DAWN PUTNAM SESSION

Putnam 2013 — Papers A and B

Putnam 120 Archive

Contents

Putnam 2013 — Paper A **5**
2013 A1 5
2013 A2 7
2013 A3 10
2013 A4 13
2013 A5 16
2013 A6 19

Putnam 2013 — Paper B **23**
2013 B1 23
2013 B2 25
2013 B3 27
2013 B4 30
2013 B5 33
2013 B6 36

Putnam 2013 — Paper A

2013 A1

2013 A1. Recall that a regular icosahedron is a convex polyhedron having 12 vertices and 20 faces; the faces are congruent equilateral triangles. On each face of a regular icosahedron is written a nonnegative integer such that the sum of all 20 integers is 39. Show that there are two faces that share a vertex and have the same integer written on them.

Fields: combinatorics — pigeonhole via double counting **Position:** A1 — the opener **Answer:** Two such faces always exist. Assume for contradiction every vertex's five incident faces carry five distinct labels; summing the resulting per-vertex lower bound of 10 over all 12 vertices forces a total of at least 120, but the true total, recomputed by counting each face's label once for each of its 3 vertices, is exactly $3 \times 39 = 117$; $120 \leq 117$ is false.

Layer 1 THE WHAT — surface analysis

The natural attempt. Twenty objects, nonnegative integer labels, total 39, and a claim that two of them (sharing a vertex) must coincide – this smells like pigeonhole, and the obvious place to apply it is locally: at any one vertex of the icosahedron exactly five faces meet (the vertex figure of an icosahedron is a pentagon of five triangles). So fix a vertex v and look at its five incident faces. If pigeonhole is going to bite here, it should say something like "five faces, few enough possible labels, two must match." But the labels are nonnegative integers with no stated upper bound – there is no fixed alphabet size to fall back on, so pigeonhole cannot be applied directly to a set of five unbounded integers.

The fallback is to argue by contradiction: suppose the five labels at v are all different. Since they are five distinct nonnegative integers, the smallest they could possibly be is $0, 1, 2, 3, 4$, so their sum is at least $0 + 1 + 2 + 3 + 4 = 10$. That is a genuine constraint – but it is a constraint on five of the twenty faces, and $10 \leq 39$ is not a contradiction. Nothing stops the other fifteen faces from soaking up the remaining 29. Trying the same bound at a second vertex doesn't obviously help either, since the two vertices' incident faces overlap (they share at least one face, if adjacent) and it isn't yet clear how to combine several such local bounds into one clean global statement. At this point a natural next move is to try small constructions by hand – attempt to actually build a labelling with sum 39 in which every vertex's five faces are pairwise distinct – and discover this is fiddly and inconclusive: some attempts seem to almost work, others clearly fail, but no pattern emerges that proves impossibility in general. The local bound is true but, used in isolation, it is inert.

The model chain. Assume for contradiction that at every vertex, the five incident faces carry five pairwise distinct labels. Then, as above, the sum of labels at any single vertex v is at least 10.

Now sum this inequality over *all twelve* vertices:

$$\sum_{v} \Big(\text{sum of labels on the 5 faces at } v \Big) \;\geq\; 12 \times 10 = 120.$$

Evaluate the left-hand side a second way, by regrouping the double sum around faces instead of vertices. Each face is an equilateral triangle, so it has exactly 3 vertices, and its label is added into the vertex-sum once for each of those 3 vertices – that is, once for every vertex it touches. So

$$\sum_v \Big(\text{sum of labels on the 5 faces at } v\Big) = \sum_{\text{faces } f} 3 \cdot \text{label}(f) = 3 \sum_{\text{faces } f} \text{label}(f) = 3 \times 39 = 117.$$

Combining the two evaluations gives $117 \geq 120$, which is false. The assumption that every vertex has five distinct incident labels is therefore impossible: some vertex has two incident faces with the same label, and those two faces share that vertex, as required.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree up to the single-vertex bound: "if all five labels at v are distinct, their sum is at least 10." The natural attempt tries to use this bound at one vertex (or a handful) directly against the total 39 and stalls, because 10 is far smaller than 39 and the bound only concerns five of the twenty faces. The model chain instead sums the bound over *every* vertex and rewrites the resulting total by regrouping around faces rather than vertices.

> *This per-vertex inequality only tells me about five of the twenty faces – to say anything about the whole configuration I should not stop at one vertex, I should sum it over every vertex; but then I need to know exactly how many times each face's label gets counted once I do that, which means I need the incidence structure between faces and vertices, not just the bound itself.*

> *Since every face is a triangle, it touches exactly three vertices – so if I sum the same face's label once for each vertex it belongs to, across all twelve vertices, I count that label exactly three times; the grand total is therefore forced to equal* $3 \times 39 = 117$ *no matter how the labels are arranged, and that fixed number is something I can directly compare against the* 120 *the distinctness assumption demands.*

Layer 3 THE WHY

The whole argument runs on a bridge between a *local* claim (two faces at one vertex share a label) and a *global* constraint (the fixed total 39 over all twenty faces). A local pigeonhole bound, by itself, only ever touches the handful of objects at one local structure; it cannot be compared against a global total unless it is aggregated across every local structure and then correctly rescaled. The rescaling factor is not free-floating – it is handed to us directly in Layer 0: "the faces are congruent equilateral triangles," which fixes the number of vertices per face at exactly 3. That single structural fact is what turns "sum the per-vertex bound over all twelve vertices" into a statement about the original global sum: it is the double-counting (handshake) identity

$$\sum_v \sum_{f \ni v} \text{label}(f) \;=\; \sum_f \text{label}(f) \cdot (\#\text{vertices of } f),$$

which holds for *any* assignment of numbers to faces of *any* polyhedron – it is pure incidence bookkeeping, true before a single number is written down, not a trick invented for this particular value 39. Once that identity is available, the only remaining work is arithmetic: force the left side up to at least 120 via the distinctness hypothesis, and pin the right side down at exactly 117 via the given total, and the two numbers must clash. Nothing about this argument needed to construct or rule out any specific labelling; it needed only the count "5 faces per vertex, 3 vertices per face" and the total 39, combined through the one identity that always relates a sum-over-vertices to a sum-over-faces.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the pairing of a fixed global total (39 over 20 faces) with a claim about two objects at a *local* structure (5 faces meeting at a vertex). On seeing it, write down: "there are two natural incidence numbers here – 5 (local structures at each global object's neighbourhood) and 3 (global objects belonging to each local structure, since faces are triangles) – and a double-counting identity almost certainly connects them."

2. That cue suggests proof by contradiction plus a global double-counting argument: assume the local pigeonhole bound fails to force a repeat everywhere, sum the resulting per-structure lower bound over every local structure, and rewrite the total via the incidence multiplicity.

3. Hesitation sets in right after computing the single-vertex bound of 10 and finding $10 \leq 39$ is no contradiction – the trap is to conclude the bound is simply too weak and abandon the pigeonhole idea altogether. What reinforces staying with it is asking: this bound holds at *every* vertex simultaneously, not just one – what happens if I add up all twelve copies of it instead of examining one in isolation?

4. The trap is treating the single-vertex "5 distinct values $\Rightarrow$ sum $\geq$ 10" bound as the entire argument, the way pigeonhole is usually taught (five pigeons, need a small fixed number of holes). Here there is no fixed number of holes – the escape is aggregation across vertices, not a bigger single bound.

5. Recognise this faster by treating "fixed global total, local claim about a repeat" as a standing instruction to find the two incidence numbers (local structures per object, objects per local structure) and write down the double-counting identity before attempting anything else, especially before trying to construct or rule out explicit labellings by hand.

6. Reusable rule: to turn a per-local-structure pigeonhole bound into a global contradiction, sum the bound over every local structure in the configuration, then rewrite that same total by regrouping around the global objects using each object's incidence multiplicity (how many local structures it belongs to); compare the two evaluations of the same sum.

▷ **Cue:** a fixed global total spread over N objects, with a claim that two objects sharing membership in some local structure (containing k of them) must coincide $\longrightarrow$ **Move:** assume all k values are distinct at every local structure to get a per-structure lower bound $(0 + 1 + \cdots + (k-1))$, sum that bound over every local structure, and rewrite the resulting double sum via each object's incidence multiplicity to compare against the true global total

▷ **Cue:** a polyhedron whose faces are triangles $\longrightarrow$ **Move:** every face touches exactly 3 vertices, so any per-face quantity, summed once for each vertex a face belongs to, is counted exactly 3 times in a vertex-indexed sum

2013 A2

2013 A2. Let S be the set of all positive integers that are *not* perfect squares. For n in S, consider choices of integers $a_1, a_2, \ldots, a_r$ such that $n < a_1 < a_2 < \cdots < a_r$ and $n \cdot a_1 \cdot a_2 \cdots a_r$ is a perfect square, and let $f(n)$ be the minimum of a_r over all such choices. For example, $2 \cdot 3 \cdot 6$ is a perfect square, while $2 \cdot 3$, $2 \cdot 4$, $2 \cdot 5$, $2 \cdot 3 \cdot 4$, $2 \cdot 3 \cdot 5$, $2 \cdot 4 \cdot 5$, and $2 \cdot 3 \cdot 4 \cdot 5$ are not, and so $f(2) = 6$. Show that the function f from S to the integers is one-to-one.

Fields: number theory — perfect squares as an $\mathbb{F}_2$-vector space **Position:** A2 **Answer:** f is injective. Given $f(n) = f(m) = t$ with $n < m$, multiplying the two minimal witnessing square-products together and cancelling every value that occurs in both leaves a smaller square witness for n whose top term is strictly below t, contradicting minimality of $f(n)$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The defining condition "$n \cdot a_1 \cdots a_r$ is a perfect square" is a statement

about every prime's exponent being even. The standard way to make that precise is to record, for each positive integer k, its *squarefree kernel* – the vector, indexed by primes, of its prime exponents reduced mod 2 – and note that a product is a perfect square exactly when the sum of these vectors (in the vector space $(\mathbb{F}_2)^{\text{primes}}$) is the zero vector. So "$n \cdot a_1 \cdots a_r$ square" becomes: $\mathbf{v}(n) + \mathbf{v}(a_1) + \cdots + \mathbf{v}(a_r) = \mathbf{0}$.

Suppose, for contradiction, $f(n) = f(m) = t$ for some $n < m$ in S. Then there are minimal witnesses: $\mathbf{v}(n) + \sum_i \mathbf{v}(a_i) = \mathbf{0}$ with $n < a_1 < \cdots < a_r = t$, and $\mathbf{v}(m) + \sum_j \mathbf{v}(b_j) = \mathbf{0}$ with $m < b_1 < \cdots < b_s = t$. Adding the two vector equations gives $\mathbf{v}(n) + \mathbf{v}(m) + \sum_i \mathbf{v}(a_i) + \sum_j \mathbf{v}(b_j) = \mathbf{0}$ – a single new relation involving n, m, and all the a_i, b_j together. The instinct at this point is to try to isolate a clean witness for n alone directly from this combined relation – perhaps by manipulating $\mathbf{v}(m)$, or by dividing through by the product involving m and its own witnesses in some way – but m is stuck in the middle of the equation and there is no obvious way to remove it and its own witness sequence $b_1, \ldots, b_s$ symmetrically without more structure. Trying to reason about which primes cancel between n's witness set and m's witness set, prime by prime, is possible in principle but heavy, and it is not clear in advance that it produces a sequence still respecting the required strict-increase, all-above-n shape that $f(n)$'s definition demands.

The model chain. Work multiplicatively rather than additively. $n \cdot a_1 \cdots a_r$ is a perfect square and $m \cdot b_1 \cdots b_s$ is a perfect square, so their product

$$(n \cdot a_1 \cdots a_r) \cdot (m \cdot b_1 \cdots b_s)$$

is also a perfect square (a product of two perfect squares is a perfect square). This product is literally the product of the multiset $\{n, a_1, \ldots, a_r, m, b_1, \ldots, b_s\}$.

Now delete from this multiset every value that occurs in it *exactly twice* – i.e., whenever some a_i equals some b_j, remove both, and since $a_r = b_s = t$, remove that matching pair too. Deleting a matched pair $\{k, k\}$ removes a factor of k^2, itself a perfect square, so the product of what remains is still a perfect square. What remains is now a set of *distinct* integers (no repeats survive), and:

- n is among them: n is strictly smaller than every a_i (given) and every b_j (since $b_j > m > n$), so n cannot have been cancelled against anything.
- every element other than n exceeds n, since each came from $\{a_i\} \cup \{b_j\}$, all of which exceed n.
- the largest remaining element is strictly less than t: the two copies of t (namely a_r and b_s) cancelled each other exactly, and every other surviving element is a distinct value $\leq t$ other than t itself, hence $< t$.

So the surviving elements, listed in increasing order, form a legitimate witness $n < c_1 < \cdots < c_p$ with $n \cdot c_1 \cdots c_p$ a perfect square and $c_p < t$. This contradicts the minimality of $t = f(n)$. Hence $f(n) = f(m)$ with $n \neq m$ is impossible, and f is injective.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree up to translating "perfect square" into the $\mathbb{F}_2$-vector-space language and combining n's witness with m's witness into one relation. The natural attempt tries to work with that combined vector relation directly and stalls on isolating a clean witness for n that still respects the strict-increase, all-above-n requirement. The model chain instead goes back to the multiplicative form of the same fact – "product of two squares is a square" – and cancels literal repeated *values* rather than manipulating a vector equation abstractly.

> *If I multiply the two square products together, I get another perfect square – and a value appearing twice in that product is a factor of $k \times k = k^2$, itself a perfect square, so I can delete every value that occurs on both sides without disturbing squareness; the vector-space cancellation $\mathbf{v}(k) + \mathbf{v}(k) = \mathbf{0}$*

> *is exactly this deletion, but stated as literal removal of repeated factors it tells me directly what the surviving witness sequence looks like.*

> *The two copies of $t - a_r$ from n's witness and b_s from m's witness – are equal by hypothesis, so they are exactly the kind of matched pair that cancels; once they're gone, nothing left in the multiset can be as large as t, so whatever survives is a witness for n with top term strictly below t, and that alone is enough to contradict minimality without needing to know anything else about which other values happened to coincide.*

Layer 3 THE WHY

The reason the multiplicative cancellation is the natural move, and not merely a trick, is that "perfect square" is definitionally an additive (mod 2) condition on prime exponents: k is a perfect square iff every prime's exponent in k is even, i.e. iff the exponent vector reduced mod 2 is $\mathbf{0}$. Multiplying integers adds their exponent vectors; multiplying two perfect squares therefore adds two zero vectors and gets zero again – "product of squares is a square" is not a separate fact to be recalled, it is the vector-space statement $\mathbf{0} + \mathbf{0} = \mathbf{0}$ read back multiplicatively. And deleting a value that appears twice is using $\mathbf{v}(k) + \mathbf{v}(k) = \mathbf{0}$, the single defining property of working mod 2 (every element is its own additive inverse). So the entire cancellation manoeuvre is just this vector space's arithmetic, translated into the language of multisets of integers, which is the language the problem is stated in.

Given that, the only remaining content is bookkeeping about *sizes*, not about which primes are involved: because the witnessing sequences for n and m are both strictly increasing and strictly above their respective bases, and because $n < m$ forces n to be strictly below everything in m's sequence, n can never be one of the cancelled values – it survives every possible cancellation for free, purely from the ordering hypotheses in the problem's own definition of a witness. And because the two claimed-minimal top terms are, by the contradiction hypothesis, literally the same number t, they are forced to cancel against each other, which is precisely the fact that produces a strictly smaller top term. Nothing about the specific prime factorizations of n, m, or any a_i, b_j ever needs to be examined: minimality of f is attacked purely through the ordering and the multiplicative-square structure, which is why the argument works uniformly for every pair $n < m$ without case analysis.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the phrase "perfect square" attached to a *product* of several integers. On seeing it, write down: "perfect square is an additive-mod-2 condition on prime exponents; products of perfect squares combine for free, and repeated factors cancel in pairs."

2. That cue suggests a proof by contradiction that multiplies two square-witnessing products together and exploits cancellation of matched values, rather than a direct prime-by-prime or GCD-based comparison of n and m.

3. Hesitation sets in right after writing the combined vector relation for n and m together and finding m stuck in the middle with no obvious way to remove it cleanly. What reinforces the switch away from vector bookkeeping is remembering that the vector relation is only a translation of a multiplicative fact – go back to the multiplicative statement and cancel literal repeated integers instead of abstract vectors.

4. The trap is trying to extract a witness for n by manipulating the exponent-vector equation symbol by symbol, treating m as an obstacle to be eliminated algebraically, rather than noticing that multiplying the two square products and physically striking out repeated values does the elimination automatically, with the ordering hypotheses supplying every fact needed about what survives.

5. Recognise this faster by treating any claim of the form "the minimal top term of a square-completing sequence is achieved twice, for two different base numbers" as a standing

invitation to multiply the two witnessing square products and cancel matched values, specifically watching for the two copies of the shared top term cancelling each other.

6. Reusable rule: to contradict the minimality of a top witnessing value shared by two different base numbers in a "complete this product to a perfect square" problem, multiply the two minimal square-witnesses together and delete every value occurring twice (this preserves squareness, since $k \cdot k$ is itself a square); if the two claimed-minimal top terms are equal, they are exactly such a matched pair and their cancellation strictly lowers the top term of the surviving witness for the smaller base number.

▷ **Cue:** a claimed minimal top element t of an increasing sequence whose product with a fixed base number is a perfect square $\longrightarrow$ **Move:** assume two different base numbers share the same minimal top element t; multiply their two witnessing square-products together and cancel every value occurring on both sides to obtain a smaller square witness for the smaller base

▷ **Cue:** a number is asserted to be, or required to be, a perfect square $\longrightarrow$ **Move:** translate to its prime-exponent vector reduced mod 2; multiplying perfect squares, or cancelling a value that occurs twice in a product, is addition in $(\mathbb{F}_2)^{\text{primes}}$, so combining two square witnesses and deleting duplicate entries preserves squareness automatically

2013 A3

2013 A3. Suppose that the real numbers $a_0, a_1, \ldots, a_n$ and x, with $0 < x < 1$, satisfy

$$\frac{a_0}{1-x} + \frac{a_1}{1-x^2} + \cdots + \frac{a_n}{1-x^{n+1}} = 0.$$

Prove that there exists a real number y with $0 < y < 1$ such that

$$a_0 + a_1 y + \cdots + a_n y^n = 0.$$

Fields: analysis — generating functions, geometric series, continuity/IVT **Position:** A3 **Answer:** Such y always exists. Writing $P(y) = a_0 + a_1 y + \cdots + a_n y^n$ and expanding each $1/(1-x^{i+1})$ as a geometric series shows the given sum equals $\sum_{m=0}^{\infty} x^m P(x^m)$; if P had no zero in $(0,1)$ it would have one constant sign there by continuity, forcing this weighted sum of infinitely many same-signed nonzero terms to be nonzero, a contradiction.

Layer 1 THE WHAT — surface analysis

The natural attempt. Write $P(y) = a_0 + a_1 y + \cdots + a_n y^n$, the polynomial whose root is wanted. Try $n = 1$ concretely first, to get a feel for the mechanism: the hypothesis reads $\frac{a_0}{1-x} + \frac{a_1}{1-x^2} = 0$. Multiplying through by $(1-x)(1-x^2)/(1-x) = (1-x^2)$ (or just clearing denominators directly) gives $a_0(1+x) + a_1 = 0$, so $a_1 = -a_0(1+x)$, and then $P(y) = a_0 + a_1 y = a_0(1-(1+x)y)$, which vanishes at $y = 1/(1+x)$. Since $0 < x < 1$, this y lies in $(1/2, 1) \subset (0,1)$, exactly as required. Good – the case $n = 1$ works, and it worked by brute algebraic clearing of denominators followed by directly solving the resulting linear equation for y.

The obvious next step is to try the same thing for general n: clear every denominator in $\frac{a_0}{1-x} + \cdots + \frac{a_n}{1-x^{n+1}} = 0$ by multiplying through by $\prod_{i=0}^{n}(1-x^{i+1})$, producing a single polynomial identity in the fixed number x that equals zero. But this identity is a statement purely about x – it doesn't obviously mention a free variable y at all, so there's no direct way to read a claim about $P(y)$ for some *other* point y off of it. Attempting a proof by induction on n runs into a similar wall: peeling off the last term $a_n/(1-x^{n+1})$ and trying to relate the remaining n-term sum to a lower-degree polynomial's root requires knowing how a_n's specific value interacts with the others, and nothing in the $n = 1$ computation suggests a clean inductive

step. The direct algebraic route that worked for $n = 1$ does not obviously generalise; something is missing that turns "one messy identity in x" into "a family of values of P that can be reasoned about."

The model chain. Suppose for contradiction that $P(y) = a_0 + a_1 y + \cdots + a_n y^n$ has no zero in $(0,1)$. Since P is a polynomial, it is continuous, so by the intermediate value theorem it cannot change sign on the interval $(0,1)$ without vanishing somewhere in it; hence P has one fixed sign throughout $(0,1)$. Without loss of generality take $P(y) > 0$ for all $y \in (0,1)$ (otherwise replace every a_i by $-a_i$).

Now use the fact that $0 < x < 1$, so every power x^m for $m \geq 1$ lies strictly inside $(0,1)$, and hence $P(x^m) > 0$ for every $m \geq 1$; by continuity, taking $y \to 1^-$ also gives $P(x^0) = P(1) \geq 0$. So

$$x^m P(x^m) \geq 0 \text{ for } m = 0, 1, 2, \ldots, \quad \text{and strictly } > 0 \text{ for } m \geq 1.$$

Expand $x^m P(x^m) = \sum_{i=0}^{n} a_i x^{m(i+1)} \cdot x^0$... more precisely, $x^m P(x^m) = \sum_{i=0}^{n} a_i x^{mi} \cdot x^m = \sum_{i=0}^{n} a_i x^{m(i+1)}$. Sum this over all $m = 0, 1, 2, \ldots$ and swap the (absolutely convergent, since finitely many i and $0 < x < 1$) order of summation:

$$\sum_{m=0}^{\infty} x^m P(x^m) = \sum_{m=0}^{\infty} \sum_{i=0}^{n} a_i x^{m(i+1)} = \sum_{i=0}^{n} a_i \sum_{m=0}^{\infty} x^{m(i+1)} = \sum_{i=0}^{n} \frac{a_i}{1 - x^{i+1}},$$

using the geometric series $\sum_{m \geq 0} r^m = \frac{1}{1-r}$ for $r = x^{i+1} \in (0,1)$. By hypothesis, the right-hand side is exactly 0. But the left-hand side is a sum of terms each ≥ 0, with every term for $m \geq 1$ strictly positive – infinitely many strictly positive terms summing to (at most) something ≥ 0 cannot total 0 unless the sum itself diverges to something strictly positive, which it does since it equals the given finite quantity 0... more precisely: the $m = 0$ term is ≥ 0 and every $m \geq 1$ term is > 0, so the partial sum up to any $M \geq 1$ already exceeds 0, and the series (which converges, being equal to the finite quantity 0 by the algebra above) is therefore strictly positive. This contradicts its being equal to 0. Hence P must have a zero $y \in (0,1)$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree on introducing $P(y) = a_0 + a_1 y + \cdots + a_n y^n$ as the object of interest and, implicitly, on the idea of assuming no root exists and deriving a contradiction (the natural attempt reaches this only after its direct algebra stalls; the model chain starts there deliberately). They split at how to use the hypothesis: the natural attempt clears denominators and treats the resulting identity as a single static fact about the fixed number x; the model chain instead reads $\frac{a_i}{1 - x^{i+1}}$ as $\sum_{m \geq 0} a_i x^{m(i+1)}$, a geometric series, and reassembles the whole hypothesis as a weighted sum $\sum_m x^m P(x^m)$ over infinitely many *sample values* of P.

> *The denominator $1 - x^{i+1}$ is not just something to clear by multiplying through – it is the sum of a geometric series $\sum_{m \geq 0} x^{m(i+1)}$, so if I swap the order of summation, the entire left-hand side of the hypothesis becomes a sum, over $m = 0, 1, 2, \ldots$, of x^m times $P(x^m)$ – the hypothesis is really a statement about infinitely many values of P, not a single algebraic identity in x alone.*

> *Every one of those sample points x^m for $m \geq 1$ lies strictly inside $(0,1)$, exactly where I assumed P has one fixed nonzero sign – so if that assumption holds, every one of those infinitely many weighted terms is already forced to be strictly positive, and a sum of infinitely many nonnegative terms that are eventually all strictly positive can never equal exactly zero.*

Layer 3 THE WHY

The geometric-series rewriting is not an imported trick; it is reading the expression $1/(1 - x^{i+1})$ in the very form it is built from. The problem hands you denominators of the shape

$1 - x^k$ for consecutive $k = 1, \ldots, n+1$ – and the single most basic identity available for such a denominator, $\frac{1}{1-r} = \sum_{m \geq 0} r^m$ for $|r| < 1$, is valid here precisely because $0 < x < 1$ makes every x^{i+1} lie in $(0, 1)$. Recognising "$1/(1 - x^{i+1})$ is a sum in disguise" is simply taking the given expression at face value rather than immediately reaching for algebraic clearing, which discards the very structure (a sum over m) that makes a family of sample points $\{x^m\}$ available in the first place.

Swapping the order of the resulting double summation from "sum over i of a sum over m" to "sum over m of a sum over i" is legitimate because the sum is absolutely convergent: there are only finitely many i (from 0 to n), and each inner geometric series converges absolutely since $0 < x^{i+1} < 1$. This is what upgrades a merely formal rearrangement into a genuine equality between two limits, and it is what produces $\sum_i a_i x^{m(i+1)} = x^m \sum_i a_i x^{mi} = x^m P(x^m)$ – the polynomial P reappearing, evaluated at the geometric points x^m, is the payoff of the swap.

The intermediate value theorem is the bridge from "P has no zero in $(0, 1)$" to "P has one constant sign on $(0, 1)$": this is nothing more than the definitional content of continuity plus the intermediate value property applied to the open interval $(0, 1)$ – a continuous real function that never hits zero on a connected interval cannot switch between positive and negative without passing through zero. Once that sign is fixed, every sampled value $P(x^m)$ for $m \geq 1$ inherits it automatically, because $x^m \in (0, 1)$ for every such m – this is exactly the role played by the hypothesis $0 < x < 1$, which is what guarantees infinitely many distinct sample points all land inside the very interval on which the sign was assumed constant. The contradiction is therefore not a coincidence of this particular problem: it is forced the moment a "sum equals zero" hypothesis is re-expressed as a positively-weighted sum of same-signed, eventually-strictly-positive terms.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is a sum of the form $\sum_i a_i/(1 - x^{i+1})$ set equal to zero, where the same $x \in (0, 1)$ appears raised to consecutive powers in every denominator. On seeing it, write down: "each denominator $1 - x^k$ is the sum of a geometric series in x^k – expand it and see what sum-over-m structure appears."

2. That cue suggests: expand every term as a geometric series, swap the order of summation (justified by absolute convergence, since $0 < x < 1$ and there are finitely many i), and recognise the resulting inner sum over i as the target polynomial P evaluated at a new point.

3. Hesitation sets in right after the $n = 1$ case is solved directly by clearing denominators – the temptation is to keep pushing that same brute-force clearing to general n, because it visibly worked once. What reinforces the switch away from it is noticing that the cleared-denominator identity for general n is a single relation in the fixed number x with no free variable y in sight, whereas the problem wants a statement about some *other* point y – so the direct route needs a fundamentally different idea to produce that second point, and geometric-series expansion is what manufactures a whole family of candidate points x^m for free.

4. The trap is treating "it worked for $n = 1$ by direct algebra" as evidence that direct algebra will scale to general n; the $n = 1$ success is really a coincidence of low degree (a linear equation can always be solved explicitly for its unique root), and it conceals the fact that no explicit root formula exists in general, which is exactly why an existence argument (contradiction plus IVT) is needed instead of a construction.

5. Recognise this faster by treating "a weighted sum of $1/(1 - x^k)$ terms across consecutive powers, hypothesised to equal zero, alongside a polynomial with the same coefficients" as a standing cue to expand geometric series and look for the polynomial reappearing evaluated at x^m for $m = 0, 1, 2, \ldots$

6. Reusable rule: whenever a hypothesis is a sum of terms of the shape $1/(1 - r_k)$ for various

$r_k \in (0,1)$, and the goal concerns a related polynomial or power series, expand each $1/(1-r_k)$ as a geometric series and swap summation order to rewrite the hypothesis as a weighted sum of the target object evaluated at the sample points r_k^m; then attack existence of a root by contradiction via IVT, using that an assumed constant sign on the sampling interval makes every sampled term the same sign, which usually collides with the hypothesis's stated value (here, zero).

- ▷ **Cue:** a sum $\sum_i a_i/(1-x^{i+1})$ (or any sum of $1/(1-r_k)$ terms with $r_k \in (0,1)$) is hypothesised to equal a fixed value $\longrightarrow$ **Move:** expand each denominator as a geometric series $\sum_{m\geq 0} r_k^m$, swap the order of summation, and look for a related function (here $P(y) = \sum a_i y^i$) reappearing evaluated at the sample points r_k^m
- ▷ **Cue:** want to show a continuous function has a zero on an open interval, and direct construction of the zero is not available $\longrightarrow$ **Move:** assume for contradiction it has no zero, use IVT to fix its sign on the interval, then find a way to sample that function at points known to lie inside the interval and derive a numerical contradiction against a given hypothesis

2013 A4

2013 A4. A finite collection of digits 0 and 1 is written around a circle. An *arc* of length $L \geq 0$ consists of L consecutive digits around the circle. For each arc w, let $Z(w)$ and $N(w)$ denote the number of 0's in w and the number of 1's in w, respectively. Assume that $|Z(w) - Z(w')| \leq 1$ for any two arcs w, w' of the same length. Suppose that some arcs $w_1, \ldots, w_k$ have the property that

$$Z = \frac{1}{k}\sum_{j=1}^{k} Z(w_j) \text{ and } N = \frac{1}{k}\sum_{j=1}^{k} N(w_j)$$

are both integers. Prove that there exists an arc w with $Z(w) = Z$ and $N(w) = N$.

Fields: combinatorics on words (balanced circular words); averaging; discrete intermediate value arguments **Position:** A4 — mid-paper. Nothing technical is needed; the difficulty is seeing that the hypothesis is really a statement about a *density*. **Answer:** Proof. Every arc of length ℓ has $|Z(w) - \ell\rho| < 1$, where ρ is the proportion of 0's on the circle; averaging this over $w_1, \ldots, w_k$ pins Z within distance 1 of $L\rho$, $L = Z + N$, and every integer that close to $L\rho$ is attained by some arc of length L.

Layer 1 THE WHAT — surface analysis

The natural attempt. Notation first. Let the circle carry m digits, z of them 0's. Write $\ell_j = Z(w_j) + N(w_j)$ for the length of w_j, and $L = Z + N$, so that $L = \frac{1}{k}\sum \ell_j$ is the average length. Since every $\ell_j \leq m$, also $L \leq m$, so arcs of length L exist.

The target is an arc of length exactly L with exactly Z zeros (then $N(w) = L - Z = N$ automatically). A strong student sees two things quickly. First, arcs of a fixed length ℓ have zero-counts in some set $\{c_\ell, c_\ell + 1\}$, by hypothesis. Second, sliding an arc of length ℓ one step round the circle changes its zero-count by at most 1, so as it slides it takes every value between its minimum and maximum. So the job is to show that Z lies in $\{c_L, c_L + 1\}$ and that the relevant value is actually hit.

Then the attack: induction on k. For $k = 1$ there is nothing to do. For $k = 2$ you try to trade w_1, w_2 for two arcs of the common length L. Put w_2' (same length as w_2) immediately after w_1; the block $w_1 w_2'$ has length $2L$ and splits into two adjacent arcs u_1, u_2 of length L. Then $Z(u_1) + Z(u_2) = Z(w_1) + Z(w_2') \in \{2Z - 1, 2Z, 2Z + 1\}$, and $Z(u_1), Z(u_2)$ differ by at most 1. Some fiddling shows one of them equals Z. Encouraging.

Now $k = 3$, and the induction breaks. You would like to replace w_1, w_2 by one arc of their average length and recurse, but $(\ell_1 + \ell_2)/2$ need not be an integer, and even when it is,

$(Z(w_1) + Z(w_2))/2$ need not be. The integrality hypothesis belongs to the whole collection and is not inherited by any sub-collection. At this point a well-read student tries to prove the structure theorem — balanced circular words are "mechanical", with the i-th zero at position $\lfloor i/\rho + \theta \rfloor$ — and that is a genuine theorem whose proof is harder than the problem. That is the wall.

The model chain. Let $\rho = z/m$ be the density of zeros.

Step 1: the average over all arcs of a given length. Fix ℓ with $0 \le \ell \le m$, and for each starting position $t = 0, 1, \ldots, m-1$ let v_t be the arc of length ℓ starting at t. Each digit of the circle lies in exactly ℓ of these m arcs, so

$$\sum_{t=0}^{m-1} Z(v_t) = \ell z, \qquad \text{i.e.} \qquad \frac{1}{m}\sum_t Z(v_t) = \ell\rho.$$

Step 2: every arc is within 1 *of the average.* By hypothesis the numbers $Z(v_t)$ lie in $\{c, c+1\}$ for some integer c (depending on ℓ). Their mean $\ell\rho$ therefore satisfies $c \le \ell\rho \le c+1$. If $\ell\rho$ is an integer, then either all $Z(v_t)$ equal c (mean c) or all equal $c+1$; in both cases every arc of length ℓ has $Z = \ell\rho$ exactly. If $\ell\rho$ is not an integer, then $c < \ell\rho < c+1$, both values occur, and each arc has $|Z - \ell\rho| < 1$. In all cases:

$$|Z(w) - \ell\rho| < 1 \quad \text{for every arc } w \text{ of length } \ell. \tag{$*$}$$

Step 3: average over the given arcs. Apply $(*)$ to each w_j and add:

$$\Big|\sum_j Z(w_j) - \rho\sum_j \ell_j\Big| \le \sum_j |Z(w_j) - \ell_j\rho| < k.$$

But $\sum_j Z(w_j) = kZ$ and $\sum_j \ell_j = kL$. Dividing by k: $|Z - L\rho| < 1$.

Step 4: that integer is hit. Apply Step 2 with $\ell = L$. If $L\rho$ is an integer, then $|Z - L\rho| < 1$ forces $Z = L\rho$, and every arc of length L has exactly Z zeros. If not, then Z is an integer strictly within 1 of $L\rho \in (c, c+1)$, so $Z \in \{c, c+1\}$, and Step 2 showed both values occur. Either way there is an arc w of length L with $Z(w) = Z$, hence $N(w) = L - Z = N$. (The case $L = 0$ is covered: the empty arc has $Z = N = 0$.)

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains reduce the goal to "Z is one of the two zero-counts that arcs of length L take". They split at $(\dagger)$: how do you compare arcs of *different* lengths ℓ_j with arcs of length L? The natural attempt compares them to each other, pair by pair, and needs integrality at every stage of an induction. The model compares every one of them to a single external yardstick, $\ell\rho$, which exists for every length at once and is linear in ℓ, so averaging commutes with it.

> *The hypothesis says all arcs of the same length have almost the same number of zeros. Almost the same as what? There has to be a number each of them is close to — and the only candidate is the average, length times the fraction of zeros on the whole circle.*

> *Once each $Z(w_j)$ is within* 1 *of $\ell_j\rho$, the errors add up to less than k, and dividing by k leaves an error less than* 1. *An integer within* 1 *of $L\rho$ is exactly what an arc of length L can have.*

Layer 3 THE WHY

Why is the density the natural object and not an inspired one? Read the hypothesis again: $|Z(w) - Z(w')| \le 1$ whenever $|w| = |w'|$. That is a statement of the form "this quantity is

nearly constant on a class". Whenever a quantity is nearly constant on a class, the thing it is nearly equal to is its average over the class, and the average is computed by double counting. Here the class is "all m rotations of an arc of length ℓ", and the double count — each digit sits in ℓ of them — is the first line anyone would write. So $\ell\rho$ is not a trick; it is the definition of "the value the hypothesis says they are all close to".

The second reason is linearity. The conclusion concerns averages of the w_j, and the one function of ℓ that averages cleanly is a linear one: $\frac{1}{k}\sum \ell_j\rho = L\rho$. The natural induction failed precisely because it tried to average the integer-valued quantities $Z(w_j)$ in stages, and integers are not closed under halving. Comparing each $Z(w_j)$ with the real number $\ell_j\rho$ lets you average in one go, in $\mathbb{R}$, and only return to integers at the end. The strictness in $(*)$ is where the argument closes: k strict inequalities sum to something strictly less than k, and dividing by k gives strictly less than 1. A non-strict bound would leave $Z = L\rho \pm 1$ open, and that is exactly the case that must be excluded.

Kedlaya and Ng's solution never mentions ρ. It lays copies $w'_1, \dots, w'_k$ (with $|w'_j| = |w_j|$, $w'_1 = w_1$) end to end, re-cuts the resulting string of length kL into k consecutive arcs $w''_1, \dots, w''_k$ of length L, and observes $|kZ - \sum Z(w''_j)| = |\sum_{j\geq 2}(Z(w_j) - Z(w'_j))| \leq k-1$. If no w''_j had exactly Z zeros, all would lie on one side of Z (their counts are in $\{c, c+1\}$), giving a gap of at least k. Here is where that comes from: it is the same argument with the average over all rotations replaced by the average over one well-chosen chain of adjacent arcs. The concatenation is a finite, explicit way of producing arcs whose total zero-count equals that of the w_j up to an error of at most $k-1$ — and $k-1 < k$ is the same strict-inequality closure as $(*)$. The fixed $w'_1 = w_1$ is what buys the "-1"; without it the error could be k and the argument would not close.

It is worth knowing the name of what you have met. Binary words with $|Z(w) - Z(w')| \leq 1$ for equal lengths are *balanced* words; the infinite aperiodic ones are the Sturmian words, and Step 2 is the first lemma of their theory: a balanced word has a well-defined frequency, and every factor's letter count is within 1 of length times frequency. You did not need the classification; you needed only the frequency.

Caveat. Verified by exhaustive computer check: every balanced circular binary word with at most 9 digits and every multiset of at most 3 arcs with integral averages Z, N (25 994 cases) has an arc attaining (Z, N). The proof above does not depend on this check.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the hypothesis "$|Z(w) - Z(w')| \leq 1$ for arcs of the same length". Write down: nearly constant on each length class, so compute the class average — $\ell z/m$ by double counting.

2. Averaging against a linear yardstick: bound each $Z(w_j)$ against $\ell_j\rho$, sum the errors, divide by k. Then a discrete intermediate-value step (both values c and $c+1$ occur when the mean is strictly between them) finishes.

3. Hesitation comes at $k \geq 3$, when pairwise combination stops preserving integrality. Reinforce with: do the averaging in the reals and return to the integers once, at the end, using a strict inequality.

4. The trap is induction on k, which looks natural because $k = 2$ yields to a direct juxtaposition. The hypothesis "Z, N integers" is global and is not inherited by sub-collections. A second trap is reaching for the Sturmian structure theorem, which is true but costs far more than the problem.

5. Slogan: *a quantity that is nearly constant on a class is nearly equal to its class average; find that average by double counting, and compare everything to it.*

6. Decision rule: **when a hypothesis says that objects of the same size have nearly equal counts, introduce the density (count divided by size) and bound each object's deviation**

from size times density; averages then pass through because size times density is linear. Companion rule: k strict inequalities of size < 1 average to strictly less than 1, which is exactly what is needed to pin an integer.

- ▷ **Cue:** a hypothesis of the form $|f(w) - f(w')| \leq 1$ whenever w, w' have the same size $\longrightarrow$ **Move:** compute the average of f over all objects of that size by double counting, and show every $f(w)$ is within 1 of it
- ▷ **Cue:** an average of several integer quantities is assumed to be an integer, and induction on the number of terms keeps losing integrality $\longrightarrow$ **Move:** compare every term to a real-valued linear yardstick, average in the reals, and use strictness to recover the integer at the end
- ▷ **Cue:** arcs or windows sliding around a cycle with counts changing by at most one per step $\longrightarrow$ **Move:** the count takes every integer value between its minimum and maximum; if the mean lies strictly between c and $c+1$, both are attained
- ▷ **Cue:** arcs of varying lengths to be compared with arcs of their average length $\longrightarrow$ **Move:** lay them end to end and re-cut into equal pieces; the total count is preserved exactly

2013 A5

2013 A5. For $m \geq 3$, a list of $\binom{m}{3}$ real numbers a_{ijk} $(1 \leq i << j < k \leq m)$ is said to be *area definite* for $\mathbb{R}^n$ if the inequality

$$\sum_{1 \leq i < j < k \leq m} a_{ijk} \cdot \text{Area}(\Delta A_i A_j A_k) \geq 0$$

holds for every choice of m points $A_1, \ldots, A_m$ in $\mathbb{R}^n$. For example, the list of four numbers $a_{123} = a_{124} = a_{134} = 1$, $a_{234} = -1$ is area definite for $\mathbb{R}^2$. Prove that if a list of $\binom{m}{3}$ numbers is area definite for $\mathbb{R}^2$, then it is area definite for $\mathbb{R}^3$.

Fields: geometry (projections, areas); integral geometry; averaging over symmetry **Position:** A5 — late paper. Very short once seen; the entire difficulty is the decision to average over all directions instead of looking for one good direction. **Answer:** Proof. Project the configuration onto the plane perpendicular to a unit vector $\hat{n}$; each triangle's area is multiplied by $|\hat{n}_{ijk} \cdot \hat{n}|$. The $\mathbb{R}^2$ hypothesis holds for every projection, and averaging over all $\hat{n}$ on the sphere multiplies every area by the same constant 2π.

Layer 1 THE WHAT — surface analysis

The natural attempt. Notation: for points $A_1, \ldots, A_m$ write $T_{ijk} = \text{Area}(\Delta A_i A_j A_k)$ and $\Sigma = \sum a_{ijk} T_{ijk}$. (The doubled "<" in the statement's range is a typo in the source; the range is $1 \leq i < j < k \leq m$.)

You are given points in $\mathbb{R}^3$ and a hypothesis about points in $\mathbb{R}^2$. The only way to use the hypothesis is to manufacture planar configurations from the spatial one. So the strong student's first attempt is: find a plane and a map of the points into it that preserves all the areas T_{ijk}. Then Σ for the spatial points equals Σ for planar points, which is ≥ 0.

That stalls at once. For four points it can sometimes be arranged (a regular tetrahedron's four equal faces are matched by the four equal triangles of a suitably sized square), but in general you are solving $\binom{m}{3}$ area equations in $2m$ planar coordinates — 20 equations in 12 unknowns already for $m = 6$ — and there is no reason to expect a solution, let alone one you can construct for arbitrary input.

Second attempt: give up on preserving areas, and hope for monotonicity. Use orthogonal projection onto a plane; projection can only shrink areas. But the coefficients a_{ijk} have both signs, so shrinking the positive terms and the negative terms by *different* factors breaks any inequality. You try to choose the plane cleverly — parallel to the largest triangle, or to the triangle with the most negative coefficient — and each choice favours some triangles and distorts others. After

half an hour you have a collection of special cases and no general argument. The wall: no single projection treats all triangles alike.

The model chain. Step 1: how projection changes one area. Let $\hat{n}$ be a unit vector and $P(v) = v - (v \cdot \hat{n})\hat{n}$ the orthogonal projection onto the plane $\Pi_{\hat{n}}$ through the origin perpendicular to $\hat{n}$. For vectors u, v, both Pu, Pv lie in $\Pi_{\hat{n}}$, so $Pu \times Pv$ is parallel to $\hat{n}$, and

$$(Pu \times Pv) \cdot \hat{n} = \det(Pu, Pv, \hat{n}) = \det(u, v, \hat{n}) = (u \times v) \cdot \hat{n},$$

because subtracting multiples of $\hat{n}$ from the first two columns does not change the determinant. Hence $|Pu \times Pv| = |(u \times v) \cdot \hat{n}|$. With $u = A_j - A_i$, $v = A_k - A_i$, and $\hat{n}_{ijk}$ a unit normal to the triangle (any unit vector if the points are collinear, when both sides are 0),

$$\text{Area}(\Delta PA_i\, PA_j\, PA_k) = \tfrac{1}{2}|(u \times v) \cdot \hat{n}| = T_{ijk}\, |\hat{n}_{ijk} \cdot \hat{n}|.$$

Step 2: use the hypothesis on every projection. $\Pi_{\hat{n}}$ is isometric to $\mathbb{R}^2$ and isometries preserve area, so area definiteness for $\mathbb{R}^2$ applied to $PA_1, \ldots, PA_m$ gives, for every unit vector $\hat{n}$,

$$\sum_{i<j<k} a_{ijk}\, T_{ijk}\, |\hat{n}_{ijk} \cdot \hat{n}| \;\geq\; 0. \tag{$*$}$$

Step 3: average over all directions. Integrate $(*)$ over the unit sphere S^2 with its surface-area measure $d\sigma$. A nonnegative function has nonnegative integral, and the sum is finite, so

$$\sum_{i<j<k} a_{ijk}\, T_{ijk} \int_{S^2} |\hat{n}_{ijk} \cdot \hat{n}|\, d\sigma(\hat{n}) \;\geq\; 0.$$

Step 4: the integral does not depend on the triangle. For a fixed unit vector u, rotate coordinates so that u is the north pole; rotations preserve $d\sigma$. With polar angle θ, $|u \cdot \hat{n}| = |\cos\theta|$ and $d\sigma = \sin\theta\, d\theta\, d\phi$, so

$$c := \int_{S^2} |u \cdot \hat{n}|\, d\sigma = \int_0^{2\pi}\!\!\int_0^{\pi} |\cos\theta| \sin\theta\, d\theta\, d\phi = 2\pi \cdot 2 \int_0^{\pi/2} \cos\theta \sin\theta\, d\theta = 2\pi.$$

Therefore $2\pi \sum a_{ijk} T_{ijk} \geq 0$, so $\Sigma \geq 0$. As $A_1, \ldots, A_m$ were arbitrary in $\mathbb{R}^3$, the list is area definite for $\mathbb{R}^3$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains decide to project $\mathbb{R}^3$ to a plane — the only way to reach the hypothesis. They split at (†): "which plane?" The natural attempt looks for *one* good plane and is defeated because each plane rescales different triangles by different factors $|\hat{n}_{ijk} \cdot \hat{n}|$. The model refuses to choose: it keeps all planes, obtains an inequality $(*)$ for each, and adds them up.

> *Every projection gives me a true inequality, just with the wrong weights: each triangle gets multiplied by the cosine of its tilt. I have infinitely many true inequalities. A sum of true ≥ 0 inequalities is true. Can I add them so that every triangle ends up with the same weight?*

> *The weight a triangle gets from averaging over directions can only depend on its normal, and by rotational symmetry it cannot even depend on that. So averaging over the whole sphere gives every triangle the same factor.*

Layer 3 THE WHY

Why is averaging the natural move rather than an inspiration? Two reasons, both structural.

First, the *hypothesis is a cone*. The set of lists (a_{ijk}) satisfying "$\sum a_{ijk}x_{ijk} \geq 0$" for a given family of vectors x is unchanged if you enlarge the family by nonnegative combinations of its members. Area definiteness for $\mathbb{R}^2$ tells you $\sum a_{ijk}x_{ijk} \geq 0$ for every vector $x =$ (planar areas); the target is the same inequality for $x =$ (spatial areas). So the question is exactly: is the vector of spatial areas a nonnegative combination (a limit of sums, i.e. an integral) of vectors of planar areas? Stated that way, the answer must be built by adding planar configurations, not by finding one. Looking for a single plane was looking for a combination with one term, and there was never a reason to expect that to suffice.

Second, *which combination* is forced by symmetry. The spatial configuration has no preferred direction; nothing in the problem singles out a plane. When a problem is invariant under a group (here, rotations of $\mathbb{R}^3$) and you are tempted to make a choice the problem does not make, the canonical replacement for the choice is the average over the group. Each individual projection breaks the symmetry and rescales triangles unevenly; the average restores the symmetry, and a quantity attached to a triangle that is invariant under all rotations cannot depend on the triangle's orientation. That is the whole content of Step 4 — the explicit value 2π is a bonus. All the proof needs is $c > 0$, which holds because the integrand is continuous, nonnegative and not identically zero.

Kedlaya and Ng's solution begins by letting $\hat{n}_{ijk}$ be a unit normal and writing the projected area as Area $\cdot\, |\hat{n}_{ijk} \cdot \hat{n}|$; it then integrates over S^2. Here is where that comes from: the moment you decide to use every projection at once, you need the factor by which each projection rescales each triangle, and that factor is the cosine of the angle between normals. Step 1 above supplies the proof of that factor, which the official solution takes as known: the cross product measures area, and projection preserves the component of the cross product along $\hat{n}$.

The remark in the official solution identifies the family: $c = 2\pi$ against total solid angle 4π means the average area of the shadow of a flat triangle is half its area, and summing over the faces of a convex body (whose shadow is covered twice by its faces) gives *Cauchy's surface area formula*: the mean shadow area of a convex body is one quarter of its surface area. The same averaging proves the statement from $\mathbb{R}^2$ to any $\mathbb{R}^n$, using 2-dimensional projections and the orthogonal group.

Caveat. Numerical check of Step 4: midpoint-rule evaluation of $\int_0^{2\pi}\int_0^{\pi} |\cos\theta| \sin\theta \, d\theta \, d\phi$ gives $6.2832\ldots = 2\pi$. The proof itself needs only $c > 0$ and independence of u.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the mismatch "hypothesis in $\mathbb{R}^2$, conclusion in $\mathbb{R}^3$". Write down: I must build planar configurations from the spatial one, and the natural map is orthogonal projection, with area factor $|\cos(\text{angle between normals})|$.

2. Integral geometry: apply the hypothesis to every projection and integrate over the sphere of directions, so that rotational symmetry makes the rescaling uniform.

3. Hesitation occurs at "which plane?". Reinforce with: a family of valid inequalities can be summed with nonnegative weights; if no member does the job, a weighted sum of members might, and the symmetric weight is the first to try.

4. The trap is optimisation: searching for a best plane (largest triangle, most negative coefficient). Mixed signs in the a_{ijk} make every single plane fail on some configuration. A smaller trap is forgetting degenerate triangles: when A_i, A_j, A_k are collinear both sides of the projection formula are 0, so any choice of $\hat{n}_{ijk}$ is fine.

5. Slogan: *when a hypothesis holds in low dimension and every low-dimensional slice or shadow distorts differently, average over all slices; symmetry makes the distortion the same for everything.*

6. Decision rule: **if the problem has a symmetry and your proof requires a choice that breaks it, replace the choice by the average over the symmetry group.** Companion rule: a linear

inequality that holds for a family of vectors holds for every nonnegative combination (sum, integral) of them.

- ▷ **Cue:** a linear inequality known in $\mathbb{R}^2$ (or for planar objects) is wanted in $\mathbb{R}^3$ $\longrightarrow$ **Move:** project onto every plane, apply the hypothesis to each projection, and integrate over directions
- ▷ **Cue:** a single choice (plane, direction, coordinate) treats different parts of the configuration unequally $\longrightarrow$ **Move:** average over all choices; by symmetry the average weight is the same for every part
- ▷ **Cue:** area of the orthogonal projection of a flat figure $\longrightarrow$ **Move:** area times $|\cos|$ of the angle between normals; prove it with $|Pu \times Pv| = |(u \times v) \cdot \hat{n}|$
- ▷ **Cue:** a family of true inequalities $\sum a_i x_i \geq 0$ for many vectors x $\longrightarrow$ **Move:** any nonnegative combination of the x's also satisfies it; ask whether the target vector is such a combination

2013 A6

2013 A6. Define a function $w : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ as follows. For $|a|, |b| \leq 2$, let $w(a,b)$ be as in the table shown; otherwise, let $w(a,b) = 0$.

$w(a,b)$		b				
		-2	-1	0	1	2
a	-2	-1	-2	2	-2	-1
	-1	-2	4	-4	4	-2
	0	2	-4	12	-4	2
	1	-2	4	-4	4	-2
	2	-1	-2	2	-2	-1

For every finite subset S of $\mathbb{Z} \times \mathbb{Z}$, define

$$A(S) = \sum_{(\mathbf{s},\mathbf{s}') \in S \times S} w(\mathbf{s} - \mathbf{s}').$$

Prove that if S is any finite nonempty subset of $\mathbb{Z} \times \mathbb{Z}$, then $A(S) > 0$. (For example, if $S = \{(0,1),(0,2),(2,0),(3,1)\}$, then the terms in $A(S)$ are $12, 12, 12, 12, 4, 4, 0, 0, 0, 0, -1, -1, -2, -2, -4, -4$.)

Fields: Fourier analysis on $\mathbb{Z}^2$; positive-definite kernels; polynomial inequalities on a square **Position:** A6 — the hardest problem of the paper, and it is hard. The key move is standard in analysis and invisible from combinatorics. **Answer:** Proof. $A(S)$ is the constant term of $f(x,y)f(x^{-1},y^{-1})g(x,y)$ with $f = \sum_S x^a y^b$ and $g = \sum w(a,b)x^a y^b$; on the torus this is $\frac{1}{4\pi^2}\iint |f|^2 g$, and $g(e^{is}, e^{it}) = 16\,G(\cos s, \cos t)$ with $G(z,w) = zw + z^2 + w^2 - z^2 w - zw^2 - z^2 w^2 \geq 0$ on $[-1,1]^2$.

Layer 1 THE WHAT — surface analysis

The natural attempt. (The source typesets the table with a two-row merged label; it is reproduced here with the same entries and headings.)

Notation: write $\mathbf{s} = (a,b)$, and note from the table that $w(-\mathbf{d}) = w(\mathbf{d})$; in fact w is even in each coordinate separately. Separate the diagonal:

$$A(S) = 12|S| + \sum_{\mathbf{s} \neq \mathbf{s}'} w(\mathbf{s} - \mathbf{s}').$$

The obvious plan is induction on $|S|$ by adding an extreme point. Let $\mathbf{p}$ be the lexicographically largest point of S and $T = S \setminus \{\mathbf{p}\}$. Then

$$A(S) = A(T) + 12 + 2\sum_{\mathbf{t} \in T} w(\mathbf{p} - \mathbf{t}),$$

and all of $\mathbf{p}-\mathbf{t}$ lie in the half-plane $\{a>0\}\cup\{a=0,b>0\}$. You would like the increment to be positive. It is not: take $S=\{(3,2),(3,4),(4,3),(4,4)\}$, $\mathbf{p}=(4,4)$. The differences are $(1,2),(1,0),(0,1)$ with weights $-2,-4,-4$, so the increment is $12+2(-10)=-8$. The point $\mathbf{p}$ is a net liability, and induction by adding points one at a time is dead.

Second idea: charge each point its "neighbourhood sum". For a point deep inside a large block of S, its contribution is $\sum_{\mathbf{d}} w(\mathbf{d})$, the sum of all 25 entries. Add the rows: $-4,0,8,0,-4$. Total 0. So interior points contribute nothing, positivity must come entirely from the boundary, and a local charging scheme has to prove a boundary inequality for arbitrary shapes. (Indeed A of an $N\times N$ square is $32N-32$ for $N\geq 2$ — linear in N, not quadratic; e.g. $32,64,128$ for $N=2,3,5$.) A student can spend the whole session on boundary bookkeeping. That is the wall.

The model chain. *Step 1: generating functions.* Let

$$f(x,y)=\sum_{(a,b)\in S} x^a y^b, \qquad g(x,y)=\sum_{(a,b)\in\mathbb{Z}^2} w(a,b)\,x^a y^b,$$

Laurent polynomials. The product $f(x,y)f(x^{-1},y^{-1})=\sum_{\mathbf{s},\mathbf{s}'}\mathbf{x}^{\mathbf{s}-\mathbf{s}'}$ (multi-index notation). Multiplying by g and extracting the coefficient of x^0y^0 picks out, for each pair, the term $w(-(\mathbf{s}-\mathbf{s}'))=w(\mathbf{s}-\mathbf{s}')$. So $A(S)$ is the constant term of $h=f(x,y)f(x^{-1},y^{-1})g(x,y)$.

Step 2: constant term as an average over the torus. For integers a,b, $\frac{1}{4\pi^2}\int_0^{2\pi}\int_0^{2\pi} e^{i(as+bt)}\,ds\,dt$ is 1 if $a=b=0$ and 0 otherwise. Put $x=e^{is}$, $y=e^{it}$. Since f has real coefficients, $f(e^{-is},e^{-it})=\overline{f(e^{is},e^{it})}$, so

$$A(S)=\frac{1}{4\pi^2}\int_0^{2\pi}\int_0^{2\pi}\left|f(e^{is},e^{it})\right|^2 g(e^{is},e^{it})\,ds\,dt.$$

Step 3: compute g on the torus. Because w is even in each coordinate, $g(e^{is},e^{it})=\sum w(a,b)\cos(as)\cos(bt)$, real. Put $u=\cos s$, $v=\cos t$, so $\cos 2s=2u^2-1$, $\cos 2t=2v^2-1$. Grouping entries by $(|a|,|b|)$:

$$\begin{aligned} g &= 12-8u-8v+16uv+4(2u^2-1)+4(2v^2-1)\\ &\quad -8u(2v^2-1)-8v(2u^2-1)-4(2u^2-1)(2v^2-1)\\ &= 16\big(uv+u^2+v^2-u^2v-uv^2-u^2v^2\big)=16\,G(u,v).\end{aligned}$$

(The constant and linear terms cancel exactly: $12-4-4-4=0$ and $-8u+8u=0$.)

Step 4: $G\geq 0$ on $[-1,1]^2$. If $uv\geq 0$:

$$G=uv(1-uv)+u^2(1-v)+v^2(1-u),$$

each term a product of nonnegatives since $0\leq uv\leq 1$ and $u,v\leq 1$. If $uv\leq 0$:

$$G=(u+v)^2-uv(1+u)(1+v),$$

and $-uv\geq 0$, $1+u\geq 0$, $1+v\geq 0$. Both identities expand to G directly.

Step 5: strict positivity. The integrand in Step 2 is continuous and ≥ 0, so $A(S)\geq 0$, and $A(S)=0$ would force $|f|^2g\equiv 0$ on the torus. But on the torus $|f|^2g$ is the restriction of the Laurent polynomial $h=f(x,y)f(x^{-1},y^{-1})g(x,y)$, and a Laurent polynomial vanishing on the whole torus has every coefficient zero (each coefficient is an integral as in Step 2). Laurent polynomials have no zero divisors, $f\neq 0$ because $S\neq\emptyset$, and $g\neq 0$; so $h\neq 0$. Hence $A(S)>0$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains write $A(S)$ as a sum over pairs. They split at (†): "$A(S)$ is a sum of $w(\mathbf{s}-\mathbf{s}')$ over pairs". The natural attempt reads this as a combinatorial count and tries to charge it to points. The model reads it as a *quadratic form* in the indicator function of S whose kernel depends only on the difference $\mathbf{s}-\mathbf{s}'$ — a convolution — and diagonalises it with the Fourier transform.

> *$A(S)$ is $\sum_{\mathbf{s},\mathbf{s}'} \mathbf{1}_S(\mathbf{s})\, w(\mathbf{s}-\mathbf{s}')\, \mathbf{1}_S(\mathbf{s}')$. That is $v^{\mathsf{T}} W v$ with a matrix that depends only on differences. I want to prove a quadratic form is positive. Difference-kernels are diagonalised by characters $e^{i(as+bt)}$: I should look at $\sum w(a,b)e^{i(as+bt)}$ and check its sign.*

> *The table is symmetric in $a \mapsto -a$ and in $b \mapsto -b$, and only goes up to ± 2. So its Fourier transform is a polynomial of degree 2 in each of $\cos s$ and $\cos t$. Proving a polynomial in two variables is nonnegative on a square is a finite job.*

Layer 3 THE WHY

Why is Fourier analysis forced rather than clever? Start from the definition. $A(S)$ is quadratic in S: every term involves two points of S. Writing $v = \mathbf{1}_S$, it is $\sum_{\mathbf{s},\mathbf{s}'} v(\mathbf{s})w(\mathbf{s}-\mathbf{s}')v(\mathbf{s}')$, and the claim is that this form is positive on every nonzero 0/1 vector. The universal tool for proving a quadratic form is positive is to diagonalise it — write it as $\sum \lambda_i (\text{something})^2$ and check $\lambda_i \geq 0$. The special feature here is that the matrix depends only on $\mathbf{s}-\mathbf{s}'$: it commutes with every translation of $\mathbb{Z}^2$. Matrices that commute with all translations share the translations' eigenvectors, and those are the characters $\mathbf{s} \mapsto e^{i(as+bt)}$. So the diagonalisation is not a choice: the eigenvalues of a translation-invariant kernel are its Fourier transform $g(e^{is}, e^{it})$, and "$A(S) \geq 0$ for all S" is essentially equivalent to "$g \geq 0$ on the torus". The generating-function identity of Step 2 is the continuous version of "$v^{\mathsf{T}} W v = \sum \lambda |\hat{v}|^2$".

The natural attempt's wall is explained by the same computation. The sum of all entries is $g(1,1) = 16G(1,1) = 0$: the kernel has eigenvalue 0 at frequency zero. Constant-ish functions (large blocks) are nearly in the kernel, which is exactly why interior points contribute nothing and why no local charging scheme can see positivity: the form is positive only barely, and only when measured frequency by frequency. The adversarial example with increment -8 is the same phenomenon at small scale.

Harm Derksen's solution (as printed by Kedlaya and Ng) states the identity $g(e^{is}, e^{it}) = 16G(\cos s, \cos t)$ and two case identities for G. Here is where they come from. The table is even in each variable, so only cosines appear; entries at distance 2 produce $\cos 2s = 2\cos^2 s - 1$, which is why G has degree 2 in each variable. Step 3 is that calculation carried out. For the positivity: $G(1,1) = 0$ and $G(0,0) = 0$ are zeros of a nonnegative function, so near them G must look like a square or a product of opposite-signed factors. On $\{uv \geq 0\}$ you pair the terms $uv - u^2v^2$, $u^2 - u^2v$, $v^2 - uv^2$ — each of the three negative monomials absorbs into the positive monomial it divides. On $\{uv \leq 0\}$ the monomial uv is now the enemy, and completing the square $u^2 + uv + v^2 = (u+v)^2 - uv$ turns it into $-uv \geq 0$ times $(1 + u + v + uv) = (1+u)(1+v)$. The case split follows the sign of the one cross term.

The official solution concludes "it is enough to check that $g \geq 0$". That gives $A(S) \geq 0$; the problem asks for > 0, and Step 5 supplies the missing line.

Caveat. Verified by computer: the expansion $g = 16G$ was checked symbolically; the given example evaluates to $A(S) = 42$ with the listed terms; $A(S) > 0$ held on 20 000 random subsets of small boxes; the negative increment -8 in the natural attempt and the square values $A = 12, 32, 64, 128$ for $N = 1, 2, 3, 5$ were computed directly. The formula $32N - 32$ for $N \geq 2$ is read off from computed values $N = 2, 3, 5, 10, 20$ and is not proved here; the proof does not use it.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "$\sum_{(\mathbf{s},\mathbf{s}') \in S \times S} w(\mathbf{s}-\mathbf{s}')$": a sum over pairs of a function of the *difference*. Write down $f = \sum_S x^a y^b$ and note that $A(S)$ is the constant term of $f(x,y)f(1/x,1/y)g(x,y)$.

2. Fourier analysis / positive-definite kernels: evaluate on the unit torus, turn the constant term into an integral of $|f|^2 g$, and reduce to $g \geq 0$.

3. Hesitation occurs after the generating functions are written down, because the torus integral feels like analysis imported into a combinatorics problem. Reinforce with: a constant term of a Laurent polynomial *is* an average over the torus — that is the orthogonality of e^{ins}, which you already use for Fourier series.

4. The trap is induction by adding extreme points, and more generally local charging: the total $\sum w = 0$ makes interior points worthless and boundary effects decisive. A second trap is stopping at $A(S) \geq 0$; strictness needs the no-zero-divisor argument.

5. Slogan: *a quadratic sum over pairs with weights depending only on the difference is a convolution; its sign is the sign of the Fourier transform of the weights.*

6. Decision rule: **to prove** $\sum_{x,y \in S} K(x-y) > 0$ **for all finite** S**, compute** $\hat{K}(\theta) = \sum_d K(d) e^{id \cdot \theta}$ **and prove** $\hat{K} \geq 0$**; then reduce the trigonometric inequality to a polynomial one in the cosines.** Companion rule: when a polynomial inequality on a box has a mixed-sign cross term, split on the sign of that term.

▷ **Cue:** a double sum $\sum_{s,s' \in S} w(s-s')$ over pairs, weights depending only on the difference $\longrightarrow$ **Move:** form $f = \sum_S x^s$; the sum is the constant term of $f(x) f(x^{-1}) g(x)$ with $g = \sum w(d) x^d$

▷ **Cue:** constant term of a Laurent polynomial is needed with a sign $\longrightarrow$ **Move:** set $x = e^{is}$ and average over the torus; $f(x) f(x^{-1})$ becomes $|f|^2$

▷ **Cue:** a kernel symmetric under $d \mapsto -d$ in each coordinate, supported in $|d_i| \leq 2$ $\longrightarrow$ **Move:** its Fourier transform is a polynomial of degree 2 in each $\cos \theta_i$; substitute $\cos 2\theta = 2\cos^2 \theta - 1$

▷ **Cue:** a two-variable polynomial to be shown nonnegative on $[-1,1]^2$ with a cross term uv of indefinite sign $\longrightarrow$ **Move:** split into $uv \geq 0$ and $uv \leq 0$; in each case group terms so every product has a visible sign

▷ **Cue:** a nonnegative integral that must be shown strictly positive, integrand a product of polynomials $\longrightarrow$ **Move:** a nonzero Laurent polynomial cannot vanish on the whole torus, and products of nonzero ones are nonzero

Putnam 2013 — Paper B

2013 B1

2013 B1. For positive integers n, let the numbers $c(n)$ be determined by the rules $c(1) = 1$, $c(2n) = c(n)$, and $c(2n+1) = (-1)^n c(n)$. Find the value of

$$\sum_{n=1}^{2013} c(n)c(n+2).$$

Fields: recursive sequences, telescoping sums, binary structure **Position:** B1 (opener of Paper B)
Answer: The sum equals -1.

Layer 1 THE WHAT — surface analysis

The natural attempt. The definition of c is recursive but concrete, so the first instinct is to make it concrete on paper. You unwind $c(2n) = c(n)$ and $c(2n+1) = (-1)^n c(n)$ far enough to see the shape of the values: $c(1) = 1$, $c(2) = 1$, $c(3) = -1$, $c(4) = 1$, $c(5) = 1$, $c(6) = -1$, $c(7) = 1$, $c(8) = 1$. You notice $c(n) \in \{-1, 1\}$ always (an easy induction: the recursion only ever multiplies by ± 1 starting from $c(1) = 1$), and you start hunting for a period. There is a temptation here to believe c is eventually periodic in n, because the first eight values look almost periodic with period 3: $1, 1, -1, 1, 1, -1, 1, 1, \ldots$. That pattern breaks at $c(9)$: $9 = 2 \cdot 4 + 1$, so $c(9) = (-1)^4 c(4) = c(4) = 1$; but $c(10) = c(5) = 1$ and $c(11) = (-1)^5 c(5) = -1$, and by $c(12) = c(6) = -1$ the period-3 guess is already dead ($1, 1, -1, 1, 1, -1, 1, 1, 1, 1, -1, -1, \ldots$ does not repeat with period 3). At this point the honest brute-force plan is: compute $c(n)$ for all n from 1 to 2015 (you need up to $c(2015)$ for the last term $c(2013)c(2015)$), multiply consecutive-plus-two pairs, and add 2013 signed terms. It is a completely correct plan. It is also a plan for spending the whole exam on arithmetic, and on a competition that punishes exactly this kind of computation, that should be the first alarm bell, not the last.

The model chain. Fix $k \geq 1$ and look at the two products $c(2k)c(2k+2)$ and $c(2k+1)c(2k+3)$ together, because $2k, 2k+2$ are both even and $2k+1, 2k+3$ are both odd, and the recursion treats evens and odds by completely different rules. By the even rule, $c(2k) = c(k)$ and $c(2k+2) = c(k+1)$, so

$$c(2k)c(2k+2) = c(k)c(k+1).$$

By the odd rule, $c(2k+1) = (-1)^k c(k)$ and $c(2k+3) = (-1)^{k+1} c(k+1)$, so

$$c(2k+1)c(2k+3) = (-1)^k (-1)^{k+1} c(k)c(k+1) = -c(k)c(k+1).$$

The two right-hand sides are negatives of each other. Hence for every $k \geq 1$,

$$c(2k)c(2k+2) + c(2k+1)c(2k+3) = 0. \qquad (*)$$

Now $(*)$ is a statement about *pairs* of consecutive integers $2k, 2k+1$, and the sum $\sum_{n=2}^{2013} c(n)c(n+2)$ runs over exactly the integers $n = 2, 3, \ldots, 2013$. Taking $k = 1, \ldots, 1006$: the even values $2k$

give $n = 2, 4, \ldots, 2012$, and the odd values $2k+1$ give $n = 3, 5, \ldots, 2013$. Together these are precisely $n = 2, \ldots, 2013$, each hit exactly once. So

$$\sum_{n=2}^{2013} c(n)c(n+2) = \sum_{k=1}^{1006} \big(c(2k)c(2k+2) + c(2k+1)c(2k+3)\big) = \sum_{k=1}^{1006} 0 = 0.$$

What remains is the one term the pairing does not reach: $n = 1$. Directly, $c(3) = (-1)^1 c(1) = -1$, so $c(1)c(3) = 1 \cdot (-1) = -1$. Therefore

$$\sum_{n=1}^{2013} c(n)c(n+2) = c(1)c(3) + \sum_{n=2}^{2013} c(n)c(n+2) = -1 + 0 = -1.$$

I checked this against a small case before trusting it: for $N = 3$, $\sum_{n=1}^{3} c(n)c(n+2) = c(1)c(3) + c(2)c(4) + c(3)c(5) = (1)(-1) + (1)(1) + (-1)(1) = -1$, and the pairing formula gives $c(1)c(3) + [c(2)c(4) + c(3)c(5)] = -1 + 0$, which matches.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths begin identically: unwind the recursion, notice values are ± 1, notice no short period. The split happens at the moment you decide what a *single term* $c(n)c(n+2)$ is worth computing versus what a *pair* of terms is worth computing. The natural attempt treats $c(n)c(n+2)$ as the atomic unit and tries to evaluate 2013 of them. The model treats $c(2k)c(2k+2)$ and $c(2k+1)c(2k+3)$ *together* as the atomic unit, because that pair — one even-indexed term, one odd-indexed term, sharing the same underlying k — is what the recursion actually reduces to something exact.

> *The recursion gives me two completely different rules depending on parity of the argument. If I want two terms to collapse against each other, I should choose the two terms so that one uses the even rule and the other uses the odd rule on the* same *pair of smaller indices $k, k+1$ — then whatever $c(k)c(k+1)$ turns out to be, it cancels between the two, and I never need to know its value.*

Layer 3 THE WHY

The definition of c is not really a definition of individual values; it is a rule for pushing every question about c down to a question about half-sized arguments, with a sign attached only on the odd branch. That sign is $(-1)^k$ for $c(2k+1)$: it depends on k, the very quantity being halved into. This means that whenever the same k appears in both an even-indexed and an odd-indexed term of the sum, the odd branch's sign is forced to be the exact opposite of what the even branch contributes, *for that k specifically* (the sign flips as $(-1)^k$ vs. $(-1)^{k+1}$ across the two consecutive odd terms in play). Once you see the recursion as "same underlying quantity, opposite forced sign," pairing consecutive integers $2k$ and $2k+1$ stops being a trick and becomes the only sensible unit to compute: it is the smallest window in which the recursion's sign convention is guaranteed to self-cancel. The remaining boundary term, $c(1)c(3)$, survives only because $n = 1$ has no partner below it in the sum — the pairing is a bijection on $\{2, \ldots, 2013\}$, not on $\{1, \ldots, 2013\}$, and a correct telescoping argument must always account for exactly which endpoint escapes the pairing.

> **Caveat.** The claim $c(n) \in \{-1, 1\}$ for all $n \geq 1$ is used implicitly (to know the products are well-defined signed unit products) and is an easy induction on the recursion, but is not spelled out step-by-step above; it follows immediately from $c(1) = 1$ and both recursive rules multiplying by ± 1.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the recursion's parity split: $c(2n) = c(n)$ versus $c(2n+1) = (-1)^n c(n)$. On seeing it, write down: "this function is defined by an even rule and an odd rule that share the same smaller argument n but differ by a sign on the odd side."

2. That cue suggests pairing terms by parity rather than evaluating them individually — specifically, looking for two terms of the sum whose recursive reductions land on the same smaller quantity, so that quantity can be eliminated without ever being computed.
3. Hesitation occurs right after listing $c(1), \ldots, c(8)$ and seeing the false period-3 pattern: it is tempting to keep extending the table looking for the real period. That is reinforced away by remembering that $c(n)$ depends on the full binary expansion of n (each bit contributes a sign depending on the bits below it), so no finite period can exist except by accident on a short prefix.
4. The trap is the false period-3 pattern in the first eight values, which is exactly long enough to look convincing and exactly short enough to be wrong — it breaks at $n = 9$. Trusting a pattern from $2 \cdot (\text{period})$ sample points without deriving *why* it should hold is the failure mode.
5. Recognise this faster next time by asking, on seeing any two-branch recursion (one rule for $2n$, one for $2n + 1$) applied inside a sum indexed consecutively, whether consecutive terms of the sum reduce to the *same* smaller argument under the two different branches. If they do, the telescoping pairing is available immediately, without computing a single table of values.
6. Reusable rule: when a sequence is defined by rules that treat even and odd indices differently but recurse to the same class of smaller arguments, do not evaluate the sequence — pair consecutive terms of the target sum so that one falls under each rule at the same underlying argument, and check whether the two branches force opposite signs. If they do, the sum collapses to its unpaired boundary terms alone.

2013 B2

2013 B2. Let $C = \bigcup_{N=1}^{\infty} C_N$, where C_N denotes the set of those 'cosine polynomials' of the form

$$f(x) = 1 + \sum_{n=1}^{N} a_n \cos(2\pi n x)$$

for which:

(i) $f(x) \geq 0$ for all real x, and

(ii) $a_n = 0$ whenever n is a multiple of 3.

Determine the maximum value of $f(0)$ as f ranges through C, and prove that this maximum is attained.

Fields: Fourier analysis, trigonometric polynomials, optimisation under a constraint **Position:** B2
Answer: The maximum value of $f(0)$ is 3, attained at $N = 2$, $a_1 = \frac{4}{3}$, $a_2 = \frac{2}{3}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. $f(0) = 1 + \sum_{n=1}^{N} a_n$, so the problem is really: maximise $S = \sum a_n$ subject to $f(x) \geq 0$ for every real x and $a_n = 0$ whenever $3 \mid n$. The natural first move is to treat this as a constrained optimisation over the coefficients $a_1, a_2, \ldots$, with constraint (i) an infinite family of inequalities (one for every x). You might try Lagrange multipliers, but the constraint set is infinite-dimensional in x and you have no obvious finite parametrisation of "which x are binding." A more concrete instinct: fix N small, write $f(x) \geq 0$ as a polynomial inequality in $c = \cos(2\pi x) \in [-1, 1]$ (using multiple-angle formulas to express $\cos(2\pi n x)$ as a Chebyshev polynomial in c), and then ask for the maximum of the linear functional S over the region of coefficient-space where that polynomial in c stays nonnegative on $[-1, 1]$. This is a real

approach — it is exactly how one would attack $N = 1$ or $N = 2$ by hand — but for $N = 1$, condition (ii) already forces $a_1 = 0$ (since $3 \nmid 1$ doesn't help, wait: 1 is not a multiple of 3, so a_1 is *allowed*; it's $a_3, a_6, \ldots$ that vanish), so $N = 1$ gives $f(x) = 1 + a_1 \cos(2\pi x) \geq 0$ for all x forces $|a_1| \leq 1$, hence $f(0) \leq 2$. Chasing this case by case, increasing N and re-deriving the Chebyshev substitution and the resulting nonnegativity region each time, is a slow climb with no guarantee you have found the true supremum over all N at once — you would need to prove a uniform bound holds for every N simultaneously, which the case-by-case method does not, by itself, deliver.

The model chain. Condition (ii) is not there to make the problem harder; it is there to create a special point where the whole infinite family of coefficients behaves identically. Evaluate f at $x = \frac{1}{3}$. For any integer n,

$$\cos\left(\tfrac{2\pi n}{3}\right) = \begin{cases} 1 & 3 \mid n \\ -\frac{1}{2} & 3 \nmid n. \end{cases}$$

Since $a_n = 0$ whenever $3 \mid n$, the value of $a_n \cos(2\pi n/3)$ is $-a_n/2$ for *every* n from 1 to N — when $3 \mid n$ this is $0 \cdot 1 = 0 = -a_n/2$ automatically, and when $3 \nmid n$ it is $a_n \cdot (-\frac{1}{2})$ directly. So

$$f\left(\tfrac{1}{3}\right) = 1 + \sum_{n=1}^{N} a_n \cos\left(\tfrac{2\pi n}{3}\right) = 1 - \tfrac{1}{2} \sum_{n=1}^{N} a_n = 1 - \tfrac{S}{2}.$$

Condition (i) applied at this one point gives $f(1/3) \geq 0$, i.e. $1 - S/2 \geq 0$, i.e. $S \leq 2$. Hence

$$f(0) = 1 + S \leq 3$$

for every $f \in C$, uniformly in N — one evaluation handles all N at once, because condition (ii) guarantees $\cos(2\pi n/3)$ contributes the same coefficient $-\frac{1}{2}$ to every surviving term regardless of how large N is.

To show 3 is attained, exhibit $f \in C_2$ with $a_1 = \frac{4}{3}$, $a_2 = \frac{2}{3}$ (note $3 \nmid 1, 3 \nmid 2$, so (ii) holds vacuously — there is no a_3 term to set to zero). Using $\cos(4\pi x) = 2\cos^2(2\pi x) - 1$ and writing $c = \cos(2\pi x)$,

$$f(x) = 1 + \tfrac{4}{3}c + \tfrac{2}{3}(2c^2 - 1) = \tfrac{1}{3} + \tfrac{4}{3}c + \tfrac{4}{3}c^2 = \tfrac{1}{3}(2c + 1)^2 \geq 0$$

for all x, since it is $\frac{1}{3}$ times a perfect square. At $x = 0$, $c = 1$, so $f(0) = \frac{1}{3}(3)^2 = 3$. I verified this by direct substitution back into the original expression: $1 + \frac{4}{3} + \frac{2}{3} = 1 + 2 = 3$, agreeing with the squared form. So the bound $f(0) \leq 3$ is tight and the maximum is attained.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths agree that the problem reduces to bounding $S = \sum a_n$. They split at the choice of *how* to use condition (i). The natural attempt treats (i) as a family of constraints to be understood globally — turning $f \geq 0$ on $[-1, 1]$ (in the variable $c = \cos 2\pi x$) into a full nonnegative-polynomial characterisation, case by case in N. The model instead asks: is there a *single* value of x at which (i) alone, with no further structure, already forces the bound I want? That question is only answerable because of condition (ii) — without it, no single evaluation point gives a clean bound, because different a_n would see different values of $\cos(2\pi n x)$ at any fixed x.

> *Condition (ii) kills exactly the frequencies* $n \equiv 0 \pmod 3$, *and* cos *evaluated at* $x = 1/3$ *takes exactly two values depending on* n mod 3*: it is* 1 *when* $n \equiv 0$ *and* $-\frac{1}{2}$ *otherwise. The hypothesis has already removed the one case where the cosine would differ from the rest — so if I evaluate at* $x = 1/3$*, every surviving coefficient gets multiplied by the* same *number,* $-\frac{1}{2}$*, and the whole sum collapses to a multiple of* S *itself.*

Layer 3 THE WHY

The reason $x = 1/3$ is not a guess but a forced choice is that $\cos(2\pi nx)$, as a function of n for fixed rational $x = p/q$, is periodic in n with period q and takes only finitely many values on each residue class mod q. Condition (ii) singles out residue class $n \equiv 0 \pmod 3$ as special (coefficient forced to zero); this is exactly the residue class at which $\cos(2\pi n/3)$ also happens to be the outlier value (1, versus $-\frac{1}{2}$ for the other two classes). The problem's hypothesis and the arithmetic of $\cos(2\pi n/3)$ are aligned on the same modulus, 3. Once you notice that the vanishing condition and a natural trigonometric identity share a modulus, evaluating at the corresponding rational point is not a trick, it is the unique way to make an infinite linear functional of the a_n collapse to a scalar multiple of their sum. Kedlaya and Ng's solution states this evaluation with no further comment; here is where that comes from: it is the only rational $x = p/3$ (other than $x = 0$ itself, which is useless) at which the two-valued nature of $\cos(2\pi n/3)$ interacts with condition (ii) to erase the dependence on n entirely. The equality case, similarly, is not found by guesswork: knowing the bound comes from forcing $f(1/3) = 0$, you look for an f that vanishes exactly at $x = 1/3$ (and, by the evenness of cosine polynomials, at $x = -1/3 \equiv 2/3$) and is a perfect square elsewhere — a sum-of-squares certificate is the standard witness for a tight nonnegativity bound, and $(2\cos 2\pi x + 1)^2$ is exactly the smallest cosine polynomial vanishing at $x = \pm 1/3$.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is condition (ii): $a_n = 0$ for n a multiple of 3. On seeing it, write down: "this constraint is stated in terms of divisibility by 3; ask what special property the number 3 has for a trigonometric identity at rational multiples of π."

2. That cue suggests evaluating f at $x = 1/3$ (or another rational point with denominator 3), since $\cos(2\pi nx)$ is periodic in n modulo the denominator of x, and the hypothesis is phrased modulo the same number.

3. Hesitation occurs right before committing to evaluate at $x = 1/3$ specifically, rather than trying $x = 1/6$ or another point: it feels arbitrary until you check that $\cos(2\pi n/3)$ genuinely takes a uniform value on the surviving frequencies. That is reinforced away by computing the two cases directly ($n \equiv 0$ gives 1, else gives $-1/2$) and observing this exactly matches the frequencies condition (ii) kills.

4. The trap is treating condition (i) as needing to be understood in full generality (as a polynomial-nonnegativity condition on all of $[-1, 1]$) before any progress can be made, when in fact a single well-chosen evaluation point extracts the entire bound in one line.

5. Recognise this faster next time by checking, whenever a problem imposes a vanishing condition modulo some integer m on the coefficients of a trigonometric or power series, whether m also governs the periodicity of the basis functions themselves ($\cos(2\pi nx)$ has period m in n at $x = k/m$). If so, evaluating at $x = k/m$ is very likely to collapse the constraint to a single inequality.

6. Reusable rule: when a linear constraint kills coefficients on a residue class mod m, and the functional you want to bound is a sum of the surviving coefficients, evaluate the generating object at the rational point of denominator m that makes the basis functions constant on each residue class. This turns an infinite-dimensional nonnegativity constraint into a single linear inequality, and the equality case is found by seeking a perfect-square certificate vanishing exactly there.

2013 B3

2013 B3. Let $\mathcal{P}$ be a nonempty collection of subsets of $\{1, \ldots, n\}$ such that:

(i) if $S, S' \in \mathcal{P}$, then $S \cup S' \in \mathcal{P}$ and $S \cap S' \in \mathcal{P}$, and

(ii) if $S \in \mathcal{P}$ and $S \neq \emptyset$, then there is a subset $T \subset S$ such that $T \in \mathcal{P}$ and T contains exactly one fewer element than S.

Suppose that $f : \mathcal{P} \to \mathbb{R}$ is a function such that $f(\emptyset) = 0$ and

$$f(S \cup S') = f(S) + f(S') - f(S \cap S') \text{ for all } S, S' \in \mathcal{P}.$$

Must there exist real numbers $f_1, \ldots, f_n$ such that

$$f(S) = \sum_{i \in S} f_i$$

for every $S \in \mathcal{P}$?

Fields: combinatorics, lattice theory, modular set functions **Position:** B3 **Answer:** Yes; such $f_1, \ldots, f_n$ always exist.

Layer 1 THE WHAT — surface analysis

The natural attempt. The target formula $f(S) = \sum_{i \in S} f_i$ says f should be additive over the elements of S. The natural guess for f_i is the marginal value of adding element i to something: try $f_i := f(\{i\})$ when $\{i\} \in \mathcal{P}$. This works beautifully on the trivial example $\mathcal{P} = 2^{\{1,\ldots,n\}}$ (the full power set), where every singleton is present and induction on $|S|$ using the modularity relation reconstructs f exactly. The trouble starts once you take $\mathcal{P}$ to be something sparser. Take $n = 2$ and $\mathcal{P} = \{\emptyset, \{1,2\}\}$: this is closed under $\cup, \cap$ (only one nonempty set), and condition (ii) requires a subset of $\{1,2\}$ one element smaller that lies in $\mathcal{P}$ — but no singleton is in $\mathcal{P}$ at all! Rechecking the definition, condition (ii) demands *some* $T \subset S$ with $T \in \mathcal{P}$ and $|T| = |S| - 1$; if $\mathcal{P} = \{\emptyset, \{1,2\}\}$ then for $S = \{1,2\}$ we'd need a one-element $T \in \mathcal{P}$, which is absent — so this $\mathcal{P}$ actually violates (ii) and is not a legal example. Chasing legal examples, you realise $\mathcal{P}$ can still omit *some* singletons while containing others tangled up inside larger sets, and there is no guarantee $\{i\}$ itself is ever in $\mathcal{P}$ for a given i that nonetheless appears inside some set of $\mathcal{P}$. So "define $f_i = f(\{i\})$" is not well-posed in general: the natural definition of f_i has to be more indirect, and once you see that, the honest question is whether *any* consistent choice of f_i can exist at all — which pushes toward trying to disprove the statement with a sparse counterexample before believing it is true.

The model chain. Fix $i \in \{1, \ldots, n\}$ and suppose i appears in *some* set of $\mathcal{P}$ (if it appears in none, set $f_i = 0$, which is clearly forced since no constraint ever touches i). Among all $S \in \mathcal{P}$ containing i, choose one, S, of minimum cardinality. By (ii), there is $T \subset S$ with $T \in \mathcal{P}$, $|T| = |S| - 1$. I claim $i \notin T$: if it were, T would be a smaller set in $\mathcal{P}$ containing i, contradicting the minimality of S. So $T = S \setminus \{i\}$, i.e. S is the disjoint union of T and $\{i\}$, both in $\mathcal{P}$. This is **Lemma 1**: every i appearing anywhere in $\mathcal{P}$ has a pair $(S, T) \in \mathcal{P}^2$ with $S = T \sqcup \{i\}$.

Define $f_i := f(S) - f(T)$ for such a pair. This is only a valid definition if it does not depend on *which* pair (S, T) was chosen — this is **Lemma 2**, and it is where condition (i) (closure under $\cup, \cap$) actually gets used. Suppose $S_1 = T_1 \sqcup \{i\}$ and $S_2 = T_2 \sqcup \{i\}$ are two such decompositions, all four sets in $\mathcal{P}$. Apply modularity to the pair (S_1, T_2) and to the pair (T_1, S_2):

$$f(T_1 \cup T_2 \cup \{i\}) = f(S_1 \cup T_2) = f(S_1) + f(T_2) - f(S_1 \cap T_2) = f(S_1) + f(T_2) - f(T_1 \cap T_2),$$

since $i \notin T_2$ forces $S_1 \cap T_2 = (T_1 \cup \{i\}) \cap T_2 = T_1 \cap T_2$. Symmetrically,

$$f(T_1 \cup T_2 \cup \{i\}) = f(T_1 \cup S_2) = f(T_1) + f(S_2) - f(T_1 \cap S_2) = f(T_1) + f(S_2) - f(T_1 \cap T_2).$$

Both left-hand sides are the same quantity $f(T_1 \cup T_2 \cup \{i\})$ (this set, and $T_1 \cap T_2$, are in $\mathcal{P}$ by (i) since $\mathcal{P}$ is closed under unions and intersections of its members, and $T_1 \cup T_2 \cup \{i\} = S_1 \cup T_2 = T_1 \cup S_2$). Equating the two right-hand sides and cancelling $-f(T_1 \cap T_2)$ from both sides:

$$f(S_1) + f(T_2) = f(T_1) + f(S_2) \implies f(S_1) - f(T_1) = f(S_2) - f(T_2).$$

So f_i is well-defined, independent of the chosen witnessing pair.

Finally, prove $f(S) = \sum_{i \in S} f_i$ for every $S \in \mathcal{P}$ by induction on $|S|$. The base case: $\emptyset \in \mathcal{P}$ (repeated application of (ii) to any nonempty member of $\mathcal{P}$ eventually reaches $\emptyset$, since cardinalities strictly decrease and (ii) can always be applied again while the set is nonempty), and $f(\emptyset) = 0$ by hypothesis, matching the empty sum. For nonempty $S \in \mathcal{P}$, (ii) gives $T \subset S$, $T \in \mathcal{P}$, $|T| = |S| - 1$; write $S = T \sqcup \{j\}$ for the unique missing element j. By definition (Lemma 1's construction applied to this particular pair) and Lemma 2 (well-definedness), $f_j = f(S) - f(T)$ regardless of which witnessing pair was used to originally define f_j, so $f(S) = f(T) + f_j$. By the induction hypothesis $f(T) = \sum_{i \in T} f_i$, so $f(S) = \sum_{i \in T} f_i + f_j = \sum_{i \in S} f_i$, completing the induction. Hence yes: such $f_1, \ldots, f_n$ always exist.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths start from the same instinct: f_i should be a marginal value, $f(S) - f(S \setminus \{i\})$ for some convenient S. They split at the question of *well-definedness*. The natural attempt picks the most convenient witness ($S = \{i\}$ when available) and never confronts what happens when different witnesses might disagree — because in the motivating example, the full power set, there is only ever one sensible witness, the singleton. The model chain is forced to confront well-definedness head-on because condition (ii) only guarantees *some* one-smaller subset exists, not a canonical one, and different elements $S \in \mathcal{P}$ containing i may witness f_i via unrelated sets T.

> *To prove f_i is well-defined I need two different (S, T) representations of "i added to something" to agree; the natural move is to combine them into one appeal to the modularity identity, by forming $T_1 \cup T_2 \cup \{i\}$ and $T_1 \cap T_2$ — both guaranteed to be back in $\mathcal{P}$ because $\mathcal{P}$ is closed under union and intersection. If I can express the same set two ways, as $S_1 \cup T_2$ and as $T_1 \cup S_2$, modularity applied to each gives two formulas for the same number, and the f_i-candidates must match.*

Layer 3 THE WHY

The modularity equation $f(S \cup S') = f(S) + f(S') - f(S \cap S')$ is exactly the inclusion–exclusion identity satisfied by a finitely additive measure on a lattice of sets, and $\mathcal{P}$ closed under $\cup, \cap$ is precisely a sublattice of $2^{\{1,\ldots,n\}}$. A function that is additive over a lattice in this sense is called a *valuation*; the question "is every valuation on this lattice a sum of coordinate values" is really asking whether the lattice is rich enough, via condition (ii), to expose every coordinate individually as a rank-1 step ($|S|$ to $|S| - 1$) somewhere inside it. Condition (ii) is the hypothesis that guarantees such a step exists for every element that appears at all — without it, $\mathcal{P}$ could consist of large sets only, with no way to isolate any single coordinate's contribution, and the claim would be false (take $\mathcal{P} = \{\emptyset, \{1, 2\}\}$, which does satisfy (i) but genuinely fails (ii), and on which one could define $f(\{1, 2\})$ to be any real number with no constraint forcing it to split as $f_1 + f_2$ for fixed f_1, f_2 independent of the set). The reason well-definedness does not need to be assumed, only proved, is that modularity is a strong enough algebraic constraint to force any two witnesses of the same rank-1 step to report the same increment — this is the archive's real payoff here: modularity plus lattice closure is *equivalent to* the existence of coordinate weights, and the proof of Lemma 2 is the general mechanism (build the join and the meet of the two witnessing pairs, apply modularity twice, cancel the shared intersection term) by which any valuation on a lattice satisfying condition (ii) is shown to decompose. This is the same

mechanism, in embryo, behind the fact that a modular function on the subsets of a matroid decomposes along a basis, and behind rank functions being additive along maximal chains.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the target formula $f(S) = \sum_{i \in S} f_i$ itself: it says f is determined by per-element weights. On seeing it, write down: "each f_i should be some marginal difference $f(S) - f(S \setminus \{i\})$ — the question is whether that difference is the same no matter which S I use."
2. That cue suggests looking at condition (ii) specifically for where a one-element-smaller witness is guaranteed to exist, and then asking whether two different witnesses for the same element could disagree — which is a well-definedness question, and well-definedness questions in the presence of a union/intersection-closed structure almost always resolve via the closure property applied to the two witnesses together.
3. Hesitation occurs at exactly the point of trying $f_i = f(\{i\})$ and discovering singletons need not be in $\mathcal{P}$. That is reinforced away by not abandoning the marginal-value idea, but relaxing it: instead of requiring the specific witness $\{i\}$, allow *any* pair (S, T) with $S = T \sqcup \{i\}$, both in $\mathcal{P}$, and prove such a pair always exists (Lemma 1) using minimality.
4. The trap is believing the sparse example $\mathcal{P} = \{\emptyset, \{1,2\}\}$ is a valid instance of the hypotheses and hunting for a value of $f(\{1,2\})$ that cannot split as $f_1 + f_2$; in fact this $\mathcal{P}$ violates condition (ii) (no one-smaller subset of $\{1,2\}$ lies in $\mathcal{P}$), so it does not test the theorem at all, and time spent on it is wasted until the violation is noticed.
5. Recognise this faster next time by checking, whenever a problem gives you closure under $\cup, \cap$ plus a rank-reduction axiom like (ii), that the natural object being described is a finite lattice and that the functional equation given is the modularity/inclusion-exclusion identity — at which point "does this decompose along coordinates" is a standard valuation-on-a-lattice question with a standard two-step answer: existence of a witness (from the rank-reduction axiom), then well-definedness (from applying the functional equation to the join and meet of two witnesses).
6. Reusable rule: to show a modular set function decomposes as a sum of per-element weights, first use the rank-reduction hypothesis to produce, for each element, at least one pair of lattice members differing by exactly that element; then prove well-definedness by taking two such pairs, forming their union and intersection (both back in the lattice by closure), and applying the modularity identity twice to the same resulting set — the shared intersection term cancels and forces the two candidate weights to agree.

2013 B4

2013 B4. For any continuous real-valued function f defined on the interval $[0,1]$, let

$$\mu(f) = \int_0^1 f(x)\,dx, \ \mathrm{Var}(f) = \int_0^1 (f(x) - \mu(f))^2\,dx,$$
$$M(f) = \max_{0 \le x \le 1} |f(x)|.$$

Show that if f and g are continuous real-valued functions defined on the interval $[0,1]$, then

$$\mathrm{Var}(fg) \le 2\mathrm{Var}(f)M(g)^2 + 2\mathrm{Var}(g)M(f)^2.$$

Fields: real analysis / inequalities; variance as a least-squares minimum; the product rule for differences
Position: B4 — mid-paper, and an honest one: the official solution is a page of expansion, the right idea is four lines **Answer:** Since $\mathrm{Var}(h) \le \int_0^1 (h-c)^2$ for every constant c, take $c = \mu(f)\mu(g)$ and split $fg - \mu(f)\mu(g) = (f - \mu(f))g + \mu(f)(g - \mu(g))$; then $(a+b)^2 \le 2a^2 + 2b^2$ and $|\mu(f)| \le M(f)$ finish it.

Layer 1 THE WHAT — surface analysis

The natural attempt. Write $f_0 = f - \mu(f)$ and $g_0 = g - \mu(g)$, so $\int f_0 = \int g_0 = 0$, $\int f_0^2 = \mathrm{Var}(f)$, $\int g_0^2 = \mathrm{Var}(g)$. (All integrals are over $[0,1]$.) The strong student's reflex is: the left side is a variance, so expand it. With $\alpha = \mu(f)$, $\beta = \mu(g)$,

$$fg = (f_0 + \alpha)(g_0 + \beta) = f_0 g_0 + \beta f_0 + \alpha g_0 + \alpha\beta, \qquad \mu(fg) = \int f_0 g_0 + \alpha\beta,$$

so, putting $\rho = \int f_0 g_0$,

$$\mathrm{Var}(fg) = \int \big(f_0 g_0 - \rho + \beta f_0 + \alpha g_0\big)^2.$$

Now expand the square: six squared-and-cross terms, $\int f_0^2 g_0^2 - \rho^2 + \beta^2 \mathrm{Var}(f) + \alpha^2 \mathrm{Var}(g) + 2\beta \int f_0^2 g_0 + 2\alpha \int f_0 g_0^2 + 2\alpha\beta\rho$. The student now has to bound a cubic moment $\int f_0^2 g_0$, a quartic moment $\int f_0^2 g_0^2$, and a cross term $\alpha\beta\rho$ of unknown sign, each against a right side that contains $M(f)$ and $M(g)$ — quantities that never appeared in the expansion. They try $|g_0| \le 2M(g)$, which costs a factor 4 and overshoots the constant 2. They try Cauchy–Schwarz on ρ. Each estimate loses something, and there is no way to see in advance which losses the constant 2 can afford. This is the official solution's route, and it does close (Kedlaya–Ng regroup to $\int (f_0^2 g^2 + f^2 g_0^2)$, subtract, and use $(\beta f_0 - \alpha g_0)^2 \ge 0$), but only after a page of bookkeeping whose every line has to be checked. It is zombie maths: correct, and not chosen.

The model chain. **Step 1: variance is a minimum.** For any integrable h and any real constant c,

$$\int_0^1 (h - c)^2 = \int_0^1 \big((h - \mu(h)) + (\mu(h) - c)\big)^2 = \mathrm{Var}(h) + (\mu(h) - c)^2,$$

because the cross term is $2(\mu(h) - c)\int (h - \mu(h)) = 0$. Hence $\mathrm{Var}(h) \le \int (h - c)^2$ for every c.

Step 2: choose c and split. Take $h = fg$ and $c = \mu(f)\mu(g)$. Then, pointwise,

$$fg - \mu(f)\mu(g) = \big(f - \mu(f)\big)g + \mu(f)\big(g - \mu(g)\big) = f_0\, g + \mu(f)\, g_0.$$

Step 3: estimate. Using $(a+b)^2 \le 2a^2 + 2b^2$ (equivalent to $(a-b)^2 \ge 0$),

$$\mathrm{Var}(fg) \le \int_0^1 \big(f_0 g + \mu(f) g_0\big)^2 \le 2\int_0^1 f_0^2 g^2 + 2\mu(f)^2 \int_0^1 g_0^2.$$

Since $g^2 \le M(g)^2$ pointwise, $\int f_0^2 g^2 \le M(g)^2 \mathrm{Var}(f)$. Since $|\mu(f)| \le \int |f| \le M(f)$, $\mu(f)^2 \int g_0^2 \le M(f)^2 \mathrm{Var}(g)$. Therefore

$$\mathrm{Var}(fg) \le 2\,\mathrm{Var}(f) M(g)^2 + 2\,\mathrm{Var}(g) M(f)^2. \qquad \square$$

A sharper form, for free. Replace $(a+b)^2 \le 2a^2 + 2b^2$ by Minkowski's inequality $\|u + v\|_2 \le \|u\|_2 + \|v\|_2$: this gives $\sqrt{\mathrm{Var}(fg)} \le M(g)\sqrt{\mathrm{Var}(f)} + M(f)\sqrt{\mathrm{Var}(g)}$, and the stated bound follows from it by $(p+q)^2 \le 2p^2 + 2q^2$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains write $f = f_0 + \mu(f)$, $g = g_0 + \mu(g)$. They split at the first line about the left-hand side.

(N) "$\mathrm{Var}(fg) = \int (fg - \mu(fg))^2$; compute $\mu(fg)$ and expand." The centring constant is forced to be the true mean, which is $\mu(f)\mu(g) + \int f_0 g_0$, and that extra $\int f_0 g_0$ drags a cross-moment through every later line.

(M) "$\mathrm{Var}(fg) \le \int (fg - c)^2$ for any c; pick the c that makes the algebra factor." With $c = \mu(f)\mu(g)$ the integrand is the product-rule difference $f_0 g + \mu(f) g_0$, two terms, each already the shape of one term on the right.

> *I don't have to subtract the exact mean of fg. Variance is the smallest mean-square distance to a constant, so any constant gives me an upper bound, and an upper bound is what I want. The constant I'd like to subtract is $\mu(f)\mu(g)$, because then $fg - \mu(f)\mu(g)$ splits like a product rule.*

> *The right side has two terms, one with $\mathrm{Var}(f)$ and one with $\mathrm{Var}(g)$, each with a factor 2. That 2 is $(a+b)^2 \le 2a^2 + 2b^2$. So I want $fg - c$ written as a sum of exactly two pieces, one carrying $f - \mu(f)$ and one carrying $g - \mu(g)$.*

- ▷ **Cue:** an upper bound is wanted for a variance, or for any quantity defined as a minimum $\longrightarrow$ **Move:** replace the optimal choice (the mean) by a convenient one; $\mathrm{Var}(h) \le \int (h-c)^2$ for every constant c
- ▷ **Cue:** a difference of products, $fg - \alpha\beta$, where $f \approx \alpha$ and $g \approx \beta$ $\longrightarrow$ **Move:** use the product-rule split $fg - \alpha\beta = (f-\alpha)g + \alpha(g-\beta)$
- ▷ **Cue:** a target of the form $2A + 2B$ $\longrightarrow$ **Move:** it is $(a+b)^2 \le 2a^2 + 2b^2$; write the quantity as a two-term sum whose squares are A and B
- ▷ **Cue:** $\max|f|$ appears on the right but not naturally on the left $\longrightarrow$ **Move:** it enters by the pointwise bound $|g| \le M(g)$ inside an integral, and by $|\mu(f)| \le M(f)$

Layer 3 THE WHY

Why is the model's first line not a trick? Go back to the definition. $\mathrm{Var}(h) = \int (h - \mu(h))^2$ is the mean-square distance from h to its mean, and the one fact every student knows about the mean is that it is the *best* constant in that sense — it is the least-squares fit by a constant, the orthogonal projection of h onto the constants. Step 1 is exactly Pythagoras for that projection: $\|h-c\|^2 = \|h-\mu(h)\|^2 + \|\mu(h)-c\|^2$. So the definition of variance, read as geometry, says "the variance is the minimum of $\int (h-c)^2$". And when a quantity is a minimum, every admissible choice bounds it from above. An upper bound on a minimum is always free; you are allowed to be suboptimal. That is the whole of the first critical thought, and it comes from the definition, not from inspiration.

Why that constant? Because the problem told you. The right-hand side is built from $\mathrm{Var}(f)$ and $\mathrm{Var}(g)$, that is from $f - \mu(f)$ and $g - \mu(g)$ separately. The left-hand side is about fg. The only natural bridge from "f close to $\mu(f)$ and g close to $\mu(g)$" to "fg close to something" is that fg is then close to $\mu(f)\mu(g)$ — the same reasoning that proves the product rule in calculus, $uv - u_0v_0 = (u-u_0)v + u_0(v-v_0)$. The true mean $\mu(fg)$ differs from $\mu(f)\mu(g)$ by the covariance $\int f_0 g_0$, and the natural attempt pays for insisting on the exact mean by carrying that covariance through every line. The model declines the payment.

Why does the constant 2 come out exactly? Because $(a+b)^2 \le 2a^2 + 2b^2$ is the cost of splitting a square into two pieces, and the problem has two pieces. When the target is a sum of two terms each multiplied by 2, that inequality is almost announced. The asymmetry of the split — $f_0\, g$ with the full g, but $\mu(f)\, g_0$ with only the mean of f — is what makes the M's land correctly: the full g is bounded by $M(g)$, and the mean $\mu(f)$ is bounded by $M(f)$. Split the other way, $f\, g_0 + \mu(g) f_0$, and you get the same bound with the roles swapped; either works.

Kedlaya and Ng's solution begins by bounding $\int f_0^2(M(g)^2 - g^2) \ge 0$ and then spends a page showing that $\int (f_0^2 g^2 + f^2 g_0^2)$ exceeds $\mathrm{Var}(fg)$ up to a covariance term. Here is where that comes from: it is the same product-rule split, but with the exact mean kept, so the covariance has to be fought separately at the end. The minimum property of variance removes that fight entirely.

Caveat. Checked numerically on 2000 random pairs of smooth functions (200-point quadrature): both the stated inequality and the sharper Minkowski form $\sqrt{\mathrm{Var}(fg)} \le$

$M(g)\sqrt{\mathrm{Var}(f)} + M(f)\sqrt{\mathrm{Var}(g)}$ held in every case. The written proof does not depend on the numerics.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the word "Var" on the side of the inequality that must be *small*. On seeing it, write $\mathrm{Var}(h) = \min_c \int (h-c)^2 \le \int (h-c)^2$.
2. The cue suggests choosing a convenient centring constant and then a two-term decomposition matched to the two terms of the right side, bounded by $(a+b)^2 \le 2a^2 + 2b^2$.
3. Hesitation occurs at the moment of writing $\mu(fg)$: it feels illegitimate to subtract anything other than the true mean. It is reinforced away by Step 1, the one-line Pythagoras identity, which shows that the true mean is only the best choice, not the only admissible one.
4. The trap is to expand $\mathrm{Var}(fg)$ with the exact mean and then bound a cubic moment, a quartic moment and a covariance separately; each crude bound (such as $|g_0| \le 2M(g)$) spends a constant factor the problem cannot afford.
5. Whenever the quantity to be bounded above is defined as a minimum or infimum — variance, distance to a set, least-squares error, an eigenvalue by Rayleigh quotient — ask first: *which non-optimal choice makes the algebra collapse?*
6. Reusable decision rule: *to bound a minimum from above, do not compute the minimiser; substitute the most convenient competitor.* For products, the convenient competitor is the product of the parts' centres, and the error splits as in the product rule.

2013 B5

2013 B5. Let $X = \{1, 2, \ldots, n\}$, and let $k \in X$. Show that there are exactly $k \cdot n^{n-1}$ functions $f : X \to X$ such that for every $x \in X$ there is a $j \ge 0$ such that $f^{(j)}(x) \le k$. [Here $f^{(j)}$ denotes the j^{th} iterate of f, so that $f^{(0)}(x) = x$ and $f^{(j+1)}(x) = f(f^{(j)}(x))$.]

Fields: combinatorics / enumeration; functional digraphs; rooted forests and Cayley's formula **Position:** B5 — hard, because the count is Cayley's formula in disguise and nothing in the statement says "tree" **Answer:** The values $f(1), \ldots, f(k)$ are irrelevant (n^k choices); the values on $\{k+1, \ldots, n\}$ form a rooted forest on X with root set exactly $\{1, \ldots, k\}$, and there are $k\,n^{n-k-1}$ of those, so the total is $n^k \cdot k\,n^{n-k-1} = k\,n^{n-1}$.

Layer 1 THE WHAT — surface analysis

Throughout, $[k] = \{1, \ldots, k\}$, and call f *good* if every $x \in X$ has some $f^{(j)}(x) \in [k]$.

The natural attempt. Test the extremes. For $k = n$ every f is good: $n^n = n \cdot n^{n-1}$. For $k = n-1$, f fails only if the orbit of n never leaves $\{n\}$, that is $f(n) = n$; so the count is $n^n - n^{n-1} = (n-1)n^{n-1}$. Encouraging. Now $k = n-2$: f fails iff some orbit stays inside $\{n-1, n\}$ for ever, which happens iff f has a cycle inside $\{n-1, n\}$: $f(n) = n$, or $f(n-1) = n-1$, or $f(n-1) = n$, $f(n) = n-1$. Inclusion–exclusion over these events gives $n^{n-1} + n^{n-1} - n^{n-2} + n^{n-2}$, and the pieces already need care. For general k the bad event is "f has a cycle avoiding $[k]$", and a union over all possible cycles in $\{k+1, \ldots, n\}$ with overlaps is not something to do in an exam.

The next reflex is induction on $n-k$: pass from $[k]$ to $[k+1]$ and count the functions that are good for $[k+1]$ but not for $[k]$. Those are the functions with exactly one cycle avoiding $[k]$, and that cycle passes through $k+1$. Summing over the cycle length ℓ gives the official first solution's

$$\sum_{\ell=1}^{n-k} (n-k-1)(n-k-2)\cdots(n-k-\ell+1)\,(k+\ell)\,n^{n-\ell-1},$$

which must be shown to equal n^{n-1} by a telescoping rearrangement. It works; it is also a sum that a student cannot be expected to know will telescope before writing it, and it proves the answer without explaining it.

The model chain. **Step 1: the values on $[k]$ do not matter.** The condition asks only whether the orbit of x ever *reaches* $[k]$. If $x \in [k]$ it has reached already ($j = 0$). If $x \notin [k]$, let j be the first time $f^{(j)}(x) \in [k]$; the points $x, f(x), \ldots, f^{(j-1)}(x)$ all lie in $\{k+1, \ldots, n\}$, so only the values of f on $\{k+1, \ldots, n\}$ were used. Hence whether f is good depends only on $f|_{\{k+1,\ldots,n\}}$, and the count is n^k times the number of good restrictions $\phi : \{k+1, \ldots, n\} \to X$.

Step 2: good restrictions are rooted forests. Draw an arrow $v \to \phi(v)$ for each $v > k$. Goodness says every $v > k$, following arrows, reaches $[k]$. That is exactly a *rooted forest on X with root set $[k]$*: every non-root has one parent, and following parents always ends at a root. Let $F(r, N)$ be the number of rooted forests on a fixed N-element vertex set with a fixed r-element root set, $r \geq 1$. We need $F(k, n)$.

Step 3: $F(r, N) = r\, N^{N-r-1}$. Induct on N. Put $m = N - r$ (the number of non-roots). If $m = 0$, $F = 1 = r \cdot r^{-1}$. If $m \geq 1$, let T be the set of non-roots whose parent is a root. $T \neq \emptyset$: from any non-root follow parents; the last non-root before a root is in T. Fix T with $|T| = j$, $1 \leq j \leq m$. The parents of vertices of T are any roots: r^j choices. Every other non-root has, by definition of T, a parent among the non-roots, and its parent chain reaches T (the vertex just before it first hits a root lies in T). So the rest of the structure is a rooted forest on the m non-roots with root set T; conversely any such forest together with parents for T gives a forest counted by $F(r, N)$. As $m < N$, induction gives $F(j, m) = j\, m^{m-j-1}$, and

$$F(r,N) = \sum_{j=1}^{m} \binom{m}{j} r^j \cdot j\, m^{m-j-1} = \sum_{j=1}^{m} \binom{m-1}{j-1} r^j m^{m-j} = r \sum_{i=0}^{m-1} \binom{m-1}{i} r^i m^{m-1-i} = r(r+m)^{m-1},$$

using $\binom{m}{j}\frac{j}{m} = \binom{m-1}{j-1}$ and the binomial theorem. Since $r + m = N$, $F(r, N) = r\, N^{N-r-1}$.

Step 4: assemble. The number of good f is $n^k \cdot F(k, n) = n^k \cdot k\, n^{n-k-1} = k\, n^{n-1}$. □

Sanity check. $k = 1$: good f are "$f(1)$ arbitrary, everything else a spanning tree rooted at 1", and $F(1, n) = n^{n-2}$ is Cayley's formula. The problem at $k = 1$ *is* Cayley's theorem.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start from "the bad functions are those with a cycle avoiding $[k]$". They split at what to do with the values $f(1), \ldots, f(k)$.

(N) Keep them in play and count cycles: the natural attempt treats f as a whole functional digraph, so cycles through $[k]$ must be tracked, and the count becomes an alternating or telescoping sum over cycle structures.

(M) Throw them away: the condition is a first-hitting condition, so f on $[k]$ is never consulted. What remains is a digraph with no out-arrows from $[k]$ — which cannot have a cycle through $[k]$ at all, and so is a forest.

> *The condition only says the orbit gets* into $[k]$ *at some point. Once it's there, I stop looking. So what f does on $[k]$ never matters, and that's where the n^k in $k \cdot n^{n-1} = n^k \cdot k\, n^{n-k-1}$ comes from.*

> *If I cut the arrows leaving $[k]$, every other point has one arrow out and has to drain into $[k]$. One arrow out, no cycles, everything drains to a root: that's a rooted forest. So I'm counting forests with k given roots, and the answer should look like Cayley's n^{n-2}.*

> *To count forests, peel off the layer next to the roots. The vertices hanging directly on a root become the new roots of a smaller forest, and the recursion closes by the binomial theorem.*

- ▷ **Cue:** a condition "for every x there is $j \geq 0$ with $f^{(j)}(x) \in A$" $\longrightarrow$ **Move:** it is a first-hitting condition: the values of f on A are irrelevant; factor out $|X|^{|A|}$ and delete the arrows leaving A
- ▷ **Cue:** a function $X \to X$ in which every non-special point must eventually reach the special set $\longrightarrow$ **Move:** draw the functional digraph; with the special set's out-arrows deleted, it is a rooted forest with the special set as roots
- ▷ **Cue:** an answer of the form n^{n-1}, n^{n-2} or $k\,n^{n-k-1}$ in a counting problem $\longrightarrow$ **Move:** suspect trees: Cayley's formula or its forest version
- ▷ **Cue:** counting rooted trees or forests with prescribed roots $\longrightarrow$ **Move:** peel the layer of vertices adjacent to the roots; they become the roots of a smaller forest; close with the binomial theorem

Layer 3 THE WHY

Why must the n^k be split off? Because of the words "there is a $j \geq 0$". An existential over time is a *hitting* condition, and hitting conditions only ever look at the path up to the first hit. This is the same logic as a stopping time in probability, or as "absorbing states" in a Markov chain: once the walk is absorbed, what the transition rule does next is not part of the event. So the definition of the event tells you the values $f(1), \ldots, f(k)$ are free, and the arithmetic of the answer confirms it: $k\,n^{n-1}$ factors as $n^k \cdot k\,n^{n-k-1}$, and the second factor no longer depends on n^k-many irrelevant choices.

Why a forest? A function is a digraph in which every vertex has out-degree exactly one. Such a digraph, in each connected piece, is a cycle with trees hanging on it — that is what "out-degree one" forces, because following arrows from any vertex must eventually repeat. Delete the out-arrows from $[k]$: now the vertices of $[k]$ have out-degree zero and every other vertex has out-degree one. If every vertex's path reaches $[k]$, no path can enter a cycle, so there are no cycles; a digraph with out-degree at most one and no cycles is, by definition, a forest of in-trees with roots at the out-degree-zero vertices. Nothing about this is a trick. It is what "function" plus "every orbit reaches $[k]$" mean when drawn.

Why peel the layer next to the roots? A recursive count of trees needs a cut that leaves a structure of the same kind, and the vertices adjacent to the roots are the only layer whose removal does that: delete the old roots, and those vertices are roots of a smaller forest of exactly the same type. That self-similarity is the whole reason the formula $F(r, N)$ carries the root count r as a parameter — you need the general r even if you only care about $r = 1$, because the peeled layer can have any size. This is the reason the problem is posed with general k: the general-k statement is the inductive strengthening that makes Cayley's formula provable.

Why does the binomial theorem appear? After peeling, the count is a sum over the size j of the new root layer of $\binom{m}{j} r^j \cdot (\text{forests on the rest})$; the factor j/m from the inductive formula converts $\binom{m}{j}$ into $\binom{m-1}{j-1}$, and what remains is $(r+m)^{m-1}$ expanded. Kedlaya–Ng's second solution is the same computation with the function recast as a "contraction"; here is where it comes from: their contraction at $\{*\} \cup S$ is exactly deleting the old roots and promoting the next layer.

> **Caveat.** Verified by exhaustive enumeration: for every $1 \leq k \leq n \leq 6$ the number of good functions equals $k\,n^{n-1}$. The forest formula $F(r, N) = r\,N^{N-r-1}$ is a standard result (the forest version of Cayley's formula); the inductive proof above is complete and self-contained.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "there is a $j \geq 0$ such that $f^{(j)}(x) \leq k$". On seeing it, write: *only*

the path until the first hit matters, so $f(1), \dots, f(k)$ are free; and factor the target as $k\,n^{n-1} = n^k \cdot k\,n^{n-k-1}$.

2. The cue suggests drawing the functional digraph, deleting the out-arrows of $[k]$, and recognising a rooted forest; then counting forests with prescribed roots by peeling the root-adjacent layer.

3. Hesitation occurs when the count reduces to forests and the student does not know the forest formula. It is reinforced away by noting that the formula need not be known: $k = 1$ must give Cayley's n^{n-2}, which suggests $k\,n^{n-k-1}$, and the peeling recursion proves the guess in four lines.

4. The trap is to count the complement — functions with a cycle avoiding $[k]$ — by inclusion–exclusion over cycles, or to induct on k through cycle lengths; both keep the irrelevant values on $[k]$ in play and lead to sums whose closed form is not visible in advance.

5. Recognise it faster by reading the answer: a power n^{n-1} or n^{n-2} in a counting problem about functions or graphs is Cayley's signature. When it appears, look for the tree.

6. Reusable decision rule: *in an eventual-hitting condition, the behaviour after the hit is free — factor it out first.* And: *to count a recursive structure, find the cut that leaves a smaller copy of the same structure, and carry whatever parameter (here the number of roots) that cut produces.*

2013 B6

2013 B6. Let $n \geq 1$ be an odd integer. Alice and Bob play the following game, taking alternating turns, with Alice playing first. The playing area consists of n spaces, arranged in a line. Initially all spaces are empty. At each turn, a player either

- places a stone in an empty space, or
- removes a stone from a nonempty space s, places a stone in the nearest empty space to the left of s (if such a space exists), and places a stone in the nearest empty space to the right of s (if such a space exists).

Furthermore, a move is permitted only if the resulting position has not occurred previously in the game. A player loses if he or she is unable to move. Assuming that both players play optimally throughout the game, what moves may Alice make on her first turn?

Fields: combinatorics / combinatorial game theory; monovariants; a modular invariant (sum of occupied labels mod $n+1$) **Position:** B6 — genuinely hard: two separate ideas are needed, an endgame parity count and a modular weight, and neither is suggested by small cases alone **Answer:** Alice's only winning first move is to place a stone in the central space $\frac{n+1}{2}$. Outside the endgame, the player to move loses exactly when the sum of the occupied positions is $\equiv \frac{n+1}{2} \pmod{n+1}$.

Layer 1 THE WHAT — surface analysis

Number the spaces $1, \dots, n$, write $c = \frac{n+1}{2}$, and let P_i be the set of positions with i stones. A move of the first kind is *type 1*, of the second kind *type 2*.

The natural attempt. Play small cases. $n = 1$: Alice must place the stone; Bob's only move would recreate the empty board, which has occurred, so Alice wins. $n = 3$: Alice has three first moves; by hand, the middle one wins and the end ones lose, but the game trees are already a few dozen positions with the repetition rule to track. The obvious conjecture is "centre, then mirror Bob", and a strong student reaches for the symmetry strategy: after Alice takes the centre, whatever Bob does, Alice does the mirror image.

The wall arrives at once. The mirror of a type-2 move is a type-2 move, but it need not be legal: if Bob's move changes the blocks near the centre, the "mirror" may interact with Bob's stones, and the repetition rule may forbid it. Worse, the problem asks for *all* winning first moves,

so the student must also prove that every off-centre move loses — and for that a symmetry strategy for Bob is useless, since the position after an off-centre move is not symmetric. The student is left with a correct guess and no invariant to prove it.

The model chain. **Step 1: stone counts.** A type-1 move from P_i lands in P_{i+1}. A type-2 move at an occupied s removes one stone and adds one for each side of s that contains an empty space. If the board is not full, some empty space lies to the left or right of s, so type 2 from P_i, $i < n$, lands in $P_i \cup P_{i+1}$. It stays in P_i exactly when only one side has an empty space, i.e. s lies in an *extremal block* (the maximal run of stones touching an end). From P_n the only moves are type 2, landing in P_{n-1}.

Step 2: no repetition before the end zone. Let $Z = P_{n-1} \cup P_n$. Take $x \in P_i$, $i \leq n-2$, with extremal blocks of lengths a (left) and b (right). A type-2 move at $s \leq a$ empties s and fills $a+1$; since there are at least two empty spaces, the largest empty space is untouched, so b is unchanged and a becomes $s-1 < a$. Symmetrically on the right. So within P_i ($i \leq n-2$) every move strictly decreases $a+b$, and the stone count never decreases outside Z. Hence no position outside Z can recur, and outside Z the repetition rule forbids nothing. The game outside Z is finite.

Step 3: the end zone decides by parity. $|Z| = n+1$. From P_n, type 2 at s gives the position with hole at s: every element of P_{n-1}. From the position with single hole h, type 1 gives P_n, and type 2 at $s \neq h$ removes s and fills h (the only empty space, on one side of s), giving the position with hole s. So from any position of Z every other position of Z is one move away, and no move leaves Z. Once some player moves into Z, the game continues until all $n+1$ positions of Z have occurred and then stops: exactly n further moves. As n is odd, the last of these is made by the opponent of the player who entered Z, and the entering player is then unable to move. *Whoever first moves into Z loses.* (For $n \geq 3$ the empty board is in $P_0 \not\subset Z$; $n = 1$ is the case above.)

Step 4: the weight. Define the weight $w(x)$ as the sum of the occupied labels, modulo $n+1$. For $x \notin Z$ look at the n possible moves, one per space t:

- t empty: type 1 at t changes w by $+t$.
- t occupied, in the maximal block $\{i+1, \ldots, j-1\}$ with i, j empty or $i = 0$, $j = n+1$: type 2 at t changes w by $i + j - t$, where a missing neighbour at 0 or $n+1$ contributes $0 \equiv n+1$.

In a block, $t \mapsto i + j - t$ is the reflection of the block onto itself. So the n weight changes are a permutation of $\{1, \ldots, n\}$: *every nonzero residue mod $n+1$ is achieved by exactly one move, and 0 by none.*

Step 5: the losing positions. Claim: for $x \notin Z$, the player to move loses iff $w(x) \equiv c$. Induct along the game (Step 2: every move out of $x \notin Z$ goes to a later position). Note that P_n cannot be reached from outside Z in one move (counts rise by at most 1), and every $y \in P_{n-1}$ with hole h has $w(y) \equiv \frac{n(n+1)}{2} - h \equiv c - h \not\equiv c$, because $\frac{n(n+1)}{2} = n \cdot c \equiv -c \equiv c \pmod{n+1}$.

- If $w(x) \not\equiv c$, play the unique move changing w by $c - w(x) \not\equiv 0$. It lands at weight c, hence not in Z, hence (by induction) at a position where the opponent loses.
- If $w(x) \equiv c$, every move changes w by a nonzero amount. It lands either in Z (the mover loses, Step 3) or outside Z at weight $\not\equiv c$ (the opponent wins, by induction).

Step 6: Alice's first move. For $n \geq 3$, from the empty board (weight $0 \not\equiv c$) only type-1 moves exist, and placing at s gives weight s. The winning moves are those reaching weight c: exactly $s = c = \frac{n+1}{2}$, the central space. For $n = 1$ the only move is the centre, and it wins. □

Caveat. For $n = 1$ the answer depends on reading "has not occurred previously" as including the initial empty board; under that reading Alice's only move wins. Kedlaya–Ng flag the same ambiguity. For $n \geq 3$ the empty board is never reachable again and the issue does not arise.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains find that the centre wins for small n. They split at the choice of what to prove.

(N) "Alice mirrors Bob." A strategy for one side from one position; it cannot show that off-centre moves lose, and it breaks on legality.

(M) "Find a function of the position that is the same for all losing positions." Clean up the rules first (Steps 1–3: repetition is irrelevant outside the end zone, and entering the end zone loses), then look for an invariant w such that from $w = c$ every move leaves c and from $w \neq c$ some move returns to c.

The second split is inside (M), at the choice of modulus. Summing labels mod n, or with no modulus, fails at the ends of the board. Mod $n + 1$ makes a missing neighbour count as $0 \equiv n + 1$, and the type-2 move becomes a reflection.

> *The repetition rule is making everything ugly. Before I think about strategy, I'll find where it can actually bite. Counting stones: they only go down when the board is full. So repetition can only matter near the full board, and there the positions form a small clique I can count.*

> *There are n spaces and every position has exactly n moves, one per space. If I can label moves so that the n moves always hit n different residues, then "weight ≡ target" is a losing set for the same reason Nim-sum zero is: from it you can't stay, from outside you can always get in. n nonzero residues means the modulus is $n + 1$.*

> *Mod $n + 1$ the ends of the board behave like empty spaces at 0 and $n + 1$, which are the same residue. So the line is really a circle of $n + 1$ cells with one permanently empty cell, and a type-2 move at t in a block changes the sum by $i + j - t$: the mirror image of t in its block.*

- ▷ **Cue:** a game with a "no position may repeat" rule ⟶ **Move:** first locate where repetition is even possible, using a monovariant (here the stone count and then the extremal block lengths); outside that region delete the rule
- ▷ **Cue:** a terminal region where every position reaches every other ⟶ **Move:** the game there is forced to enumerate the region; the parity of its size decides who entered it and lost
- ▷ **Cue:** every position has exactly N available moves, one per cell ⟶ **Move:** look for a weight mod $N + 1$ under which the N moves hit the N nonzero residues; the losing positions are then one residue class
- ▷ **Cue:** an operation on a line that treats the two ends like empty cells ⟶ **Move:** work mod (length +1), so both ends become the same residue and the line becomes a circle with a fixed hole
- ▷ **Cue:** a problem asking for *all* winning first moves ⟶ **Move:** a strategy for one player is not enough; you need a characterisation of all losing positions

Layer 3 THE WHY

Why must the rules be cleaned up first? Because the repetition clause makes a position's outcome depend on history, and every tool of game theory — P-positions, N-positions, invariants — needs the outcome to depend on the position alone. The only way to recover that

is to prove the history irrelevant. The natural monovariant is the one the moves themselves suggest: a type-1 move adds a stone; a type-2 move removes one and adds up to two. So the count can only go down when there is nowhere to put a stone on either side, that is, on the full board. That is where repetition can live, so that is the region to isolate. The block-length argument in Step 2 then covers the one remaining case, count-preserving moves, and it is forced: a count-preserving move is one where only one side has an empty space, which means the move is in an extremal block, and pushing stones out of a block towards the middle shrinks it.

Why does the end zone reduce to parity? Because its positions form a clique: from any of them, every other is reachable, so the players have no choice about *how many* moves remain, only about the order. A game with no real choice is decided by counting. $|Z| = n + 1$ is even; n further moves after entry, odd; entering loses. This is why the problem insists that n is odd — the official remark notes that for even n the parity flips, and there the answer is only conjectured.

Why the weight, and why mod $n + 1$? Here the definition of a losing position does the work. A set L of positions is the losing set exactly when (i) no move goes from L to L and (ii) from outside L some move reaches L. The cheapest way to get (i) and (ii) simultaneously is a function w with the property "the available moves change w by every nonzero value, and never by zero": then $L = \{w = \text{const}\}$ satisfies both. That is the Nim-sum principle, and it tells you what to look for before you know what w is. Here each of the n spaces gives exactly one move, so you need n distinct nonzero changes: the modulus is $n + 1$. With that modulus, a type-1 move at t changes the sum by t, and the n type-1 moves on an empty board already cover $\{1, \ldots, n\}$; the check that type-2 moves fill in the remaining residues is the reflection $t \mapsto i + j - t$, which works at the edges only because $0 \equiv n + 1$. The modulus is not guessed; it is counted.

Why the centre? Because the full-board weight $\frac{n(n+1)}{2}$ is $\equiv c$, and the end-zone entry positions P_{n-1} have weights $c - h \neq c$. The target residue is the one the endgame forces, and on the empty board it is reached by exactly one placement. Kedlaya–Ng state the weight and verify it; they say nothing about where it comes from. It comes from "n moves, n residues" plus "the ends are empty cells".

Caveat. Verified by computer: with the full repetition rule and history-dependent search, the only winning first move is the centre for $n = 1, 3, 5$; for the reduced game (no repetition outside Z, entering Z loses), the criterion "player to move loses iff weight $\equiv \frac{n+1}{2}$" matches exhaustive search on every position outside Z for $n = 3, 5, 7, 9, 11$. The proof above does not rely on the computation.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the repetition clause. On seeing it, write the stone-count bookkeeping of each move type and ask where the count can fall; that isolates the end zone $Z = P_{n-1} \cup P_n$.

2. The cue suggests a two-stage method: a monovariant to show the game is history-free outside a small region, then a parity count inside that region, then an invariant (a weight mod $n + 1$) to classify every other position.

3. Hesitation occurs twice: at the end zone, where it looks as if the order of moves matters (it does not — the region is a clique, so only its size matters), and at the choice of modulus, where mod n and plain sums fail at the edges. The second is reinforced away by counting: n moves per position must hit n distinct nonzero residues, so the modulus is $n + 1$.

4. The trap is the symmetry strategy. It proves (with some trouble over legality) that the centre wins, and says nothing about whether other first moves might also win — which is what the question asks.

5. Recognise it faster by counting the moves available from a typical position. When that number is the same for every position, a modular weight whose modulus is one more than that number is the first thing to try.

6. Reusable decision rule: *in any impartial game, first make the position determine the outcome (dispose of history rules with a monovariant), then look for a weight under which the available moves realise every nonzero change exactly once; the losing positions are a single level set.*