DAWN PUTNAM SESSION

Putnam 2023 — Papers A and B

Putnam 120 Archive

Contents

Putnam 2023 — Paper A **5**
2023 A1 5
2023 A2 7
2023 A3 10
2023 A4 13
2023 A5 17
2023 A6 22

Putnam 2023 — Paper B **29**
2023 B1 29
2023 B2 31
2023 B3 32
2023 B4 35
2023 B5 39
2023 B6 43

Putnam 2023 — Paper A

2023 A1

2023 A1. For a positive integer n, let $f_n(x) = \cos(x)\cos(2x)\cos(3x)\cdots\cos(nx)$. Find the smallest n such that $|f_n''(0)| > 2023$.

Fields: Calculus (Taylor/Maclaurin expansion, parity of even functions) **Position:** A1 – the opener
Answer: $n = 18$, because $f_n''(0) = -\sum_{k=1}^{n} k^2 = -\frac{n(n+1)(2n+1)}{6}$, and 18 is the smallest n for which this exceeds 2023 in absolute value.

Layer 1 THE WHAT — surface analysis

The natural attempt. The problem hands you a product of n cosines and asks for a second derivative at a point, so the honest first move is the product rule. Write $f_n(x) = \prod_{k=1}^{n}\cos(kx)$ and differentiate once:

$$f_n'(x) = -\sum_{i=1}^{n} i\sin(ix)\prod_{k\neq i}\cos(kx).$$

This much is routine. Now you need f_n'', so you differentiate the sum term by term, and each term is itself a product of n factors, so each one needs its own product rule. Differentiating the i-th summand $-i\sin(ix)\prod_{k\neq i}\cos(kx)$ produces one term from differentiating $\sin(ix)$ (giving $-i^2\cos(ix)\prod_{k\neq i}\cos(kx)$) and $n-1$ further terms, one for each $k \neq i$, from differentiating that factor's own $\cos(kx)$ (each giving something with a $\sin(kx)$ in it). You now have, in principle, n terms of the first kind and $n(n-1)$ terms of the second kind — a genuine double sum over pairs (i,k). You can push through this bookkeeping (it is exactly the "two types of terms" the official solution names), but doing it by brute differentiation, without first asking what happens when you plug in $x = 0$, means carrying the full $O(n^2)$-term expression through the algebra before any of it simplifies. For n as large as 17 or 18 this is a lot of paper to manage correctly under exam pressure, and every sign error in the product rule is fatal to the final count.

The model chain. f_n is a product of the functions $\cos(kx)$, each of which is even, so f_n itself is even: $f_n(-x) = f_n(x)$. An even function has a Maclaurin series in even powers of x only, so

$$f_n(x) = f_n(0) + \tfrac{1}{2}f_n''(0)x^2 + O(x^4),$$

with no linear term to worry about. Now expand each factor to the order that matters and no further:

$$\cos(kx) = 1 - \frac{(kx)^2}{2} + O(x^4).$$

Multiplying the n factors together and discarding anything of order x^4 or higher (cross terms between two different factors' x^2 pieces are already $O(x^4)$, so they never contribute to the x^2 coefficient),

$$f_n(x) = \left(1 - \tfrac{1^2x^2}{2}\right)\left(1 - \tfrac{2^2x^2}{2}\right)\cdots\left(1 - \tfrac{n^2x^2}{2}\right) = 1 - \frac{x^2}{2}\sum_{k=1}^{n} k^2 + O(x^4).$$

Comparing coefficients of x^2 with the Maclaurin form above,

$$\frac{1}{2}f_n''(0) = -\frac{1}{2}\sum_{k=1}^{n} k^2 \implies f_n''(0) = -\sum_{k=1}^{n} k^2 = -\frac{n(n+1)(2n+1)}{6}.$$

The function $g(n) = \frac{n(n+1)(2n+1)}{6}$ is increasing, with $g(17) = \frac{17\cdot 18\cdot 35}{6} = 1785$ and $g(18) = \frac{18\cdot 19\cdot 37}{6} = 2109$. Since $1785 \leq 2023 < 2109$, the smallest n with $|f_n''(0)| > 2023$ is $n = 18$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree that f_n is even and that $f_n''(0)$ is ultimately a sum of squares – that structural fact is forced and shows up in the official solution too, when it separates "terms where two distinct factors have each been differentiated once" (which vanish at $x = 0$ because $\sin(0) = 0$) from "terms where a single factor has been differentiated twice" (which survive). Both paths reach the same destination. The split is in *how* you get there: at the moment you need $f_n''(0)$, the natural attempt commits to producing the function $f_n''(x)$ in full generality via the product rule, then substitutes $x = 0$ afterward; the model substitutes the smallness of x *first*, truncating each factor's Taylor series before multiplying, and never differentiates a product at all.

> *I don't need the formula for $f_n''(x)$ as a function – I need one number, its value at $x = 0$ – so instead of differentiating a product of n terms twice and then plugging in 0, I should Taylor-expand every factor to the lowest order that can possibly affect the x^2 coefficient and multiply those truncated pieces together; anything of order x^4 or higher is invisible to $f_n''(0)$ and never needs to be computed.*

> *f_n is manifestly even, being a product of even functions, so its Maclaurin series has no odd powers at all – I get $f_n'(0) = 0$ for free, without differentiating, and I know in advance that the only information in $f_n''(0)$ is the coefficient of x^2, which tells me exactly which order of expansion to stop at.*

- ▷ **Cue:** a function is presented as a product of several even (or several odd) functions and a low-order derivative at 0 is wanted $\longrightarrow$ **Move:** Taylor-expand each factor to the lowest order needed and multiply the truncated series, rather than applying the product rule and then substituting the point
- ▷ **Cue:** only one specific derivative value at one specific point is required, not the derivative as a function $\longrightarrow$ **Move:** stop differentiating symbolically; extract the relevant coefficient from a local series expansion instead, since a k-th derivative at a point is $k!$ times the coefficient of x^k in the Maclaurin series

Layer 3 THE WHY

Why must the reroute to series expansion be the thought here, rather than a shortcut that merely happens to be faster? Because of what a derivative *at a point* actually is: $f_n''(0)$ is, by definition, 2! times the coefficient of x^2 in the Maclaurin series of f_n about 0. The product rule is a tool for producing the function $f_n''(x)$ in closed form; it is strictly more information than the problem asks for. Once you recognise that you only want a single coefficient, computing the full derivative function and then evaluating it is doing surplus work – work whose extra content (the x-dependence away from 0) is thrown away immediately after being computed.

The truncation to order x^2 is similarly forced, not chosen for convenience: a product of n series each of the form $1 + O(x^2)$ has cross terms between any two factors' $O(x^2)$ pieces that are automatically $O(x^4)$, so no cross term can ever reach the x^2 coefficient. The only contributions to that coefficient are the n individual $-\frac{k^2x^2}{2}$ terms, taken one at a time. This is exactly the same partition the official solution makes when it separates "two factors differentiated once" (the cross terms, which vanish at 0 because $\sin 0 = 0$ – the series-expansion mirror of "these are $O(x^4)$")

from "one factor differentiated twice" (the surviving $-k^2$ terms). The series route reaches that partition automatically, as a consequence of order-counting, instead of as an observation made after wading through the full product-rule expansion.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that f_n is a product of even functions, hence even – this should be written down the instant the problem is read, before any differentiation, because it tells you $f_n'(0) = 0$ for free and that only even-order Maclaurin coefficients matter.
2. That cue suggests working with the Maclaurin series of f_n directly rather than the product rule: expand each $\cos(kx)$ to order x^2, multiply, and read off $f_n''(0)$ as twice the coefficient of x^2.
3. Hesitation occurs at the moment you need a second derivative and your instinct is to reach for the product rule, because that is the tool taught for "differentiate a product." It is reinforced away by asking what, specifically, the problem wants: not a formula for $f_n''(x)$, but a single number, which a truncated series supplies with far less algebra.
4. The trap is grinding through the full double product rule for a product of n cosines, generating $O(n^2)$ terms, when only n of them (the "differentiate one factor twice" terms) survive evaluation at $x = 0$ – the other $O(n^2) - n$ terms are wasted effort that vanishes identically at the point in question.
5. Recognition would be faster next time by treating "product of several functions, derivative wanted only at a single point" as an automatic trigger for local Taylor expansion rather than symbolic differentiation, especially when the functions involved (sine, cosine, exponentials) have standard series.
6. The reusable decision rule: when a derivative of any order is wanted only at a single point, and the function is built from a product (or sum) of pieces with known local series, expand each piece to the minimum order that can affect the target coefficient and multiply; never produce the full derivative function symbolically and substitute afterward, since that computes and discards information the problem never asked for.

2023 A2

2023 A2. Let n be an even positive integer. Let p be a monic, real polynomial of degree $2n$; that is to say, $p(x) = x^{2n} + a_{2n-1}x^{2n-1} + \cdots + a_1 x + a_0$ for some real coefficients $a_0, \ldots, a_{2n-1}$. Suppose that $p(1/k) = k^2$ for all integers k such that $1 \leq |k| \leq n$. Find all other real numbers x for which $p(1/x) = x^2$.

Fields: Algebra (polynomials, substitution to clear denominators, coefficient matching) **Position:** A2 – second on the paper **Answer:** $x = \pm\dfrac{1}{n!}$, the only other real solutions, obtained by turning $p(1/x) = x^2$ into a degree-$(2n+2)$ polynomial equation with $2n$ of its roots already known.

Layer 1 THE WHAT — surface analysis

The natural attempt. The problem gives you $2n$ conditions – $p(1/k) = k^2$ for $k = \pm 1, \ldots, \pm n$ – on a monic polynomial with exactly $2n$ unknown coefficients $a_0, \ldots, a_{2n-1}$. That looks like a determined linear system: $2n$ equations, $2n$ unknowns, so the honest first instinct is that p is pinned down completely, and you should be able to recover it, say by Lagrange interpolation on the $2n$ points $(1/k, k^2)$, and then just solve $p(1/x) = x^2$ directly once you have p explicitly. You start setting this up: each condition $p(1/k) = k^2$, after multiplying through by k^{2n}, becomes a genuine polynomial identity in k, and Lagrange interpolation in principle produces p as an explicit (if ugly) combination of $2n$ basis polynomials. But two things go wrong quickly. First, this is a construction for one specific p, whereas the problem's phrasing ("Find all other real numbers x") suggests the answer should be uniform in n – and an explicit Lagrange formula

gives you a mess of binomial-coefficient sums that does not obviously simplify to anything as clean as $\pm 1/n!$ without further insight. Second, and more fundamentally, this path reconstructs $p(x)$ as a polynomial in x, but the equation you actually need to solve, $p(1/x) = x^2$, is not polynomial in x at all – it has negative powers of x built into $p(1/x)$ – so even with p in hand you still have to do the same denominator-clearing step the model chain does immediately, except now on top of an unwieldy interpolated formula instead of the clean unknowns $a_0, \ldots, a_{2n-1}$.

The model chain. Multiply the target equation by x^{2n} before doing anything else. Define

$$q(x) := x^{2n+2} - x^{2n} p(1/x).$$

Since $p(y) = y^{2n} + a_{2n-1}y^{2n-1} + \cdots + a_1 y + a_0$, substituting $y = 1/x$ and multiplying by x^{2n} gives

$$x^{2n} p(1/x) = 1 + a_{2n-1}x + a_{2n-2}x^2 + \cdots + a_1 x^{2n-1} + a_0 x^{2n},$$

which is a genuine polynomial in x of degree at most $2n$. Hence $q(x) = x^{2n+2} - a_0 x^{2n} - a_1 x^{2n-1} - \cdots - a_{2n-1}x - 1$ is a monic polynomial in x of degree exactly $2n+2$, with constant term -1. Crucially, for $x \neq 0$, $p(1/x) = x^2$ is equivalent to $q(x) = 0$: this converts a non-polynomial equation in x into a polynomial one without ever needing p's coefficients individually.

The hypothesis says $p(1/k) = k^2$ for $k = \pm 1, \ldots, \pm n$, i.e. $q(k) = 0$ for these $2n$ values. Since q has degree $2n+2$ and these $2n$ roots are already known and distinct, we may write

$$q(x) = (x^2 + ax + b) \prod_{k=1}^{n} (x^2 - k^2)$$

for some monic quadratic $x^2 + ax + b$ still to be determined – only two unknowns remain, not $2n$. The product $\prod_{k=1}^{n}(x^2 - k^2)$ is an even polynomial in x (only even powers), monic of degree $2n$. Comparing the coefficient of x^{2n+1} on both sides: on the left it is 0 (nothing of that degree appears in q); on the right, since $\prod(x^2 - k^2)$ has no x^{2n-1} term (it is even), the only source of x^{2n+1} is $a \cdot x \cdot x^{2n}$, giving coefficient a. So $a = 0$.

Comparing constant terms: on the left it is -1; on the right it is $b \cdot \prod_{k=1}^{n}(-k^2) = b \cdot (-1)^n (n!)^2$. Since n is even, $(-1)^n = 1$, so $b = -1/(n!)^2$. Hence

$$q(x) = \left(x^2 - \frac{1}{(n!)^2}\right) \prod_{k=1}^{n} (x^2 - k^2),$$

and the two roots of q beyond $\pm 1, \ldots, \pm n$ are $x = \pm \dfrac{1}{n!}$. These are genuinely *other* values since $n > 1$ forces $1/n! \neq 1, \ldots, n$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains recognise that the problem, stated as it is, over-determines p if approached coefficient by coefficient – that much is visible on a first read. The split is at the very first move after reading the problem: "Both paths see $2n$ conditions and a degree-$2n$ unknown p. At this point the natural attempt tries to solve *for* p – recover $a_0, \ldots, a_{2n-1}$ explicitly, e.g. via Lagrange interpolation. The model instead multiplies the target equation $p(1/x) = x^2$ by x^{2n} and studies the resulting polynomial $q(x)$ directly, never solving for p's coefficients as a named list at all."

> *I don't need to know p – I need the solution set of $p(1/x) = x^2$, and that equation is not polynomial in x as written, so before doing anything else I should clear the denominators by multiplying by x^{2n}; whatever comes out will be an honest polynomial in x whose roots are exactly the x-values I want, and I can reason about its roots without ever naming $a_0, \ldots, a_{2n-1}$ individually.*

> *Once I see $q(x)$ has degree $2n+2$ and I already know $2n$ of its roots from the hypothesis, the remaining unknown is a single monic quadratic factor – two coefficients, not $2n$ – so the interpolation problem I was worried about was never a $2n$-parameter problem in disguise; it collapses to two coefficient-matching equations the instant I stop trying to reconstruct p itself.*

- ▷ **Cue:** an equation involves $p(1/x)$ for a polynomial p and is not itself polynomial in x $\longrightarrow$ **Move:** multiply through by $x^{\deg p}$ to clear the negative powers and turn it into a genuine polynomial equation in x, whose roots are the solutions sought
- ▷ **Cue:** a polynomial equation of degree d has $d-2$ (or more generally $d-m$) of its roots already known from the problem's hypotheses $\longrightarrow$ **Move:** factor out the known roots explicitly and solve only for the remaining low-degree factor by matching a small number of coefficients (leading/near-leading and constant are usually cheapest), rather than reconstructing the whole polynomial from scratch

Layer 3 THE WHY

Why must clearing denominators via x^{2n} be the first move, rather than a trick noticed after the fact? Because of what "$p(1/x) = x^2$" structurally is: it mixes a polynomial evaluated at a reciprocal argument with a polynomial in the original variable, and these live in different rings until you force them together. $p(1/x)$ genuinely has negative powers of x down to x^{-2n}; there is no way to compare it to x^2 as polynomials without first multiplying by enough of a power of x to clear them. The exponent $2n$ is not a free choice – it is exactly $\deg p$, which is exactly what the problem set up (a degree-$2n$ monic p) to make this multiplication land you back in polynomial territory of a controlled degree, $2n+2$. Once you see that the *only* obstruction to comparing the two sides algebraically is the negative powers, multiplying by x^{2n} stops being a trick and becomes the unique way to make the equation legible.

Why is it then legitimate to skip reconstructing p and go straight for q's known roots? Because a monic polynomial of degree $2n+2$ is *determined* by $2n+2$ of its roots (with multiplicity) up to nothing – it is exactly those roots, as a product of linear factors, once you have that many. The hypothesis was never really "$2n$ conditions on p"; restated as conditions on q, it is "$2n$ of q's $2n+2$ roots are handed to you for free." The interpolation problem the natural attempt sets up is solving the same information through the wrong unknowns (p's coefficients) instead of the right ones (the two coefficients of q's remaining quadratic factor). And the parity hypothesis "n even" is not decorative: it is exactly what forces $(-1)^n = 1$ in the constant-term comparison, which is what guarantees $b < 0$ and hence a real (rather than complex) quadratic factor $x^2 - 1/(n!)^2$ – the problem's evenness condition on n is the hidden reason the "other real numbers" asked for actually exist.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that $p(1/x) = x^2$ is not polynomial in x as written – it should be flagged immediately as "clear denominators before doing anything else," by multiplying through by x^{2n} (the degree of p).

2. That cue suggests defining $q(x) = x^{2n+2} - x^{2n}p(1/x)$ and working entirely with its roots, since the hypothesis translates directly into $2n$ known roots of q, rather than attempting to reconstruct p's coefficients via interpolation.

3. Hesitation occurs at seeing "$2n$ conditions, $2n$ unknown coefficients" and reflexively treating it as a full interpolation problem for p. It is reinforced away by noticing that the actual question is about roots of an equation, not about p itself, and that q's degree ($2n+2$) exceeds the number of unknowns in p ($2n$) by exactly 2 – so only 2 genuinely new unknowns remain once the known roots are factored out.

4. The trap is committing to full Lagrange interpolation for p, which produces a correct but unwieldy closed form that must then itself be substituted back into $p(1/x) = x^2$ and cleared

of denominators anyway – all the real work of the problem, done twice, with the second pass now burdened by an ugly formula instead of clean symbols $a_0, \ldots, a_{2n-1}$.

5. Recognition would be faster next time by treating "$p(1/x)$ compared with a polynomial in x" as an automatic trigger for "multiply by $x^{\deg p}$ first," and by treating "$2n$ hypotheses on a degree-$2n$ object, but the question is about a related equation of higher degree" as a cue to count how many roots of the higher-degree equation are already known, rather than how many unknowns the original object has.

6. The reusable decision rule: when an equation compares $p(1/x)$ (or any reciprocal/composed evaluation of a polynomial) to something else, clear denominators by multiplying by the appropriate power of x before analysing roots; then, if the resulting polynomial's degree exceeds the number of unknowns by a small amount because some of its roots are given by hypothesis, factor those out explicitly and solve only for the remaining low-degree factor via a handful of coefficient comparisons – never reconstruct the full original object as an intermediate step if the question only needs its roots.

2023 A3

2023 A3. Determine the smallest positive real number r such that there exist differentiable functions $f\colon \mathbb{R} \to \mathbb{R}$ and $g\colon \mathbb{R} \to \mathbb{R}$ satisfying

(a) $f(0) > 0$,

(b) $g(0) = 0$,

(c) $|f'(x)| \leq |g(x)|$ for all x,

(d) $|g'(x)| \leq |f(x)|$ for all x, and

(e) $f(r) = 0$.

Fields: Analysis (differential inequalities, a planar/polar viewpoint on coupled functions) **Position:** A3 – third on the paper, the first real analysis problem **Answer:** $r = \pi/2$, achieved exactly by $f = \cos$, $g = \sin$; no smaller r works because the coupled bounds $|f'| \leq |g|$, $|g'| \leq |f|$ force the point $(f(x), g(x))$ to rotate around the origin at speed at most 1, and reaching $f = 0$ from $f(0) > 0, g(0) = 0$ requires a quarter turn.

Layer 1 THE WHAT — surface analysis

The natural attempt. The instinct on seeing $f(r) = 0$ with $f(0) > 0$ is to try to bound f away from zero directly, using the two differential inequalities as a growth-control system, the way you would bound a solution of $|y'| \leq C$ by $|y(x) - y(0)| \leq Cx$. You write $f(x) = f(0) + \int_0^x f'(t)\,dt$, so $f(x) \geq f(0) - \int_0^x |f'(t)|\,dt \geq f(0) - \int_0^x |g(t)|\,dt$, and since $g(0) = 0$, $|g(t)| \leq \int_0^t |g'(s)|\,ds \leq \int_0^t |f(s)|\,ds$. This is a genuine Gronwall-type setup: you have $f(x) \geq f(0) - \int_0^x \int_0^t |f(s)|\,ds\,dt$, and if you are careful you can turn this into a differential inequality of the shape $f'' \geq -f$ (informally) and try to solve it, or you iterate the integral inequality to get $f(0)$ minus a power series that looks suspiciously like $f(0)\cos x$ once you push it far enough. You can, with real effort, make this rigorous and recover $r \geq \pi/2$ – but you have to notice, mid-computation, that the iterated bound you are building *is* a cosine series, essentially rediscovering the trig identity by brute Taylor-coefficient chasing rather than by structural insight. Worse, the bound naturally wants to track $|f|$ and $|g|$ as separate quantities with separate integral inequalities, which multiplies the bookkeeping and makes it easy to lose the sharp constant along the way – a sloppy version of this approach gives some exponential-type bound on how slowly f can reach 0, not the exact threshold $\pi/2$.

The model chain. Treat $(f(x), g(x))$ as a single point moving in the plane and track two things about it: how big it is, and how fast it turns.

It never returns to the origin. Let $h(x) = f(x)^2 + g(x)^2$. Then $h'(x) = 2f(x)f'(x) + 2g(x)g'(x)$, so by the triangle inequality and the hypotheses,

$$|h'(x)| \le 2|f(x)||f'(x)| + 2|g(x)||g'(x)| \le 2|f(x)||g(x)| + 2|g(x)||f(x)| = 4|f(x)||g(x)|.$$

By AM–GM, $2|f(x)||g(x)| \le f(x)^2 + g(x)^2 = h(x)$, so $|h'(x)| \le 2h(x)$. Since $h(0) = f(0)^2 > 0$, this bound (via $|\frac{d}{dx} \log h| \le 2$ wherever $h > 0$, and continuity) forces $h(x) > 0$ for every x where a hypothetical zero has not yet occurred; in particular $f(r) = 0$ forces $g(r) \neq 0$, since $h(r) > 0$.

It turns at speed at most 1**.** Since $(f(x), g(x))$ is never the origin, define the angle $\theta(x) = \tan^{-1}(g(x)/f(x))$ on any interval where $f \neq 0$ (more carefully, using $h(x) \neq 0$, an angle function exists and is differentiable throughout). Then

$$\theta'(x) = \frac{f(x)g'(x) - g(x)f'(x)}{f(x)^2 + g(x)^2},$$

so

$$|\theta'(x)| \le \frac{|f(x)||g'(x)| + |g(x)||f'(x)|}{f(x)^2 + g(x)^2} \le \frac{|f(x)|^2 + |g(x)|^2}{f(x)^2 + g(x)^2} = 1,$$

using $|g'| \le |f|$ and $|f'| \le |g|$ in the numerator. Since $g(0) = 0$ and $f(0) > 0$, $\theta(0) = 0$. Reaching $f(r) = 0$ (with $g(r) \neq 0$) means $\theta(r)$ is an odd multiple of $\pi/2$; since $|\theta(r) - \theta(0)| \le r$ (the angle can change by at most 1 per unit x), the smallest possible r for which $|\theta(r)|$ can reach $\pi/2$ is $r = \pi/2$ itself.

Equality. $f(x) = \cos x$, $g(x) = \sin x$ satisfy $f' = -\sin x = -g$, $g' = \cos x = f$, so $|f'| = |g|$ and $|g'| = |f|$ with equality everywhere – the angle turns at the maximum possible rate, exactly 1, and reaches $\pi/2$ exactly at $x = r = \pi/2$, where $f(\pi/2) = 0$. So $r = \pi/2$ is achieved, and no smaller value is possible.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains reach for the same two building blocks – the hypotheses bound f' by g and g' by f, and both recognise this is a coupled-growth situation calling for an integral or Gronwall-style estimate. The split is in what quantity gets bounded. "Both paths agree that $|f'| \le |g|$, $|g'| \le |f|$ must be exploited via integration. At this point the natural attempt bounds f (and separately g) as individual real-valued quantities, chasing how large they can get or how fast they can shrink; the model instead bounds the *angle* of the vector $(f(x), g(x))$, treating f and g jointly as coordinates of a single point that cannot sit at the origin and cannot rotate faster than speed 1."

> *I don't need to know how large or small f and g individually get – I need to know how the* pair *$(f(x), g(x))$ moves, because $f(0) > 0$, $g(0) = 0$ says this point starts on the positive f-axis, and $f(r) = 0$ says it has to reach the g-axis; that is a statement about the point crossing a quarter-turn, not about either coordinate's magnitude on its own.*

> *The two bounds are symmetric under swapping $f \leftrightarrow g$ – $|f'| \le |g|$ and $|g'| \le |f|$ are the same shape with the roles reversed – and that symmetry is exactly what a rotation argument uses and a one-sided bound on f alone throws away; if I only bound f, I am discarding half the given information.*

> *Before trusting any bound on the angle, I first need (f, g) to never be $(0, 0)$, or "angle" does not make sense – so the AM–GM step $2|f||g| \le f^2 + g^2$ giving $|h'| \le 2h$ is not a detour, it is the fact that makes the whole angular argument legitimate in the first place.*

- ▷ **Cue:** two functions f, g satisfy symmetric coupled bounds of the form $|f'| \le |g|$, $|g'| \le |f|$ $\longrightarrow$ **Move:** treat (f, g) as coordinates of a single planar point and bound the rate at which its *angle* $\tan^{-1}(g/f)$ changes, rather than bounding f and g separately
- ▷ **Cue:** a bound on f' and g' each in terms of the other, and $f(0), g(0)$ are given with one zero and one nonzero $\longrightarrow$ **Move:** first check $f^2 + g^2$ cannot reach 0 (via a log-derivative/Gronwall bound), since only then does an angle function exist and stay well-defined for the whole argument
- ▷ **Cue:** a claimed sharp constant matches $\pi/2$, or the setup resembles $f' = -g, g' = f$ $\longrightarrow$ **Move:** test $f = \cos x$, $g = \sin x$ as the extremal case immediately; equality in every inequality used should hold there, confirming the bound is tight rather than merely valid

Layer 3 THE WHY

Why must the reroute to an angle be the thought here, rather than a trick pulled from nowhere? Because of what the two conditions $|f'| \le |g|$ and $|g'| \le |f|$ actually encode: they say the rate of change of each coordinate is controlled by the size of the *other* coordinate, which is precisely the differential structure of rotation. The pair $(\cos x, \sin x)$ satisfies this with equality by definition of what a rotation at unit angular speed is; the hypotheses are, structurally, "the point (f, g) turns, and turns no faster than the point on the unit circle does." Once that is seen, asking "how fast can f shrink" is asking the wrong question – shrinking is not an intrinsic property of the point's motion, it is a coordinate artifact that depends on which direction you happen to be facing. The intrinsic, coordinate-free quantity attached to a rotating point is the angle it has swept, and that is exactly the quantity the hypotheses bound directly and symmetrically (the Cauchy–Schwarz-shaped numerator $|fg' - gf'| \le f^2 + g^2$ collapses to 1 only because the two bounds are dual to each other; a one-sided bound on f' alone would not produce a dimension-free constant).

This is the Blueprint's "centre of the circle" standard, transposed to motion rather than shape: a circle is the set of points equidistant from a centre, so the centre is the natural object to study; a system whose derivatives are cross-bounded by each other's own sizes is a system whose *natural object* is the point's own angle, because angle is the one quantity whose rate of change these particular bounds can control absolutely (bounded by 1, with no dependence on how large or small f, g currently are — the scaling $f^2 + g^2$ cancels top and bottom). That cancellation is not a coincidence to be marvelled at; it is forced by the fact that both derivative bounds have exactly the same "the other coordinate" shape, so any quantity built symmetrically in f, g that is homogeneous of degree 0 (a ratio, an angle) will see the $f^2 + g^2$ in numerator and denominator meet in exactly this way. The problem's target $f(r) = 0$ is then legible immediately as "reach the axis where the angle is $\pm\pi/2$," and the answer $r = \pi/2$ is nothing more than "distance $\pi/2$, top speed 1, minimum time $\pi/2$."

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that the two hypotheses $|f'| \le |g|$, $|g'| \le |f|$ are symmetric under swapping f and g – this should be flagged immediately, because a symmetric coupling between two functions' derivatives and each other's values is the signature of a rotation, not of independent growth bounds.
2. That cue suggests treating $(f(x), g(x))$ as a single point in the plane and bounding the rate of change of its angle $\theta(x) = \tan^{-1}(g(x)/f(x))$, rather than separately bounding how fast f or g individually can grow or shrink.
3. Hesitation occurs at the instinct to apply a Gronwall-style bound directly to f, since that is the standard tool for "derivative bounded by another quantity." It is reinforced away by

noticing the bound produced this way is not obviously sharp, and by checking the extremal example $f = \cos, g = \sin$: since equality holds there in *both* hypotheses simultaneously, the sharp bound must come from a quantity for which $(\cos, \sin)$ also achieves equality throughout – which points at angle, since $(\cos x, \sin x)$ sweeps angle at exactly rate 1.

4. The trap is bounding f and g as two separate scalar quantities via independent integral inequalities; this produces *some* valid bound but loses the sharp constant $\pi/2$, because it never uses the fact that the two bounds interact multiplicatively (via $fg' - gf'$) rather than additively.
5. Recognition would be faster next time by treating "$|f'| \leq |g|$ and $|g'| \leq |f|$" (or any symmetric cross-bound between two functions' derivatives and each other's values) as an automatic trigger for a planar/polar reformulation, and by always testing whether the boundary case of the inequalities (equality throughout) is literally $(\cos, \sin)$ or a rescaling of it.
6. The reusable decision rule: when two real-valued functions satisfy derivative bounds that are symmetric cross-bounds in terms of each other's magnitudes, do not bound each function separately; form the combined quantity $f^2 + g^2$ to rule out a simultaneous zero, then form the angle $\tan^{-1}(g/f)$ and bound its derivative using both hypotheses at once (the numerator $fg' - gf'$ uses one bound on each term) – the resulting bound on angular speed is typically dimension-free and sharp, whereas separate bounds on f and g generally are not.

2023 A4

2023 A4. Let $v_1, \ldots, v_{12}$ be unit vectors in $\mathbb{R}^3$ from the origin to the vertices of a regular icosahedron. Show that for every vector $v \in \mathbb{R}^3$ and every $\varepsilon > 0$, there exist integers $a_1, \ldots, a_{12}$ such that $\|a_1 v_1 + \cdots + a_{12} v_{12} - v\| < \varepsilon$.

Fields: geometry (Platonic solids), algebra (finitely generated subgroups of $\mathbb{R}^n$), number theory (irrationality and Kronecker density) **Position:** A4 — the pivot of the paper. Nothing to compute, nothing to optimise: a one-line statement that is either instant or impenetrable depending on one reformulation **Answer:** The subgroup of $\mathbb{R}^3$ generated by the twelve vertex vectors is dense; the golden ratio ϕ is what makes it so.

Layer 1 THE WHAT — surface analysis

The natural attempt. The statement is an approximation statement, so the honest first move is to try to approximate. You have twelve vectors, you want to hit an arbitrary target to within ε, and twelve vectors in three dimensions feels like a great deal of room. So you try greedy: take the vertex closest in direction to v, step towards v, overshoot, correct with another vertex, and hope the residual shrinks.

It does not shrink. Each step you take has length 1 — these are unit vectors — so after any number of moves you are sitting somewhere in a set of points that might perfectly well be spaced out at distance $1/10$ or $1/100$, but might equally well be spaced at distance 1. Greedy approximation cannot distinguish those two worlds, because greedy approximation is a local procedure and the question is global.

That is the moment the strong student re-reads the problem and realises what it actually says. The set of all $a_1 v_1 + \cdots + a_{12} v_{12}$ with $a_i \in \mathbb{Z}$ is not a random collection of points: it is closed under addition and negation. It is the subgroup

$$H = \mathbb{Z}v_1 + \cdots + \mathbb{Z}v_{12} \leq (\mathbb{R}^3, +),$$

and the problem is exactly "prove H is dense".

Now the wall is visible and it is the right wall. A finitely generated subgroup of $\mathbb{R}^3$ is very often *not* dense: $\mathbb{Z}^3$ is generated by three vectors and is as far from dense as possible, and if

you take any twelve vectors all of whose coordinates are rational, the group they generate is contained in $\frac{1}{N}\mathbb{Z}^3$ for some N and is again a lattice. So "twelve vectors, only three dimensions" proves nothing whatsoever. Whether H is dense is a question about the *arithmetic* of the coordinates of the icosahedron, and until you go and look at those coordinates you have no information at all.

Many scripts stall here, having correctly reduced the problem and then treated the reduction as the answer.

The model chain. *Step 0: the centre is at the origin.* All twelve vectors have length 1, so all twelve vertices lie on the unit sphere about the origin. A regular icosahedron is inscribed in a unique sphere, whose centre is the centre of the icosahedron. Hence the origin is the centre. (This is not pedantry: without it the vertex vectors need not have the symmetry we are about to use.)

Step 1: coordinates, up to rotation and scale. Every regular icosahedron centred at the origin is $R\Delta$ for a fixed model Δ and some $R \in O(3)$. Density is preserved by R, since R is a homeomorphism and a group automorphism of $\mathbb{R}^3$. Density is also preserved by scaling: if $u_i = \lambda w_i$ with $\lambda > 0$, then $\mathbb{Z}u_1 + \cdots = \lambda(\mathbb{Z}w_1 + \cdots)$, and a positive dilation of a dense set is dense. So we may work with the standard model, whose twelve vertices are the cyclic permutations of

$$(0,\ \pm 1,\ \pm\phi), \qquad \phi = \frac{1+\sqrt{5}}{2},$$

and prove that $H := \mathbb{Z}w_1 + \cdots + \mathbb{Z}w_{12}$ is dense.

Step 2: H contains a product. Let $G = \mathbb{Z} + \mathbb{Z}\phi \leq \mathbb{R}$. The four vertices $(0, \pm 1, \pm\phi)$ give, by one addition and one subtraction,

$$(0,1,\phi) + (0,1,-\phi) = (0,2,0), \qquad (0,1,\phi) - (0,1,-\phi) = (0,0,2\phi),$$

and the pair $(0,1,\phi), (0,-1,\phi)$ gives $(0,0,2\phi)$ and $(0,2,0)$ likewise. So H contains $2e_2$ and $2\phi e_3$. The vertices $(\pm 1, \pm\phi, 0)$ give in the same way $(2,0,0)$ and $(0,2\phi,0)$, i.e. $2e_1$ and $2\phi e_2$; the vertices $(\pm\phi, 0, \pm 1)$ give $(2\phi,0,0)$ and $(0,0,2)$, i.e. $2\phi e_1$ and $2e_3$. Collecting, H contains $2e_1, 2\phi e_1, 2e_2, 2\phi e_2, 2e_3, 2\phi e_3$, hence contains

$$2G \times 2G \times 2G.$$

Step 3: G is dense in $\mathbb{R}$. This is the only genuinely arithmetic input. Let $A \leq (\mathbb{R}, +)$ be a subgroup and let $\alpha = \inf\{a \in A : a > 0\}$ (with $\alpha = \infty$ if $A = \{0\}$). If $\alpha > 0$ then $\alpha \in A$: otherwise there are $a, a' \in A$ with $\alpha < a' < a < 2\alpha$, and $a - a' \in A$ is positive and less than α. Then $A = \alpha\mathbb{Z}$, since any $a \in A$ has $a - \alpha\lfloor a/\alpha \rfloor \in A \cap [0, \alpha)$, forcing it to be 0. If $\alpha = 0$ then A has nonzero elements of arbitrarily small modulus, and the integer multiples of such an element are α-dense, so A is dense. Conclusion: *every subgroup of $\mathbb{R}$ is either cyclic or dense.*

Now $G = \mathbb{Z} + \mathbb{Z}\phi$. If $G = \alpha\mathbb{Z}$ then $1 = m\alpha$ and $\phi = k\alpha$ for integers m, k, so $\phi = k/m \in \mathbb{Q}$. But $\phi = (1+\sqrt{5})/2$ and $\sqrt{5}$ is irrational (5 is not a perfect square), so $\phi \notin \mathbb{Q}$. Hence G is dense in $\mathbb{R}$.

Step 4: assemble. $2G$ is dense in $\mathbb{R}$, so $2G \times 2G \times 2G$ is dense in $\mathbb{R}^3$ (approximate each coordinate separately; if $|x_i - g_i| < \varepsilon/\sqrt{3}$ for each i then $\|x - g\| < \varepsilon$). Since $H \supseteq 2G \times 2G \times 2G$, H is dense in $\mathbb{R}^3$. Undoing the rotation and the scaling of Step 1, the group generated by the original unit vectors $v_1, \ldots, v_{12}$ is dense, which is precisely the assertion to be shown. ■

Layer 2 THE HOW — the bifurcation

Bifurcating point. The two paths agree on the reading of the problem for about thirty seconds, and split at the question "what object am I actually studying?". At (†) — the set of all $\sum a_i v_i$ with $a_i \in \mathbb{Z}$ — the natural attempt treats it as a *search space* and starts searching; the model names it as an *algebraic object*, a subgroup, and immediately asks the only question there is to ask about a subgroup of $\mathbb{R}^n$: is it discrete, or is it dense? Everything after that is coordinates.

> *The coefficients are integers and they range over all integers. That means the set I am approximating with is closed under addition and under negation — it is a group, not a list. I should stop trying to reach v and start asking what the group looks like.*

> *A finitely generated subgroup of $\mathbb{R}^3$ generated by vectors with rational coordinates is a lattice, and lattices are the opposite of dense. So the whole problem lives in the irrationality of the coordinates of an icosahedron. Where does an irrational number enter a regular icosahedron? The golden ratio. That is not a decoration of the figure, it is the content of the problem.*

> *I do not need to understand the group H exactly. I only need to trap something dense* inside *it. So I should look for the cheapest dense subgroup I can manufacture from sums of two vertices — and the cheapest dense subgroup of $\mathbb{R}^3$ is $G \times G \times G$ with G dense in $\mathbb{R}$.*

Layer 3 THE WHY

Why must ϕ be irrational, and why is that the whole problem? Drive it to the axioms.

A regular icosahedron has a five-fold rotational symmetry about each vertex. Five-fold symmetry is exactly the symmetry that is *incompatible with a lattice*: the crystallographic restriction says a lattice in the plane or in space admits rotations of order $1, 2, 3, 4, 6$ only. So the moment you are told "regular icosahedron" you have been told "this figure cannot sit inside a lattice with its symmetry intact", and a group generated by vectors that refuse to be lattice-compatible is exactly the kind of group that has a chance of being dense. The golden ratio is not an accident of a coordinate choice; ϕ is the number $2\cos(\pi/5)$ in disguise ($2\cos 36^\circ = \phi$), and it is irrational for the same reason that the pentagon is not crystallographic.

Here is the same fact stated so that it can actually be used, and this is the part the official solution does not say. *Let w be any vertex vector and let $w^{(1)}, \dots, w^{(5)}$ be its five neighbours. Then*

$$w^{(1)} + \cdots + w^{(5)} = \sqrt{5}\, w.$$

This is forced before computation: the five-fold rotation about the axis $\mathbb{R}w$ permutes the neighbours, so their sum is fixed by that rotation, so the sum lies on the axis, so it is $c\,w$ for some scalar c. Dotting with w gives $c = 5\langle w, w^{(1)}\rangle / \|w\|^2$, and in the model coordinates $\langle (0,1,\phi), (0,-1,\phi)\rangle = \phi^2 - 1 = \phi$ while $\|w\|^2 = 1 + \phi^2 = 2 + \phi$, so $c = 5\phi/(2+\phi) = \sqrt{5}$. (Direct check with $w = (0,1,\phi)$ and neighbours $(0,-1,\phi), (1,\phi,0), (-1,\phi,0), (\phi,0,1), (-\phi,0,1)$: the sum is $(0,\, 2\phi - 1,\, \phi + 2) = (0, \sqrt{5}, \sqrt{5}\phi) = \sqrt{5}\, w$, using $2\phi - 1 = \sqrt{5}$ and $\phi + 2 = \sqrt{5}\,\phi$.)

That single identity re-proves the problem without any coordinates at all, and it is worth carrying because it explains the mechanism. Since $\sqrt{5}\, w_i \in H$ for every generator w_i, multiplication by $\sqrt{5}$ maps H into H; so H is a module over $\mathbb{Z}[\sqrt{5}]$. In particular $(\sqrt{5} - 2)^m w_1 \in H$ for every $m \geq 0$, and $0 < \sqrt{5} - 2 = 0.2360\ldots < 1$, so H contains nonzero vectors of arbitrarily small length along the line $\mathbb{R}w_1$. The closure $\overline{H}$ is a closed subgroup; a closed subgroup containing arbitrarily short nonzero vectors on a line contains that whole line; doing this for three vertices that span (say $(0,1,\phi), (1,\phi,0), (\phi,0,1)$, whose determinant is $-2\phi - 2 \neq 0$) gives $\overline{H} \supseteq \mathbb{R}w_1 + \mathbb{R}w_2 + \mathbb{R}w_3 = \mathbb{R}^3$. Done.

So the deep reason is: *the icosahedron carries a self-similarity of ratio $\sqrt{5} - 2 < 1$ built out of its own vertices.* A figure that reproduces a shrunken copy of itself inside its own integer span cannot generate a discrete group, because discreteness means a positive minimum distance and self-similarity of ratio < 1 destroys any positive minimum. The golden ratio, the five-fold symmetry, the irrationality and the density are four faces of one fact.

The solution of Bhargava–Kedlaya–Ng begins "we may choose coordinates so that the vertices are the cyclic permutations of $(\pm\frac{1}{2}, \pm\frac{1}{2}\phi, 0)$". Here is where that comes from: it is the

smallest coordinate model in which the two generators 1 and ϕ of the relevant subgroup of $\mathbb{R}$ appear *separately on separate axes*, so that the three-dimensional problem factors into three copies of a one-dimensional problem you already know how to do. Choosing coordinates is not bookkeeping here; it is the act that converts a question about $\mathbb{R}^3$ into three questions about $\mathbb{R}$.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "there exist *integers* $a_1, \ldots, a_{12}$" together with "for every $\varepsilon > 0$". Write down at once: $H = \mathbb{Z}v_1 + \cdots + \mathbb{Z}v_{12}$ *is a subgroup of* $\mathbb{R}^3$*, and the problem is '*H *is dense'.* The second cue is "regular icosahedron": write down $\phi = \frac{1+\sqrt{5}}{2}$ before doing anything else.

2. The cue suggests the dichotomy for subgroups of $\mathbb{R}^n$: discrete (a lattice, possibly in a proper subspace) or not. For $n = 1$ the dichotomy is sharp and elementary — cyclic or dense — so the method is to reduce the 3-dimensional question to three 1-dimensional ones by finding a product subgroup $G \times G \times G$ inside H.

3. Hesitation occurs at exactly one place: after reducing to "show H is dense", you do not know what to do next and are tempted to reach for a structure theorem for closed subgroups of $\mathbb{R}^n$ that you may not be able to state correctly under pressure. What reinforces it away is the realisation that you never need to determine H — you only need to trap something dense inside it, and containment is proved by exhibiting explicit sums of two vertices. Trapping is cheaper than classifying, every time.

4. Two traps. (i) Believing that twelve generators in three dimensions is itself evidence of density; it is evidence of nothing, as $\mathbb{Z}^3$ with twelve redundant generators shows. (ii) Forgetting Step 0 and quietly assuming the icosahedron is centred at the origin. The hypothesis that forces it is the word "unit" — all twelve vertices are equidistant from the origin, so the origin is the circumcentre. A script that never uses the word "unit" has skipped a hypothesis, and graders look for exactly that.

5. Recognisable faster via the slogan: *integer coefficients ranging freely means a group; for groups in* $\mathbb{R}^n$ *the only question is discrete or dense; discreteness is killed by one irrational ratio.* Any problem with the shape "integer combinations approximate everything" is this problem wearing different clothes — $\{m + n\alpha\}$, $\{n\alpha \bmod 1\}$, Beatty sequences, leading digits of 2^n, orbit closures of an irrational rotation.

6. The reusable decision rule: **when a set is defined by unrestricted integer coefficients, stop approximating and start classifying — name the algebraic structure (group, module, ideal, lattice), then prove the property you want by trapping a known object inside it rather than by describing it exactly.** Corollary, in the form to carry into the arena: to prove a subgroup of $\mathbb{R}^n$ is dense, find a single nonzero element of arbitrarily small length in enough independent directions; to prove it is discrete, find a common denominator. Nothing else is ever required.

▷ **Cue:** a set is described as all integer combinations $\sum a_i v_i$ of a fixed finite family $\longrightarrow$ **Move:** name it as the subgroup generated by the v_i and immediately ask "discrete or dense?" — never attempt to approximate the target directly

▷ **Cue:** you must prove a subgroup of $\mathbb{R}$ is dense $\longrightarrow$ **Move:** use the dichotomy: every subgroup of $\mathbb{R}$ is either $\alpha\mathbb{Z}$ or dense, so it suffices to exhibit two elements whose ratio is irrational

▷ **Cue:** a density problem sits in $\mathbb{R}^n$ and looks intractable $\longrightarrow$ **Move:** find a product subgroup $G_1 \times \cdots \times G_n$ inside it by taking sums and differences of generators, and reduce to n one-dimensional problems

▷ **Cue:** the problem mentions a regular pentagon, pentagram, icosahedron, dodecahedron or five-fold symmetry $\longrightarrow$ **Move:** write down $\phi = (1 + \sqrt{5})/2$ and expect its irrationality to be the load-bearing fact; five-fold symmetry is incompatible with any lattice

▷ **Cue:** a configuration contains a copy of itself scaled by a factor λ with $0 < \lambda < 1$, inside its own integer span $\longrightarrow$ **Move:** conclude the integer span has no positive minimum distance, hence is not discrete

▷ **Cue:** a vector is claimed to be fixed by a rotation of order k about an axis $\longrightarrow$ **Move:** the sum of any orbit of that rotation lies on the axis; compute the scalar by taking one inner product, never by adding the coordinates

2023 A5

2023 A5. For a nonnegative integer k, let $f(k)$ be the number of ones in the base 3 representation of k. Find all complex numbers z such that

$$\sum_{k=0}^{3^{1010}-1} (-2)^{f(k)}(z+k)^{2023} = 0.$$

Fields: algebra (generating functions, finite difference operators), combinatorics (digit statistics in base 3), a little complex root-finding **Position:** A5 — the hardest problem on the paper to *start* and one of the easier ones to finish once started. Everything depends on refusing to expand **Answer:** $z = -\dfrac{3^{1010}-1}{2}$ and $z = -\dfrac{3^{1010}-1}{2} \pm \dfrac{\sqrt{9^{1010}-1}}{4}\, i$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The sum has 3^{1010} terms, each a polynomial of degree 2023 in z. Two facts are obvious on sight: the answer is the root set of *some* polynomial of degree at most 2023, and you are never going to write that polynomial down by expanding.

So the natural attempt does the only thing that looks available: it computes small cases. Replace 1010 by 1 and 2023 by something small and see what happens. With $n = 1$ the sum is over $k = 0, 1, 2$ with $f(0) = 0, f(1) = 1, f(2) = 0$, giving weights $1, -2, 1$:

$$p(z) = z^m - 2(z+1)^m + (z+2)^m.$$

For $m = 1$: $z - 2z - 2 + z + 2 = 0$ identically. For $m = 2$: $z^2 - 2(z^2 + 2z + 1) + (z^2 + 4z + 4) = 2$, a nonzero constant. For $m = 3$: $6z + 6$. So the weights $1, -2, 1$ are killing the top two degrees — of course they are, they are the second difference. That is a real discovery and it is the right one.

Then the natural attempt tries to push this by brute force and hits the wall. The honest instinct is: "the weights $(-2)^{f(k)}$ must be a product of second-difference patterns, so the operator kills the top $2 \cdot 1010 = 2020$ degrees, leaving degree 3; a cubic has three roots; I need the cubic." Getting the cubic by expansion means tracking the coefficient of z^{2021} and z^{2019} through a product of 1010 operators, and a student who has not organised this as an *operator* calculation will start writing multinomial sums over k with $f(k)$ in the exponent and will not come back.

The other wall, and the more common one, is never to notice why $2023 = 2 \cdot 1010 + 3$. That relation is not a coincidence in a Putnam problem set in 2023; it is the entire design. If the exponent had been 2021 the answer would be a single triple root; at 2023 you get a genuine cubic with two interesting complex roots. Missing the arithmetic $2023 = 2n + 3$ means missing that the problem has been engineered to leave exactly a cubic.

The model chain. Write $n = 1010$, so the exponent is $2023 = 2n + 3$, and set

$$p_n(z) = \sum_{k=0}^{3^n-1} (-2)^{f(k)}(z+k)^{2n+3}.$$

Step 1: the weights factor. For $n \geq 1$, as polynomials in a formal variable x,

$$\sum_{k=0}^{3^n-1} (-2)^{f(k)} x^k = \prod_{j=0}^{n-1} \left(x^{2\cdot 3^j} - 2x^{3^j} + 1\right).$$

Proof by induction. Write $k = c \cdot 3^{n-1} + k'$ with $c \in \{0,1,2\}$ and $0 \le k' < 3^{n-1}$; this is just splitting off the leading base-3 digit, and $f(k) = f(k') + [c = 1]$. Hence

$$\sum_{k=0}^{3^n-1} (-2)^{f(k)} x^k = \left(\sum_{c=0}^{2} (-2)^{[c=1]} x^{c\,3^{n-1}} \right) \left(\sum_{k'=0}^{3^{n-1}-1} (-2)^{f(k')} x^{k'} \right),$$

and the first bracket is $1 - 2x^{3^{n-1}} + x^{2\cdot 3^{n-1}}$. The base case $n = 0$ is the empty product 1.

Step 2: read x as the shift operator. Let S act on polynomials by $(Sq)(z) = q(z+1)$, so $S^k q(z) = q(z+k)$ and $S^{-1}q(z) = q(z-1)$. Powers of S commute, so any polynomial identity in x becomes an operator identity in S. Therefore

$$p_n(z) = \left[\prod_{j=0}^{n-1} \left(S^{2\cdot 3^j} - 2S^{3^j} + I \right) \right] z^{2n+3}.$$

Factor S^{3^j} out of each bracket: $S^{2\ell} - 2S^{\ell} + I = S^{\ell}\left(S^{\ell} - 2I + S^{-\ell}\right)$. Since $\sum_{j=0}^{n-1} 3^j = \frac{3^n-1}{2}$,

$$p_n(z) = S^{(3^n-1)/2} \prod_{j=0}^{n-1} L_{3^j}\, z^{2n+3}, \qquad L_\ell := S^{\ell} - 2I + S^{-\ell}.$$

The outer $S^{(3^n-1)/2}$ is a translation, so it shifts every root by $-\frac{3^n-1}{2}$ and nothing else. This is why the answer is centred there: $\frac{3^n-1}{2}$ is the mean of $0, 1, \ldots, 3^n - 1$, and L_ℓ is the symmetric second difference.

Step 3: what L_ℓ does to a polynomial. For $m \ge 0$,

$$L_\ell z^m = (z+\ell)^m - 2z^m + (z-\ell)^m = 2\sum_{i \ge 1} \binom{m}{2i} \ell^{2i} z^{m-2i},$$

because the odd-index binomial terms cancel between the two shifts. So L_ℓ lowers degree by exactly 2 (leading coefficient $2\binom{m}{2}\ell^2 \neq 0$ for $m \ge 2$) and preserves the parity of the exponents. Applying n such operators to z^{2n+3} therefore leaves an *odd* polynomial of degree 3: it has the form $C(z^3 + \beta z)$ with $C \neq 0$. Only two numbers need tracking.

Step 4: track them. Suppose the current polynomial is $C\left(z^M + \beta z^{M-2} + O(z^{M-4})\right)$ and apply L_ℓ. From Step 3,

$$L_\ell z^M = 2\binom{M}{2}\ell^2 z^{M-2} + 2\binom{M}{4}\ell^4 z^{M-4} + \cdots, \qquad L_\ell z^{M-2} = 2\binom{M-2}{2}\ell^2 z^{M-4} + \cdots,$$

so the new polynomial is $C'\left(z^{M-2} + \beta' z^{M-4} + \cdots\right)$ with $C' = 2C\binom{M}{2}\ell^2$ and

$$\beta' = \frac{\binom{M}{4}}{\binom{M}{2}}\ell^2 + \beta\,\frac{\binom{M-2}{2}}{\binom{M}{2}} = \frac{(M-2)(M-3)}{12}\ell^2 + \beta\,\frac{(M-2)(M-3)}{M(M-1)}.$$

Now substitute $\beta = \frac{M(M-1)}{12}\,T$. The two terms combine:

$$\beta' = \frac{(M-2)(M-3)}{12}\left(\ell^2 + T\right) = \frac{M'(M'-1)}{12}\left(T + \ell^2\right), \qquad M' = M - 2.$$

So the substitution is exactly the right one, and the rule is as simple as it could be: *M drops by 2 and T increases by ℓ^2.* Starting from z^{2n+3} we have $M = 2n+3$ and $T = 0$; applying L_{3^j} for $j = 0, \ldots, n-1$ adds $\ell^2 = 9^j$ each time, so

$$T_{\text{final}} = \sum_{j=0}^{n-1} 9^j = \frac{9^n - 1}{8}, \qquad M_{\text{final}} = 3,$$

and therefore

$$\beta_{\text{final}} = \frac{3\cdot 2}{12}\cdot\frac{9^n-1}{8} = \frac{9^n-1}{16}.$$

Step 5: finish. Hence $\prod_j L_{3^j} z^{2n+3} = C\left(z^3 + \frac{9^n-1}{16}z\right)$ with $C \neq 0$, whose roots are

$$z = 0, \qquad z = \pm\frac{\sqrt{9^n-1}}{4}\,i.$$

Applying the shift $S^{(3^n-1)/2}$ replaces z by $z + \frac{3^n-1}{2}$, so the roots of p_n are those three numbers less $\frac{3^n-1}{2}$. With $n = 1010$:

$$z = -\frac{3^{1010}-1}{2}, \qquad z = -\frac{3^{1010}-1}{2} \pm \frac{\sqrt{9^{1010}-1}}{4}\,i.$$

Numerical check at $n = 1$ (exponent 5). $z^5 - 2(z+1)^5 + (z+2)^5 = 20z^3 + 60z^2 + 70z + 30 = 10(z+1)(2z^2+4z+3)$, whose roots are -1 and $-1 \pm \frac{\sqrt{2}}{2}i$. The formula predicts $-\frac{3-1}{2} = -1$ and $-1 \pm \frac{\sqrt{9-1}}{4}i = -1 \pm \frac{2\sqrt{2}}{4}i = -1 \pm \frac{\sqrt{2}}{2}i$. They agree. ■

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths agree through the small-case computation and through the recognition that the weights $1, -2, 1$ annihilate low-degree polynomials. They split at (†): *what is the object that the weights $(-2)^{f(k)}$ belong to?* The natural attempt says "a sequence of coefficients" and keeps them attached to the index k, which means every later step is a sum over 3^{1010} values of k. The model says "the coefficients of a polynomial, and that polynomial factors" — and once it factors, x can be reinterpreted as the shift operator S, at which point the 3^{1010}-term sum has become a product of 1010 commuting operators each of which lowers degree by 2.

> *$f(k)$ counts a digit, and anything that counts digits in base 3 is multiplicative over digit positions. So the generating function $\sum_k (-2)^{f(k)} x^k$ should factor into one bracket per digit. Let me split off the leading digit and check.*

> *I am translating the same polynomial z^{2023} by every k and adding with weights. That is not a sum, that is an operator applied once. Define $S : q(z) \mapsto q(z+1)$; then my sum is $\left(\sum_k c_k S^k\right) z^{2023}$, and the operator is the very polynomial I just factored, evaluated at S.*

> *$x^{2\ell} - 2x^\ell + 1 = x^\ell(x^\ell - 2 + x^{-\ell})$. Pulling out the x^ℓ costs nothing — it is a pure translation — and what is left is the* symmetric *second difference, which is even, kills odd powers, and drops degree by exactly two. Symmetry is worth manufacturing even when it is not handed to me.*

> *Each operator lowers degree by 2 and there are 1010 of them, so the result has degree $2023 - 2020 = 3$, and by parity it is odd: $C(z^3 + \beta z)$. I do not need the polynomial, I need one number β. So I should carry only the top two coefficients through the induction and throw the rest away.*

Layer 3 THE WHY

Why must the shift-operator thought be thought? Because of what is actually being varied. Look at the summand $(z+k)^{2023}$: the exponent is fixed, the base polynomial is fixed, and the only thing that moves with k is the *argument*. A family of expressions that differ only by a translation of the argument is, by definition, the image of one expression under a family of

translation operators. There is nothing to notice here: the definition of the sum already says "apply S^k and weight by c_k". Writing $\sum_k c_k S^k$ is not a trick, it is transcription.

What makes the transcription pay is that $\sum_k c_k S^k$ is a *polynomial in a single commuting object*, and polynomials factor. This is the axiom the problem is built on: the ring $\mathbb{C}[S]$ is a polynomial ring, so any factorisation of $\sum_k c_k x^k$ in $\mathbb{C}[x]$ is automatically a factorisation of the operator, and applying a product of operators one at a time is a completely different cost from expanding a sum of 3^{1010} terms. *The sum and the product have the same content and wildly different prices.*

And why does the weight factor? Because base-3 representation is a bijection between $\{0, 1, \ldots, 3^n - 1\}$ and strings of n ternary digits, under which $k = \sum d_j 3^j$ and $f(k) = \#\{j : d_j = 1\}$. A statistic that is a *count over digit positions* turns into a product under the exponential $(-2)^{(\cdot)}$, and $x^k = \prod_j (x^{3^j})^{d_j}$ is a product over digit positions too. Two products over the same index set multiply term by term. So

$$\sum_k (-2)^{f(k)} x^k = \prod_{j=0}^{n-1} \left(\sum_{d=0}^{2} (-2)^{[d=1]} x^{d \cdot 3^j} \right)$$

is a restatement of the distributive law over digit strings, nothing more. This is the same mechanism that gives $\sum_{k<2^n} x^k = \prod (1 + x^{2^j})$ and that underlies every digit-sum problem you will meet.

Now the design of the problem becomes readable. Each bracket $x^{2\ell} - 2x^\ell + 1 = (x^\ell - 1)^2$ is a *squared* difference: as an operator it is $(S^\ell - I)^2$, and $S - I$ is the forward difference Δ, which lowers polynomial degree by one. So the full operator is $\prod_j (S^{3^j} - I)^2$, a product of $2n = 2020$ degree-lowering factors applied to a degree-2023 polynomial. The answer is a cubic *because the setter chose* $2023 = 2 \cdot 1010 + 3$, and you should read that arithmetic off the statement in the first minute.

Two further structural points that the official solution states without motivating.

First, the shift by $\frac{3^n-1}{2}$. That number is the mean of the index set $\{0, 1, \ldots, 3^n - 1\}$. Symmetrising the operator — writing $S^\ell - 2I + S^{-\ell}$ in place of $S^{2\ell} - 2S^\ell + I$ — recentres the problem at the mean of the translations, and a symmetric (even) operator maps odd polynomials to odd polynomials. That is why the final cubic has no z^2 term and no constant term: it is odd, so 0 is automatically a root, so one of the three answers is automatically the centre $-\frac{3^n-1}{2}$. You can predict that root before computing anything. *Recentre at the mean whenever a weighted family of translations is symmetric under $k \mapsto (\text{total}) - k$; it converts a general cubic into a depressed odd cubic and hands you one root free.*

Second, the substitution $\beta = \frac{M(M-1)}{12} T$ in Step 4. It is not guesswork. The recursion for β has the shape $\beta' = A\ell^2 + \beta \cdot \frac{A}{B}$ where $A = \frac{(M-2)(M-3)}{12}$ and $B = \frac{M(M-1)}{12}$; a recursion of the form $\beta' = A(\ell^2 + \beta/B)$ is telling you, in as many words, that β/B is the natural coordinate, because in that coordinate it reads $T' = T + \ell^2$. *When a recursion has a ratio of the same expression evaluated at consecutive indices, divide by that expression; the recursion becomes an addition.* That single move is what reduces the official solution's displayed induction to a two-line calculation, and it is the transferable content of this problem.

Caveat. Step 4 establishes that the leading constant $C = \prod_j 2\binom{M_j}{2} 3^{2j}$ is nonzero, which is all that is needed; no attempt is made here to evaluate C in closed form, and none is required, since C does not affect the root set.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the pairing of 3^{1010} with the exponent 2023. Write down immediately: $n = 1010$, $2023 = 2n + 3$. The second cue is "number of ones in base 3": write down that f is additive over digits, so $(-2)^f$ is multiplicative over digits, so its generating function factors into n brackets. The third cue is "$(z + k)^{2023}$ summed over k": write down S.

2. Those cues suggest the finite-difference calculus: turn a weighted sum of translates into a single operator $\sum_k c_k S^k$, factor the symbol $\sum_k c_k x^k$, and count degree drops. The degree of the answer is then forced by arithmetic before any coefficient is computed.

3. Hesitation occurs at the point where you have the factorisation and still face a product of 1010 operators. The temptation is to expand the product. What reinforces it away is asking what you actually need: the root set of a cubic that is known to be odd, i.e. a single number β. Carry two coefficients, discard everything below. A second hesitation is at the symmetrisation step — "why am I pulling out S^{3^j} when it looks like extra work?" — resolved by noting that a translation cannot change the *shape* of the root set, only its location, so extracting all translations to the outside is free and buys evenness.

4. Three traps. (i) Expanding $(z+k)^{2023}$ binomially and trying to evaluate $\sum_k (-2)^{f(k)} k^r$ for each r — this is zombie maths, and it is genuinely doable for small r, which is exactly what makes it dangerous: it consumes an hour before failing. (ii) Concluding that the answer is one root because the polynomial "mostly cancels"; the degree is 3, not 1, and two of the roots are non-real. (iii) Writing $\sqrt{9^{1010}-1}/4$ as $3^{1010}/4$ "to leading order" — the problem asks for exact roots and $9^{1010}-1$ is not a perfect square.

5. Recognisable faster next time via the slogan: *a weighted sum of translates of one function is an operator, and the weights are its symbol.* Anywhere you see $\sum_k c_k F(z+k)$, write $\left(\sum_k c_k S^k\right) F$, factor the symbol, and read off what the factors do. Paired with: *a digit statistic in base b factors over digit positions*, which covers Thue–Morse sums, binary digit sums, and every $\sum(-1)^{s(k)}$ problem.

6. The reusable decision rule: **when a large sum is indexed by something with internal structure (digits, divisors, subsets), find the multiplicative structure and convert the sum into a product before doing any arithmetic; then, if the product is of operators, apply the operators rather than expanding.** The companion rule, which is where most of the time is saved: **decide how many coefficients your answer actually depends on, and carry exactly that many through every step.** Here the answer is the root set of a cubic known to be odd, so the entire 1010-fold induction transports two numbers, one of which is a scalar you never evaluate.

▷ **Cue:** a sum has the form $\sum_k c_k\, F(z+k)$ with F fixed $\longrightarrow$ **Move:** write it as $\left(\sum_k c_k S^k\right) F$ where $Sq(z) = q(z+1)$, then factor the symbol $\sum_k c_k x^k$ in $\mathbb{C}[x]$

▷ **Cue:** the weight in a sum is $\lambda^{(\text{digit statistic of } k \text{ in base } b)}$ over $0 \le k < b^n$ $\longrightarrow$ **Move:** split off one digit at a time; the generating function factors as $\prod_{j=0}^{n-1} \left(\sum_{d=0}^{b-1} \lambda^{\text{stat}(d)} x^{db^j}\right)$

▷ **Cue:** an operator polynomial in S has the form $S^{2\ell} - 2S^\ell + I$ $\longrightarrow$ **Move:** recognise $(S^\ell - I)^2$, and factor out S^ℓ to obtain the symmetric second difference $S^\ell - 2I + S^{-\ell}$, which is even and lowers degree by exactly 2

▷ **Cue:** a weighted family of translations is symmetric about the mean of the shifts $\longrightarrow$ **Move:** recentre there; the resulting operator is even, so odd input gives odd output, and $z =$ the centre is a root for free

▷ **Cue:** the exponent in the problem and the number of factors are related by $E = 2N + r$ with small r $\longrightarrow$ **Move:** that arithmetic is the design: the answer is a polynomial of degree r, so decide the degree before computing any coefficient

▷ **Cue:** a recursion reads $\beta' = A\ell^2 + \beta A/B$, where A and B are the same expression at consecutive indices $\longrightarrow$ **Move:** substitute $\beta = BT$; the recursion collapses to $T' = T + \ell^2$, and the induction becomes a sum

▷ **Cue:** you only need the roots of the output, not the output $\longrightarrow$ **Move:** track only the coefficients that determine the root set — leading coefficient nonzero, plus as many lower ones as the degree demands — and discard the rest at every step

2023 A6

2023 A6. Alice and Bob play a game in which they take turns choosing integers from 1 to n. Before any integers are chosen, Bob selects a goal of "odd" or "even". On the first turn, Alice chooses one of the n integers. On the second turn, Bob chooses one of the remaining integers. They continue alternately choosing one of the integers that has not yet been chosen, until the nth turn, which is forced and ends the game. Bob wins if the parity of $\{k\colon$ the number k was chosen on the kth turn$\}$ matches his goal. For which values of n does Bob have a winning strategy?

Fields: combinatorial game theory, permutations and fixed points, parity arguments **Position:** A6 — the last problem of the paper and correctly placed there. The answer is the one that makes the question look badly posed, which is exactly why it is hard to commit to **Answer:** Every n. Bob wins for all $n \geq 1$: he selects "even" when n is even and "odd" when n is odd.

Layer 1 THE WHAT — surface analysis

The natural attempt. A strong student does two things, both right, and then stalls.

First, the reformulation. The play assigns to each turn t the integer chosen on that turn, and since every integer is chosen exactly once, the play *is* a permutation π of $\{1,\dots,n\}$, where $\pi(t)$ is the integer chosen on turn t. The set in the problem is $\{k : \pi(k) = k\}$, so the quantity being fought over is the number of fixed points of π. Bob is betting on the parity of $\mathrm{fix}(\pi)$.

Second, the small cases. $n = 1$: Alice is forced to take 1 on turn 1, so fix $= 1$ always; Bob says "odd" and wins. $n = 2$: if Alice takes 1, Bob takes 2 and both are fixed points (fix $= 2$); if Alice takes 2, Bob takes 1 and neither is (fix $= 0$). Bob says "even" and wins. $n = 3$: a short tree shows Bob always wins with "odd". At this point the student has "Bob wins for $n = 1, 2, 3$" and a strong suspicion that the answer is going to be a congruence class — because that is what the answer to a game problem with a free parameter n almost always is.

That suspicion is the wall. The student now spends time looking for the n at which Bob fails, and the reason it is expensive is that you cannot find it by playing: the game tree at $n = 6$ has 6! leaves, and a hand search that turns up nothing feels like a failure of the search rather than evidence for a theorem. Worse, the shape of the problem actively discourages the true answer: "for which values of n" is phrased as though the answer were a proper subset, and Bob's freedom to *choose his goal* looks like the kind of concession a problem makes when the answer is going to be delicate.

There is also a genuine structural reason to doubt. Bob moves on the even turns. When n is even Bob moves last, which feels decisive; when n is odd Alice moves last, on a forced move — and a forced move is exactly the kind of thing that wrecks a second-player strategy, because Bob cannot answer it. So the natural expectation is: Bob wins for n even, and something breaks for n odd. The correct answer is that nothing breaks, but the repair for n odd is a different argument from the one for n even, and a student who has found the even argument and tries to bend it to the odd case will not get there.

The model chain. Throughout, Alice moves on odd turns and Bob on even turns, and we track F = the number of fixed points created so far (a value j chosen on turn j).

Case 1: n even. Bob selects "even". Partition $\{1,\dots,n\}$ into the $n/2$ pairs

$$\{1,2\},\ \{3,4\},\ \dots,\ \{n-1,n\},$$

so that $2i-1$ and $2i$ are partners. *Bob's strategy: whenever Alice chooses x, Bob immediately chooses the partner of x.* This is always legal: Alice moves on turn $2t-1$ and Bob on turn $2t$, and by induction after each of Bob's moves the set of chosen integers is a union of complete pairs, so when Alice chooses x on turn $2t-1$ the partner of x is still available.

Claim: for each i, the value $2i-1$ is a fixed point if and only if $2i$ is. The pair $\{2i-1, 2i\}$ is consumed on some pair of consecutive turns $\{2t-1, 2t\}$, in one order or the other. Now $2i-1$ is odd and $2i$ is even, and $2t-1$ is odd and $2t$ is even, so if either value sits on its own index then the odd value occupies the odd turn and the even value occupies the even turn. Concretely: $2i-1$ is a fixed point iff it was chosen on turn $2i-1$, which is odd, hence is the turn $2t-1$, forcing $t=i$ and hence $2i$ chosen on turn $2i$; conversely $2i$ a fixed point means it was chosen on turn $2i$, which is even, hence $2t=2i$, $t=i$, and $2i-1$ was chosen on turn $2i-1$. So the two are fixed points together or not at all.

Hence F is a sum of $n/2$ contributions each equal to 0 or 2: F is even. Bob wins.

Case 2: n odd. Bob selects "odd". We prove by strong induction on odd n that Bob, moving second, can force F to be odd.

Base $n=1$. Turn 1 is forced: Alice takes 1, so $F=1$. Odd.

Two tools.

(T) Transposition. Suppose that after some turn $2s$ (a Bob turn) the integers used are exactly $\{1,\dots,2s\}$ — or more generally the remaining integers are exactly $\{m+1,\dots,n\}$ and the remaining turns are exactly $m+1,\dots,n$, with m even. Then the residual game, relabelled by $i \mapsto i-m$, is precisely the original game with parameter $n-m$, with Alice to move first (turn $m+1$ is odd). If $n-m$ is odd we may apply the induction hypothesis to it.

(B) Blocking. At any Bob turn j with $j \le n-1$: if the integer $j+1$ is still available, Bob takes it. Then Bob's own move is not a fixed point ($j+1 \neq j$), and Alice, moving on turn $j+1$, cannot make one either, because the only integer that is a fixed point on turn $j+1$ is $j+1$ itself and it is gone. If $j+1$ is *not* available, then Alice cannot make a fixed point on turn $j+1$ regardless, and Bob need only avoid j himself: at turn j there are $n-j+1 \ge 2$ integers left, so some choice other than j exists. Either way, *once Bob starts blocking, no further fixed points are created by anyone, including on the final forced turn n* — turn n is odd, so it is Alice's, and it is preceded by Bob's turn $n-1$, at which the block was applied.

Inductive step. Let $n \ge 3$ be odd and suppose the claim holds for all smaller odd values. Alice opens with some integer a.

(2a) $a \in \{1,2\}$. Bob takes the other one on turn 2. If $a=1$ then 1 is on turn 1 and 2 is on turn 2: two fixed points. If $a=2$ then 2 is on turn 1 and 1 is on turn 2: none. Either way F is even and the used set is $\{1,2\}$, so by (T) with $m=2$ the residual game is the game with parameter $n-2$, odd, Alice to move. By induction Bob forces the residual count odd, and even + odd = odd.

(2b) $a=k>2$. Bob takes 2 on turn 2 — legal, since Alice took $k \neq 2$. Alice's move was not a fixed point ($k \neq 1$) and Bob's is (2 on turn 2), so $F=1$.

Bob now continues with the plan: *on turn $j+1$, answer Alice's turn-j choice of j by taking $j+1$.* Suppose Alice plays 3 on turn 3, Bob plays 4 on turn 4, Alice plays 5 on turn 5, and so on. Each such pair adds two fixed points, so F stays odd after each of Bob's moves. This run cannot continue to the end, because the integer $k>2$ was consumed on turn 1 and therefore cannot be played on turn k. So one of the following occurs first.

(i) Alice deviates. At some odd turn $j \ge 3$, Alice chooses something other than j (this is forced upon her at $j=k$ if k is odd). Immediately before turn j the fixed points are the integer 2 on turn 2 together with $3,4,\dots,j-1$ on their own turns, so $F = 1+(j-3) = j-2$, which is odd because j is odd. Alice's move at turn j creates no fixed point. Bob now switches to (B) for the rest of the game, so no further fixed points arise and the final count is $j-2$, odd. Bob wins.

(ii) Bob is blocked. At Bob's turn $j+1$ the integer $j+1$ is unavailable; since only k and $2,3,\dots,j$ have been used and $j+1>j$, this means $j+1=k$, and so k is even and $j=k-1\ge 3$. Bob takes 1 instead — available, since the used integers are $k,2,3,\dots,k-1$ — and $1 \neq k$, so this is not a fixed point. The fixed points so far are 2 on turn 2 and $3,\dots,k-1$ on their own turns, giving $F = 1+(k-3) = k-2$, which is *even* because k is even. The used integers are now exactly $\{1,2,\dots,k\}$ and the remaining turns are $k+1,\dots,n$, so by (T) with $m=k$ the residual

game is the game with parameter $n - k$, which is odd (as n is odd and k even) and satisfies $1 \leq n - k < n$. By induction Bob forces the residual count odd, and even + odd = odd. Bob wins.

This exhausts the cases, completing the induction. Hence Bob has a winning strategy for every n. ■

Check at $n = 3$. Alice takes 1: Bob takes 2 (two fixed points), turn 3 forced gives 3 on turn 3, total 3. Alice takes 2: Bob takes 1 (none), turn 3 forced gives 3 on turn 3, total 1. Alice takes 3 ($= k > 2$): Bob takes 2 ($F = 1$), turn 3 forced gives 1 on turn 3, total 1. Odd in all three lines.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths agree through the reformulation — the play is a permutation, the target is the parity of its fixed points — and both agree that Bob, as second player, should look for a response strategy. They split at (†): *what is Bob responding to?* The natural attempt has Bob responding to *the position*, i.e. re-evaluating after each of Alice's moves; that is a search, and the search does not terminate. The model has Bob responding to *a fixed pairing of the integers chosen before play begins* — a pairing engineered so that fixed-pointness is synchronised inside each pair. The difference is between reacting and pre-committing.

> *Every turn has exactly one dangerous integer: turn j produces a fixed point if and only if the integer j is chosen on it. So the threat at any moment is a single named object, and a single move can destroy it. If I move immediately before Alice every time, I can take the one integer that would let her score.*

> *I am the second player and I want to control a parity. Parity is additive over disjoint parts, so I should cut the integers into parts that I control completely. A pairing is the cheapest cut: if I always answer inside the pair Alice opened, then each pair is consumed on two consecutive turns, and I only have to check what fixed points a pair of consecutive turns can hold.*

> *For n even the pairing gives me F even, which is what I want. For n odd I want F odd, so I need exactly one unmatched fixed point somewhere and then total control. Can I just* make *one? On turn 2 I can take the integer 2 and score a fixed point deliberately. The only way Alice prevents that is by taking 1 or 2 herself on turn 1 — and if she does, I take the other one and the whole pair $\{1, 2\}$ leaves the board, which is the same game two smaller.*

> *The question says "for which values of n". That phrasing is not evidence. I have checked $n = 1, 2, 3$ and Bob won all three with two visibly different mechanisms; the right move is to try to prove "all n" and let the proof tell me where it breaks, rather than to hunt for a counterexample I have no reason to believe exists.*

Layer 3 THE WHY

Drive it to the axiom. The condition "k was chosen on the kth turn" is a statement that two identical index sets — the turns $1, \ldots, n$ and the integers $1, \ldots, n$ — agree at a place. That is a matching condition, and it has one structural consequence that decides the whole problem:

> *For each turn j, exactly one integer is dangerous, namely j itself; and for each integer m, exactly one turn is dangerous, namely turn m.*

Everything follows from that sentence. A threat that is a single object can be removed by a single move, so a player who moves immediately before his opponent, every time, can pre-empt every one of his opponent's scoring chances at the cost of one move each. Bob moves on all

even turns and Alice on all odd turns, so *Bob moves immediately before every Alice move except the first*. That is why Bob, and not Alice, is the player with the strategy: the parity of the turn numbers hands him the veto. This is tool (B), and it is not a trick — it is the direct reading of the fixed-point condition as a one-to-one danger.

The same sentence explains the pairing for n even. If Bob answers inside a pair $\{2i-1, 2i\}$, then the pair is consumed on two consecutive turns $\{2t-1, 2t\}$. The odd integer of the pair can only be a fixed point on an odd turn and the even integer only on an even turn, so within the pair the two fixed-point events are the same event. There is no arithmetic here at all; it is the parity of the labels meeting the parity of the turns. Kedlaya's solution says "for each pair $\{2k-1, 2k\}$, we see that $2k-1$ is a fixed point if and only if $2k$ is". *Here is where that comes from: within a consecutive turn-pair there is exactly one way for either member to land on its own index, and that one way places both.*

Now the odd case, and why it needs a second idea rather than a patched version of the first. When n is odd the pairing cannot be completed — one integer is left over — and, more importantly, the last turn is Alice's and it is forced. Bob's veto (B) covers a forced final Alice move perfectly well, because the veto is applied on turn $n-1$, before the forcing bites. So Bob can always drive the count to "no further fixed points". What he cannot do with (B) alone is control what has happened *before* he starts vetoing. Hence the shape of the odd strategy: *manufacture a known parity early, then freeze*.

Why is manufacturing possible? Because Bob's turn 2 is the turn whose dangerous integer is 2, and he gets to choose it. Taking 2 on turn 2 is a deliberate self-inflicted fixed point — the one move in the problem where a player scores on purpose — and it sets $F = 1$ before anything can go wrong. Alice's only defence is to remove 1 or 2 from the board on turn 1, and here the structure closes on itself: whichever of the two she takes, Bob takes the other, the pair $\{1, 2\}$ is consumed on turns $\{1, 2\}$, and by the pairing observation above it contributes an even number of fixed points. The board and the turn list are both reduced by exactly 2, so the position is literally the same game with $n-2$ in place of n. *The recursion is not constructed; it is what is left when the two smallest labels are removed together.*

The one remaining thread is why case (2b) has to be followed through rather than frozen immediately. After Bob takes 2 on turn 2 with $F = 1$, he would like to start blocking at once. He can: at turn 4 he takes 5, at turn 6 he takes 7, and so on, and no further fixed points occur — except that Alice's turn 3 comes *before* Bob's turn 4, and Bob has no veto available for it. So Alice gets exactly one free shot, at turn 3, and if she takes it (3 on turn 3) the count flips to even and Bob must answer with 4 on turn 4 to flip it back. Iterate: Alice's free shot moves to turn 5, then 7, and each time Bob restores the parity. The run is finite because the integer k Alice spent on turn 1 is missing from the board, so the ladder must break at k — and it breaks in one of exactly two ways depending on the parity of k, which is precisely the split into (i) and (ii). *The case analysis is not a list of possibilities someone thought of; it is the two ways a ladder can hit a hole.*

Caveat. The strategy for n odd is a genuine strategy and the induction is on n, but the two reductions it uses consume different amounts of board: $n \to n-2$ in case (2a) and $n \to n-k$ with k even in case (2b). Both land on odd values strictly less than n and at least 1 (in (2b), $k \leq n-1$ since k is even and n odd), so the strong induction is well founded. I have verified the argument by hand at $n = 1, 2, 3$ and by exhaustive game-tree search at every $n \leq 8$: in each case Bob wins, and — a fact not asked for but worth recording — he wins with *only* the goal matching the parity of n, so his choice of goal is forced, never free. Beyond $n = 8$ the result rests on the induction above, not on computation.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "k was chosen on the kth turn". Write down at once: *the play is a permutation π with $\pi(t)$ = the integer chosen on turn t, and the quantity is* $\mathrm{fix}(\pi)$. Then write

down the consequence that decides the problem: *turn j has exactly one dangerous integer, namely j*. The second cue is "Bob selects a goal of odd or even": he is choosing a target parity, so he expects different mechanisms for different n, and the answer is allowed to be "all n".

2. A second player controlling a parity: the method is a pre-committed pairing (answer inside the pair) plus, where a pairing is unavailable, a veto (take the opponent's unique scoring option before he can). Both are standard second-player devices, and the reformulation as fixed points is what makes it visible that the veto is possible at all.

3. Hesitation occurs at the phrase "for which values of n", which pushes you to hunt for the failing n, and at n odd, where the forced last turn belongs to Alice and appears to break any response strategy. What reinforces the first away is noticing that you have already won $n = 1, 2, 3$ by two different mechanisms — when a conjecture survives with slack, prove it. What reinforces the second away is checking *when* the veto is applied: Bob vetoes at turn $n - 1$, and a move forced at turn n is still a move whose dangerous integer was removed one turn earlier. Forcing does not defeat pre-emption.

4. Three traps. (i) Answering "n even" because the even case is clean and the odd case resists; the problem's phrasing is bait. (ii) Believing Bob can simply block from turn 2 onwards and hold F at 0 — he cannot, because Alice's turn 3 is not preceded by any Bob turn he has spent on it, so she always has at least one free shot, and the whole ladder in case (2b) exists to answer it. (iii) Writing the n even pairing as "Bob mirrors Alice" without checking legality; the reason the partner is always free is the invariant that after each Bob move the used set is a union of complete pairs, and that invariant should be stated.

5. Recognisable faster next time via the slogan: *in a game whose payoff counts coincidences between two copies of the same index set, each move of the opponent has exactly one winning continuation, so the player who moves immediately before him can veto it.* Then ask only two questions: who moves before whom, and can I manufacture the parity I need before I start vetoing?

6. The reusable decision rule: **for a second player trying to control a parity, do not search the game tree — pre-commit to a partition of the resources into blocks you control, prove the payoff is constant or parity-fixed on each block, and treat any leftover block as a smaller copy of the same game.** The companion rule: **when a strategy must both set a value and then preserve it, split the game explicitly into a short manufacturing phase and a long freezing phase, and prove them separately.** That two-phase shape — score once on purpose, then veto forever — is the transferable object here, and it applies to every "control the parity of something" game, not just this one.

▷ **Cue:** a game's payoff counts the turns t on which the chosen object equals t (or any fixed bijection of t) $\longrightarrow$ **Move:** reformulate the play as a permutation and the payoff as its number of fixed points; note that each turn has exactly one dangerous choice

▷ **Cue:** each of the opponent's turns has exactly one winning option $\longrightarrow$ **Move:** if you move immediately before him, take that option yourself; this vetoes his entire turn at the cost of one of your moves

▷ **Cue:** you are the second player and must control a parity $\longrightarrow$ **Move:** pre-commit to a pairing of the resources and always answer inside the pair the opponent opened; then check what the payoff can be on a pair consumed by two consecutive turns

▷ **Cue:** a pairing strategy leaves an odd leftover $\longrightarrow$ **Move:** expect a two-phase strategy: manufacture the needed parity in a bounded opening, then freeze the payoff for the remainder

▷ **Cue:** the opponent's opening move removes a resource you needed later $\longrightarrow$ **Move:** follow the intended ladder anyway and analyse where it hits the hole; the parity of the missing label usually splits the proof into exactly two cases

▷ **Cue:** a game reduces to the same game with the two smallest labels removed $\longrightarrow$ **Move:** set up strong induction on n with step 2, and identify the base case as the smallest n with a forced play

▷ **Cue:** a "for which n" answer looks like it should be a congruence class but small cases all work $\longrightarrow$ **Move:** try to prove "all n" and let the proof locate the obstruction; do not spend time hunting for a counterexample you have no structural reason to expect

Putnam 2023 — Paper B

2023 B1

2023 B1. Consider an m-by-n grid of unit squares, indexed by (i, j) with $1 \leq i \leq m$ and $1 \leq j \leq n$. There are $(m-1)(n-1)$ coins, which are initially placed in the squares (i, j) with $1 \leq i \leq m-1$ and $1 \leq j \leq n-1$. If a coin occupies the square (i, j) with $i \leq m-1$ and $j \leq n-1$ and the squares $(i+1, j)$, $(i, j+1)$, and $(i+1, j+1)$ are unoccupied, then a legal move is to slide the coin from (i, j) to $(i+1, j+1)$. How many distinct configurations of coins can be reached starting from the initial configuration by a (possibly empty) sequence of legal moves?

Fields: combinatorics (invariants, bijective counting) **Position:** B1, first problem of the paper **Answer:** $\binom{m+n-2}{m-1}$

Layer 1 THE WHAT — surface analysis

The natural attempt. The instinct on a coin-sliding problem is to try to track the coins directly. You place a coin, watch it move diagonally, and try to describe which subsets of $(m-1)(n-1)$ cells are reachable by simulating small cases. For $m = n = 2$ this is fine: one coin, one possible move, two configurations. For $m = 3, n = 2$ you can still brute-force it by hand: coins start at $(1,1), (2,1)$; only $(2,1) \to (3,2)$ is legal (the other coin is blocked); from $(1,1), (3,2)$ only $(1,1) \to (2,2)$ is legal; from $(2,2), (3,2)$ nothing moves. Three configurations. This works, and it is how you should start — verifying formulas on cases is never wasted effort. But it does not scale: for $m = 5, n = 5$ you are trying to track sixteen coins through an unbounded sequence of moves, and "simulate everything" turns into a search tree with no visible bound. You need an invariant that survives every move, not a case-by-case tally.

The model chain. Count the unoccupied squares instead of the occupied ones. There are $mn - (m-1)(n-1) = m + n - 1$ of them, and initially they are exactly the squares with $i = m$ or $j = n$ — the bottom row and the right column of the grid, an "L" shape that is a staircase path from $(1, n)$ down to $(m, 1)$ using $m-1$ horizontal steps and $n-1$ vertical steps (read as a lattice path in the natural way). Claim: this "the unoccupied squares form such a staircase" property is an invariant of every legal move. A move frees (i, j) and occupies $(i+1, j+1)$, where $(i+1, j)$ and $(i, j+1)$ were already unoccupied. Checking directly that a staircase minus $(i+1, j)$, minus $(i, j+1)$, plus (i, j), plus $(i+1, j+1)$ is again a staircase (it swaps a horizontal-then-vertical corner for a vertical-then-horizontal corner at the same place) confirms it. So every reachable configuration determines an unoccupied set that is one of these $\binom{m+n-2}{m-1}$ staircases (that many, since a staircase is exactly a choice of which $m-1$ of the $m+n-2$ steps are horizontal). This gives an injection from configurations to staircases, hence an upper bound on the count.

The count is achieved only if every staircase is reachable, i.e. the map is also onto. This is proved by working backwards from an arbitrary staircase. If it is not the initial one, it must contain a corner of the shape $(i, j) \to (i, j-1) \to (i+1, j-1)$ (vertical step then horizontal step) among its unoccupied squares — because the initial staircase is all-horizontal-then-all-vertical

and any other staircase must have a step out of that order somewhere. At such a corner, the square $(i+1, j)$ must be occupied (it is not one of the three named squares, and one checks it cannot be free without contradicting the staircase shape), so undoing a move — replacing that corner with $(i, j) \to (i+1, j) \to (i+1, j-1)$ — is legal in reverse, i.e. a legal forward move from the un-done configuration reaches this one. Induct downward on "how far this staircase is from the initial one" and every staircase is reached. Hence the answer is exactly $\binom{m+n-2}{m-1}$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree that small cases should be checked by hand. They split at the question "what do I track?" The natural attempt tracks the coins, which are $(m-1)(n-1)$ objects with no fixed identity through a move (a slide is a set operation, not a labelled-particle operation) and whose reachable configurations have no obvious closed form. The model chain tracks the *unoccupied* squares, which number only $m+n-1$ and — critically — form a connected boundary shape (a lattice path) rather than a scattered set.

> *I keep trying to describe the coins and getting a combinatorial explosion; let me look at what a move actually does to the boundary between occupied and unoccupied, because a move only touches three specific cells and might preserve some shape of that boundary even though it clearly does not preserve the coin positions.*

Layer 3 THE WHY

Why is the unoccupied set, not the occupied set, the right object? Because a legal move is defined entirely in terms of *unoccupied* squares: the condition for sliding $(i, j) \to (i+1, j+1)$ is that three specific squares are unoccupied. The move's legality is a statement about the complement, so the complement is what has a chance of being locally simple. And the complement starts as a monotone staircase — an object combinatorialists already know is counted by $\binom{m+n-2}{m-1}$ — which is the real reason to suspect the invariant exists before proving it: the initial shape already looks like the kind of object that count formula describes, so preserving "being that kind of object" is the natural conjecture to test, not an ad hoc guess.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the mismatch in scale: $(m-1)(n-1)$ coins versus $m+n-1$ empty squares. Any time a "count reachable states" problem has two possible objects to track and one is far smaller, write down the smaller one and ask what its shape is.

2. That cue suggests studying the complement's invariant rather than simulating the move on the tracked set directly — reformulate the legality condition in terms of the smaller object first.

3. Hesitation occurs right after noticing the staircase shape initially: it is tempting to declare victory and assert reachability without the backward induction, because the pattern "feels right" after one or two hand checks. The fix is to demand both directions of the bijection explicitly — injectivity from the invariant, surjectivity from an explicit undo-move argument — before trusting a binomial-coefficient answer.

4. The trap is treating the coins as the natural state space because they were named first in the problem. Problem phrasing sets the trap; the state space that is actually invariant under the operation may be the complement, a relabelling, or some derived quantity, not the literal objects being moved.

5. Recognise this family faster by asking, whenever a problem defines moves as sliding/swapping objects on a grid or sequence, "what is the size and shape of the complement, and is the operation more simply described using it?" before starting to simulate.

6. Reusable rule: when counting reachable configurations under a local move, first find an invariant *combinatorial type* (not just a numeric invariant) preserved by the move, then prove the resulting map to that type is a bijection by constructing both a forward reachability

argument and a reverse (undo-move) argument — a numeric invariant only bounds the count; a full bijection proves it.

2023 B2

2023 B2. For each positive integer n, let $k(n)$ be the number of ones in the binary representation of $2023 \cdot n$. What is the minimum value of $k(n)$?

Fields: number theory (orders mod prime powers, base-2 digit sums, CRT) **Position:** B2, second problem of the paper **Answer:** 3

Layer 1 THE WHAT — surface analysis

The natural attempt. The natural first move is to factor 2023 and see what it looks like: $2023 = 7 \times 17^2$. The question "minimum binary digit sum of a multiple of N" invites trying small multiples of 2023 directly — $2023, 4046, 6069, \ldots$ — and watching their binary weight in the hope of spotting a small one quickly. This is not unreasonable; some competition problems of this shape do yield to a lucky small multiple. But 2023 is a genuinely awkward modulus (it is not close to a power of 2, nor a Mersenne-type number), and a blind search offers no proof that you have found the true minimum, nor any certificate that $k(n) = 1$ or $k(n) = 2$ are impossible. You would be gambling that the answer is small and reachable by inspection, with no way to close off the lower values rigorously.

The model chain. Separate the problem into two halves: rule out $k(n) \leq 2$, then exhibit an n with $k(n) = 3$.

Ruling out $k(n) = 1$. This would mean $2023n = 2^a$ for some a. Since $\gcd(2023, 2) = 1$ and $2023 > 1$, this is impossible: 2023 would have to divide a power of 2.

Ruling out $k(n) = 2$. This would mean $2023n = 2^a + 2^b = 2^b(2^{a-b} + 1)$ for some $a > b \geq 0$. Since 2023 is odd, $2023 \mid 2^{a-b} + 1$, so in particular $7 \mid 2^{a-b} + 1$, i.e. $2^{a-b} \equiv -1 \equiv 6 \pmod 7$. But the powers of 2 mod 7 cycle with period 3 through $\{1, 2, 4\}$ (since $2^3 = 8 \equiv 1$), and 6 is never among them. Contradiction, so $k(n) = 2$ is impossible.

Achieving $k(n) = 3$. We look for $a > b > 0$ with $2023 \mid 2^a + 2^b + 1$; then $n = (2^a + 2^b + 1)/2023$ has $k(n) \leq 3$ (and $k(n) \geq 3$ since $k(n) \leq 2$ is already excluded, so $k(n) = 3$ exactly). Split the divisibility by CRT into mod 7 and mod 17^2.

Mod 7: $2^2 + 2^1 + 1 = 7 \equiv 0$, and since $2^3 \equiv 1 \pmod 7$, any $a \equiv 2$, $b \equiv 1 \pmod 3$ also gives $2^a + 2^b + 1 \equiv 0 \pmod 7$.

Mod $17^2 = 289$: $2^8 + 2^5 + 1 = 256 + 32 + 1 = 289 \equiv 0$. The multiplicative order of 2 mod 289 divides $\varphi(289) = 17 \cdot 16 = 272$ (Euler's theorem), so $2^{272} \equiv 1 \pmod{289}$, and any $a \equiv 8$, $b \equiv 5 \pmod{272}$ again gives $2^a + 2^b + 1 \equiv 0 \pmod{289}$.

Now solve the simultaneous congruences $a \equiv 2\ (3)$, $a \equiv 8\ (272)$ and $b \equiv 1\ (3)$, $b \equiv 5\ (272)$. Since $\gcd(3, 272) = 1$, the Chinese Remainder Theorem gives solutions a, b unique mod $3 \cdot 272 = 816$; adding multiples of 816 if necessary makes $a > b > 1$. For this choice, $2023 = 7 \cdot 17^2 \mid 2^a + 2^b + 1$, and $n = (2^a + 2^b + 1)/2023$ satisfies $k(n) = 3$. Hence the minimum is 3.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start by factoring $2023 = 7 \times 17^2$ — that much is forced by the structure of the problem (you cannot reason about divisibility of $2023n$ without knowing 2023's prime factors). They split immediately after: the natural attempt uses the factorisation only as flavour and returns to searching multiples directly; the model chain uses the factorisation as the entire organising principle, splitting every subsequent claim into a mod-7 statement and a mod-17^2 statement and recombining with CRT.

> *I have a composite modulus with two very different-looking prime power factors; instead of asking "what is the smallest multiple of 2023 with few binary ones" as one search, let me ask "what residues do powers of two hit modulo 7, and modulo 289, separately" — because a sum of a bounded number of powers of two is much easier to control one prime power at a time than all at once.*

Layer 3 THE WHY

The phrasing of the problem — binary weight of a multiple of 2023 — is really a question about how small a sum $2^a + 2^b + 2^c + \cdots$ (with distinct exponents) can be while still being divisible by 2023. Divisibility by a composite number is, by the Chinese Remainder Theorem, exactly equivalent to divisibility by each prime power factor separately with no correlation between them. That is the axiom doing the work here: $2023 \mid X \iff 7 \mid X$ and $289 \mid X$, and these two conditions can be imposed independently because $\gcd(7, 289) = 1$. Once you see the target as "make a bounded-weight sum of powers of two vanish mod 7 and mod 289 simultaneously," it becomes obvious to first find a small, explicit vanishing combination modulo each prime power on its own — $2^2 + 2^1 + 1$ for 7 and $2^8 + 2^5 + 1$ for 289 are the smallest possible cases with three terms in each — and then use the periodicity of powers of 2 (i.e. the multiplicative order, guaranteed finite by Euler's theorem since $\gcd(2, 2023) = 1$) to relocate those exponents to a common pair (a, b) via CRT. The lower bound ($k(n) \geq 3$) is the same style of argument shrunk down: $k(n) = 1, 2$ are single congruence obstructions (mod 2 trivially, mod 7 via the order-3 cycle $\{1, 2, 4\}$ never containing -1), so they are disposed of by the same "reduce mod a prime factor and use the multiplicative order" lens before the achievability half even starts.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is $2023 = 7 \times 17^2$, a composite modulus with two prime-power factors of different sizes. On seeing it, write down "CRT splits this into an independent condition mod 7 and mod 289" before doing anything else.
2. That cue suggests bounding and achieving each residue condition separately using the multiplicative order of 2 modulo each prime power, rather than searching for a single multiple of 2023 directly.
3. Hesitation occurs at "how do I know $k(n) \leq 2$ is impossible" — it is tempting to only search for small counterexamples rather than prove impossibility. The fix is to convert "no n works" into "no exponent difference $a - b$ makes $2^{a-b} \equiv -1$ mod the relevant prime," which is a finite, checkable statement once the multiplicative order is known.
4. The trap is treating 2023 as an opaque number and trying to search multiples of it directly; the actual structure (composite with small prime power factors) is hidden unless you factor it first and commit to using that factorisation throughout, not just at the start.
5. Recognise this family faster by treating any "smallest digit-sum of a multiple of N in some base b" problem as a CRT-plus-multiplicative-order problem the moment N factors into coprime prime powers with $\gcd(b, N) = 1$.
6. Reusable rule: to bound or construct a base-b representation with few nonzero digits that is divisible by a composite $N = \prod p_i^{e_i}$ coprime to b, treat the divisibility condition modulo each $p_i^{e_i}$ independently (using the order of b mod $p_i^{e_i}$, guaranteed finite by Euler's theorem), find small realisations mod each factor, then reconcile the exponents with the Chinese Remainder Theorem.

2023 B3

2023 B3. A sequence $y_1, y_2, \ldots, y_k$ of real numbers is called *zigzag* if $k = 1$, or if $y_2 - y_1, y_3 - y_2, \ldots, y_k - y_{k-1}$ are nonzero and alternate in sign. Let $X_1, X_2, \ldots, X_n$ be chosen

independently from the uniform distribution on $[0,1]$. Let $a(X_1, X_2, \ldots, X_n)$ be the largest value of k for which there exists an increasing sequence of integers $i_1, i_2, \ldots, i_k$ such that $X_{i_1}, X_{i_2}, \ldots, X_{i_k}$ is zigzag. Find the expected value of $a(X_1, X_2, \ldots, X_n)$ for $n \geq 2$.

Fields: probability, combinatorics (linearity of expectation, alternating-run decomposition) **Position:** B3, third problem of the paper **Answer:** $\dfrac{2n+2}{3}$

Layer 1 THE WHAT — surface analysis

The natural attempt. The natural instinct is to try to build the longest zigzag subsequence greedily — scan left to right, keep the current run going up as long as possible, then switch to going down, and count how many "turns" this produces — and this is basically correct, but doing it as a live simulation on a random sequence, term by term, obscures rather than reveals the expectation. A second natural instinct, once "longest alternating subsequence" is recognised as a named-feeling object, is to try to compute the full distribution of $a(X_1, \ldots, X_n)$ by conditioning on the relative order (the permutation type) of all n values and summing over all $n!$ orderings weighted by the length of their longest zigzag subsequence. This is completely rigorous in principle — a only depends on the relative order of the X_i, since it is defined by sign patterns of differences — but computing $\mathbf{E}[a]$ this way requires knowing, for a uniformly random permutation of n elements, the exact distribution of the longest alternating subsequence, which is itself a nontrivial enumeration problem (related to Euler zigzag numbers) that Putnam gives you no time to solve from scratch in this way.

The model chain. First pin down what $a(X_1, \ldots, X_n)$ actually equals as a function of the sequence, non-probabilistically. Decompose $X_1, \ldots, X_n$ into maximal monotone runs — its "alternating run" decomposition — and let N be the number of runs (so $N-1$ is the number of local extrema, the points where the sequence switches from increasing to decreasing or vice versa). Claim: $a(X_1, \ldots, X_n) = N+1$. In one direction, the sequence of run-endpoints (the value starting each run, together with the very last value) is itself zigzag of length $N+1$, since consecutive endpoints by definition have differences of alternating sign. In the other direction, any zigzag subsequence extracted from $X_1, \ldots, X_n$ is, when viewed as a sequence obtained from the original by deleting some terms, itself decomposed into (its own) alternating runs — and deleting a term from a sequence can never increase its number of runs (deleting an interior point of a run leaves the run count unchanged; deleting a point that separates two runs merges them, decreasing or preserving the count). So a zigzag subsequence of length m has exactly $m-1$ runs by its own definition, and this cannot exceed N (the run count cannot be created by deletion, only preserved or destroyed). Hence $m-1 \leq N$, i.e. $m \leq N+1$, matching the construction above. So $a(X_1, \ldots, X_n) = N+1$ exactly, where N is the number of monotone runs of the actual sequence.

Now compute $\mathbf{E}[a(X_1, \ldots, X_n)]$ by induction using linearity of expectation, one term at a time. Consider $a(X_1, \ldots, X_n) - a(X_1, \ldots, X_{n-1})$: appending X_n either extends the current run (leaving the run count, hence a, unchanged) or creates a new local extremum at X_{n-1} (increasing a by 1). Whether this happens is determined entirely by whether X_{n-2}, X_{n-1}, X_n is monotone (extends) or not (a genuine bend at X_{n-1}) — this needs $n \geq 3$. Since X_{n-2}, X_{n-1}, X_n are i.i.d. continuous, all $3! = 6$ relative orderings are equally likely, and exactly 2 of them (fully increasing, fully decreasing) are monotone. So

$$\mathbf{P}(a(X_1, \ldots, X_n) - a(X_1, \ldots, X_{n-1}) = 1) = \frac{4}{6} = \frac{2}{3}, \qquad \mathbf{E}[a(X_1, \ldots, X_n) - a(X_1, \ldots, X_{n-1})] = \frac{2}{3} \quad (n \geq 3).$$

The base case is $n = 2$: any two distinct reals form a zigzag of length 2 trivially (there is only one difference, and it is automatically "alternating" with nothing to clash against), and $X_1 \neq X_2$ with probability 1, so $a(X_1, X_2) = 2$ always, giving $\mathbf{E}[a(X_1, X_2)] = 2$. Telescoping,

$$\mathbf{E}[a(X_1, \ldots, X_n)] = 2 + (n-2) \cdot \frac{2}{3} = \frac{2n+2}{3}.$$

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains recognise that a depends only on the relative order of the X_i, not their exact values — that much is forced by the definition. They split at the very next step: the natural attempt tries to handle the whole random sequence's combinatorics simultaneously (summing over all orderings, or simulating a greedy scan and hoping the expectation falls out), while the model chain first strips the randomness away entirely and finds a deterministic, purely combinatorial identity $a = N + 1$, and only afterwards reintroduces the randomness — and even then only three values at a time.

> *Before I touch probability at all, what* is $a(X_1, \ldots, X_n)$ *as a plain function of a fixed sequence of numbers — can I express it as something I can count without reference to the deletion process in the definition, because if I can turn "the length of the longest zigzag subsequence" into "the number of alternating runs plus one," then the hard combinatorial object has vanished and what's left is a straightforward random-walk-style expectation.*

Layer 3 THE WHY

The definition of a as "largest k such that some subsequence is zigzag" is an extremal definition over an exponentially large family of subsequences — exactly the kind of definition that is unworkable to attack directly, because you cannot enumerate all subsequences of a length-n sequence and check each one. The move that dissolves this is recognising that deletion of a term from a sequence acts monotonically on the run count: it can never increase it. That single monotonicity fact is what turns an extremal problem ("the best subsequence") into an equality with a directly computable deterministic quantity ("the run count of the sequence itself"), because it simultaneously gives the upper bound (no subsequence can beat $N + 1$ runs' worth of endpoints) and the construction meeting it (the run endpoints, taken as-is, already are a zigzag of that length). Once $a = N + 1$ is established, the problem stops being about "longest subsequence with a global property" and becomes "expected number of local sign changes in a sequence of i.i.d. continuous random variables," which is amenable to linearity of expectation precisely because "is there a local extremum at position $n - 1$" is a statement about only three consecutive values, X_{n-2}, X_{n-1}, X_n, and by symmetry (all 3! orderings of three i.i.d. continuous variables are equally likely) that statement has a clean, position-independent probability. This is the same lens as any "expected number of local maxima/records in a random sequence" problem: reduce a global extremal count to a sum of local indicator variables, each of which only depends on a bounded window of the sequence and is thus exchangeable/symmetric.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the definitional phrase "largest k such that a subsequence has property P" combined with a property P (zigzag) that is itself about local sign changes. On seeing it, write down: "try to re-express this extremal count as a count of local features of the whole sequence" before attempting to enumerate subsequences.
2. That cue suggests the monotone-run decomposition and the claim $a = N + 1$, which should be conjectured and checked on a tiny hand example (e.g. a 4-term sequence with one bend) before being proved in general.
3. Hesitation occurs at proving the "$m \leq N$" direction — it feels like it needs a delicate induction on the subsequence itself. The fix is the single fact that deleting a term from any sequence cannot increase its run count, applied to the process of obtaining the zigzag subsequence from the original by deletions; this collapses the induction to one sentence.
4. The trap is committing to full-strength permutation-distribution machinery (Euler zigzag numbers, generating functions for alternating permutations) the moment the phrase "longest alternating subsequence" is recognised as a named object — that machinery answers a harder question (the whole distribution) than the one asked (only the expectation), and linearity of expectation via local three-term windows is far cheaper.

5. Recognise this family faster by treating "expected number of local extrema / alternations / records in a sequence of i.i.d. continuous random variables" as always reducible to summing $\mathbf{P}(\text{local event at position } i)$ over i, where the local event depends on a fixed small window and its probability is computed by symmetry over that window's orderings.

6. Reusable rule: when an extremal quantity over subsequences equals a directly countable feature of the whole sequence (here, run count), prove the equality first as a pure combinatorial fact, independent of any randomness; only then take expectations, and take them via linearity over small, symmetric local windows rather than over the full joint distribution.

2023 B4

2023 B4. For a nonnegative integer n and a strictly increasing sequence of real numbers $t_0, t_1, \ldots, t_n$, let $f(t)$ be the corresponding real-valued function defined for $t \geq t_0$ by the following properties:

(a) $f(t)$ is continuous for $t \geq t_0$, and is twice differentiable for all $t > t_0$ other than $t_1, \ldots, t_n$;

(b) $f(t_0) = 1/2$;

(c) $\lim_{t \to t_k^+} f'(t) = 0$ for $0 \leq k \leq n$;

(d) For $0 \leq k \leq n-1$, we have $f''(t) = k+1$ when $t_k < t < t_{k+1}$, and $f''(t) = n+1$ when $t > t_n$.

Considering all choices of n and $t_0, t_1, \ldots, t_n$ such that $t_k \geq t_{k-1} + 1$ for $1 \leq k \leq n$, what is the least possible value of T for which $f(t_0 + T) = 2023$?

Fields: analysis (piecewise ODE bookkeeping); constrained optimisation over a variable number of variables **Position:** B4 — the last problem of the approachable half, and the one where the setup is three lines and the optimisation is the whole exam **Answer:** $T = 29$, attained at $n = 9$ with $t_k = t_0 + k$ for $k \leq 9$ and a final stretch of length 20.

Layer 1 THE WHAT — surface analysis

The natural attempt. The setup translates itself. Condition (c) says the velocity is reset to zero at the start of every interval; condition (d) says the acceleration on the k-th interval is constant. So on $[t_{k-1}, t_k]$ we have $f''(t) = k$, and integrating from t_{k-1} with $f'(t_{k-1}^+) = 0$,

$$f'(t) = k(t - t_{k-1}), \qquad f(t_k) - f(t_{k-1}) = \tfrac{k}{2} s_k^2,$$

where $s_k := t_k - t_{k-1}$. Set $t_{n+1} := t_0 + T$ and $s_{n+1} := T - (t_n - t_0)$, so that the final open-ended stretch is just one more interval, with acceleration $n+1$. Summing telescopes:

$$f(t_0 + T) - f(t_0) = \tfrac{1}{2} \sum_{k=1}^{n+1} k\, s_k^2 = 2023 - \tfrac{1}{2} = \tfrac{4045}{2}.$$

So the problem is now bare:

$$\text{minimise } T = \sum_{k=1}^{n+1} s_k \quad \text{subject to} \quad \sum_{k=1}^{n+1} k\, s_k^2 = 4045, \;\; s_k \geq 1 \; (k \leq n),$$

over all $n \geq 0$ as well.

Here is where the natural attempt starts paying. The instinct is Cauchy–Schwarz or Lagrange multipliers, because this is a quadratic constraint with a linear objective and those are the two

tools for that shape. Lagrange gives $1 = 2\lambda k s_k$, i.e. $s_k \propto 1/k$, which violates $s_k \geq 1$ for large k and so only describes an interior critical point that is not in the feasible region — the constraint $s_k \geq 1$ is active, and Lagrange multipliers with active inequality constraints means KKT, which means case analysis on which constraints bind, which is exactly the work you were trying to avoid. Cauchy–Schwarz gives

$$\Big(\sum s_k\Big)^2 \leq \Big(\sum \tfrac{1}{k}\Big)\Big(\sum k s_k^2\Big) = 4045 H_{n+1},$$

an *upper* bound on T. Wrong direction. Reversing it, $\sum k s_k^2 \leq (n+1)\sum s_k^2 \leq (n+1)T^2$ gives $T \geq \sqrt{4045/(n+1)}$, which tends to 0 as n grows and is therefore useless on its own: it has thrown away the $s_k \geq 1$ constraint, which is the only thing stopping n from running to infinity.

The wall is this. There are two opposing pressures — large n makes the final acceleration $n+1$ large, so the last stretch covers distance cheaply; but $s_k \geq 1$ means every extra interval costs at least one unit of time up front. No single inequality sees both pressures, because one is about the constraint and the other about the objective.

The model chain. Do not optimise. *Exchange.*

First, a minimum exists. The constraint $\sum k s_k^2 = 4045$ with $s_k \geq 1$ for $k \leq n$ forces $\sum_{k=1}^{n} k \leq 4045$, so $n \leq 89$; for each fixed n the constraint bounds each s_k above, so the feasible set is a compact subset of $\mathbb{R}^{n+1}$ and the continuous function $\sum s_k$ attains its minimum on it. Taking the least over the finitely many n gives a global minimiser $(s_1^0, \ldots, s_{n+1}^0)$ with value T_0.

Second — and this is the engine — record the following:

$(\star)$ There is no feasible tuple with the *same* total $\sum s_k = T_0$ and $\sum k s_k^2 > 4045$.

Why: such a tuple gives a function with $f(t_0 + T_0) > 2023$, and f is continuous and strictly increasing on (t_0, ∞), so by the intermediate value theorem there is $T < T_0$ with $f(t_0 + T) = 2023$ — for that very same choice of n and $t_0, \ldots, t_n$, which is legal. That contradicts minimality of T_0. So $(\star)$ holds, and it converts the whole optimisation into: *any rearrangement that keeps the total fixed must not increase* $\sum k s_k^2$.

Now two exchanges.

The last stretch is at least 1. Suppose $s_{n+1}^0 < 1$. Merge the last two intervals: pass to $n-1$ with segments $s_1^0, \ldots, s_{n-1}^0$, $s_n^0 + s_{n+1}^0$. The total is unchanged and

$$n(s_n^0 + s_{n+1}^0)^2 - n(s_n^0)^2 - (n+1)(s_{n+1}^0)^2 = s_{n+1}^0\big(2n s_n^0 - s_{n+1}^0\big) > 0,$$

since $s_n^0 \geq 1 > s_{n+1}^0 > 0$. That violates $(\star)$. (If $s_{n+1}^0 = 0$ the merge is an equality and we simply relabel with one fewer interval.) So $s_{n+1}^0 \geq 1$.

Every constrained interval is exactly 1. Suppose $s_k^0 > 1$ for some $k \leq n$. Replace $(s_k^0,\, s_{n+1}^0)$ by $(1,\, s_{n+1}^0 + s_k^0 - 1)$: still feasible, same total, and the change in $\sum k s_k^2$ is

$$(s_k^0 - 1)\Big((n+1-k)(s_k^0 + 1) + 2(n+1)(s_{n+1}^0 - 1)\Big) > 0,$$

because $k \leq n$ gives $n + 1 - k \geq 1$ and we have just shown $s_{n+1}^0 \geq 1$. Again $(\star)$ is violated. So $s_k^0 = 1$ for all $k \leq n$.

Third, evaluate. With $s_k = 1$ for $k \leq n$ and $s_{n+1} = T - n$,

$$\sum_{k=1}^{n+1} k s_k^2 = \frac{n(n+1)}{2} + (n+1)(T-n)^2 = 4045,$$

so

$$T = T(n) = n + \sqrt{\frac{4045}{n+1} - \frac{n}{2}}.$$

At $n = 9$: $4045/10 - 9/2 = 400$, so $T(9) = 9 + 20 = 29$.

Finally, $T(n) \geq 29$ for every n. For $n \geq 30$ this is immediate from $T(n) \geq n$. For $n \leq 29$, the claim $\sqrt{4045/(n+1)} - n/2 \geq 29 - n$ may be squared (both sides nonnegative) and cleared:

$$2 \cdot 4045 - n(n+1) \geq 2(n+1)(29-n)^2,$$

i.e. $-2n^3 + 113n^2 - 1567n + 6408 \geq 0$, i.e.

$$2(9-n)\left(n^2 - \tfrac{95}{2}n + 356\right) \geq 0.$$

The quadratic $Q(n) = n^2 - \frac{95}{2}n + 356$ has roots ≈ 9.33 and ≈ 38.17, so $Q(n) > 0$ for $n \leq 9$ and $Q(n) < 0$ for $10 \leq n \leq 29$; in both ranges the product with $(9-n)$ is ≥ 0, with equality exactly at $n = 9$. Hence $T \geq 29$, attained. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths agree through the reduction

$$\text{minimise } \sum_{k=1}^{n+1} s_k \text{ s.t. } \sum_{k=1}^{n+1} k s_k^2 = 4045,\ s_k \geq 1\ (k \leq n). \tag{†}$$

At (†) the natural attempt reaches for an *inequality* — Cauchy–Schwarz, AM–GM, Lagrange — because the shape "linear objective, quadratic constraint" is the shape those tools were built for. The model chain reaches for an *exchange*: it fixes the objective and perturbs towards a worse constraint value, using minimality of the objective as the contradiction. The split is between optimising the constraint given the objective and optimising the objective given the constraint, and only one of those two is a finite check.

> *The number of variables is itself a variable. n is being optimised over too. So no fixed-dimensional tool — no Lagrange system, no Cauchy–Schwarz with a fixed weight vector — can be the whole argument, because whatever I write down has to be compared across different n. I need a move that changes n.*

> *I have a linear objective and a quadratic constraint. Rather than minimise the linear thing subject to the quadratic, let me freeze the linear thing at its optimum and ask what the quadratic can do. If I can push $\sum k s_k^2$ above 4045 while keeping $\sum s_k$ fixed, I have overshot 2023 at time T_0, so I hit 2023 strictly earlier — and that is a contradiction I can actually reach by moving one variable.*

> *Which direction do I push? The weights k are increasing, so mass is worth more in later slots. Any length sitting in slot $k \leq n$ above its floor of 1 is mass in a cheap slot; move it to slot $n+1$, the most expensive slot. The constraints $s_k \geq 1$ tell me exactly how much I am allowed to move, and therefore exactly where the optimum sits: everything on the floor except the last.*

> *Now n is the only freedom left and T is an explicit function of it. This is now a one-variable inequality, and $T(9) = 29$ is a clean integer — that is the answer and the factor $(9-n)$ must come out of the polynomial. If it does not, I have made an arithmetic slip.*

▷ **Cue:** a minimisation whose number of variables is itself being minimised over ⟶ **Move:** do not write a Lagrange system; look for an exchange that merges or splits variables, so that different dimensions can be compared directly

▷ **Cue:** linear objective, quadratic equality constraint, one-sided box constraints $s_k \geq c$ $\longrightarrow$ **Move:** the optimum has every box constraint active except in the slot with the largest weight; verify by a two-variable exchange, not by KKT

▷ **Cue:** "least T such that $f(T) = L$" with f continuous and strictly increasing $\longrightarrow$ **Move:** overshooting L at time T is the same as reaching L before T; so the minimality of T is a statement that $f(T)$ cannot be increased

Layer 3 THE WHY

Why must the exchange be thought of? Because of what "least T" actually means when f is monotone.

Here is the axiom. f is continuous on $[t_0, \infty)$ and, since $f'(t) = k(t - t_{k-1}) \geq 0$ on each interval with strict positivity off the breakpoints, f is strictly increasing. For a strictly increasing continuous function, the statement "$f(t_0 + T) = 2023$ and no smaller T works for any configuration" is *equivalent* to the statement "no configuration with total time T achieves $f(t_0 + T) > 2023$". These are not two facts; they are one fact read in two directions, and the second reading is the useful one because it holds the time budget fixed and lets the achieved height vary.

That matters because of what was making the problem hard. The objective $\sum s_k$ and the constraint $\sum k s_k^2$ are both sums over intervals, but the constraint carries a weight that depends on the interval's *index*, and the index changes when you merge or split intervals. You cannot compare an n-interval configuration with an $(n-1)$-interval one inside any fixed-dimension optimisation framework; the variables do not correspond. But you can compare them if you fix the total time and ask which achieves more height. Then merging two intervals is a legitimate, single, checkable move — and a move that changes n is precisely what was missing.

Kedlaya's solution begins with the sentence "Observe that there cannot be another $(s_1, \ldots, s_{n'+1})$ with the same sum satisfying $g(s_1, \ldots, s_{n'+1}) > 4045$." That sentence looks like a lemma pulled from nowhere. Here is where it comes from: it is the monotone-inverse-function statement above, and it is the only device in the problem that can compare configurations of different lengths. Everything after it is arithmetic.

Now the second axiom, which explains the shape of the answer. The constraint is $\sum k s_k^2 = 4045$. In the objective every second of time counts the same; in the constraint a second spent in slot k counts k times as much (per unit, and quadratically in the slot length). So time is strictly more productive the later it is spent, and the whole problem is the tension between "spend your time late" and "you cannot get to slot $n+1$ without paying one second for each of slots $1, \ldots, n$". That reading makes the extremal configuration obvious before any algebra: pay the minimum toll — one second per slot — for as many slots as are worth buying, then dump everything remaining into the final slot where acceleration is highest. The exchange argument is just the proof that this intuition is correct, and it is a two-variable computation because moving time between two slots is a two-variable operation.

Finally, why does $n = 9$ win, and is the integrality an accident? Treat $T(n) = n + \sqrt{4045/(n+1) - n/2}$ as a real function. Buying one more slot costs 1 second of toll and multiplies the effectiveness of the final stretch; the marginal value of the extra acceleration decays like $n^{-3/2}$ while the toll is constant, so there is a unique crossover and it is near $n \approx \sqrt[3]{2 \cdot 4045}$-ish in order of magnitude. The value 9 is where that crossover lands for the specific number $4045 = 2 \cdot 2023 - 1$; the fact that $4045/10 - 9/2 = 400$ is exactly a square is the problem-setter reverse-engineering 2023 from $n = 9$, $s_{10} = 20$. That is worth knowing in the arena: when a competition constant is one less than twice the year, the intended configuration is almost certainly the one that makes the surd disappear, and you should test small integer n numerically *before* proving anything.

Caveat. I verified the exchange inequalities symbolically and the closed form $T(n) = n + \sqrt{4045/(n+1) - n/2}$, and I checked by direct numerical minimisation over $0 \le n \le 89$ that $T(n) \ge 29$ with equality only at $n = 9$. The existence-of-a-minimum argument I give (compactness for fixed n, plus a finite bound on n) is complete, but I have not verified that the minimiser is unique; the argument shows only that every minimiser has $s_k = 1$ for $k \le n$, which combined with the strictness of $(9-n)Q(n) > 0$ off $n = 9$ does in fact force uniqueness up to the degenerate relabelling $s_{n+1} = 0$.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is condition (c): f' is reset to 0 at the start of every interval. That is not a smoothness condition, it is a decoupling condition — it means each interval's contribution to f depends on that interval alone. On seeing it, write down $f(t_k) - f(t_{k-1}) = \frac{k}{2}s_k^2$ immediately and forget f entirely.

2. The cue reduces the problem to constrained optimisation: minimise $\sum s_k$ subject to $\sum k s_k^2 = 4045$ and $s_k \ge 1$. The method it suggests — Lagrange or Cauchy–Schwarz — is the wrong one, because the dimension is not fixed.

3. Hesitation occurs when the Lagrange stationarity condition $s_k \propto 1/k$ turns out to be infeasible. The reinforcement away from it is the observation that infeasibility of the interior critical point is *information*: it says every constraint $s_k \ge 1$ that the stationary point violates is active at the optimum. Rather than run KKT, test that guess directly by an exchange, which is one line of algebra per exchange.

4. The trap is treating n as fixed. Every tool that comes to mind operates at fixed dimension, and at fixed n the answer is a surd that decreases with n, suggesting — falsely — that large n is good. The $s_k \ge 1$ floor is the counterweight and it only bites when you compare across n.

5. Recognisable faster next time by the pairing: *a monotone quantity plus a request for the least time to reach a level*. That pairing always means the problem is secretly about maximising the level at fixed time, which is a better-behaved optimisation because the budget is a simple linear constraint. Flip it before doing anything else.

6. Reusable decision rule: *when the number of variables is one of the variables, optimise by exchange, not by calculus.* Find a move that transfers a unit of the resource between two slots (or merges two slots into one), compute the sign of the change in the objective, and let the sign tell you which slots sit at their boundary. Exchange arguments are dimension-agnostic; stationarity conditions are not. This is the same rule that governs rearrangement inequalities, greedy scheduling proofs, and every "smoothing" argument in olympiad inequalities.

2023 B5

2023 B5. Determine which positive integers n have the following property: For all integers m that are relatively prime to n, there exists a permutation $\pi\colon \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ such that $\pi(\pi(k)) \equiv mk \pmod{n}$ for all $k \in \{1, 2, \ldots, n\}$.

Fields: group theory; number theory (structure of $(\mathbb{Z}/n)^\times$); permutation cycle structure **Position:** B5 — the second-hardest problem of the half, and the one that dissolves the instant you stop looking for π
Answer: The property holds exactly for $n = 1$ and $n \equiv 2 \pmod 4$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The equation $\pi(\pi(k)) \equiv mk$ invites you to guess π. The obvious guess is that π is itself multiplication: $\pi(k) = ck$ for some c with $c^2 \equiv m \pmod n$. That works precisely when m is a quadratic residue modulo n in the strong sense of being a square in $(\mathbb{Z}/n)^\times$ — and it fails constantly. For $n = 5$, $m = 2$ is not a square mod 5, so no multiplicative

π exists; yet the problem does not require π to be multiplicative, and indeed one checks by hand that no π at all exists for $n = 5$. For $n = 6$, $m = 5 \equiv -1$: -1 is not a square mod 6, yet $\pi = (1\,5)(2\,4)(3)(6)$ — let us not guess. The guessing route gives a sufficient condition that is not necessary and a necessary condition that is not sufficient, and it burns twenty minutes.

The second natural attempt is to compute. Tabulate small n and test every m by brute force. You find the property holds for $n = 1, 2, 6, 10, 14, 18$ and fails for $3, 4, 5, 7, 8, 9, 11, 12$. The pattern $n \equiv 2 \pmod 4$ (plus $n = 1$) is visible after about six minutes of arithmetic and is worth having — but a conjecture is not a proof, and the brute force gives no handle on *why* $4 \mid n$ and odd $n > 1$ both fail, which are apparently unrelated obstructions.

The wall is that you are searching a space of $n!$ permutations for each of $\varphi(n)$ values of m. Nothing about the search is structured until you stop searching.

The model chain. Reindex first. Identify $\{1, 2, \ldots, n\}$ with $\mathbb{Z}/n\mathbb{Z}$ by sending $n \mapsto 0$; the map $k \mapsto mk$ is then a genuine permutation $\sigma_{n,m}$ of $\mathbb{Z}/n\mathbb{Z}$, since $\gcd(m, n) = 1$. The question is: *for which n does $\sigma_{n,m}$ admit a square root in the symmetric group, for every m coprime to n?*

Lemma. A permutation $\sigma \in S_N$ is a square if and only if, for every even positive integer ℓ, the number of ℓ-cycles of σ is even.

Proof. ($\Rightarrow$) Let $\sigma = \tau^2$. Squaring a τ-cycle of odd length ℓ yields a single σ-cycle of length ℓ; squaring a τ-cycle of even length ℓ splits it into two σ-cycles of length $\ell/2$. So every even-length σ-cycle arises as one of a *pair* coming from a single even τ-cycle, and even lengths occur in pairs. ($\Leftarrow$) Conversely, pair up the even-length cycles of σ with equal lengths and interleave each pair into one cycle of twice the length; and for an odd-length cycle $(c_0\, c_1 \cdots c_{\ell-1})$, take $\tau = (c_0\, c_{(\ell+1)/2}\, c_1\, c_{(\ell+3)/2} \cdots)$, i.e. the unique odd power $\sigma^{(\ell+1)/2}$ restricted to that cycle. Multiply the pieces. □

So the whole problem is the cycle structure of multiplication by m on $\mathbb{Z}/n\mathbb{Z}$.

$n = 1$**.** $\sigma_{1,m}$ is the identity on one point. It is a square.

$n \equiv 2 \pmod 4$**: the property holds.** Write $n = 2k$ with k odd, so by the Chinese remainder theorem $\mathbb{Z}/n\mathbb{Z} \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/k\mathbb{Z}$ as rings. Since $\gcd(m, n) = 1$, m is odd, so $m \equiv 1 \pmod 2$ and multiplication by m acts as the identity on the first coordinate. Hence $\sigma_{n,m}$ preserves each of the two sets $\{0\} \times \mathbb{Z}/k\mathbb{Z}$ and $\{1\} \times \mathbb{Z}/k\mathbb{Z}$, and acts on each by multiplication by m on $\mathbb{Z}/k\mathbb{Z}$ — the *same* permutation, twice. Therefore every cycle length of $\sigma_{n,m}$ occurs an even number of times, in particular every even one. By the Lemma, $\sigma_{n,m}$ is a square. This holds for every m, so n has the property.

n **odd,** $n > 1$**: the property fails.** Let p be an odd prime with $n = p^e k$, $e \geq 1$, $\gcd(p, k) = 1$. Since $(\mathbb{Z}/p^e\mathbb{Z})^\times$ is cyclic, choose (by CRT) an integer m reducing to a generator g of $(\mathbb{Z}/p^e\mathbb{Z})^\times$ and to 1 in $\mathbb{Z}/k\mathbb{Z}$. Then $\gcd(m, n) = 1$, and under $\mathbb{Z}/n\mathbb{Z} \cong \mathbb{Z}/p^e\mathbb{Z} \times \mathbb{Z}/k\mathbb{Z}$, $\sigma_{n,m}$ sends $(u, v) \mapsto (gu, v)$. Fix v. On the $\varphi(p^e)$ units u, the orbit of u is $\{g^t u\}$, of length $\operatorname{ord}(g) = \varphi(p^e) = p^{e-1}(p-1)$; so the units form exactly one cycle, for each of the k values of v. That is k cycles of length $p^{e-1}(p-1)$, and k is odd while $p^{e-1}(p-1)$ is even because p is odd. No other element has a cycle of this length: if $u = p^j u'$ with $1 \leq j \leq e$ and $\gcd(u', p) = 1$, its orbit length is the order of g modulo p^{e-j}, namely $\varphi(p^{e-j}) < \varphi(p^e)$. So the number of cycles of that even length is exactly k, odd. By the Lemma, $\sigma_{n,m}$ is not a square.

$4 \mid n$**: the property fails.** Take $m = -1$, which is coprime to n. Then $\sigma_{n,-1}$ fixes x exactly when $2x \equiv 0$, i.e. $x \in \{0, n/2\}$: two fixed points. The remaining $n - 2$ elements pair into $\frac{n-2}{2} = \frac{n}{2} - 1$ transpositions. Since $4 \mid n$, $n/2$ is even, so $\frac{n}{2} - 1$ is odd: an odd number of 2-cycles. By the Lemma, $\sigma_{n,-1}$ is not a square.

Hence the property holds exactly for $n = 1$ and $n \equiv 2 \pmod 4$. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths agree on the reindexing $\{1, \ldots, n\} \leftrightarrow \mathbb{Z}/n\mathbb{Z}$ and on the observation that the problem asks whether a certain permutation $\sigma_{n,m}$ has a square root. The split is

at the very next line:

$$\text{“find } \pi \text{ with } \pi^2 = \sigma_{n,m}\text{”}. \qquad (\dagger)$$

At $(\dagger)$ the natural attempt tries to *construct* π, and constructs it from the only material it has, namely arithmetic: $\pi(k) = ck$. The model chain declines to construct anything and instead asks what invariant of σ determines the answer — which is the same as asking what squaring does to cycle type. The first route is number theory about m; the second is group theory about σ, and the number theory then enters only as a way to compute one cycle type.

> *I am being asked whether an equation $\pi^2 = \sigma$ is solvable, not for the solution. Solvability questions in a group are answered by invariants, not by formulas. So: what is the complete invariant of a permutation up to conjugacy? Cycle type. What does squaring do to cycle type? That is a finite, checkable question about a single cycle, and I can answer it in thirty seconds.*

> *Squaring an ℓ-cycle: if ℓ is odd, 2 is invertible mod ℓ and the cycle survives whole; if ℓ is even it falls apart into two halves. So squares are exactly the permutations whose even-length cycles can be paired off. Every even length must occur an even number of times. That is the entire content of the problem and it has nothing to do with n yet.*

> *Now I need cycle types of multiplication by m on $\mathbb{Z}/n$. CRT splits $\mathbb{Z}/n$ into prime-power pieces and multiplication respects the splitting, so I can engineer m one prime at a time — I get to* choose *the worst m, which means I should choose it to make some even cycle length occur an odd number of times. The cheapest way to make a count odd is to make it equal to an odd factor of n.*

> *Where does the parity of the count come from? If m acts trivially on one CRT factor, the whole permutation is a disjoint union of copies of the same smaller permutation, one copy per element of that factor — so the number of copies is the size of that factor. Make that factor $\mathbb{Z}/2$ and every count doubles: the property holds. Make it odd and I have a free odd multiplier on every cycle count: the property is in danger. That is the* 2 mod 4 *dichotomy, and I can see it before doing any of the prime-power arithmetic.*

- ▷ **Cue:** “does there exist π with $\pi^2 = \sigma$” (or $\pi^d = \sigma$) for a permutation σ $\longrightarrow$ **Move:** do not construct π; classify d-th powers by cycle type. For $d = 2$: σ is a square iff every even cycle length occurs an even number of times
- ▷ **Cue:** a permutation of $\mathbb{Z}/n\mathbb{Z}$ given by a ring operation, with a condition required for *all* m coprime to n $\longrightarrow$ **Move:** use CRT to split into prime powers, then choose m to be a generator on one factor and the identity on the rest — this isolates a single cyclic group and gives you the extremal cycle type
- ▷ **Cue:** a permutation that acts identically on each fibre of a projection $\longrightarrow$ **Move:** every cycle length occurs a multiple of (fibre count) times; so a projection onto $\mathbb{Z}/2$ makes all multiplicities even for free

Layer 3 THE WHY

Why must the first thought be “stop constructing”?

Because of what the problem is quantified over. It says: *for all* m coprime to n, *there exists* π. The inner existential is over $n!$ objects and the outer universal is over $\varphi(n)$ objects. A property with that shape is never settled by exhibiting objects; it is settled by a criterion — something you can evaluate on $\sigma_{n,m}$ directly and which is equivalent to solvability. And in a finite group there is exactly one general mechanism for turning “is x a d-th power?” into a computation: conjugacy invariants. For symmetric groups the conjugacy invariant is cycle type, and it is

complete — two permutations are conjugate if and only if they have the same cycle type. The set of squares is closed under conjugation (if $\sigma = \tau^2$ then $g\sigma g^{-1} = (g\tau g^{-1})^2$), so membership in the set of squares depends only on cycle type. That is not a lucky feature of this problem; it is a theorem-shaped fact about S_N that the problem is built on, and once you notice that the set of squares is conjugation-closed, you *know* the answer is a statement about cycle type before you know what statement it is.

Then the cycle-type computation is forced. Take one cycle of τ, of length ℓ, and square it. τ restricted to that cycle is a cyclic shift by 1 on $\mathbb{Z}/\ell$; τ^2 is a shift by 2; and the orbits of "add 2" on $\mathbb{Z}/\ell$ are the cosets of the subgroup $\langle 2 \rangle$, which is everything when ℓ is odd and index 2 when ℓ is even. This is the definition of the subgroup generated by 2 in a cyclic group, nothing more. So an odd cycle stays whole and an even cycle halves into two. Reading it backwards — which is the direction you need — the even cycles of σ must be assembled into pairs, so their multiplicities must be even, and the odd ones are unconstrained because an odd cycle is already a square of itself (specifically of $\sigma^{(\ell+1)/2}$, since $2 \cdot \frac{\ell+1}{2} \equiv 1 \pmod{\ell}$). Every clause of the Lemma is a restatement of "is 2 invertible modulo the cycle length".

Now the second axiom, the one that explains why 2 mod 4 and not some messier condition.

The Lemma is about *parity of multiplicities*. So the question becomes: where can a parity possibly come from? The answer is the CRT factorisation, and it comes from exactly one place. If $\mathbb{Z}/n\mathbb{Z} \cong R_1 \times R_2$ and m acts as the identity on R_1, then $\sigma_{n,m}$ is $|R_1|$ disjoint copies of the same permutation of R_2. Every multiplicity is multiplied by $|R_1|$. So:

- If n has a factor of exactly 2 — i.e. $n = 2k$ with k odd — then *every* unit m is automatically trivial mod 2, because the only unit mod 2 is 1. You do not get to choose; the structure hands you a factor of 2 in every multiplicity, for free, for all m. The property holds.
- If n is odd and $n > 1$, there is an odd prime power $p^e \| n$ and the group $(\mathbb{Z}/p^e)^\times$ is cyclic of *even* order $p^{e-1}(p-1)$. Pick m to generate it and to be trivial elsewhere: the units of $\mathbb{Z}/p^e$ form one long even cycle, replicated n/p^e times, and n/p^e is odd. One even length, odd multiplicity. Dead.
- If $4 \mid n$, the unit group mod 4 contains $-1 \neq 1$, so you are no longer forced into triviality on the 2-part; take $m = -1$ globally. Multiplication by -1 on $\mathbb{Z}/n$ has fixed points exactly at the elements of order dividing 2, of which there are 2 when n is even, and everything else is a transposition: $\frac{n}{2} - 1$ of them, odd precisely when $4 \mid n$. Dead.

So the three cases are not three unrelated arguments. They are one question — "is the multiplicity of some even cycle length forced odd?" — asked of the three possible behaviours of the 2-part of n: absent, exactly 2, or at least 4. The condition $n \equiv 2 \pmod{4}$ is the statement "the 2-part of n is exactly 2", which is the unique case where the unit group mod the 2-part is trivial and therefore doubles everything without costing you a choice of m.

Kedlaya's solution states the Lemma without motivation and then runs three cases. Here is where the Lemma comes from: it is the only way to answer an existence question in S_N, because cycle type is the complete conjugacy invariant and the squares form a conjugation-closed set. And here is where the case division comes from: parity of multiplicity can only be manufactured by the size of a CRT factor on which m acts trivially, so the answer must be a condition on the 2-part of n, and only three things can happen to it.

Caveat. I verified the answer by exhaustive computation for $1 \leq n \leq 39$: the property holds exactly at $n = 1, 2, 6, 10, 14, 18, 22, 26, 30, 34, 38$, testing every m coprime to n against the cycle-parity criterion of the Lemma. The "if" direction of the Lemma (assembling τ from paired even cycles and from odd cycles) I have stated with an explicit construction but have not written out the verification that the interleaving of two equal even cycles

squares correctly; it is routine but it is a place where a written-up solution must not wave.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the composite $\pi \circ \pi$ on the left of a condition that must hold *for all* m. On seeing $\pi(\pi(k))$, write down "$\pi^2 = \sigma$, so this is a square-root question in S_n" and reindex $\{1, \dots, n\}$ as $\mathbb{Z}/n\mathbb{Z}$ so that σ is literally multiplication by m.

2. That cue suggests the cycle-type criterion for squares, not a construction of π. Derive the criterion on the spot: squaring a cycle of length ℓ preserves it if ℓ is odd and halves it if ℓ is even, because the orbits of "add 2" on $\mathbb{Z}/\ell$ are the cosets of $\langle 2 \rangle$.

3. Hesitation occurs at the transition from "criterion" to "which n", because you now need cycle types of multiplication by m and there are $\varphi(n)$ values of m. What reinforces away from the hesitation is remembering that m is universally quantified and you are looking for *one* bad m in the failure cases — so you get to build m by CRT, one prime at a time, and the two cheapest choices are "generator on one factor" and "-1 everywhere".

4. The trap is the multiplicative ansatz $\pi(k) = ck$, which reduces the problem to "is every unit a square mod n" — a condition satisfied only by $n \in \{1, 2\}$ and a handful of others, and not the right condition. It is a sufficient condition masquerading as the problem. For $n = 6$, $m = 5$ the property holds even though 5 is not a square modulo 6; noticing that single counterexample early is worth the two minutes it costs, because it kills the ansatz outright.

5. Recognisable faster next time by the marker *an equation in a group whose unknown is an element, quantified existentially*. Such problems are solved by invariants that are constant on conjugacy classes, never by construction. In S_N the invariant is cycle type; in GL_n it is rational canonical form; in an abelian group it is the index of the d-th power subgroup.

6. Reusable decision rule: *when a universally quantified parameter ranges over a group that factors, choose the parameter to be extreme on one factor and trivial on the rest.* Triviality on a factor of size r multiplies every orbit multiplicity by r, and extremality on the remaining factor produces the longest possible orbit there. That single manoeuvre — generator here, identity there — is how almost every "for all m coprime to n" problem is broken, and it works because CRT makes the action a product action.

2023 B6

2023 B6. Let n be a positive integer. For i and j in $\{1, 2, \dots, n\}$, let $s(i, j)$ be the number of pairs (a, b) of nonnegative integers satisfying $ai + bj = n$. Let S be the n-by-n matrix whose (i, j) entry is $s(i, j)$. For example, when $n = 5$, we have $S = \begin{bmatrix} 6 & 3 & 2 & 2 & 2 \\ 3 & 0 & 1 & 0 & 1 \\ 2 & 1 & 0 & 0 & 1 \\ 2 & 0 & 0 & 0 & 1 \\ 2 & 1 & 1 & 1 & 2 \end{bmatrix}$. Compute the determinant of S.

Fields: linear algebra (Cauchy–Binet); number theory (Möbius inversion, Mertens function) **Position:** B6 — the hardest problem of the half; a determinant of an arithmetic matrix, which is a genre with exactly one entry point **Answer:** $\det S = (-1)^{\lceil n/2 \rceil - 1} 2 \lceil n/2 \rceil$. In particular $|\det S|$ grows linearly in n, which is the tell.

Layer 1 THE WHAT — surface analysis

The natural attempt. The matrix is given by a counting formula, so the honest first move is to compute small cases and look.

$$n = 1 : \ \det = 2; \quad n = 2 : \ \det = 2; \quad n = 3 : \ \det = -4; \quad n = 4 : \ \det = -4;$$

$$n = 5 : \ \det = 6; \quad n = 6 : \ \det = 6; \quad n = 7 : \ \det = -8.$$

The pattern is immediate: $|\det S| = 2\lceil n/2 \rceil$, with sign alternating in blocks of two. So the answer is $(-1)^{\lceil n/2 \rceil - 1} 2\lceil n/2 \rceil$, and you should write that down at minute five and spend the rest of the problem proving it.

Now the wall. The entries $s(i, j)$ are not multiplicative, not polynomial, and not even monotone; they depend on divisibility relations between i, j and n in a way that makes direct expansion hopeless. Row reduction is available and in fact works — one can eliminate most of the matrix using the structure of rows $i > n/2$, where $s(i, j) \in \{0, 1, 2\}$ is governed by whether $j \mid n - i$ — but the elimination has to be set up carefully, involves striking out rows and columns in a prescribed order, and leaves a 3×3 or 4×4 residue with a case split on the parity of n. That is Kedlaya's first solution. It is correct and it is a long bookkeeping exercise; every struck entry is a sign to track.

The real obstacle is that you are staring at $\det S$ with no reason to believe S is anything other than an arbitrary integer matrix. A determinant that grows only *linearly* in n, for an $n \times n$ matrix with entries of size up to $n + 1$, is wildly small — generic such matrices have determinants of size $n^{n/2}$ or worse. Something is collapsing, and row reduction discovers the collapse without ever explaining it.

The model chain. Factor the matrix.

The entry is a count, so write it as a sum of indicator products. If $ai + bj = n$ with $a, b \geq 0$, put $k := ai$. Then k ranges over $\{0, 1, \ldots, n\}$, $i \mid k$, and $j \mid n - k$; and conversely each such k gives exactly one pair $(a, b) = (k/i, (n - k)/j)$. Hence

$$s(i, j) = \sum_{k=0}^{n} [\, i \mid k \,] \, [\, j \mid n - k \,],$$

which says $S = AB$, where A is $n \times (n + 1)$ with $A_{ik} = [\, i \mid k \,]$ for $i = 1, \ldots, n$ and $k = 0, \ldots, n$, and B is $(n + 1) \times n$ with $B_{kj} = [\, j \mid n - k \,]$. (Here $i \mid 0$ is true for every i, so A's column 0 is all ones.) Note $B^{\mathsf{T}}_{jk} = [\, j \mid n - k \,] = A_{j, n-k}$: transposing B gives A with its columns reversed.

A is $n \times (n + 1)$: one column too many. That is exactly the hypothesis of the Cauchy–Binet formula,

$$\det(AB) = \sum_{k=0}^{n} \det\left(A^{(k)}\right) \det\left(B^{(k)}\right),$$

the sum over which single column of A (equivalently, which row of B) is deleted.

Let $v_0, v_1, \ldots, v_n \in \mathbb{R}^n$ be the columns of A, and let M_k denote A with column k deleted. Reversing the n columns of B^{T} costs $(-1)^{n(n-1)/2}$, so $\det(B^{(k)}) = (-1)^{n(n-1)/2} \det(M_{n-k})$, and

$$\det S = (-1)^{n(n-1)/2} \sum_{k=0}^{n} \det(M_k) \det(M_{n-k}). \tag{1}$$

Computing $\det(M_k)$. With $e_1, \ldots, e_n$ the standard basis, $v_k = \sum_{d \mid k} e_d$ for $k \geq 1$, and $v_0 = \sum_{d=1}^{n} e_d$. Since $d \mid k$ forces $d \leq k$ and $k \mid k$, the matrix $M_0 = [v_1 \cdots v_n]$ is upper triangular with 1s on the diagonal:

$$\det(M_0) = 1.$$

For $k \geq 1$, Möbius inversion of $v_k = \sum_{d \mid k} e_d$ gives $e_k = \sum_{d \mid k} \mu(d)\, v_{k/d}$, whence

$$v_0 = \sum_{m=1}^{n} e_m = \sum_{m=1}^{n} \sum_{d \mid m} \mu(d) v_{m/d} = \sum_{j=1}^{n} C_j v_j, \qquad C_j := \sum_{d=1}^{\lfloor n/j \rfloor} \mu(d),$$

the reindexing being over pairs (d, j) with $dj \leq n$. (So $C_j = M(\lfloor n/j \rfloor)$, the Mertens function.) Now $M_k = [v_0, v_1, \ldots, \widehat{v_k}, \ldots, v_n]$; expanding v_0 in that basis, every term but $C_k v_k$ repeats an existing column, so

$$\det(M_k) = C_k \det[v_k, v_1, \ldots, \widehat{v_k}, \ldots, v_n] = C_k (-1)^{k-1} \det(M_0) = (-1)^{k-1} C_k,$$

the sign coming from moving v_k past $v_1, \dots, v_{k-1}$. Setting $C_0 := -1$ makes this formula hold at $k = 0$ too.

Assembling. Substituting into (1),

$$\det S = (-1)^{n(n-1)/2} \sum_{k=0}^{n} (-1)^{k-1}(-1)^{n-k-1} C_k C_{n-k} = (-1)^{n(n-1)/2}(-1)^n \sum_{k=0}^{n} C_k C_{n-k}.$$

For the inner sum: $C_n = M(1) = 1$, so the two end terms $k = 0, n$ contribute $2C_0C_n = -2$. For $1 \le k \le n-1$, if $k > n/2$ then $\lfloor n/k \rfloor = 1$ and $C_k = 1$; if n is even, $C_{n/2} = M(2) = \mu(1) + \mu(2) = 0$. So pairing k with $n-k$,

$$\sum_{k=0}^{n} C_k C_{n-k} = -2 + 2 \sum_{k=1}^{\lfloor n/2 \rfloor} C_k.$$

And that last sum telescopes arithmetically:

$$\sum_{k=1}^{\lfloor n/2 \rfloor} C_k = \sum_{\substack{d,k \ge 1 \\ dk \le n}} \mu(d) \; - \sum_{\substack{d,k \ge 1,\ dk \le n \\ k > \lfloor n/2 \rfloor}} \mu(d).$$

The first sum is $\sum_{f=1}^{n} \sum_{d \mid f} \mu(d) = \sum_{f=1}^{n} [f = 1] = 1$. In the second, $k > n/2$ and $dk \le n$ force $d = 1$, so it is $\mu(1)$ counted $n - \lfloor n/2 \rfloor = \lceil n/2 \rceil$ times. Hence $\sum_{k \le \lfloor n/2 \rfloor} C_k = 1 - \lceil n/2 \rceil$ and

$$\sum_{k=0}^{n} C_k C_{n-k} = -2 + 2\big(1 - \lceil \tfrac{n}{2} \rceil\big) = -2\lceil \tfrac{n}{2} \rceil.$$

Therefore

$$\det S = (-1)^{n(n-1)/2+n+1}\, 2\lceil \tfrac{n}{2} \rceil = (-1)^{n(n+1)/2+1}\, 2\lceil \tfrac{n}{2} \rceil.$$

Finally $n(n+1)/2 + 1$ and $\lceil n/2 \rceil - 1$ have the same parity: both are periodic in n with period 4, so checking $n = 1, 2, 3, 4$ suffices ($n(n+1)/2 + 1 = 2, 4, 7, 11$ and $\lceil n/2 \rceil - 1 = 0, 0, 1, 1$). Hence

$$\boxed{\det S = (-1)^{\lceil n/2 \rceil - 1}\, 2\lceil \tfrac{n}{2} \rceil.} \qquad \square$$

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths agree on the numerics — compute $n \le 7$, read off $(-1)^{\lceil n/2 \rceil - 1} 2\lceil n/2 \rceil$ — and both then face the same line:

$$s(i, j) = \#\{(a, b) \in \mathbb{Z}^2_{\ge 0} : ai + bj = n\}. \tag{†}$$

At (†) the natural attempt reads the entry as a *number* and starts doing row operations on the matrix of numbers. The model chain reads the entry as a *sum over an index it has not yet named*, introduces $k = ai$, and thereby turns S into a product of two 0/1 matrices. Everything else follows: once $S = AB$ with A of size $n \times (n+1)$, Cauchy–Binet is not a choice, it is the only formula in existence with that signature.

> *An $n \times n$ integer matrix with entries up to $n+1$ has determinant of order $n^{n/2}$ if the rows are anything like independent. Mine has determinant of order n. The matrix is therefore very nearly singular in a structured way — its rank is effectively being supplied by something much smaller than n. Small determinant means low-rank factorisation, so I should be looking for $S = AB$ with A, B thin, not for row operations on S.*

> *The entry is a count of solutions of $ai + bj = n$. A count is a sum of* 1*s. What am I summing over? The solutions — but each solution is determined by the single number $k = ai$. So $s(i,j) = \sum_{k=0}^{n} [i \mid k][j \mid n-k]$, and the i-dependence and the j-dependence have separated. That is a matrix product with inner dimension $n+1$.*

> *Inner dimension $n+1$, outer dimension n: one too many. That is precisely the Cauchy–Binet configuration, and Cauchy–Binet will express* det S *as a sum of $n+1$ products of determinants of $n \times n$ minors. Those minors are all the same matrix with one column dropped, so I need to understand a single one-parameter family, not $\binom{2n}{n}$ objects.*

> *A's columns are $v_k = \sum_{d \mid k} e_d$. Dropping column* 0 *leaves a unitriangular matrix, so* det $= 1$ *— and every other minor is measured against that one by expressing the extra column v_0 in the basis $\{v_k\}$. Expressing e's in terms of v's is Möbius inversion. I do not need to invert anything; the coefficient of v_j in v_0 is a partial sum of μ, which is the Mertens function, and Mertens is* 1 *for arguments $\geq$ nothing much — it is* 1 *at* 1 *and* 0 *at* 2*, which is exactly the collapse I predicted.*

- ▷ **Cue:** a matrix entry defined as a count of solutions to an equation $\longrightarrow$ **Move:** write the count as a sum of indicator products over the natural intermediate variable; this factors the matrix as a product, usually with an off-square inner dimension
- ▷ **Cue:** a matrix product AB with A of size $n \times m$ and B of size $m \times n$, $m > n$ $\longrightarrow$ **Move:** Cauchy–Binet: $\det(AB) = \sum_{|T|=n} \det(A_T) \det(B_T)$ over n-subsets T of the inner index; it is useful exactly when $m - n$ is small, so that the sum has few terms
- ▷ **Cue:** a determinant whose numerical value grows far more slowly in n than a generic determinant of that size $\longrightarrow$ **Move:** stop row-reducing; the matrix is a product of thin or highly structured factors and the smallness is the factorisation's fingerprint
- ▷ **Cue:** columns of the form $v_k = \sum_{d \mid k} e_d$ (divisor-indicator vectors) $\longrightarrow$ **Move:** they are a basis, unitriangular against the standard one, and the change of basis is Möbius inversion: $e_k = \sum_{d \mid k} \mu(d) v_{k/d}$

Layer 3 THE WHY

Why is the factorisation forced, and not found?

Start from the definition, and take it absolutely literally. $s(i,j)$ is defined as the *cardinality of a set*. A cardinality is $\sum_{x \in X} 1$, and the only question is how to index X. Here $X = \{(a,b) : ai + bj = n,\ a,b \geq 0\}$, and the defining equation has already told you how to index it: the equation reads "ai plus bj equals n", so the natural coordinate on X is the value of the first summand, $k := ai$. Once k is fixed, a is determined (as k/i, which exists iff $i \mid k$) and b is determined (as $(n-k)/j$, which exists iff $j \mid n-k$). That is not a manipulation; it is what the equation says. And the consequence is that the condition on i and the condition on j are attached to *different* factors of the same index k. A sum over a shared index of a product of a thing depending on (i,k) and a thing depending on (k,j) is, by the definition of matrix multiplication, an entry of a matrix product. So $S = AB$ is not a discovery about this matrix; it is the definition of s rewritten in the notation that matches its own shape.

That is the first axiom. The second explains why the answer is small, which is the fact that should have driven you to look for the factorisation in the first place.

Here is the real content. A has n rows and $n+1$ columns. Its rank is at most n, so the $n+1$ columns satisfy exactly one linear relation — and by the unitriangularity of $[v_1 \cdots v_n]$ there is precisely one, expressing v_0 in terms of the rest. That relation is the *only* arithmetic input the whole problem has. Its coefficients $C_j = \sum_{d \leq n/j} \mu(d)$ are partial sums of the Möbius function, and the reason the final answer is small is that these partial sums are overwhelmingly boring: $C_j = 1$ for every $j > n/2$ (because $\lfloor n/j \rfloor = 1$ and $\mu(1) = 1$), and $C_{n/2} = 0$ (because

$\mu(1) + \mu(2) = 0$). So in the convolution $\sum_k C_k C_{n-k}$, every single term has at least one factor equal to 1 — one of k, $n-k$ always exceeds $n/2$ — and the convolution degenerates into a plain sum $\sum_{k \le n/2} C_k$. And that plain sum is not a hard number-theoretic quantity either: swapping the order of summation turns it into $\sum_{f \le n} \sum_{d \mid f} \mu(d)$, which is 1 by the defining property of μ, minus a trivial count. The Mertens function, which is genuinely erratic, never gets a chance to be erratic, because every argument it is ever evaluated at is 1 or 2 except in a sum that collapses by Möbius's own identity.

This is worth saying plainly, because it is the transferable content: *the determinant is small because the only arithmetic in the matrix is a single linear relation, and the coefficients of that relation are asymptotically constant.* A determinant that grows linearly in n is the signature of exactly this situation. If you internalise that, then the moment you compute $\det S$ for $n \le 7$ and see $2, 2, -4, -4, 6, 6, -8$, you know before you know anything else that S is a product of a thin pair of factors with one relation between them, and you go looking for it.

Kedlaya's second solution (attributed to Razzi Masroor) opens with "We first observe that $S = AB$ where A is the $n \times (n+1)$ matrix $A_{ik} = [i \mid k]$." Here is where that comes from: it is the sentence "a count is a sum over the intermediate variable ai", written in matrix notation. And the appeal to Cauchy–Binet immediately afterwards is forced by a dimension count — the formula exists precisely to compute determinants of products with mismatched inner dimension, and it is useful precisely when the mismatch is 1, because then the sum has $n+1$ terms rather than exponentially many.

One more structural remark, which is the check that the argument is not lucky. The relation $B^{\mathsf{T}} = A$ with columns reversed is not decoration: it is the statement that $ai + bj = n$ is symmetric in $(a, i) \leftrightarrow (b, j)$, i.e. that S is symmetric. That symmetry is what makes the Cauchy–Binet sum a *convolution* $\sum_k C_k C_{n-k}$ rather than a sum of $n+1$ unrelated products, and a convolution with one factor always equal to 1 is a sum. Every simplification in the endgame is powered by the symmetry visible in the problem statement.

Caveat. I verified $\det S = (-1)^{\lceil n/2 \rceil - 1} 2 \lceil n/2 \rceil$ by exact rational Gaussian elimination for $1 \le n \le 14$, and verified the given $n = 5$ matrix against my own computation of $s(i, j)$. I re-derived the Cauchy–Binet chain independently and note that the published solution's write-up contains several typographical slips in this passage — the definition of C_k is printed as $\sum_{1 \le d \lfloor n/k \rfloor}$, a line reads $v_0 = \sum_{k=1}^{n} e_n$, the sign is printed as $(-1)^{n-2}$ where $(-1)^n$ is meant, and the final display carries $(-1)^{-2}$ and a $\sum_{k=1}^{\lfloor n/2 \rfloor}$ with no summand. The chain above is my own and each step is checked; the final answer agrees with the published one and with the numerics.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the phrase "the *number* of pairs (a, b)". On seeing a matrix entry defined as a count, write the count as $\sum_k [\,\cdot\,][\,\cdot\,]$ over whatever intermediate quantity the defining equation names — here $k = ai$. The second cue, equally important, is the numerical one: compute $\det S$ for $n \le 7$ and notice it grows linearly.

2. The first cue suggests factoring S as a product of indicator matrices; the second confirms it, because a linearly growing determinant cannot come from a generic matrix. Together they suggest Cauchy–Binet, which is the tool for $\det(AB)$ when the inner dimension exceeds the outer one.

3. Hesitation occurs at the minors. Cauchy–Binet in general sums over $\binom{m}{n}$ subsets, which sounds catastrophic. What reinforces away from that fear is counting: the inner dimension is $n+1$ and the outer is n, so the excess is 1 and there are only $n+1$ subsets — one per deleted column. Always check the excess before flinching.

4. The trap is row reduction. It works — Kedlaya's first solution is exactly that — but it requires a prescribed order of strikings, sign tracking through every one, and a case split on the parity

of n at the end. It is the zombie-maths branch: the most primal manipulation available, undertaken without first asking why the determinant is so small. Asking that question is what buys the factorisation.

5. Recognisable faster next time by *the size of the answer*. Before proving anything about a determinant, estimate how big it ought to be. A generic $n \times n$ matrix with entries bounded by n has $|\det|$ around $n^{n/2}$; if the actual value is polynomial in n, the matrix has a low-complexity factorisation and finding it is easier than eliminating. This test costs one minute and redirects the entire attack.

6. Reusable decision rule: *when an object is defined as a count, name the summation index before doing anything else, and check whether the two parameters of the object attach to different halves of that index.* If they do, you have a matrix product, a bilinear form, or a convolution — and each of those has a large toolkit attached, whereas the count itself has none. This applies far beyond determinants: it is the same move that turns a convolution identity into a generating-function product, and a counting problem with two constraints into a double sum that can be reversed.