon}$ that requires $x^{r-1}\sqrt{\varepsilon} \le 1$, which holds near 0 because $r > 1$ (in fact $r \ge 1$ would do for this argument).

> **Caveat.** I have not investigated whether the dichotomy survives for $0 < r < 1$, where the optimal step $x^r\sqrt{\varepsilon}$ can exceed x and the argument above breaks; nothing in this analysis should be read as a claim either way about that range.

Kedlaya's first solution (Derksen's) introduces an auxiliary nondecreasing h "inserted between $g(x)$ and x^r", namely $h(x) = x^r\sqrt{f(x)}$ with $f(x) = \sup_{z<x} |z^{-r}g(z)|$. Here is where that comes from: it is precisely the optimal Taylor step $\sqrt{A/B}$ with $A = f(x)x^r$ and $B = x^{-r}$, written as a function of x. The "inserted function" is the geometric mean of the two scales, which is what balancing produces. Solution 2 argues by contradiction with a step $\epsilon_0 x_n^r/(2M)$; that is the same balance run the other way, with the lower bound on $|g'|$ playing the role of $\sqrt{\varepsilon}$.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the pairing of x^r in one hypothesis with x^r multiplying g'' in the other. Write down: size of $g \approx \varepsilon x^r$, size of $g'' \approx Mx^{-r}$, and g' should be about $\sqrt{\varepsilon x^r \cdot Mx^{-r}} = \sqrt{\varepsilon M}$.
2. Contrapositive, then Taylor's theorem with Lagrange remainder on $[x, x+h]$, then optimise h.
3. Hesitation occurs when $h = x^r$ gives a bounded but non-vanishing error. What reinforces it away is the $A/h + Bh$ reflex: never fix a free parameter by eye when the bound is a sum of a decreasing and an increasing term — equate them.
4. Two traps. (i) Using the pointwise $o(x^r)$ with a different, unnamed ε at x and at $x+h$; the fix is the monotone supremum $\varepsilon(y)$. (ii) Forgetting that ξ is unknown, so $g''(\xi)$ must be bounded uniformly over $[x, x+h]$; stepping to the right ($\xi > x$) makes $\xi^{-r} \le x^{-r}$ free, whereas stepping left needs $h \le x/2$ to keep ξ comparable to x.
5. Recognise faster via the slogan: *a bound on f and a bound on f'' together bound f' by their geometric mean, via Taylor with an optimised step.* It appears in proofs that $f \to 0$ and f'' bounded imply $f' \to 0$ at infinity, in Landau's inequality, and in any "derivative of a small, not-too-curved function is small" statement.

6. Decision rule: **when an estimate contains a free parameter and splits as (term decreasing in the parameter) + (term increasing in it), choose the parameter to make the two terms equal.** Corollary: when you have $o(\cdot)$ information at several points, replace it by one monotone modulus before estimating.

- ▷ **Cue:** "prove either P or Q" $\longrightarrow$ **Move:** assume $\neg Q$ and prove P; the negation of an unboundedness condition hands you an explicit bound M
- ▷ **Cue:** smallness of f and a bound on f'', asked about f' $\longrightarrow$ **Move:** Taylor on $[x, x+h]$: $|f'(x)| \le 2\sup|f|/h + \frac{h}{2}\sup|f''|$, then choose $h = \sqrt{\sup|f|/\sup|f''|}$
- ▷ **Cue:** a bound of the form $A/h + Bh$ with h free $\longrightarrow$ **Move:** take $h = \sqrt{A/B}$, giving $2\sqrt{AB}$; check that this h is admissible (here $h \le x$)
- ▷ **Cue:** $f(z) = o(\phi(z))$ must be used at two or more nearby points $\longrightarrow$ **Move:** define $\varepsilon(y) = \sup_{z \le y} |f(z)|/\phi(z)$, which is monotone and tends to 0
- ▷ **Cue:** hypotheses with matching powers x^r and x^{-r} on quantities two derivatives apart $\longrightarrow$ **Move:** their product is scale-free; the intermediate derivative is controlled by the geometric mean

Putnam 2019 — Paper B

2019 B1

2019 B1. Denote by $\mathbb{Z}^2$ the set of all points (x, y) in the plane with integer coordinates. For each integer $n \geq 0$, let P_n be the subset of $\mathbb{Z}^2$ consisting of the point $(0, 0)$ together with all points (x, y) such that $x^2 + y^2 = 2^k$ for some integer $k \leq n$. Determine, as a function of n, the number of four-point subsets of P_n whose elements are the vertices of a square.

Fields: lattice geometry, elementary number theory **Position:** B1, first problem of the paper **Answer:** $5n + 1$

Layer 1 THE WHAT — surface analysis

The natural attempt. Start by finding P_n for small n directly. $P_0 = \{(0,0), (\pm 1, 0), (0, \pm 1)\}$: five points, and the only square you can make is the diamond through the four unit points, so the count at $n = 0$ is 1. Add $k = 1$: $x^2 + y^2 = 2$ gives $(\pm 1, \pm 1)$, nine points now, and a second square appears, the diamond through the four new points. Add $k = 2$: $x^2 + y^2 = 4$ gives $(\pm 2, 0), (0, \pm 2)$; now there is a third axis-aligned diamond, but also new squares that mix layers — for instance $(0,0), (2,0), (2,2), (0,2)$ is not a square with these vertices in general position, but $(1,1), (2,0), (1,-1), (0,0)$ *is* a square (check: side length $\sqrt{2}$, consecutive sides perpendicular). Counting by hand at $n = 0, 1, 2$ gives $1, 6, 11$ — consistent with a guess of $5n + 1$ — but this is pattern-spotting from three data points, not a proof, and every point added at level k creates new candidate squares with *every* existing point, so the number of pairs to check grows quadratically in $|P_n|$ while the true count grows linearly. You would need to either verify $5n + 1$ by strong induction with an explicit description of what changes from $n - 1$ to n (itself requiring the classification you were trying to avoid), or search all $\binom{|P_n|}{2}$ pairs as candidate diagonals for every n — neither is a finish, only a very well-supported conjecture.

The model chain. First pin down P_n exactly. Every perfect square is $\equiv 0$ or $1 \pmod 4$, so if $x^2 + y^2 \equiv 0 \pmod 4$ then x, y are both even or both odd; both-odd gives $x^2 + y^2 \equiv 2 \pmod 4$, impossible, so both are even. Hence for $k \geq 2$, $x^2 + y^2 = 2^k$ forces $(x/2, y/2)$ to satisfy $(x/2)^2 + (y/2)^2 = 2^{k-2}$ — a point of the *same shape* two levels down. Checking $k = 0, 1$ directly ($x^2 + y^2 = 1$ gives the four unit points; $x^2 + y^2 = 2$ gives $(\pm 1, \pm 1)$) and inducting on this halving relation gives, for every n,

$$P_n = \{(0,0)\} \cup Q_{\lfloor n/2 \rfloor} \cup R_{\lfloor (n-1)/2 \rfloor},$$

where $Q_m = \{(0, \pm 2^m), (\pm 2^m, 0)\}$ (the axis points, one full set of four per even level $k = 2m \leq n$) and $R_m = \{(\pm 2^m, \pm 2^m)\}$ (the diagonal points, one full set of four per odd level $k = 2m + 1 \leq n$). Direct count: $|P_n| = 1 + 4(\lfloor n/2 \rfloor + 1) + 4\lfloor (n+1)/2 \rfloor = 1 + 4(n+1) = 4n + 5$, using $\lfloor n/2 \rfloor + \lfloor (n+1)/2 \rfloor = n$.

Now classify squares not by searching pairs but by picking, for each square, *one* diagonal: if (a, b) and (c, d) are opposite vertices, the other two vertices are forced by rotating the diagonal 90° about its midpoint,

$$\left(\tfrac{a-b+c+d}{2}, \tfrac{a+b-c+d}{2}\right), \qquad \left(\tfrac{a+b+c-d}{2}, \tfrac{-a+b+c+d}{2}\right).$$

So "does this pair of points extend to a square inside P_n" is now a single arithmetic check on the pair's coordinates, not a geometric search. Sort by which of the three types (origin, some Q_m, some R_m) each of $(a,b),(c,d)$ belongs to.

Diagonal from the origin. Take $(a,b) = (0,0)$; the formula gives other vertices $\left(\frac{c+d}{2}, \frac{-c+d}{2}\right)$ and $\left(\frac{c-d}{2}, \frac{c+d}{2}\right)$. I checked directly: if $(c,d) \in Q_0 = \{(\pm 1, 0), (0, \pm 1)\}$ these come out non-integer (e.g. $(c,d) = (1,0)$ gives $(0.5, -0.5)$), so Q_0 never pairs with the origin. But if $(c,d) \in R_0 = \{(\pm 1, \pm 1)\}$, say $(1,1)$, the other vertices are $(1,0),(0,1)$ — both in $Q_0 \subset P_n$ for $n \geq 1$ — a genuine square. And if $(c,d) \in Q_m$ for $m \geq 1$, say $(2^m, 0)$, the other vertices are $(2^{m-1}, -2^{m-1}), (2^{m-1}, 2^{m-1})$, which lie in $R_{m-1} \subset P_n$ whenever $k = 2m-1 \leq n$, i.e. whenever $2m \leq n+1$, which holds since $2m \leq n$ was already required. So exactly the nonzero points *outside* Q_0 pair validly with the origin: that is $(4n+5-1) - 4 = 4n$ points, each yielding one distinct square (the diagonal $(0,0)$–(c,d) determines its square uniquely, and distinct (c,d) give distinct diagonals). This matches $|P_n| - 5 = 4n$ exactly, and I verified it against the $m = 1$ instance by hand above.

Both endpoints in the same Q_m. Take $(a,b) = (0, 2^m), (c,d) = (0, -2^m)$: the formula returns $(2^m, 0), (-2^m, 0)$ — the whole layer Q_m forms one square, for each $m = 0, \ldots, \lfloor n/2 \rfloor$: that is $\lfloor n/2 \rfloor + 1$ squares, one per axis layer.

Both endpoints in the same R_m. Take $(a,b) = (2^m, 2^m), (c,d) = (-2^m, -2^m)$: the formula returns $(2^m, -2^m), (-2^m, 2^m)$ — the whole layer R_m forms one square, for each $m = 0, \ldots, \lfloor (n-1)/2 \rfloor$: that is $\lfloor (n-1)/2 \rfloor + 1 = \lfloor (n+1)/2 \rfloor$ squares, one per diagonal layer.

> **Caveat.** I have verified directly that these three families are genuine squares and that the origin-diagonal family has exactly $4n$ members (matching the point count exactly, and spot-checked on two instances). I have not independently re-derived, from scratch, the exhaustive argument that *no other* pairing of layers (two different Q layers, two different R layers, or a Q layer mixed with an R layer, neither containing the origin) ever produces a fourth valid square; Kedlaya-Ng rule these out by a short triangle-inequality/parity argument on the derived-vertex coordinates, which I have spot-checked on individual instances but not reverified as a complete case sweep. I present the negative half of the classification on their authority rather than as independently reconstructed.

Summing: $4n + (\lfloor n/2 \rfloor + 1) + \lfloor (n+1)/2 \rfloor = 4n + (n+1) = 5n+1$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree that the answer looks like $5n+1$ after computing a few cases. They split at the question of *how to search*: "each square is a 4-point subset, so check subsets" (natural) versus "each square is determined by one diagonal via a fixed rotation formula, so check ordered pairs by which of finitely many layer-types they belong to" (model). The natural attempt's search space is all $\binom{4n+5}{2}$ pairs of points, growing quadratically with n; the model's search space is a finite, n-independent list of layer-type combinations (origin/Q/R), checked once and for all.

> *I don't need to search all pairs of points in P_n for squares — a square is completely determined by a single diagonal through the explicit rotate-90°-about-the-midpoint formula, so the only question is which combinations of* types *(origin, some Q_m, some R_m) a diagonal's two endpoints can have and still land the other two vertices back in P_n; since P_n has only three types of points regardless of how large n is, this turns an n-growing search into a fixed, finite classification I only have to do once.*

Layer 3 THE WHY

A square is not merely "four points with the right pairwise distances" — it is the orbit of one vertex under repeated 90° rotation about the centre, and the centre is the midpoint of either

diagonal. That means fixing one diagonal $\{(a,b),(c,d)\}$ pins down the entire square through an explicit, computable affine formula; nothing about the other two vertices needs to be searched for, only checked. Once that is seen, the natural unit of work stops being "a 4-point subset of P_n" and becomes "a pair of points of P_n", and the only thing that makes this problem tractable rather than merely well-defined is that P_n's own structure — forced by the mod-4 argument, since 2^k for $k \geq 2$ can only be a sum of two squares when both are even — collapses into finitely many congruence-classes-of-points (origin, one axis-type per even level, one diagonal-type per odd level) independent of n. The rotation formula's output type is then a function only of the two input types, not of which specific level within a type you chose (the origin-to-Q_m computation above works identically for every $m \geq 1$). This is the same structural move as recognising that a problem's search space is secretly finite even though it is presented as ranging over an unbounded parameter: the halving relation is what converts "2^k, k unbounded" into "finitely many residue types, repeating".

Layer $+\alpha$ BACKTESTING

1. The first useful cue is $x^2+y^2 \equiv 0 \pmod 4 \Rightarrow x,y$ both even, the moment $k \geq 2$ appears in the exponent; write immediately that P_n must reduce to P_{n-2} under halving, before attempting any square-count.
2. That cue suggests classifying P_n into finitely many types by parity of k (axis points for even k, diagonal points for odd k), which in turn suggests classifying squares by the *types* of their diagonal's endpoints rather than searching arbitrary point-subsets.
3. Hesitation occurs right after you have a formula for "the other two vertices given a diagonal" and are tempted to plug in specific numbers case by case forever; it is reinforced away by asking "what varies is the level m within a type — does the formula's *output type* actually depend on m, or only on the two input types?" (it does not, once you check one representative of each type).
4. The trap is treating the problem as a search over 4-point subsets, whose count grows with n, when the real classification is over pairs of *types*, which does not.
5. Recognise this family faster by asking, whenever a problem defines a set via a modular/self-similar rule (here, halving under $x^2+y^2=2^k$), whether that rule forces the set into finitely many repeating congruence classes — if so, any subsequent counting problem on that set should be attacked type-by-type, not instance-by-instance.
6. Reusable rule: to count squares (or other rigid figures) inside a structured point set, fix one diagonal (or side) via an explicit rotation formula rather than searching 4-point subsets, then classify by which finitely-many structural "layers" or "types" of the ambient set the two chosen points belong to — the search collapses from $\binom{|S|}{2}$ instances to (number of types)^2 cases.

2019 B2

2019 B2. For all $n \geq 1$, let

$$a_n = \sum_{k=1}^{n-1} \frac{\sin\left(\frac{(2k-1)\pi}{2n}\right)}{\cos^2\left(\frac{(k-1)\pi}{2n}\right)\cos^2\left(\frac{k\pi}{2n}\right)}.$$

Determine

$$\lim_{n\to\infty} \frac{a_n}{n^3}.$$

Fields: trigonometric identities, asymptotic analysis **Position:** B2, second problem of the paper **Answer:** $\frac{8}{\pi^3}$

Layer 1 THE WHAT — surface analysis

The natural attempt. Write $\theta_k = \dfrac{k\pi}{2n}$, so the term is $\dfrac{\sin(\theta_{k-1}+\theta_k)}{\cos^2\theta_{k-1}\cos^2\theta_k}$ (since $\theta_{k-1}+\theta_k = \frac{(2k-1)\pi}{2n}$). The instinct, faced with a sum of $n-1$ terms divided by n^3 and asked for a limit, is to Riemann-sum it: for fixed $k/n = t$, as $n\to\infty$, $\theta_k \to 0$, so approximate $\sin\theta_{k-1}\approx\theta_{k-1}$, $\cos\theta_{k-1}\approx 1-\theta_{k-1}^2/2\approx 1$, giving each term $\approx \theta_{k-1}+\theta_k = \frac{(2k-1)\pi}{2n} = O(k/n)$, and summing $O(k/n)$ over $k=1,\dots,n-1$ gives $O(n)$ — nowhere near n^3. Something in the naive approximation throws away too much: the terms with k near $n-1$ have θ_k near $\pi/2$, where $\cos\theta_k\to 0$ and the small-angle approximation for cos fails badly — those terms blow up and dominate. So you re-substitute $k\mapsto n-k$ (as Kedlaya's second official solution does) to push the singular behaviour to small angles, then Taylor-expand $\sin x/x = 1+O(x^2)$ carefully, track that the error terms stay $o(n^3)$ after summing $O(1/k)$-type residuals over $k=1,\dots,n-1$ (which introduces a $\log n$ that must itself be shown to vanish against n^2), and only then evaluate the leading sum $\sum\frac{8(2k+1)}{k^2(k+1)^2\pi^3}$ — which conveniently telescopes to $1-1/n^2$. This works, but it costs an error-term bookkeeping pass on top of a second telescoping identity you still had to find at the end anyway.

The model chain. Do not approximate; look for exact structure in the numerator first. Apply the double-angle identity $2\cos^2 x = 1+\cos 2x$ to the two cosines in the denominator's arguments:

$$2\cos^2\theta_{k-1} - 2\cos^2\theta_k = \cos 2\theta_{k-1} - \cos 2\theta_k = \cos\left(\tfrac{(k-1)\pi}{n}\right) - \cos\left(\tfrac{k\pi}{n}\right).$$

Apply sum-to-product, $\cos A-\cos B = -2\sin\frac{A+B}{2}\sin\frac{A-B}{2}$, with $A=\frac{(k-1)\pi}{n}$, $B=\frac{k\pi}{n}$: then $\frac{A+B}{2} = \theta_{k-1}+\theta_k = \frac{(2k-1)\pi}{2n}$ (exactly the numerator's angle) and $\frac{A-B}{2} = -\frac{\pi}{2n}$. So

$$2\cos^2\theta_{k-1} - 2\cos^2\theta_k = 2\sin\left(\tfrac{(2k-1)\pi}{2n}\right)\sin\left(\tfrac{\pi}{2n}\right),$$

i.e. $\sin\left(\frac{(2k-1)\pi}{2n}\right) = \dfrac{\cos^2\theta_{k-1}-\cos^2\theta_k}{\sin(\pi/2n)}$. Substituting into the k-th term of a_n:

$$\frac{\sin\left(\frac{(2k-1)\pi}{2n}\right)}{\cos^2\theta_{k-1}\cos^2\theta_k} = \frac{1}{\sin(\pi/2n)}\left(\frac{1}{\cos^2\theta_k} - \frac{1}{\cos^2\theta_{k-1}}\right).$$

This telescopes exactly. Summing $k=1$ to $n-1$:

$$a_n = \frac{1}{\sin(\pi/2n)}\left(\frac{1}{\cos^2\theta_{n-1}} - \frac{1}{\cos^2\theta_0}\right).$$

$\theta_0 = 0$ so $\cos^2\theta_0 = 1$. $\theta_{n-1} = \frac{(n-1)\pi}{2n} = \frac{\pi}{2}-\frac{\pi}{2n}$, so $\cos\theta_{n-1} = \sin(\pi/2n)$, giving $\frac{1}{\cos^2\theta_{n-1}} = \frac{1}{\sin^2(\pi/2n)}$. Hence

$$a_n = \frac{1}{\sin^3(\pi/2n)} - \frac{1}{\sin(\pi/2n)}.$$

This is now an exact closed form, valid for every n, with no error terms anywhere. Only now take the limit: as $n\to\infty$, $\sin(\pi/2n)\sim\pi/2n$, so $\dfrac{1}{\sin^3(\pi/2n)}\sim\dfrac{8n^3}{\pi^3}$, while $\dfrac{1}{\sin(\pi/2n)}\sim\dfrac{2n}{\pi} = o(n^3)$. So $\dfrac{a_n}{n^3}\to\dfrac{8}{\pi^3}$. (I carried out every step of this derivation myself, checked the sum-to-product identity by direct expansion, and confirmed $a_n = \csc^3(\pi/2n)-\csc(\pi/2n)$ is exact by testing $n=2$ directly against the original sum: for $n=2$ the sum has one term, $k=1$, giving $\sin(\pi/4)/\cos^2(0)\cos^2(\pi/4) = \frac{\sqrt{2}/2}{1\cdot 1/2} = \sqrt{2}$; the closed form gives $\csc^3(\pi/4)-\csc(\pi/4) = 2\sqrt{2}-\sqrt{2} = \sqrt{2}$ — matches.)

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains recognise that individual terms blow up near $k = n - 1$ and that some care with small angles is needed. They split at the very first move on the numerator: "approximate sin and cos by their small-angle series and sum the resulting rational approximation, tracking errors" (natural) versus "notice the numerator's angle is exactly the average of the two denominator angles, and test whether that signals an exact telescoping identity before approximating anything" (model). The natural chain reaches a correct answer but only after building an asymptotic estimate with explicit $O(\cdot)$ error control; the model chain never approximates until the single final limit at the very end.

> *The angle in the numerator, $\frac{(2k-1)\pi}{2n}$, is exactly the arithmetic mean of the two angles $\theta_{k-1} = \frac{(k-1)\pi}{2n}$ and $\theta_k = \frac{k\pi}{2n}$ that already appear in the denominator — and an averaged angle appearing next to a difference is precisely the signature of the sum-to-product identity $\cos A - \cos B = -2\sin\frac{A+B}{2}\sin\frac{A-B}{2}$, so before I try to estimate this term asymptotically I should check whether $\cos^2\theta_{k-1} - \cos^2\theta_k$, divided by that same sine, reproduces my numerator exactly — because if it does, the whole sum telescopes and I never need an asymptotic approximation at all.*

Layer 3 THE WHY

A sum telescopes when its k-th term is exactly $f(k) - f(k+1)$ for some function f; recognising that in advance, rather than after exhausting other methods, comes down to reading the term's structure for the fingerprint of a difference identity. Here the fingerprint is specific: the numerator's argument is the *average* of two angles that both appear, squared, in the denominator, and $2\cos^2 x = 1 + \cos 2x$ means "cosine squared at two nearby points" is secretly "cosine at double those points" — exactly the setting where the sum-to-product formula converts a difference of cosines into a product involving the sine of the half-sum (the averaged angle) times the sine of the half-difference (here the constant $\pi/2n$, independent of k — the reason the whole thing telescopes rather than merely simplifying termwise). This is the general principle: whenever a sum's numerator angle equals the mean of two angles that recur elsewhere in the same term, that structural coincidence should be tested for an exact telescoping identity before any asymptotic or Taylor approximation is attempted, because an exact closed form is strictly less work (no error bookkeeping) and gives the precise value of a_n for every n, not merely its leading-order growth.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that the numerator's angle $\frac{(2k-1)\pi}{2n}$ is exactly the arithmetic mean of the two denominator angles θ_{k-1}, θ_k; write down immediately, on seeing this, "test $\cos^2\theta_{k-1} - \cos^2\theta_k$ against a sum-to-product identity" before doing anything asymptotic.
2. That cue suggests the double-angle identity $2\cos^2 x = 1 + \cos 2x$ followed by the sum-to-product identity for a difference of cosines, which together produce the numerator exactly (up to the constant factor $\sin(\pi/2n)$) and hence an exact telescoping sum.
3. Hesitation occurs at the moment the terms near $k = n - 1$ are seen to blow up and the reflex is to reach for small-angle Taylor approximations; it is reinforced away by asking "is there an exact identity that makes this term a difference $f(k) - f(k+1)$, so I never have to approximate at all?"
4. The trap is treating "the answer is only a limit, so approximation is enough" as license to skip looking for an exact closed form — the approximate route works here but forces you to separately re-derive a telescoping identity for the leading-order sum anyway, so you do the same work twice.
5. Recognise this family faster by asking, of any sum whose numerator's argument is an average of two arguments appearing (possibly squared or otherwise transformed) elsewhere in the same term, whether a sum-to-product or difference identity converts the numerator into

exactly that difference — a mean-of-two-angles numerator is close to a universal signal for telescoping.

6. Reusable rule: before estimating a sum asymptotically, check whether its term is an exact difference $f(k) - f(k+1)$ by testing standard trigonometric (or algebraic) difference identities against the term's own structure; an exact telescoping closed form is always preferable to an asymptotic estimate when the term's arguments show a mean/difference pattern, because it needs no error control and answers every finite n, not just the limit.

2019 B3

2019 B3. Let Q be an n-by-n real orthogonal matrix, and let $u \in \mathbb{R}^n$ be a unit column vector (that is, $u^T u = 1$). Let $P = I - 2uu^T$, where I is the n-by-n identity matrix. Show that if 1 is not an eigenvalue of Q, then 1 is an eigenvalue of PQ.

Fields: linear algebra, orthogonal matrices **Position:** B3, third problem of the paper **Answer:** proof: $\det Q = \det(PQ)$ would force a contradiction if neither had eigenvalue 1; hence 1 is an eigenvalue of PQ

Layer 1 THE WHAT — surface analysis

The natural attempt. The problem asks you to show 1 is an eigenvalue of PQ, which superficially means: exhibit a nonzero w with $PQw = w$. So start constructing one. $P = I - 2uu^T$ is the Householder reflection in the hyperplane $u^\perp$: $Pu = -u$, and $Pv = v$ for $v \perp u$. Write $(I - PQ)w = 0$ as $w - Qw + 2u(u^T Qw) = 0$. If $u^T Qw = 0$ this reduces to $w = Qw$, i.e. w is a fixed vector of Q itself — but Q has none, by hypothesis, so $w \neq 0$ forces $u^T Qw \neq 0$ for any candidate. So try the Cayley-transform route: since 1 is not an eigenvalue of Q, $I - Q$ is invertible, and $A = (I - Q)(I + Q)^{-1}$ is skew-symmetric ($A^T = -A$) — a genuinely useful fact, but proving skew-symmetry itself takes a short computation, and you still have to exhibit w: setting $w = (I - A)u$ and checking $(I - QP)w = 0$ (working with QP, which shares PQ's characteristic polynomial) requires first showing $Au \in u^\perp$, via $\langle u, Au\rangle = \langle A^T u, u\rangle = -\langle Au, u\rangle$. This is a genuine, correct route to the answer (it is literally the official second solution), but notice what it costs: an explicit change of variables (the Cayley transform), a lemma that it produces a skew-symmetric matrix, and a specific vector construction whose defining property ($Au \perp u$) has to be checked from the skew-symmetry before you can even verify it lands in the right eigenspace.

The model chain. Step back from constructing a vector: "1 is an eigenvalue of M" is equivalent to $\det(M - I) = 0$, a single number, and determinants of orthogonal matrices are governed by a clean global fact you can establish once and reuse.

Lemma. If M is an $n \times n$ orthogonal matrix with no eigenvalue equal to 1, then $\det M = (-1)^n$. *Proof.* The eigenvalues of M (over $\mathbb{C}$) all have modulus 1, since M is orthogonal ($\|Mv\| = \|v\|$ for eigenvectors, and this extends to the complex eigenvalues via the real characteristic polynomial). Because M is a real matrix, its characteristic polynomial has real coefficients, so non-real eigenvalues occur in conjugate pairs $z, \bar{z}$ with $z\bar{z} = |z|^2 = 1$ — each such pair contributes a factor of exactly 1 to $\det M = \prod(\text{eigenvalues})$. The only eigenvalues that can be real are the fixed points of complex conjugation on the unit circle, namely ± 1. By hypothesis 1 occurs with multiplicity 0; let m be the multiplicity of -1. Then $n = m + 2(\text{number of conjugate pairs})$, so $m \equiv n \pmod 2$. Hence $\det M = (-1)^m = (-1)^n$. ■ (I derived and checked every step of this lemma myself; it uses only that eigenvalues of an orthogonal matrix lie on the unit circle and that non-real roots of a real polynomial pair off.)

Now apply the lemma twice. P is itself orthogonal: $P^T = I - 2uu^T = P$ (symmetric) and $P^2 = I - 4uu^T + 4u(u^T u)u^T = I - 4uu^T + 4uu^T = I$, so $P^T P = P^2 = I$. Its eigenvalues are -1 (eigenvector u, since $Pu = u - 2u(u^T u) = -u$) with multiplicity 1, and $+1$ on the $(n-1)$-dimensional space $u^\perp$; hence $\det P = (-1) \cdot 1^{n-1} = -1$.

Suppose, for contradiction, that PQ also has no eigenvalue 1. PQ is a product of orthogonal matrices, hence orthogonal, so the lemma applies to it too: $\det(PQ) = (-1)^n$. But also Q has no eigenvalue 1 by hypothesis, so the lemma applies to Q: $\det Q = (-1)^n$. Then

$$\det(PQ) = \det P \cdot \det Q = (-1) \cdot (-1)^n = (-1)^{n+1} \neq (-1)^n,$$

contradicting $\det(PQ) = (-1)^n$ from the lemma. So the assumption was false: PQ *must* have 1 as an eigenvalue. No vector is ever exhibited; the proof is a single determinant inequality, $(-1)^n \neq (-1)^{n+1}$, forced by a hypothesis that eliminates one specific eigenvalue from two related matrices.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains use the same starting fact, $Pu = -u$, $Pv = v$ for $v \in u^\perp$. They split immediately after: the natural attempt reads "show 1 is an eigenvalue of PQ" as an instruction to produce an eigenvector, and commits to building one (via the Cayley transform or a direct case-split on $u^T Q w$); the model chain reads the same sentence as "show $\det(PQ - I) = 0$", a scalar statement, and looks for a global invariant — the determinant — that the hypothesis "Q has no eigenvalue 1" already constrains.

> *The problem only asks me to show that* some *nonzero vector is fixed by PQ — that is exactly the statement* $\det(PQ - I) = 0$ *— and I don't have to build that vector by hand to prove a determinant vanishes; determinants are a single global number I can pin down for an entire orthogonal matrix at once, from the cheap fact that its eigenvalues sit on the unit circle and non-real ones pair up, so before hunting for an explicit fixed vector I should ask what the hypothesis "no eigenvalue* 1*" forces the determinant to be, for both Q and PQ.*

Layer 3 THE WHY

The phrase "1 is an eigenvalue of M" and the phrase "$\det(M - I) = 0$" say the same thing, but they invite different kinds of proof: the first invites you to construct a witness (a vector), the second invites you to compute a number. For a general matrix that number is expensive to get at without eigenvectors — but M orthogonal is exactly the situation where the number is cheap, because orthogonality forces every eigenvalue onto the unit circle, and a real characteristic polynomial forces non-real eigenvalues into conjugate pairs whose product is always 1. That leaves only ± 1 as the eigenvalues capable of being "left over" and affecting the determinant; if the hypothesis excludes $+1$ entirely, the determinant's sign is pinned to the parity of n through -1's multiplicity alone, with no freedom left. This is the general principle: when a hypothesis says a specific eigenvalue is *absent*, ask first what global scalar invariant (determinant, trace, or similar) that absence pins down, because "eigenvalue absent" is a strong global constraint precisely for matrices (orthogonal, unitary, permutation, and similar) whose spectra are already constrained to a small, symmetric set — and only reach for an explicit eigenvector construction once that global route has been checked and found wanting.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that Q is orthogonal and a specific eigenvalue (1) is hypothesised absent; write down immediately "orthogonal $\Rightarrow$ eigenvalues on the unit circle, non-real ones pair up, so 'no eigenvalue 1' should pin down $\det Q$" before attempting to construct any vector.
2. That cue suggests proving a lemma (absence of eigenvalue 1 in an orthogonal matrix forces $\det = (-1)^n$) and applying it to both Q and PQ, reducing the whole problem to comparing $\det P \cdot \det Q$ against $(-1)^n$.
3. Hesitation occurs at "show 1 is an eigenvalue" being read literally as "construct an eigenvector"; it is reinforced away by rewriting the target as $\det(PQ - I) = 0$ and asking which global invariant of PQ that statement actually depends on.

4. The trap is the word "eigenvalue" itself, which primes a search for eigenvectors even when the cheaper route is a scalar (determinant) argument; a hypothesis phrased as an eigenvalue's absence is often more efficiently used via the constraint it puts on a global invariant than via the eigenspace it excludes.

5. Recognise this family faster by asking, whenever a hypothesis excludes one specific eigenvalue from a matrix with a constrained spectrum (orthogonal: unit circle; permutation: roots of unity; stochastic: modulus ≤ 1), what that exclusion forces about the determinant, trace, or other symmetric function of the eigenvalues, before reaching for an explicit eigenvector.

6. Reusable rule: to prove a specific eigenvalue λ_0 *must* occur in a matrix from a constrained family (orthogonal, unitary, etc.), first try proving instead that its absence would force a global invariant (determinant is the standard one) to a value contradicted by a companion fact about a related matrix in the problem — this needs no eigenvector construction and often reduces to one inequality between two signs or magnitudes.

2019 B4

2019 B4. Let $\mathcal{F}$ be the set of functions $f(x,y)$ that are twice continuously differentiable for $x \geq 1$, $y \geq 1$ and that satisfy the following two equations (where subscripts denote partial derivatives):

$$xf_x + yf_y = xy\ln(xy),$$
$$x^2 f_{xx} + y^2 f_{yy} = xy.$$

For each $f \in \mathcal{F}$, let

$$m(f) = \min_{s \geq 1} \left(f(s+1,s+1) - f(s+1,s) - f(s,s+1) + f(s,s)\right).$$

Determine $m(f)$, and show that it is independent of the choice of f.

Fields: analysis (partial differential equations, Euler operators); multivariable calculus **Position:** B4 — mid-paper, and much softer than its position suggests once you read what the question is actually asking for **Answer:** $m(f) = 2\ln 2 - \frac{1}{2}$ for every $f \in \mathcal{F}$, attained at $s = 1$.

Layer 1 THE WHAT — surface analysis

The natural attempt. Two PDEs, one unknown function: the reflex is to solve them. That is a perfectly respectable instinct and it does work, so let us carry it through honestly.

Equation (1), $xf_x + yf_y = xy\ln(xy)$, is first order, and the operator $x\partial_x + y\partial_y$ is the derivative along rays from the origin: if $\gamma(t) = (tx_0, ty_0)$ then $\frac{d}{dt} f(\gamma(t)) = \frac{1}{t}\left(xf_x + yf_y\right)$. So along a ray, with $P = x_0 y_0$,

$$t\frac{d}{dt} f(tx_0, ty_0) = t^2 P \ln(t^2 P).$$

You spot (or verify by differentiating) that $h(xy)$ with $h(u) = \frac{1}{2}(u\ln u - u)$ does the job, since $x\partial_x h(xy) + y\partial_y h(xy) = 2xy\,h'(xy) = xy\ln(xy)$. Everything else in the kernel of $x\partial_x + y\partial_y$ is homogeneous of degree 0, so the general solution of (1) is

$$f(x,y) = \tfrac{1}{2}\big(xy\ln(xy) - xy\big) + \varphi(y/x)$$

for a C^2 function φ. Now feed this into (2). The particular part contributes $x^2 \cdot \frac{y}{2x} + y^2 \cdot \frac{x}{2y} = xy$, exactly the right-hand side, so $g = \varphi(y/x)$ must satisfy $x^2 g_{xx} + y^2 g_{yy} = 0$. With $r = y/x$ a two-line chain-rule computation gives

$$x^2 g_{xx} + y^2 g_{yy} = 2r\varphi'(r) + 2r^2\varphi''(r) = 2r\big(r\varphi'(r)\big)',$$

so $r\varphi' = a$ and $\varphi = a\ln r + b$. Hence $\mathcal{F} = \{\frac{1}{2}(xy\ln(xy) - xy) + a\ln(y/x) + b\}$. Plug into the four-point expression: the $a\ln(y/x) + b$ part cancels (check it: $\ln 1 - \ln\frac{s}{s+1} - \ln\frac{s+1}{s} + \ln 1 = 0$), and the rest is a one-variable function of s that you then minimise by differentiating.

It works. It is also about three times longer than it needs to be, and it involves a guess ($h(xy)$), a characteristic argument that needs care about the domain $x, y \geq 1$ (every ray from the origin that meets the quadrant does meet it in a half-line, fine, but you have to say so), and an ODE. On a Putnam clock that is where the careless sign creeps in.

The model chain. Write $E = x\partial_x + y\partial_y$, and label the equations (1) $Ef = xy\ln(xy)$ and (2) $x^2 f_{xx} + y^2 f_{yy} = xy$.

Step 1: the equations determine f_{xy}. Apply E to (1). On the left, using $f \in C^2$ so that $f_{xy} = f_{yx}$,

$$E(Ef) = x(f_x + xf_{xx} + yf_{xy}) + y(xf_{xy} + f_y + yf_{yy}) = Ef + (x^2 f_{xx} + y^2 f_{yy}) + 2xyf_{xy}.$$

On the right, $E(xy\ln(xy)) = x(y\ln(xy) + y) + y(x\ln(xy) + x) = 2xy\ln(xy) + 2xy$. Substituting (1) and (2) for the first two terms on the left,

$$xy\ln(xy) + xy + 2xyf_{xy} = 2xy\ln(xy) + 2xy, \qquad \text{so} \qquad f_{xy} = \tfrac{1}{2}\big(\ln x + \ln y + 1\big).$$

Step 2: the four-point expression is the integral of f_{xy}. Let $\Delta(s) = f(s+1, s+1) - f(s+1, s) - f(s, s+1) + f(s, s)$. By the fundamental theorem of calculus twice (legitimate because f_{xy} is continuous),

$$\Delta(s) = \int_s^{s+1} \big(f_x(x, s+1) - f_x(x, s)\big)\,dx = \int_s^{s+1}\int_s^{s+1} f_{xy}(x, y)\,dy\,dx.$$

Using Step 1 and the symmetry of the square in x and y,

$$\Delta(s) = \tfrac{1}{2} + \tfrac{1}{2}\int_s^{s+1} \ln x\,dx + \tfrac{1}{2}\int_s^{s+1} \ln y\,dy = \tfrac{1}{2} + \int_s^{s+1} \ln x\,dx.$$

This depends on nothing about f except that $f \in \mathcal{F}$.

Step 3: minimise. $\Delta'(s) = \ln(s+1) - \ln s > 0$, so Δ is strictly increasing on $[1, \infty)$ and the minimum is at $s = 1$:

$$m(f) = \tfrac{1}{2} + \int_1^2 \ln x\,dx = \tfrac{1}{2} + \big[x\ln x - x\big]_1^2 = \tfrac{1}{2} + (2\ln 2 - 1) = 2\ln 2 - \tfrac{1}{2}.$$

Step 4: $\mathcal{F} \neq \varnothing$. The minimum is only meaningful if some f exists. $f_0 = \frac{1}{2}(xy\ln(xy) - xy)$ works: $xf_{0,x} = yf_{0,y} = \frac{1}{2}xy\ln(xy)$ and $x^2 f_{0,xx} = y^2 f_{0,yy} = \frac{1}{2}xy$. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths read the same two equations. They split at the first thing written after reading the *question*, not the equations.

(N) "Find all f satisfying (1) and (2)" — the question is treated as "solve the system, then evaluate".

(M) "The quantity $f(s+1, s+1) - f(s+1, s) - f(s, s+1) + f(s, s)$ is $\iint f_{xy}$, so I only need f_{xy}" — the question is read first, and it names exactly one derivative.

From that point (M) has a single target, f_{xy}, and asks which combination of (1), (2) and their derivatives produces it. (N) has the whole of f as its target and has to earn information it will then throw away.

> *The thing I am asked to minimise is a second difference in both variables at once — plus, minus, minus, plus on the corners of a square. That is the discrete version of $\partial_x\partial_y$, and it is exactly $\iint f_{xy}$ over the square. So I do not need f. I need f_{xy}, and the question has told me so.*

> *The problem says "show that it is independent of the choice of f". So $\mathcal{F}$ has more than one element and the four-point expression kills the difference between any two of them. A mixed second difference kills anything of the form $u(x)+v(y)$. That is a strong hint that the equations pin down f_{xy} and leave $u(x)+v(y)$ free.*

> *The equations are in the operator $E = x\partial_x + y\partial_y$ and in $x^2\partial_{xx} + y^2\partial_{yy}$. The cross term $xy\partial_{xy}$ is missing from the second one, and E^2 contains it: $E^2 = x^2\partial_{xx} + y^2\partial_{yy} + 2xy\partial_{xy} + E$. So applying E to equation (1) is the one move that produces the missing cross term.*

- ▷ **Cue:** a quantity of the form $f(b,d) - f(b,c) - f(a,d) + f(a,c)$ $\longrightarrow$ **Move:** recognise it as $\int_a^b \int_c^d f_{xy}$; it annihilates every $u(x)+v(y)$, so aim only at f_{xy}
- ▷ **Cue:** "show that the answer is independent of the choice of f" where f ranges over solutions of equations $\longrightarrow$ **Move:** identify the ambiguity in the solution set and the functional that kills it; compute only what the functional sees
- ▷ **Cue:** equations built from $x\partial_x$, $y\partial_y$ (Euler operators) $\longrightarrow$ **Move:** switch mentally to $u = \ln x$, $v = \ln y$ where $x\partial_x = \partial_u$; the equations become constant-coefficient and squaring the operator is legitimate
- ▷ **Cue:** you have Pf and Qf and need Rf, where R is a missing cross term of P^2 $\longrightarrow$ **Move:** apply P to the equation for Pf and subtract Qf and the lower-order terms

Layer 3 THE WHY

Why is the model chain natural rather than lucky? Two separate "of course"s.

First: the target is named by the question. The expression $f(s+1,s+1) - f(s+1,s) - f(s,s+1) + f(s,s)$ is not a random four-term combination. Write it as $\big[f(s+1,s+1) - f(s,s+1)\big] - \big[f(s+1,s) - f(s,s)\big]$: a difference in x taken at two heights, then differenced in y. A single difference in x is $\int f_x\,dx$ by the fundamental theorem of calculus; differencing that in y is $\iint f_{xy}$. That is the definition of the fundamental theorem, applied twice, nothing more. Once it is written down, the problem is no longer "find f", it is "find f_{xy}", and a Putnam problem that asks you to prove independence from f is telling you that the full f is not determined — which is exactly why hunting for the full f is the wrong target.

Second: the operators tell you how to get f_{xy}. What does $x\partial_x$ *mean*? Put $x = e^u$. Then $\frac{\partial}{\partial u} = \frac{dx}{du}\frac{\partial}{\partial x} = x\partial_x$. So $x\partial_x$ is plain differentiation in the logarithmic variable, and equation (1) says $(\partial_u + \partial_v)f = e^{u+v}(u+v)$. Also $x^2\partial_{xx} = \partial_u^2 - \partial_u$ (because $\partial_u^2 = x\partial_x(x\partial_x) = x^2\partial_{xx} + x\partial_x$), and $xy\partial_{xy} = \partial_u\partial_v$. In these coordinates the system reads

$$(\partial_u + \partial_v)f = e^w w, \qquad (\partial_u^2 + \partial_v^2)f - (\partial_u + \partial_v)f = e^w, \qquad w = u + v,$$

and the target is $\partial_u\partial_v f$. Now it is A-level algebra: $(\partial_u + \partial_v)^2 = \partial_u^2 + \partial_v^2 + 2\partial_u\partial_v$. You know the first-order sum, so you know its square (apply it again); you know $\partial_u^2 + \partial_v^2$; subtract, halve. This is the identity $2ab = (a+b)^2 - (a^2+b^2)$, with a, b the two derivatives. Kedlaya's first solution writes the combination as "$x\frac{\partial}{\partial x}(1) + y\frac{\partial}{\partial y}(1) - (1) - (2)$" without comment. Here is where that comes from: it is $2ab = (a+b)^2 - a^2 - b^2$ in logarithmic coordinates, with the $-(1)$ accounting for the $-\partial_u$ in $x^2\partial_{xx} = \partial_u^2 - \partial_u$.

That is also why the answer is independent of f in the deepest sense. In (u,v) coordinates the homogeneous system is $(\partial_u + \partial_v)g = 0$, $(\partial_u^2 + \partial_v^2)g = 0$, whose solutions are $g = a(u-v) + b$ — a function of the form $u(x) + v(y)$, precisely what the mixed difference annihilates. The

problem-setter chose the functional to be blind to exactly the freedom the equations leave. Recognising that is recognising the design.

Finally, the minimisation. $\Delta(s) = \frac{1}{2} + \int_s^{s+1} \ln x \, dx$ is an average of an increasing function over a window of fixed width sliding right, so it increases; the minimum sits at the left edge of the domain. The domain restriction $x, y \geq 1$ in the statement is there to make $s = 1$ the answer.

Caveat. Verified symbolically: $f_0 = \frac{1}{2}(xy\ln(xy) - xy)$ and $f_0 + 3\ln(y/x) + 7$ both satisfy (1) and (2); for the latter $\Delta'(s) = \ln\frac{s+1}{s}$ and $\Delta(1) = 0.886294\ldots = 2\ln 2 - \frac{1}{2}$. The derivation of f_{xy} uses $f_{xy} = f_{yx}$, which needs $f \in C^2$ — supplied by the hypothesis. As Kedlaya–Ng remark, existence of some $f \in \mathcal{F}$ is logically required for $m(f)$ to mean anything; Step 4 supplies it.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the plus-minus-minus-plus pattern on the four corners of the unit square $[s, s+1]^2$. On seeing it, write $\Delta(s) = \int_s^{s+1}\int_s^{s+1} f_{xy}\, dy\, dx$ and underline f_{xy} as the only unknown you need.

2. The cue suggests elimination rather than solution: combine the given equations, and derivatives of them, to isolate f_{xy}. Because both equations are written in Euler operators, the combination is found by treating $x\partial_x, y\partial_y$ as commuting symbols a, b and using $2ab = (a+b)^2 - a^2 - b^2$, with the correction $x^2\partial_{xx} = (x\partial_x)^2 - x\partial_x$.

3. Hesitation occurs at "can I differentiate equation (1)?" — students worry that differentiating an equation loses information or needs more smoothness. It doesn't: (1) holds identically on an open set, so its derivatives hold identically too, and $f \in C^2$ is exactly what is needed to differentiate once. Second hesitation: whether Δ is minimised at $s = 1$ or somewhere interior. Compute $\Delta'(s) = \ln(s+1) - \ln s$ and the hesitation is gone.

4. The trap is zombie maths: solving the PDE system completely by characteristics, guessing $h(xy)$, and then solving an ODE for the homogeneous part. Every step of it is correct and most of it is wasted, since the answer only needs f_{xy}. The subtler trap is forgetting the $+Ef$ term in $E^2 = x^2\partial_{xx} + y^2\partial_{yy} + 2xy\partial_{xy} + E$; drop it and you get $f_{xy} = \ln(xy) + \frac{1}{2}$ and the answer $4\ln 2 - \frac{3}{2}$, which looks as plausible as the right one.

5. Read the object the question asks for before reading the hypotheses as a to-do list. If the asked-for quantity is a linear functional of the unknown (a difference, an integral, a coefficient, a trace), work out which derivative or component that functional actually sees, and make that — not the unknown — your target. "Independent of the choice of" is a flashing sign that the unknown is underdetermined and the functional is designed to be blind to the freedom.

6. Reusable decision rule: *when you are asked for a functional of a solution, not the solution, find the smallest piece of the solution the functional depends on and derive only that.* For derivatives the elimination tool is algebra in the operators themselves — Euler operators $x\partial_x$ are ordinary derivatives in $\ln x$, and polynomial identities like $2ab = (a+b)^2 - a^2 - b^2$ hold for commuting operators, so you can build the missing derivative out of the ones you are given.

2019 B5

2019 B5. Let F_m be the mth Fibonacci number, defined by $F_1 = F_2 = 1$ and $F_m = F_{m-1} + F_{m-2}$ for all $m \geq 3$. Let $p(x)$ be the polynomial of degree 1008 such that $p(2n+1) = F_{2n+1}$ for $n = 0, 1, 2, \ldots, 1008$. Find integers j and k such that $p(2019) = F_j - F_k$.

Fields: algebra (polynomial interpolation, finite differences); recurrences **Position:** B5 — hard for anyone who reaches for Lagrange, a two-line problem for anyone who reaches for differences **An-**

swer: $p(2019) = F_{2019} - F_{1010}$, i.e. $(j,k) = (2019, 1010)$; in general a degree-N interpolant through $F_1, F_3, \ldots, F_{2N+1}$ gives $F_{2N+3} - F_{N+2}$ at the next odd point.

Layer 1 THE WHAT — surface analysis

The natural attempt. You are told a polynomial through 1009 points and asked for its value at the 1010th. The reflex is Lagrange interpolation, and before committing you sensibly test small cases. Let N be the degree, so p passes through $(1, F_1), (3, F_3), \ldots, (2N+1, F_{2N+1})$ and you want $p(2N+3)$.

- $N = 0$: $p \equiv 1$, $p(3) = 1 = F_3 - F_2$.
- $N = 1$: the line through $(1,1), (3,2)$ gives $p(5) = 3 = F_5 - F_3$.
- $N = 2$: through $(1,1), (3,2), (5,5)$: first differences $1, 3$, second difference 2, so the next first difference is 5 and $p(7) = 10 = F_7 - F_4$.

The pattern $p(2N+3) = F_{2N+3} - F_{N+2}$ is visible, giving $F_{2019} - F_{1010}$ at $N = 1008$. Now prove it. Lagrange at the equally spaced nodes $x_i = 2i+1$, evaluated at $x = 2N+3$, gives

$$\prod_{j \neq i} \frac{(2N+3) - (2j+1)}{(2i+1) - (2j+1)} = \prod_{j \neq i} \frac{N+1-j}{i-j} = (-1)^{N-i} \binom{N+1}{i},$$

so

$$p(2N+3) = \sum_{i=0}^{N} (-1)^{N-i} \binom{N+1}{i} F_{2i+1}.$$

Here most strong students stall. An alternating binomial sum of *odd-indexed* Fibonacci numbers does not look like anything. You can push on with Binet's formula $F_m = (\alpha^m - \beta^m)/\sqrt{5}$, where $\alpha, \beta = \frac{1 \pm \sqrt{5}}{2}$ are the roots of $t^2 = t + 1$: for $\gamma \in \{\alpha, \beta\}$,

$$\sum_{i=0}^{N} (-1)^{N-i} \binom{N+1}{i} \gamma^{2i+1} = -\gamma \Big[(\gamma^2 - 1)^{N+1} - \gamma^{2N+2} \Big] = \gamma^{2N+3} - \gamma^{N+2},$$

using $\gamma^2 - 1 = \gamma$. It lands, and it is Kedlaya–Ng's second solution. But count the steps at which a sign or an index can slip: the Lagrange product, the $(-1)^{N-i}$, the missing $i = N+1$ term of the binomial expansion, the conjugate root. It is a long corridor with no lights on.

The model chain. Put $a(n) = F_{2n+1}$ for all $n \geq 0$ (the true sequence) and $q(n) = p(2n+1)$ (the polynomial, a polynomial in n of degree $N = 1008$). They agree for $n = 0, 1, \ldots, N$. Let Δ be the forward difference, $\Delta b(n) = b(n+1) - b(n)$.

Fact 1 (Fibonacci is its own difference). $\Delta a(n) = F_{2n+3} - F_{2n+1} = F_{2n+2}$ by the recurrence. More generally, for the shifted sequences $a_k(n) = F_{2n+1+k}$ we have $\Delta a_k = a_{k+1}$, since $F_{m+2} - F_m = F_{m+1}$. By induction on k, $\Delta^k a(n) = F_{2n+1+k}$; in particular

$$\Delta^{N+1} a(0) = F_{N+2}.$$

Fact 2 (a degree-N polynomial has vanishing $(N+1)$st difference). Δ lowers the degree of a polynomial by at least one, so $\Delta^{N+1} q \equiv 0$.

Fact 3 (the top coefficient of a difference is 1). Expanding $\Delta = E - 1$ with E the shift, $\Delta^{N+1} b(0) = \sum_{i=0}^{N+1} (-1)^{N+1-i} \binom{N+1}{i} b(i)$; the coefficient of $b(N+1)$ is $+1$.

Conclusion. a and q agree at $0, \ldots, N$, so in $\Delta^{N+1} a(0) - \Delta^{N+1} q(0)$ every term cancels except the last:

$$a(N+1) - q(N+1) = \Delta^{N+1} a(0) - \Delta^{N+1} q(0) = F_{N+2} - 0.$$

Hence $q(N+1) = F_{2N+3} - F_{N+2}$, and at $N = 1008$,

$$p(2019) = F_{2019} - F_{1010}, \qquad (j,k) = (2019, 1010). \qquad \square$$

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree on the setup: a degree-N polynomial through $N+1$ equally spaced points, one more point wanted. They split at the choice of tool.

(N) "Write p by Lagrange interpolation and evaluate at 2019" — a formula for p everywhere, specialised afterwards.

(M) "The $(N+1)$st difference of p at the nodes is 0" — a single linear relation among the $N+2$ values $p(1), \ldots, p(2N+3)$, which is all you need to recover the one unknown.

(M) then asks a second question that (N) never asks: what does the *same* difference operator do to the Fibonacci data? And the answer is that it returns Fibonacci numbers.

> *I only want one new value, and the nodes are equally spaced. I don't need the polynomial; I need the one linear relation that pins down the next value. For a degree-*1008 *polynomial on equally spaced points that relation is: the* 1009*th difference is zero.*

> *What does a step-*2 *difference do to odd-indexed Fibonacci numbers?* $F_{m+2} - F_m = F_{m+1}$. *So differencing the data does nothing but shift the index — the difference table of* $F_1, F_3, F_5, \ldots$ *has* $F_2, F_4, \ldots$ *in the next row, then* $F_3, F_5, \ldots$, *and so on. The diagonal entries are* $F_1, F_2, F_3, \ldots$. *The polynomial and the Fibonacci sequence have identical difference tables except that the polynomial's* $(N+1)$*st row is zero where Fibonacci's is* F_{N+2}.

> *The answer has the shape "true value minus a correction".* F_{2019} *is what the Fibonacci sequence would give at the next point, so the correction* F_{1010} *must be exactly what the polynomial is missing — the one term the degree cap throws away.*

- ▷ **Cue:** a polynomial of degree N specified at $N+1$ equally spaced points, and one more value requested $\longrightarrow$ **Move:** write the $(N+1)$st finite difference and set it to zero; never build the polynomial
- ▷ **Cue:** data given by a linear recurrence, sampled at equally spaced indices $\longrightarrow$ **Move:** compute what the difference operator does to the data; for Fibonacci with step 2, $F_{m+2} - F_m = F_{m+1}$, so differencing shifts the index
- ▷ **Cue:** an answer of the form (true value) − (something) $\longrightarrow$ **Move:** interpret the subtracted term as the error of the approximation, and look for the single place where the approximation and the truth diverge
- ▷ **Cue:** an alternating binomial sum $\sum(-1)^{N-i}\binom{N}{i}b_i$ $\longrightarrow$ **Move:** recognise it as $\Delta^N b(0)$ and ask what Δ does to b

Layer 3 THE WHY

Why is the difference operator the natural tool rather than a trick?

Start from what "a polynomial of degree N" means on equally spaced points. On the integers, the discrete analogue of "the $(N+1)$st derivative vanishes" is "the $(N+1)$st difference vanishes", and it holds for the same reason: $\Delta n^d = (n+1)^d - n^d$ has degree $d-1$. So the set of degree-$\leq N$ polynomials, restricted to equally spaced points, is *exactly* the set of sequences whose $(N+1)$st difference is identically zero. That is not a technique; it is a characterisation. The problem gives you $N+1$ values and asks for the $(N+2)$nd, and the characterisation

supplies precisely one linear equation linking $N+2$ consecutive values. One unknown, one equation. The Lagrange route builds the entire polynomial (infinitely many values) to extract one of them, which is why it is long.

Now why does Fibonacci cooperate with Δ? Because of what the recurrence *says*. $F_{m+2} = F_{m+1} + F_m$ rearranges to $F_{m+2} - F_m = F_{m+1}$: the step-2 difference of the sequence is the sequence itself, shifted by one. That is the whole reason the problem samples at odd arguments $2n+1$. With step 1 you would get $F_{m+1} - F_m = F_{m-1}$ — also a shift, but backwards — and the same method goes through, except that the diagonal of the difference table runs down into negative indices ($F_0 = 0$, $F_{-1} = 1$, $F_{-2} = -1, \ldots$), and the interpolant through $F_1, \ldots, F_{N+1}$ at $1, \ldots, N+1$ gives $F_{N+2} + (-1)^N F_N$ at $N+2$ (check $N = 2$: through $1, 1, 2$ the next value is $4 = F_4 + F_2$). The problem-setter chose step 2 so that the shift is forwards and the diagonal of the difference table reads $F_1, F_2, F_3, \ldots$ with no gaps. In operator language: on Fibonacci sequences the shift E satisfies $E^2 = E + 1$, so $E^2 - 1 = E$; the step-2 difference $E^2 - 1$ *is* the shift E. The characteristic equation $t^2 = t + 1$ is doing all the work, which is also why Binet's formula (built from the roots of that equation) works in the Lagrange route — it is the same fact, seen through the roots instead of through the operator.

Once both facts are in hand the conclusion is forced. The polynomial and Fibonacci share the values at $n = 0, \ldots, N$, so they share the first $N+1$ entries of the diagonal of the difference table: $F_1, \ldots, F_{N+1}$. The next diagonal entry is where they part company: Fibonacci has F_{N+2}, the polynomial has 0. Because the $(N+1)$st difference involves the new value with coefficient exactly $+1$, the gap F_{N+2} passes straight through to the new value. The answer $F_{2019} - F_{1010}$ is "what Fibonacci does at the next point, minus the first diagonal entry the polynomial is forbidden to have". The 1010 is $N + 2 = 1008 + 2$; nothing more mysterious than that.

Kedlaya–Ng's first solution defines $p_k(x) = p_{k-1}(x) - p_{k-1}(x+2)$ and claims $p_k(2n+1) = F_{2n+1+k}$. Here is where that comes from: p_k is the kth difference polynomial, and the claim is Fact 1 above. Their second step, $p_{N-k}(2k+3) = p_{N-k}(2k+1) + p_{N-k+1}(2k+1)$, is reading the difference table back up to reconstruct the new value — the same argument, done row by row instead of in one line.

Caveat. Verified by exact integer arithmetic: for $N = 0, 1, \ldots, 7$ and for $N = 1008$, the value forced by $\Delta^{N+1} q(0) = 0$ equals $F_{2N+3} - F_{N+2}$. Three misprints in the official source, for anyone comparing: (i) in Solution 1 the definition should read $p_k(x) = p_{k-1}(x+2) - p_{k-1}(x)$ — as printed, $p_1(1) = F_1 - F_3 = -1 \neq F_2$, while the rest of their argument uses the corrected sign; (ii) in Solution 2 Binet's formula is printed with β^{-n} for β^n; (iii) the closing remark states the unique solution as $(2019, 2010)$, which should be $(2019, 1010)$. None affects the answer.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "polynomial of degree 1008" next to "$n = 0, 1, \ldots, 1008$" — exactly degree-plus-one equally spaced nodes, and one more point asked for. On seeing it, write $\Delta^{1009} q \equiv 0$ where $q(n) = p(2n+1)$.
2. The cue suggests finite differences: build (mentally) the difference table of the data and read the relation off its diagonal. The second cue, Fibonacci data sampled at step 2, suggests checking Δ on the data first: $F_{m+2} - F_m = F_{m+1}$.
3. Hesitation occurs after the small cases give the conjecture $F_{2N+3} - F_{N+2}$: the solver believes it and cannot see how to prove it without Lagrange. It is reinforced away by writing the $N = 2$ difference table in full — rows $1, 2, 5$; $1, 3$; 2 — and noticing the rows are F_1, F_3, F_5; F_2, F_4; F_3. The proof is that table, in general.
4. The trap is zombie maths via Lagrange: it produces a correct but unfamiliar alternating binomial sum and then needs Binet, a binomial expansion with a missing term, and a conjugate-root argument, with four places to lose a sign. The subtler trap is overfitting to

the small cases: $N = 0$ gives $F_3 - F_2$, which also equals F_1 or F_2, so small cases alone do not fix the form $F_j - F_k$; you need the general structure to be sure of $k = N + 2$.

5. Whenever you are asked to extrapolate one value of a low-degree polynomial fitted to known data, stop thinking about the polynomial and think about the data's difference table. If the data come from a linear recurrence, the difference table will itself be made of the recurrence's terms, and the answer will be "true value minus the first difference the polynomial cannot see".

6. Reusable decision rule: *to get one value, find one equation; do not construct the whole object.* For polynomials on equally spaced points that equation is a vanishing finite difference. For data satisfying a linear recurrence with characteristic polynomial $\chi(E)$, compute what the difference operator becomes modulo χ — here $E^2 - 1 \equiv E$ — and the difference table collapses to the sequence itself.

2019 B6

2019 B6. Let $\mathbb{Z}^n$ be the integer lattice in $\mathbb{R}^n$. Two points in $\mathbb{Z}^n$ are called *neighbors* if they differ by exactly 1 in one coordinate and are equal in all other coordinates. For which integers $n \geq 1$ does there exist a set of points $S \subset \mathbb{Z}^n$ satisfying the following two conditions?

(1) If p is in S, then none of the neighbors of p is in S.

(2) If $p \in \mathbb{Z}^n$ is not in S, then exactly one of the neighbors of p is in S.

Fields: combinatorics (tilings, perfect codes); elementary group theory (homomorphisms $\mathbb{Z}^n \to \mathbb{Z}/m\mathbb{Z}$)
Position: B6 — the last problem of the paper, but one whose difficulty is entirely in choosing the right frame; once the frame is chosen the construction and its verification take a paragraph **Answer:** Such a set exists for every $n \geq 1$: take $S = \{x \in \mathbb{Z}^n : x_1 + 2x_2 + \cdots + nx_n \equiv 0 \pmod{2n+1}\}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. "For which n" on a B6 smells like "only small n" or "only n of a special form", so the natural first move is to try small cases and look for an obstruction.

$n = 1$: neighbours of x are $x \pm 1$. $S = 3\mathbb{Z}$ works: a multiple of 3 has no neighbour that is a multiple of 3, and each non-multiple is adjacent to exactly one multiple ($3k+1$ to $3k$, $3k+2$ to $3k+3$).

$n = 2$: by hand on graph paper, placing a point and then being forced to place the next, you find the knight's-move pattern: S generated by $(1, 2)$ and $(2, -1)$, which you may or may not recognise as $\{x + 2y \equiv 0 \pmod 5\}$. Each point of S with its four neighbours forms a plus-shaped pentomino, and these pentominoes tile the plane.

$n = 3$: now the hand search is expensive, and the tempting idea is induction by layers. Build $S \subset \mathbb{Z}^{n+1}$ by stacking copies of an n-dimensional solution, suitably translated, in the layers $x_{n+1} = t$. Let $S_t \subset \mathbb{Z}^n$ be the layer at height t and suppose every S_t is itself an n-dimensional solution. Take any $p \in \mathbb{Z}^n$ and look at (p, t):

- if $p \in S_t$, condition (1) forbids $(p, t+1) \in S$, so $p \notin S_{t+1}$;
- if $p \notin S_t$, then (p, t) already has exactly one S-neighbour inside its own layer, so condition (2) forbids $(p, t+1) \in S$, so again $p \notin S_{t+1}$.

Hence $S_{t+1} = \varnothing$, which is not an n-dimensional solution. That is the wall, and it is decisive: a cross-section of an $(n+1)$-dimensional solution is *never* an n-dimensional solution, so any inductive construction has to build layers that are individually defective and repair each other. The solver now either sinks time into designing such layers, or — worse — takes the wall as

evidence that no solution exists for $n \geq 3$ and starts hunting for a parity or counting obstruction that is not there.

The model chain. **Reformulation.** For $p \in \mathbb{Z}^n$ let $B(p) = \{p\} \cup \{p \pm e_i : 1 \leq i \leq n\}$, the point with its $2n$ neighbours; $|B(p)| = 2n + 1$. Conditions (1) and (2) say exactly that every point of $\mathbb{Z}^n$ lies in $B(p)$ for exactly one $p \in S$: a point of S lies in its own B and, by (1), in no other; a point outside S lies in $B(p)$ precisely for its S-neighbours p, of which (2) says there is one. So we want translates of the cross $B(0)$ to tile $\mathbb{Z}^n$. Such a tiling has density $\frac{1}{2n+1}$, which suggests looking for S as the kernel of a homomorphism onto a group of order $2n + 1$.

Construction. Let $m = 2n + 1$ and $f(x_1, \ldots, x_n) = x_1 + 2x_2 + \cdots + nx_n \bmod m$, a homomorphism $\mathbb{Z}^n \to \mathbb{Z}/m\mathbb{Z}$. Put $S = f^{-1}(0)$. The key fact is:

> $(*)$ The $2n$ residues $\pm 1, \pm 2, \ldots, \pm n$ are distinct and are exactly the nonzero residues mod m.

They are nonzero since $1 \leq i \leq n < m$. If $i \equiv j$ with $1 \leq i, j \leq n$ then $i = j$, as $|i - j| < m$. If $i \equiv -j$ then $m \mid i + j$, impossible as $2 \leq i + j \leq 2n < m$. So they are $2n$ distinct nonzero residues, and there are exactly $2n$ nonzero residues.

Condition (1). If $p \in S$ and $q = p \pm e_i$, then $f(q) = f(p) \pm i = \pm i \not\equiv 0$, so $q \notin S$.

Condition (2). If $p \notin S$ then $c = f(p) \not\equiv 0$. The neighbour $p + \varepsilon e_i$ ($\varepsilon = \pm 1$) lies in S iff $c + \varepsilon i \equiv 0$, i.e. iff $-\varepsilon i \equiv c$. By $(*)$ there is exactly one pair (i, ε) with $-\varepsilon i \equiv c$. So p has exactly one neighbour in S.

Hence a set S exists for every $n \geq 1$. □

(For $n = 1$ this is $3\mathbb{Z}$; for $n = 2$ it is $\{x + 2y \equiv 0 \pmod 5\}$, the knight's-move lattice from the hand search.)

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths do $n = 1$ and $n = 2$. They split at what is extracted from those cases.

(N) "How do I get from dimension n to dimension $n + 1$?" — the small cases are treated as base cases for an induction.

(M) "What do the small cases have in common as *sets*?" — $3\mathbb{Z}$ and the knight lattice are both kernels of a map to $\mathbb{Z}/(2n + 1)$, and $2n + 1$ is the size of the cross. The small cases are treated as instances of a formula.

(N) runs into the layer wall. (M) never meets it, because a formula in all n coordinates at once does not care about layers.

> *Conditions (1) and (2) together say every point is within distance one of exactly one point of S. So the crosses centred at points of S tile the lattice. I'm not choosing a set, I'm tiling space with a fixed shape of size* $2n + 1$.

> *A tile of size* $2n + 1$ *that tiles by translation — the first thing to try is a* lattice *of translates, and a lattice of index* $2n + 1$ *is the kernel of a homomorphism* $\mathbb{Z}^n \to G$ *with* $|G| = 2n + 1$. *The tiling condition is then that the homomorphism maps the cross bijectively onto G. Try* $G = \mathbb{Z}/(2n + 1)$*: the cross is* 0 *and* $\pm a_1, \ldots, \pm a_n$ *where* a_i *is the image of* e_i*, so I need* $\{\pm a_i\}$ *to be all the nonzero residues. Take* $a_i = i$.

> *The layer argument shows cross-sections are never solutions, so induction on n is the wrong shape. That is not evidence against existence; it is evidence against building the set one coordinate at a time.*

- ▷ **Cue:** "every point is either in S or adjacent to exactly one point of S, and points of S are not adjacent" $\longrightarrow$ **Move:** rewrite as: the closed neighbourhoods of points of S partition the space; it is a tiling (a perfect code)
- ▷ **Cue:** a tiling of $\mathbb{Z}^n$ by translates of a finite tile T $\longrightarrow$ **Move:** look for S as the kernel of a homomorphism $\phi : \mathbb{Z}^n \to G$ with $|G| = |T|$ and $\phi|_T$ a bijection
- ▷ **Cue:** need $2n$ elements that are $\pm$ pairs covering all nonzero residues mod an odd m $\longrightarrow$ **Move:** use $\pm 1, \ldots, \pm\frac{m-1}{2}$; the nonzero residues of an odd modulus split into exactly those $\frac{m-1}{2}$ pairs
- ▷ **Cue:** small cases of a "for which n" problem all work $\longrightarrow$ **Move:** extract a uniform formula from the small cases before attempting any induction on dimension; check whether the inductive step is even compatible with the conditions

Layer 3 THE WHY

Why is the homomorphism frame forced, rather than a lucky guess?

Begin with the definition of the conditions. Condition (1) says the points of S are pairwise at distance at least 2 in the graph metric (no two adjacent), which, since two crosses $B(p), B(p')$ overlap iff p, p' are equal, adjacent, or at distance 2, is not quite enough for disjointness by itself — but condition (2) supplies the rest: a point outside S with two S-neighbours would sit in two crosses, and (2) forbids that. Together they say every point is covered exactly once. That is the *definition* of a partition of $\mathbb{Z}^n$ into crosses, i.e. a tiling. The statement never says the word "tiling", but it is the circle-and-centre situation of the Blueprint: once you see the object is a partition into congruent pieces, the centres of the pieces are the object of interest, and the question is about how the centres can be arranged.

Now ask what the simplest possible arrangement of centres is. $\mathbb{Z}^n$ is a group, and translation by any lattice vector preserves the cross shape. The most symmetric possible set of centres is therefore a *subgroup* L: then the tiling is invariant under every translation in L, and checking it reduces to checking one tile. For a subgroup L, "the translates $B(0) + \ell$, $\ell \in L$, partition $\mathbb{Z}^n$" is word for word the statement "$B(0)$ contains exactly one element of each coset of L", which is "the quotient map $\mathbb{Z}^n \to \mathbb{Z}^n/L$ restricted to $B(0)$ is a bijection". That forces $|\mathbb{Z}^n/L| = 2n + 1$. You are no longer searching for a set; you are searching for a homomorphism onto a group of order $2n + 1$ that is injective on $2n + 1$ specific points. The group $\mathbb{Z}/(2n + 1)$ is the first one to try, and a homomorphism $\mathbb{Z}^n \to \mathbb{Z}/m$ is nothing but a choice of $a_i = \phi(e_i)$, with $\phi(x) = \sum a_i x_i$. Since $\phi(\pm e_i) = \pm a_i$, injectivity on the cross says: $0, \pm a_1, \ldots, \pm a_n$ are all distinct. That is a pairing of the $2n$ nonzero residues into n pairs $\{r, -r\}$ — and for an odd modulus the nonzero residues *already are* paired this way, since $r \neq -r$ when m is odd. Choose one element from each pair. The choice $1, 2, \ldots, n$ is merely the most legible one; any choice of representatives works equally well.

This is exactly Kedlaya–Ng's solution, which states $f = x_1 + 2x_2 + \cdots + nx_n$ without motivation. Here is where it comes from: the coefficients are one representative from each $\pm$ pair of nonzero residues mod $2n + 1$, because the cross has one centre and n opposite pairs of arms, and the tiling condition says the arms must land on distinct residues.

Last, why does the layer induction fail while the formula succeeds? The formula's value on a layer $x_n = t$ is $x_1 + \cdots + (n-1)x_{n-1} + nt \bmod 2n + 1$ — a function valued in $\mathbb{Z}/(2n + 1)$, not in $\mathbb{Z}/(2n - 1)$, so the layer sets are not $(n-1)$-dimensional solutions (those need modulus $2n - 1$). The tiling is intrinsically n-dimensional: the nth pair of arms is needed to fill the residues $\pm n$ that the lower-dimensional crosses cannot reach. The layer wall in the natural attempt is the combinatorial shadow of that arithmetic fact.

Caveat. Verified by exhaustive check on the torus $(\mathbb{Z}/(2n+1))^n$ for $n = 1, \ldots, 5$: every point of S has no S-neighbour and every other point has exactly one. This suffices for all of $\mathbb{Z}^n$, because $f(x + (2n+1)e_i) = f(x) + i(2n+1) \equiv f(x)$, so S and the neighbour relation are periodic with period $2n+1$ in each coordinate. The layer argument in the natural attempt is proved in full there. Context, not needed for the proof: S is a perfect 1-error-correcting code in the Lee metric on $\mathbb{Z}^n$; the question for larger radius is the subject of the Golomb–Welch conjecture, and Kedlaya–Ng note the problem was a 1985 IMO submission.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the pair of conditions read together: "none of the neighbours" for points of S and "exactly one of the neighbours" for points outside. On seeing it write: "the sets $\{p\} \cup \{\text{neighbours of } p\}$, $p \in S$, partition $\mathbb{Z}^n$; each has $2n+1$ points".

2. The cue suggests a lattice tiling, i.e. $S = \ker \phi$ for a homomorphism $\phi : \mathbb{Z}^n \to G$ with $|G| = 2n+1$, bijective on the cross. With $G = \mathbb{Z}/(2n+1)$ the method reduces to choosing $\phi(e_i)$.

3. Hesitation occurs at "for which n" — the phrasing suggests some n fail, so the solver looks for an obstruction. It is reinforced away by the density count: $2n+1$ is always a positive integer, so the counting obstruction that kills many tiling problems (tile size not dividing something) never arises here. Second hesitation: whether the choice $a_i = i$ really gives distinct $\pm a_i$; the three-line proof of $(*)$ settles it.

4. The trap is induction on dimension by stacking layers — natural because the problem is phrased for general n and $n = 1, 2$ fall quickly by hand. The layer argument shows no layer of a solution can itself be a solution, so this path cannot succeed as stated, and a solver who misreads the wall may conclude that no solution exists for $n \geq 3$. The secondary trap is treating $n = 2$ as a picture (a knight's-move pattern) rather than as a formula ($x + 2y \equiv 0 \bmod 5$); the picture does not generalise, the formula does.

5. When a covering condition says "exactly one", rewrite it as a partition and count the piece. A piece of size m that tiles a group by translation suggests a homomorphism onto a group of order m that is injective on the piece; then the problem becomes choosing images of generators, which is almost always a small arithmetic question.

6. Reusable decision rule: *in a "for which n" existence question where the small cases all work, look for a uniform algebraic construction before looking for an obstruction or an induction.* For tilings of $\mathbb{Z}^n$ (and perfect codes, and colourings where each colour class has a fixed neighbourhood pattern) the uniform construction is a linear form modulo the tile size; the coefficients are dictated by requiring the tile to hit each residue once.