DAWN PUTNAM SESSION

Putnam 2021 — Papers A and B

Putnam 120 Archive

Contents

Putnam 2021 — Paper A **5**
2021 A1 5
2021 A2 7
2021 A3 9
2021 A4 12
2021 A5 17
2021 A6 21

Putnam 2021 — Paper B **27**
2021 B1 27
2021 B2 29
2021 B3 31
2021 B4 34
2021 B5 38
2021 B6 43

Putnam 2021 — Paper A

2021 A1

2021 A1. A grasshopper starts at the origin in the coordinate plane and makes a sequence of hops. Each hop has length 5, and after each hop the grasshopper is at a point whose coordinates are both integers; thus, there are 12 possible locations for the grasshopper after the first hop. What is the smallest number of hops needed for the grasshopper to reach the point $(2021, 2021)$?

Fields: Combinatorics / elementary number theory (integer vectors of fixed length, a linear monovariant used as a lower bound) **Position:** A1 – the opener **Answer:** 578 hops: the grasshopper needs at least 578 because no linear quantity can move fast enough in fewer, and 578 is achievable by an explicit combination of hops.

Layer 1 THE WHAT — surface analysis

The natural attempt. The twelve length-5 integer vectors are $(\pm 5, 0)$, $(0, \pm 5)$, $(\pm 3, \pm 4)$, $(\pm 4, \pm 3)$ – this much is forced by Pythagorean triples and takes no real thought. The target $(2021, 2021)$ sits on the line $y = x$, so the honest first move is to reach for the vectors that point most nearly along that line: $(3, 4)$ and $(4, 3)$, used in some alternating combination, since either one alone drifts the grasshopper off the diagonal and $(3, 4)$ then $(4, 3)$ together add to $(7, 7)$, which is exactly diagonal. Dividing, $2021/7 = 288.71\ldots$, so 288 copies of $(3, 4)$ and 288 copies of $(4, 3)$ land at $(288 \cdot 7, 288 \cdot 7) = (2016, 2016)$, using 576 hops and leaving a gap of $(5, 5)$ to close. One more hop of $(3, 4)$ or $(4, 3)$ overshoots one coordinate past 2021 and undershoots the other; the gap $(5, 5)$ is not itself one of the twelve vectors, so it takes two more hops – $(0, 5)$ and $(5, 0)$ – to finish, for a total of 578. This is a perfectly good construction, found by direct search along the obvious diagonal-hugging strategy. But it only proves 578 hops *suffice*. It says nothing about whether 577, or fewer, might also work via some cleverer, non-diagonal combination – and a natural next instinct is to keep searching for a shorter combination, which is expensive and inconclusive: absence of a shorter combination found by trial is not a proof that none exists.

The model chain. Consider the quantity $z = x + y$, the sum of the grasshopper's two coordinates. It starts at $z = 0$ and must end at $z = 2021 + 2021 = 4042$. Compute, for each of the twelve hop vectors (a, b), the change $a + b$ it causes to z: this is ± 5 for the four axis vectors, and for $(\pm 3, \pm 4)$ and $(\pm 4, \pm 3)$ the sixteen sign combinations give values in $\{\pm 7, \pm 1\}$. The maximum possible increase to z in a single hop is therefore 7, achieved only by $(3, 4)$ or $(4, 3)$. After n hops, z has increased by at most $7n$, so reaching $z = 4042$ requires $7n \geq 4042$, i.e. $n \geq 4042/7 = 577.43\ldots$, so $n \geq 578$. This is a clean lower bound, independent of any search. The construction in the natural attempt – 288 hops of $(3, 4)$, 288 of $(4, 3)$, one of $(0, 5)$, one of $(5, 0)$ – totals exactly 578 hops and lands exactly on $(2021, 2021)$, since $288(3, 4) + 288(4, 3) + (0, 5) + (5, 0) = (864 + 1152 + 0 + 5,\ 1152 + 864 + 5 + 0) = (2021, 2021)$. The lower bound and the construction meet at 578, so that is the answer.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains reach the same 578-hop construction by the same diagonal-

hugging search – that part is not where they differ. They part company on the question the natural attempt never asks: "how do I know 577 is impossible?" At that point the model chain stops searching for hop-sequences altogether and asks instead what single number changes by a bounded amount on every hop, so that a lower bound can be read off without searching.

> *Finding a construction that works is not the same as proving nothing shorter works – I have only produced an upper bound, and an upper bound by itself answers half the question. I need a quantity attached to position that I can bound* per hop, *so that I get a lower bound on hops without having to rule out every other combination by hand.*

> *The target lies on $y = x$, and the vector achieving the largest possible movement along that direction combines an x-step and a y-step that are* both *large and of the same sign – so the quantity to track is not x or y separately but $x + y$, since that is the one linear functional every hop is naturally graded by, and it is exactly the functional that the target's diagonal position makes relevant.*

- ▷ **Cue:** a problem gives a finite list of fixed-length moves and asks for the minimum number of moves to a target $\longrightarrow$ **Move:** find a linear functional (here $x + y$) whose per-move change is bounded above by some constant c; this instantly yields the lower bound $n \geq (\text{functional at target})/c$, to be matched against an explicit construction
- ▷ **Cue:** the target lies along a specific direction (here the diagonal $y = x$) $\longrightarrow$ **Move:** test the functional aligned with that direction first, since the move that maximizes it is likely also the move that was already the natural first choice in construction – the lower bound and the construction are found by the same instinct, applied to two different questions

Layer 3 THE WHY

Why is $x + y$ the correct quantity to bound, rather than x alone, or the true Euclidean distance to the target, or something else? Because a lower bound on a hop-count needs a quantity that is *additive under concatenation of hops* and for which every hop has a *provable ceiling* on its contribution. Euclidean distance to the target does not have a clean per-hop bound in this form: after a hop the grasshopper could in principle move closer to $(2021, 2021)$ by an amount depending on where it currently is, so "distance decreases by at most 5" is true but far too weak (it would only give $n \geq 2021\sqrt{2}/5$, a much worse bound, and it is not even the binding constraint here). What actually pins the twelve vectors down to a maximum of 7 is that $a + b$ for $a^2 + b^2 = 25$ with integers a, b is maximized, over the finite list of Pythagorean pairs $(0, 5), (3, 4), (4, 3), (5, 0)$ (up to sign), at $a + b = 7$ – and this maximum is achieved precisely by the vectors that also point most nearly along $y = x$, the same vectors the natural attempt reached for on instinct. So the choice of $x + y$ is not an arbitrary trick pulled to make the arithmetic work: it is the unique linear functional that (a) is additive across hops, (b) has a finite, computable per-hop maximum over the twelve allowed vectors, and (c) is aligned with the target, so that the bound it produces has a chance of being tight. The remark appended to the official solution even names this functional as the L^1 (taxicab) distance in disguise – once the target sits on the diagonal, L^1 distance and $x + y$ coincide up to the fixed origin, and L^1 distance is exactly the metric a bounded-step-size lower bound argument is built for, because it decomposes additively along coordinate directions the way Euclidean distance does not.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that the target $(2021, 2021)$ lies exactly on $y = x$; this should immediately be written down as "the relevant coordinate is $x + y$, not x and y separately."

2. That cue suggests computing, for all twelve hop vectors, the change each induces in $x + y$, and taking the maximum – turning a search problem into a one-line arithmetic bound.

3. Hesitation occurs right after the 578-hop construction is found: it is tempting to declare victory, since the number looks clean and the construction was not hard to find. The

hesitation is reinforced away by the discipline of always asking, after any "it works" answer to a minimum/maximum problem, "have I also shown nothing smaller/larger works?"

4. The trap is treating a successful construction as a complete answer to a minimization problem. A construction only ever proves an upper bound; the problem asks for the *minimum*, which needs a matching lower bound, and Putnam graders withhold marks for exactly this gap even when the number quoted is correct.

5. Recognition would be faster next time by treating "minimum number of fixed-length moves to a target" as an automatic trigger for two separate sub-tasks done in either order: (i) find a linear functional bounded per move, to get a lower bound; (ii) find an explicit sequence achieving that bound, to close the gap. Neither sub-task alone answers the question.

6. The reusable decision rule: whenever moves are drawn from a finite list of fixed vectors and the goal is a minimum hop-count to a target, look for a linear functional aligned with the target's direction, compute its maximum per-move change over the finite list, divide the target's value by that maximum and round up for the lower bound, then separately construct a sequence meeting it. A minimum is only proved when both halves are shown; a construction alone is only ever an upper bound.

2021 A2

2021 A2. For every positive real number x, let

$$g(x) = \lim_{r\to 0}((x+1)^{r+1} - x^{r+1})^{\frac{1}{r}}.$$

Find $\lim_{x\to\infty} \frac{g(x)}{x}$.

Fields: Analysis (indeterminate forms, l'Hôpital's rule applied inside a limit definition) **Position:** A2 – second problem, first calculus problem of the paper **Answer:** The limit is e: $g(x)$ itself works out to the closed form $(x+1)^{x+1}/x^x$, and $g(x)/x \to e$ because that closed form rearranges into $(1+1/x)^{x+1}$, the compound-interest limit in disguise.

Layer 1 THE WHAT — surface analysis

The natural attempt. The instinct on meeting $((x+1)^{r+1} - x^{r+1})^{1/r}$ is to try to simplify the base directly – binomial-expand $(x+1)^{r+1}$ in powers of r around $r = 0$, since $r+1$ is close to 1 there. Write $(x+1)^{r+1} = (x+1)\cdot(x+1)^r = (x+1)e^{r\ln(x+1)}$ and Taylor-expand the exponential: $(x+1)^{r+1} \approx (x+1)(1 + r\ln(x+1) + O(r^2)) = (x+1) + r(x+1)\ln(x+1) + O(r^2)$. Doing the same for $x^{r+1} \approx x + rx\ln x + O(r^2)$, the difference is

$$(x+1)^{r+1} - x^{r+1} \approx 1 + r\big[(x+1)\ln(x+1) - x\ln x\big] + O(r^2).$$

This is a base of the form $1 + r\cdot c + O(r^2)$ raised to the power $1/r$, and the standard limit $(1+rc)^{1/r} \to e^c$ as $r \to 0$ is exactly the compound-interest limit, so the computation lands on $g(x) = e^{(x+1)\ln(x+1) - x\ln x}$ without difficulty – the Taylor expansion is the honest, natural tool for a base tending to 1 raised to a power tending to ∞. Where this natural attempt gets expensive is in justifying itself: it silently assumes the $O(r^2)$ term can be discarded inside a limit that is also being exponentiated by $1/r$ – an operation that can *amplify* an $O(r^2)$ error into something of order r, which is exactly the same order as the term being kept. Making the Taylor-expansion route rigorous requires tracking the $O(r^2)$ term explicitly through the exponentiation, which is a real amount of extra bookkeeping the natural attempt has swept under the "$O(r^2)$" notation.

The model chain. Take logarithms first, before doing any expansion. Let $N(r) = (x+1)^{r+1} - x^{r+1}$, so $N(0) = (x+1) - x = 1$, and

$$\ln g(x) = \lim_{r\to 0} \frac{\ln N(r)}{r}.$$

Since $N(0) = 1$ so $\ln N(0) = 0$, this is a genuine $0/0$ form in r, licensing l'Hôpital's rule directly – no expansion, no error terms to track:

$$\lim_{r\to 0} \frac{\ln N(r)}{r} = \lim_{r\to 0} \frac{N'(r)}{N(r)} = \frac{N'(0)}{N(0)} = N'(0),$$

using $N(0) = 1$. Differentiating $N(r) = (x+1)^{r+1} - x^{r+1} = (x+1)e^{(r+1)\ln(x+1)} - xe^{(r+1)\ln x}$... more directly, $N'(r) = (x+1)^{r+1}\ln(x+1) - x^{r+1}\ln x$, so

$$N'(0) = (x+1)\ln(x+1) - x\ln x.$$

Hence $g(x) = e^{(x+1)\ln(x+1) - x\ln x} = \dfrac{(x+1)^{x+1}}{x^x}$, exactly the closed form the natural attempt also reached, but obtained with one clean application of l'Hôpital's rule on an honest $0/0$ limit, rather than a Taylor expansion whose error term needs separate justification. Now

$$\frac{g(x)}{x} = \frac{(x+1)^{x+1}}{x^{x+1}} = \left(\frac{x+1}{x}\right)^{x+1} = \left(1+\frac{1}{x}\right)^{x+1} = \left(1+\frac{1}{x}\right)^{x} \cdot \left(1+\frac{1}{x}\right).$$

As $x \to \infty$, the first factor $\to e$ (the defining limit of e) and the second factor $\to 1$, so $g(x)/x \to e$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains reach exactly the same closed form, $g(x) = (x+1)^{x+1}/x^x$ – the mathematics they extract is identical. The split is not in the destination but in the route to $N'(0)$: at the line "$N(r) = (x+1)^{r+1} - x^{r+1}$ is a $1 + O(r)$ quantity being raised to the power $1/r$", the natural attempt reaches for Taylor expansion of $N(r)$ itself and then exponentiates the approximation; the model chain instead takes ln of the whole expression *first*, converting the power $1/r$ into a quotient $\ln N(r)/r$ that is already a textbook $0/0$ form, and only then differentiates.

> *I have a quantity of the form* $(\text{base} \to 1)^{(\text{exponent}\to\infty)}$ *– this is the classic* 1^∞ *indeterminate form, and the standard fix for* 1^∞ *is never to expand the base and multiply out powers; it is to take logs, because* ln *turns the power into a product and the product* $\frac{1}{r}\ln N(r)$ *is a single ratio I can attack with l'Hôpital the moment I check* $N(0) = 1$ *makes it* $0/0$.

> *If I expand* $N(r)$ *in r and then raise the truncated expansion to the power* $1/r$*, I am exponentiating an approximation – and exponentiation is exactly the operation that can turn a negligible-looking* $O(r^2)$ *error into an* $O(r)$ *contribution to the final answer, because* $(1 + rc + O(r^2))^{1/r}$ *has the* $O(r^2)$ *term multiplied by* $1/r$ *inside the exponent before I even get to discard it. Taking the log first avoids ever exponentiating anything I have not already computed exactly.*

- ▷ **Cue:** a limit has the shape (something $\to 1$) raised to a power $\to \infty$ (a 1^∞ form) $\longrightarrow$ **Move:** take logarithms first to convert the power into a product/quotient, reducing to an ordinary $0/0$ or ∞/∞ form that l'Hôpital's rule applies to directly, rather than Taylor-expanding the base and then exponentiating the approximation
- ▷ **Cue:** after simplifying, an expression reduces to $(1 + 1/x)^{x+k}$ for a fixed constant k as $x \to \infty$ $\longrightarrow$ **Move:** split off the factor $(1 + 1/x)^k \to 1$ and recognise the remaining $(1 + 1/x)^x$ as the defining limit of e, rather than re-deriving the limit from scratch

Layer 3 THE WHY

Why must logs come before expansion here, rather than being one equally valid option among several? Because of what a Taylor expansion actually promises: it guarantees $N(r) = 1 + rc + O(r^2)$ is a good *additive* approximation of $N(r)$ for small r – the error is small compared

to 1. But the operation being applied to $N(r)$ is not addition, it is exponentiation by $1/r$, an operation whose sensitivity to error grows without bound as $r \to 0$ exactly because the exponent $1/r \to \infty$. An additive approximation is simply the wrong tool to feed into a multiplicatively (exponentially) sensitive operation without extra work to control it. Logarithm is the operation that converts exponentiation into multiplication and multiplication by $1/r$ into a plain quotient $\ln N(r)/r$ – and a quotient of two functions both vanishing at $r = 0$ is precisely the situation l'Hôpital's rule is built to handle exactly, with no approximation at all. So taking logs is not a trick chosen because it happens to work; it is the operation that matches the tool (l'Hôpital, an exact statement about ratios) to the structure of the problem (a 1^∞ form), whereas Taylor expansion matches an additive tool to a problem that is not additive. This is also why the official solution's third approach (writing $(x+1)^{r+1} - x^{r+1}$ as $\int_x^{x+1}(r+1)u^r\,du$ and squeezing $\left(\int_x^{x+1} u^r\,du\right)^{1/r}$ between x and $x+1$) works too: it sidesteps the 1^∞ issue entirely by trapping the base between two exact bounds rather than approximating it, which is the same underlying principle – never approximate a quantity you are about to raise to an unbounded power; either compute it exactly (log route) or bound it exactly on both sides (squeeze route).

Once $g(x) = (x+1)^{x+1}/x^x$ is known, why does $g(x)/x$ reduce to the e-limit specifically? Because dividing by x and regrouping exponents is forced by matching units: $(x+1)^{x+1}/x^{x+1}$ is a pure ratio raised to a power, and $\left(\frac{x+1}{x}\right)^{x+1} = (1+1/x)^{x+1}$ is definitionally the same object (up to the harmless extra factor of $1 + 1/x \to 1$) as the standard definition of e as $\lim(1+1/n)^n$. The problem is constructed so that the algebra, once the log step is taken, walks directly into the definition of e rather than requiring it to be rediscovered.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the shape $(\text{something} \to 1)^{1/r}$ as $r \to 0$ – this should be flagged immediately as a 1^∞ indeterminate form, the moment the limit is read, before any algebra is attempted.

2. That cue suggests taking ln of the whole expression first, converting the power into $\frac{1}{r}\ln(\text{base})$, which is then attacked with l'Hôpital's rule once it is confirmed to be $0/0$.

3. Hesitation occurs at the moment of choosing between Taylor-expanding $N(r)$ directly versus taking logs – both look equally like "the next step." It is reinforced away by asking what operation is about to be applied to the approximation: since it is exponentiation by an unbounded power, an additive (Taylor) approximation is the wrong choice, and the log route is preferred on that structural ground, not by trial.

4. The trap is Taylor-expanding the base and discarding the $O(r^2)$ term before exponentiating by $1/r$, silently assuming the discarded term stays negligible under an operation that is precisely designed to amplify small differences in the base into $O(1)$ differences in the result.

5. Recognition would be faster next time by treating any limit of the exact shape $(\text{base}(r))^{1/r}$ (or more generally $f(r)^{h(r)}$ with $f \to 1$, $h \to \infty$) as an automatic trigger for "take logs first," before any other manipulation is attempted.

6. The reusable decision rule: for any 1^∞-type limit, never approximate the base before exponentiating; take logarithms to convert the exponentiation into a product or quotient, reduce that to a genuine $0/0$ or ∞/∞ form, and apply l'Hôpital's rule (or an exact squeeze) only at that stage, where approximation is no longer required because the form is honest.

2021 A3

2021 A3. Determine all positive integers N for which the sphere

$$x^2 + y^2 + z^2 = N$$

has an inscribed regular tetrahedron whose vertices have integer coordinates.

Fields: Number theory / solid geometry (regular tetrahedra, the tetrahedron-cube duality, rational quantities forced to be perfect squares) **Position:** A3 – third problem, the paper's first genuinely 3D question **Answer:** $N = 3m^2$ for a positive integer m: sufficiency by the explicit construction $(\pm m, \pm m, \pm m)$ (four alternate vertices of a cube), necessity because any such tetrahedron's antipodal completion to a cube forces $N/3$ to be a rational perfect square, hence an integer perfect square.

Layer 1 THE WHAT — surface analysis

The natural attempt. The first honest move is to try small, symmetric cases directly: pick four lattice points that look like they might be equidistant from each other and from the origin. The four sign-patterns $(1,1,1), (1,-1,-1), (-1,1,-1), (-1,-1,1)$ are an obvious first guess, since they are literally half the vertices of the unit cube, chosen so no two share more than one coordinate sign in common. Checking: each has norm $\sqrt{3}$, so they lie on $x^2+y^2+z^2 = 3$, and the pairwise distance between any two, e.g. $(1,1,1)$ and $(1,-1,-1)$, is $\sqrt{0+4+4} = 2\sqrt{2}$ – the same for every pair by the symmetry of the sign pattern, so this is indeed regular. Scaling by an integer m gives $N = 3m^2$ works, for every m. This construction is quick and correct as far as it goes. The trouble is what comes next: the problem does not ask only for a construction, it asks to *determine all* N, which means proving the converse – that no other N admits such a tetrahedron. The natural instinct here is to try to rule out other N by more casework: attempt small N not of the form $3m^2$ (like $N = 1, 2, 4, 5, 6, \ldots$) and try to show by hand that no lattice-point regular tetrahedron fits. This is hopeless as a general method – there is no bound on how large or oddly-oriented such a tetrahedron could be for a given N, so no finite amount of case-checking closes the problem. Something structural about *every* inscribed regular tetrahedron, not just the symmetric one found by guessing, is needed.

The model chain. Let P_1, P_2, P_3, P_4 be the vertices of any regular tetrahedron with integer coordinates, inscribed in $x^2+y^2+z^2 = N$ (so each $|P_i|^2 = N$). Four non-coplanar points determine a unique circumscribed sphere, so that sphere must be the given one, centred at the origin; and for a *regular* tetrahedron the centroid coincides with the circumcentre (the point equidistant from all four vertices), so $P_1 + P_2 + P_3 + P_4 = 0$. Consequently the four reflected points $Q_i = -P_i$ also lie on the sphere and also have integer coordinates. A regular tetrahedron together with its point-reflection through its own centre is a standard configuration: the eight points $P_1, \ldots, P_4, Q_1, \ldots, Q_4$ are exactly the eight vertices of a cube, with each P_i adjacent (via a cube edge) to the three Q_j other than its own antipode Q_i, and each pair P_i, Q_i a cube space-diagonal. (This is visible directly in the construction above: $(\pm m, \pm m, \pm m)$ with an even number of minus signs are the P_i, with an odd number the Q_i, and together they are the full cube of side $2m$.)

The space diagonal of this cube joins an antipodal pair, e.g. P_1 and $Q_1 = -P_1$, and has length $|P_1 - Q_1| = 2|P_1| = 2\sqrt{N}$. For a cube of side length s, the space diagonal has length $s\sqrt{3}$, so $s\sqrt{3} = 2\sqrt{N}$, giving $s^2 = \dfrac{4N}{3}$ – this holds for *any* such inscribed tetrahedron, not just the symmetric example.

Now use integrality: the cube's volume s^3 equals the absolute value of the determinant of the integer vectors from one cube vertex to its three cube-edge-neighbours (this is the standard formula for the volume of a parallelepiped spanned by three edge vectors from a common vertex, and here the parallelepiped is exactly the cube itself). Since Q_1's three edge-neighbours are P_2, P_3, P_4, this determinant is built entirely from integer coordinates, so s^3 is a positive integer.

Caveat. The published solutions text asserts the cube's side length is $(N/3)^{1/2}$ and its volume $(N/3)^{3/2}$, computed as the determinant of $Q_2 - Q_1$, $Q_3 - Q_1$, $Q_4 - Q_1$; working this determinant out explicitly on the symmetric example $m = 1$ (so $N = 3$) gives 16, not

$1 = (N/3)^{3/2}$ there, and the side-length relation I derive above, $s^2 = 4N/3$ (so $s = 2\sqrt{N/3}$, matching $s = 2$ when $m = 1$, the true cube side), differs from the published figure by the same factor. I believe the published write-up understates the cube's side length by a factor of 2 and its volume by a factor of 8 (using face-diagonal vectors of the cube where edge vectors were intended), but I have verified my own relation $s^2 = 4N/3$ directly against the $m = 1$ example and it is internally consistent, so I proceed on my own corrected constant; this does not change which N are claimed to work, only the intermediate arithmetic used to get there.

So $s^3 = \left(\frac{4N}{3}\right)^{3/2}$ is a positive integer. Write $q = N/3$, a positive rational number since N is a positive integer; then $s^3 = (4q)^{3/2} = 8q^{3/2}$, so $q^{3/2}$ is rational (indeed $s^3/8$). For a positive rational q, if $q^{3/2}$ is rational then so is $\sqrt{q} = q^{3/2}/q$. Write $\sqrt{q} = a/b$ in lowest terms, a, b positive integers with $\gcd(a, b) = 1$; then $q = a^2/b^2$, so $N = 3q = 3a^2/b^2$. Since N is an integer and $\gcd(a^2, b^2) = 1$, b^2 must divide 3; as 3 is squarefree this forces $b = 1$. Hence $q = N/3 = a^2$ is a perfect square of a positive integer, i.e. $N = 3a^2$. This is exactly the family exhibited by the construction, so it is the complete answer.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains find the same working example, $(\pm m, \pm m, \pm m)$ with $N = 3m^2$, by the same reasonable guess. They split at the point where the construction is finished and the "does this attain every N of this form, and only this form?" question is asked. "Both paths agree on the construction. At that point the natural attempt tries to rule out other N one value at a time; the model chain instead asks what is true of *every* inscribed regular tetrahedron, symmetric or not, and finds that the point reflection through the centroid always completes it to a cube."

> *Checking specific small N by hand can only ever produce a finite list of failures – it can never prove a universal negative over all possible orientations and sizes of tetrahedron. I need a structural fact true of* any *lattice-point regular tetrahedron on this sphere, not just the one I happened to construct.*

> *A regular tetrahedron's centroid is also its circumcentre – that is what "regular" buys me, and the sphere already fixes the circumcentre at the origin. So the four vertices sum to zero* automatically, *with no extra hypothesis, which means their negatives are also lattice points on the same sphere – and eight points with this much forced symmetry are exactly a cube's worth of vertices, a shape I already know volume formulas for.*

- ▷ **Cue:** a regular polygon or polytope is inscribed in a sphere/circle centred at the origin $\longrightarrow$ **Move:** use that the centroid of a regular figure coincides with its circumcentre, forcing the vertex sum to zero; this often lets a smaller symmetric configuration be completed to a larger, more tractable one (here: tetrahedron completed to cube by point-reflection)
- ▷ **Cue:** a geometric quantity built from integer-coordinate points must satisfy an algebraic identity (here $s^3 = (4N/3)^{3/2}$) and is simultaneously known to be an integer (as a determinant) $\longrightarrow$ **Move:** isolate the "must be a perfect k-th power of a rational" consequence, then use gcd/lowest-terms bookkeeping to upgrade "rational perfect square" to "integer perfect square"

Layer 3 THE WHY

Why is completing the tetrahedron to a cube the natural move here, rather than a lucky trick? Because "regular tetrahedron" and "sphere centred at the origin" together force a piece of structure the problem is handing the solver for free: regularity means the centroid is automatically the circumcentre (the point equidistant from all vertices is unique for a non-degenerate simplex,

and for a regular one it coincides with the centroid by the symmetry group acting transitively on vertices), and the sphere's centre *is* that circumcentre because the four vertices lie on it. There is no assumption to add – the vertex sum being zero is not a special property of the symmetric example, it is forced by the hypotheses of the problem for *every* valid configuration. Once the vertex sum is zero, the reflected points are lattice points on the same sphere for free (reflection through the origin preserves both integrality and the sphere), and a set of eight points that are a regular tetrahedron together with its own point-reflection is, up to the labelling, always a cube – this is a fixed fact about the symmetry group of the regular tetrahedron (order 12, index 2 in the cube's rotation group of order 24), not a coincidence of this problem. So the cube is not an external object being cleverly introduced; it is the object already implicit in "regular tetrahedron plus its forced antipodal image," and recognising it is what converts an unbounded geometric search into a bounded algebraic one: cubes have a volume formula in terms of the side length, and the side length is pinned down exactly by the one Euclidean fact that survives from any orientation – the space diagonal equals twice the sphere's radius, because P_1 and Q_1 are always antipodal points on the sphere. Everything after that is routine number theory (rational cube root of a rational number forces the base itself to be rational, then lowest-terms plus 3 squarefree forces an integer square) – routine once the cube is seen, but invisible without it.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the phrase "regular tetrahedron inscribed in a sphere centred at the origin"; the consequence to write down immediately is that the centroid equals the circumcentre equals the origin, so the four vertices sum to zero with no further argument.

2. That cue suggests reflecting the tetrahedron through the origin to get four more lattice points on the same sphere, and recognising the resulting eight points as the vertices of a cube – converting a 3D existence/classification question into an algebraic one about a cube's side length.

3. Hesitation occurs right after finding the symmetric construction $(\pm m, \pm m, \pm m)$ works: it is tempting to stop, since a working example for infinitely many N feels like most of the answer. It is reinforced away by rereading the problem's verb – "determine all" – which demands a converse, and by recognising that no amount of checking further small N by hand can supply one.

4. The trap is trying to prove the converse by exhausting small counterexamples, which can never terminate in a proof, or by assuming (without justification) that the symmetric construction is the *only shape* of inscribed regular tetrahedron up to rotation, which is not obviously true and is not what the argument above actually needs – the argument works for an arbitrary orientation, which is the point.

5. Recognition would be faster next time by treating "regular [polytope] inscribed in a sphere/circle about the origin" as an automatic trigger for "vertex sum is zero, for free," and by keeping on hand the fact that a regular tetrahedron plus its point-reflection through its centroid is always a cube.

6. The reusable decision rule: when a problem asks to classify all N (or all integer parameters) for which a regular, highly symmetric lattice configuration exists on a sphere or circle, do not attempt case-by-case elimination. Use the forced centroid-equals-centre property to generate a larger, more standard configuration (often by reflection or by taking a related regular polytope one dimension up), extract an exact algebraic relation for a natural linear measurement of that configuration, and combine "this measurement is forced to be a rational k-th root of a rational" with "it is also an integer by a determinant/volume argument" to force the parameter into a perfect-power form via lowest-terms bookkeeping.

2021 A4

2021 A4. Let

$$I(R) = \iint_{x^2+y^2\le R^2} \left(\frac{1+2x^2}{1+x^4+6x^2y^2+y^4} - \frac{1+y^2}{2+x^4+y^4} \right) dx\,dy.$$

Find

$$\lim_{R\to\infty} I(R),$$

or show that this limit does not exist.

Fields: multivariable calculus: improper double integrals, conditional convergence, linear change of variables, polar coordinates; a hidden scaling symmetry **Position:** A4 — the pivot of the paper. It looks like a computation and is not one; the arithmetic at the end is routine, the structural move at the start is everything **Answer:** The limit exists and equals $\dfrac{\sqrt{2}}{2}\pi\log 2 = \dfrac{\pi\log 2}{\sqrt{2}} \approx 1.53979$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You see a double integral over a disc with two rational integrands, and the trained reflex is: polar coordinates, then compute. Do it honestly and watch what happens.

Put $x = r\cos\theta$, $y = r\sin\theta$. The second denominator becomes

$$2 + x^4 + y^4 = 2 + r^4(\cos^4\theta + \sin^4\theta) = 2 + r^4\left(1 - \tfrac{1}{2}\sin^2 2\theta\right),$$

which is at least clean. The first denominator becomes

$$1 + x^4 + 6x^2y^2 + y^4 = 1 + r^4\left(1 + \sin^2 2\theta\right),$$

also clean. So the two terms are both of the shape $(\text{quadratic in } r)/(\text{quartic in } r)$ with θ-dependent coefficients, and the area element contributes another r. Each term therefore behaves like r^{-1} as $r\to\infty$, and $\int^R r^{-1}\,dr = \log R$.

Each of the two integrals diverges. Separately, they are both $\sim c(\theta)\log R$. So you cannot compute them one at a time and subtract at the end: you would be subtracting $\infty - \infty$. This is the wall, and it is a good wall to hit, because it tells you the entire problem is about a cancellation of two logarithms.

The natural attempt now does the obvious thing: keep them together. Write

$$I(R) = \int_0^{2\pi}\int_0^R \left(\frac{1+2r^2\cos^2\theta}{1+r^4(1+\sin^2 2\theta)} - \frac{1+r^2\sin^2\theta}{2+r^4(1-\frac{1}{2}\sin^2 2\theta)} \right) r\,dr\,d\theta$$

and try to do the r-integral. Each piece is a rational function of r with a quartic denominator; partial fractions over $\mathbb{R}$ splits each quartic into two quadratics with θ-dependent coefficients whose roots are fourth roots of a θ-dependent constant. You are now looking at nested radicals in θ inside an arctangent and a logarithm, to be integrated over θ afterwards. This is zombie maths in its purest form: every step is legal, every step is a step you know how to take, and the whole thing is unfinishable in the time available. Twenty minutes in you have half a page of $\left(1+\sin^2 2\theta\right)^{1/4}$ and no route to a closed form.

The natural attempt fails not because the student lacks a technique but because it never asked the question "why did the setter put a 6 there?"

The model chain. **Step 1: symmetrise.** The domain $x^2+y^2\le R^2$ is invariant under $(x,y)\mapsto(y,x)$, and both denominators are symmetric in x and y. Only the numerators are not. So swapping the names of the variables gives a second expression for the same number:

$$I(R) = \iint_{x^2+y^2\le R^2} \left(\frac{1+2y^2}{1+x^4+6x^2y^2+y^4} - \frac{1+x^2}{2+x^4+y^4} \right) dx\,dy.$$

Average the two. Writing $I(R) = \iint_{x^2+y^2\le R^2}\big(f-g\big)$, we get

$$f(x,y) = \frac{1+x^2+y^2}{1+x^4+6x^2y^2+y^4}, \qquad g(x,y) = \frac{1+\frac12 x^2+\frac12 y^2}{2+x^4+y^4}.$$

Both are now fully symmetric, and — this is the point — both are functions of x^2+y^2 and of the quartic alone.

Step 2: identify the two integrands. Set $u = x+y$, $v = x-y$. Then

$$u^2+v^2 = 2(x^2+y^2), \qquad u^4+v^4 = 2x^4+12x^2y^2+2y^4,$$

the second because with $a = x^2+y^2$ and $c = 2xy$ we have $u^2 = a+c$, $v^2 = a-c$, so $u^4+v^4 = (a+c)^2+(a-c)^2 = 2a^2+2c^2 = 2(x^4+2x^2y^2+y^4)+8x^2y^2$. Therefore

$$g(u,v) = \frac{1+(x^2+y^2)}{2+2\,(x^4+6x^2y^2+y^4)} = \tfrac12 f(x,y),$$

that is,

$$\boxed{f(x,y) = 2\,g(x+y,\,x-y).}$$

Step 3: turn the identity into a change of domain. The map $T(x,y) = (x+y,\,x-y)$ is a rotation by -45° composed with a dilation by $\sqrt2$; its Jacobian determinant is -2, so $du\,dv = 2\,dx\,dy$, and it carries the disc of radius R onto the disc of radius $R\sqrt2$. Hence

$$\iint_{x^2+y^2\le R^2} f = \iint_{x^2+y^2\le R^2} 2\,g(T(x,y))\,dx\,dy = \iint_{u^2+v^2\le 2R^2} g(u,v)\,du\,dv.$$

Subtracting the g-integral over the disc of radius R leaves an *annulus*:

$$I(R) = \iint_{R^2\le x^2+y^2\le 2R^2} g(x,y)\,dx\,dy.$$

The $\infty-\infty$ has been converted into a single convergent object.

Step 4: evaluate. In polar coordinates, $x^4+y^4 = r^4\big(1-\frac12\sin^2 2\theta\big)$, so with $D(\theta) := 1-\frac12\sin^2 2\theta \in [\frac12, 1]$,

$$I(R) = \int_R^{R\sqrt2}\int_0^{2\pi} \frac{1+\frac12 r^2}{2+r^4 D(\theta)}\, r\,d\theta\,dr.$$

Rescale $r = Rs$ so the limits stop moving:

$$I(R) = \int_1^{\sqrt2}\int_0^{2\pi} \frac{R^2 s\,\big(1+\frac12 R^2 s^2\big)}{2+R^4 s^4 D(\theta)}\,d\theta\,ds.$$

For $R\ge 1$, $s\in[1,\sqrt2]$ the denominator is at least $R^4s^4/2$, so the integrand is at most $2R^{-2}s^{-3}+s^{-1}\le 3$: uniformly bounded, on a domain of finite measure. Dominated convergence lets the limit pass inside, and pointwise the integrand tends to $\dfrac{s^3/2}{s^4 D(\theta)} = \dfrac{1}{2sD(\theta)}$. Hence

$$\lim_{R\to\infty} I(R) = \int_1^{\sqrt2}\frac{ds}{s}\int_0^{2\pi}\frac{d\theta}{2D(\theta)} = \log\sqrt2\int_0^{2\pi}\frac{d\theta}{2-\sin^2 2\theta}.$$

Now $2-\sin^2 2\theta = 1+\cos^2 2\theta = \frac12(3+\cos 4\theta)$, so the angular integral is $\int_0^{2\pi}\frac{2\,d\theta}{3+\cos 4\theta}$. Substituting $\varphi = 4\theta$ runs through four full periods, and since the integrand has period 2π in φ this equals $2\int_0^{2\pi}\frac{d\varphi}{3+\cos\varphi}$. The half-angle substitution $t = \tan(\varphi/2)$ gives

$$\int_0^{2\pi}\frac{d\varphi}{3+\cos\varphi} = \int_{-\infty}^{\infty}\frac{2\,dt}{3(1+t^2)+(1-t^2)} = \int_{-\infty}^{\infty}\frac{dt}{2+t^2} = \frac{\pi}{\sqrt2},$$

so the angular integral is $2 \cdot \pi/\sqrt{2} = \pi\sqrt{2}$. Therefore

$$\lim_{R\to\infty} I(R) = \tfrac{1}{2}\log 2 \cdot \pi\sqrt{2} = \frac{\sqrt{2}}{2}\pi \log 2.$$

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths agree through polar coordinates and through the discovery that each half diverges logarithmically. They split at (†): "the two terms must be handled together." At (†) the natural attempt keeps the *difference of the integrands* and tries to integrate it; the model looks for a *map sending one integrand to the other* and integrates only one of them, over the region between the two images of the disc. One treats the minus sign as an algebraic operation to be performed; the other treats it as a signal that two copies of the same object are being compared.

> *Both halves diverge like* $\log R$ *and the difference is finite. That is not luck — it means the two halves are the same thing measured on two different scales. So my job is not to integrate; my job is to find the map that turns one integrand into the other, and then the answer is whatever is left between the old domain and the new one.*

> *The numerators* $1 + 2x^2$ *and* $1 + y^2$ *are not symmetric but the domain and both denominators are. A number that is unchanged when I rename the variables can be replaced by the average of the two namings, and averaging is free. Do that first, before anything clever, because it costs nothing and it may reveal a symmetry that the setter deliberately obscured.*

> *Why a 6? Because* $x^4 + 6x^2y^2 + y^4 = \frac{1}{2}\left((x+y)^4 + (x-y)^4\right)$. *That coefficient is a fingerprint of the substitution* $u = x + y$, $v = x - y$; *nothing else produces a 6 in that position.*

Layer 3 THE WHY

Why must the scaling thought be thought? Because of a fact about homogeneity that the divergence already told you.

Far from the origin, the integrand of g behaves like $\frac{1}{2}r^2/(r^4 D(\theta)) = \frac{1}{2r^2 D(\theta)}$: it is asymptotically homogeneous of degree -2 in the plane. Degree -2 in dimension 2 is exactly the critical exponent. Integrating a function homogeneous of degree -2 over the annulus $R \le |\mathbf{x}| \le \lambda R$ gives

$$\int_R^{\lambda R}\int_0^{2\pi} \frac{h(\theta)}{r^2}\, r\, d\theta\, dr = \log\lambda \int_0^{2\pi} h(\theta)\, d\theta,$$

which is *independent of* R. The scale-invariance of dr/r is the whole engine of the problem. Log-divergence at infinity and log-divergence at the origin are the same phenomenon: a critical exponent makes every dyadic shell contribute equally.

So the structure is forced. If a difference of two individually log-divergent integrals converges, the leading $\log R$ terms must cancel exactly, which means the two angular profiles $h(\theta)$ agree *after some transformation of the plane*. And the transformations that preserve the class of degree-(-2) behaviour while changing the domain are the linear ones. The setter has therefore hidden a linear map, and the only question is which one. The 6 answers it.

This is also where the archive earns its keep. Kedlaya's solution opens by interchanging x and y, averaging, and then *asserting* $f(x,y) = 2g(x+y, x-y)$. Read cold, that identity is a rabbit from a hat. Here is where it comes from. Take the two quartics you are given:

$$Q_1 = x^4 + 6x^2y^2 + y^4, \qquad Q_2 = x^4 + y^4.$$

A linear substitution $x \mapsto ax + by$, $y \mapsto bx + ay$ applied to Q_2 produces $(a^4 + b^4)(x^4 + y^4) +$ (cross terms); with $a = b = 1$ it gives $(x + y)^4 + (x - y)^4 = 2x^4 + 12x^2y^2 + 2y^4 = 2Q_1$. You do not have to guess: the coefficient ratio $6 : 1$ tells you the substitution is ± 1, and ± 1 substitutions are the only ones worth trying because the problem's constants (1 against 2 in the denominators, $1 + 2x^2$ against $1 + y^2$ in the numerators) are all small integers. Small integers in a problem mean the setter built it backwards from an integer substitution.

The symmetrisation step has its own "of course". A definite integral is a number, and a number does not know what the variables were called. If the region and the denominators are symmetric under $x \leftrightarrow y$ but the numerators are not, then the asymmetry is an artefact of presentation, and presentation is the one thing you are allowed to change for free. Replacing an expression by the average of its images under a symmetry of the problem is never a gamble — it cannot lose information that the integral retains. *Symmetrise before you compute* is not a trick; it is a consequence of the integral being a number.

Finally, the constant $\log \sqrt{2}$ is not an accident of the arithmetic: it is the logarithm of the dilation factor of T, and it is where the divergent $\log R$ went. The problem's answer is literally the residue of a cancelled infinity.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the 6 in $x^4 + 6x^2y^2 + y^4$ sitting next to a bare $x^4 + y^4$. Write down immediately: $x^4 + 6x^2y^2 + y^4 = \frac{1}{2}\left((x + y)^4 + (x - y)^4\right)$. Second cue, found within a minute of starting: in polar coordinates each term decays like $r^{-1}\,dr$, so *each half diverges*. Write that down too — it is a fact about the problem, not a failure.

2. Two divergent halves whose difference converges suggests: find a measure-preserving-up-to-scale map carrying one integrand to the other, so that the difference collapses to an integral over the region between a domain and its image. The supporting techniques are symmetrisation under a symmetry of the domain, linear change of variables with its Jacobian, and polar coordinates for the final evaluation.

3. Hesitation occurs at the moment you realise polar coordinates have not made the problem finite. The instinct is to press on with partial fractions in r. What reinforces it away is asking the diagnostic question: *what would have to be true for this limit to exist at all?* The answer — an exact cancellation of two logarithms — immediately implies the two integrands are related, not merely being subtracted. A second hesitation comes at the interchange of $\lim_{R\to\infty}$ with the integral; that is reinforced away by rescaling $r = Rs$ so that the domain becomes fixed and a one-line uniform bound applies.

4. The trap is set twice. (i) The obvious trap: computing the two integrals separately and getting a finite-looking answer by cancelling the $\log R$'s informally. Both halves are genuinely infinite; any manipulation that treats them as separate numbers is wrong, and a grader will look for exactly that. (ii) The subtle trap: after reaching the annulus, dropping the 1 and the 2 from the integrand "because they are lower order" without justification. They are lower order, but the justification is a dominated-convergence bound and it must appear — this is the only analytic content in the second half of the problem.

5. Recognisable faster next time by two flags. *Flag one: unsymmetric numerator, symmetric everything else ⇒ average over the symmetry group. Flag two: a difference of two objects that are individually divergent ⇒ do not subtract, transform.* A conspicuous integer coefficient in a homogeneous form (6 in $x^4 + 6x^2y^2 + y^4$, 3 in $x^3 + 3xy^2$, 4 in $x^4 + 4y^4$) is a fingerprint of a specific substitution; factor the form to read it off.

6. The reusable decision rule: **when a limit is asked for and the naive computation diverges, stop computing and ask what cancellation could make it finite; the structure of that cancellation is the solution.** A convergent difference of divergent pieces means the pieces are the same object seen at two scales, and the answer is the logarithm of the scale ratio times an angular average. Corollary in any dimension n: an integrand asymptotically homogeneous of degree $-n$ is log-divergent, and its integral over $\{R \le |\mathbf{x}| \le \lambda R\}$ is $\log \lambda$

times the integral of the angular profile over the sphere — independent of R. That single fact converts many "does the limit exist" integrals into one-line answers.

- ▷ **Cue:** a homogeneous form appears with a conspicuous middle coefficient, e.g. $x^4 + 6x^2y^2 + y^4$ ⟶ **Move:** factor it as a combination of shifted forms — here $\frac{1}{2}\left((x+y)^4 + (x-y)^4\right)$ — and read off the linear substitution the setter used to build the problem
- ▷ **Cue:** the domain of integration and the denominators are symmetric under $x \leftrightarrow y$ but the numerator is not ⟶ **Move:** replace the integrand by the average of itself and its swap; this is free and usually reveals the intended symmetric form
- ▷ **Cue:** a limit of a difference is requested and each piece separately diverges ⟶ **Move:** do not subtract the integrands — look for a transformation carrying one piece to the other, so that the difference becomes an integral over the region between a domain and its image
- ▷ **Cue:** an integrand in the plane decays like $1/r^2$ at infinity ⟶ **Move:** recognise the critical exponent: the integral is log-divergent, every dyadic shell contributes equally, and over an annulus $R \le r \le \lambda R$ the value is $\log \lambda$ times the angular average, independent of R
- ▷ **Cue:** $\int_0^{2\pi} d\theta/(a + b\cos k\theta)$ appears after polar coordinates ⟶ **Move:** collapse the k by periodicity to k copies of the $k = 1$ case, then use $t = \tan(\varphi/2)$ to get $\int dt/(\text{quadratic})$ and an arctangent; the value is $2\pi/\sqrt{a^2 - b^2}$
- ▷ **Cue:** you want to pass a limit inside an integral whose domain depends on the limit parameter ⟶ **Move:** rescale the variable so the domain is fixed first, then find a single uniform bound on the integrand and cite dominated convergence

2021 A5

2021 A5. Let A be the set of all integers n such that $1 \le n \le 2021$ and $\gcd(n, 2021) = 1$. For every nonnegative integer j, let

$$S(j) = \sum_{n \in A} n^j.$$

Determine all values of j such that $S(j)$ is a multiple of 2021.

Fields: number theory: power sums modulo a prime, primitive roots, the Chinese Remainder Theorem, the group structure of $(\mathbb{Z}/m)^\times$ **Position:** A5 — positioned as one of the two hardest on the paper, but structurally the most forgiving problem in the half. Everything turns on one lemma and one factorisation **Answer:** Exactly those j that are divisible by neither 42 nor 46. (In particular $j = 0$ fails, since 0 is divisible by both.)

Layer 1 THE WHAT — surface analysis

The natural attempt. The first thing a well-prepared student does is factor: $2021 = 43 \times 47$, two primes, both close to 45. Good. The second thing is to reach for the Chinese Remainder Theorem, because 43 and 47 are coprime and "$2021 \mid S(j)$" splits into "$43 \mid S(j)$ and $47 \mid S(j)$". Also good.

Then the trouble starts. The natural instinct is to compute. Take $j = 1$: $\sum_{n \in A} n$. You can get this by inclusion–exclusion,

$$\sum_{n=1}^{2021} n - 43\sum_{k=1}^{47} k - 47\sum_{k=1}^{43} k + 2021,$$

and with some care it comes out divisible by 2021. Take $j = 2$: the same inclusion–exclusion now needs $\sum k^2 = \frac{1}{6}m(m+1)(2m+1)$, and it still works, but it took five minutes. Take $j = 3$. Take $j = 42$. You cannot compute $\sum_{n \in A} n^{42}$ this way, and even if you could, "determine all j" is

a statement about infinitely many exponents. Each new j needs a new Faulhaber polynomial and the Bernoulli numbers start appearing. That is the wall.

There is a second, subtler trap in the natural attempt. Inclusion–exclusion on $\{1,\dots,2021\}$ is a statement about *integers in a range*, and the answer is going to be a statement about *residues in a group*. Every minute spent on Faulhaber's formula is a minute spent in the wrong category. The set A is not really a set of 1932 integers; it is a copy of the unit group $(\mathbb{Z}/2021)^\times$ that happens to have been written out in lowest positive representatives. The natural attempt never notices this because the problem was phrased to conceal it: $1 \le n \le 2021$ looks like a range, not a group.

Carry the natural attempt one step further and it does eventually work, provided you switch categories: note that $n \mapsto n \bmod 43$ maps A onto $\{1,\dots,42\}$ with every fibre of the same size, and then you need $\sum_{n=1}^{p-1} n^j \bmod p$. So the natural attempt is not wrong. It is just that everything before that observation was wasted.

The model chain. **Lemma.** For p prime and $j \ge 0$,

$$\sum_{n=1}^{p-1} n^j \equiv \begin{cases} -1 \pmod p & \text{if } (p-1) \mid j, \\ 0 \pmod p & \text{otherwise.} \end{cases}$$

Proof. If $(p-1) \mid j$ then by Fermat's little theorem $n^j \equiv 1$ for every $n \not\equiv 0$, so the sum is $p-1 \equiv -1$.

If $(p-1) \nmid j$, pick a primitive root m modulo p, i.e. a generator of the cyclic group $(\mathbb{Z}/p)^\times$, which has order $p-1$. Then $m^j \equiv 1$ would force $(p-1) \mid j$; so $m^j \not\equiv 1$. Multiplication by m permutes $(\mathbb{Z}/p)^\times$, hence

$$T := \sum_{n=1}^{p-1} n^j \equiv \sum_{n=1}^{p-1} (mn)^j = m^j T \pmod p.$$

Thus $(m^j - 1)T \equiv 0$ with $m^j - 1 \not\equiv 0$, and since p is prime, $T \equiv 0$. □

Reduction of $S(j)$. By the Chinese Remainder Theorem the map $n \mapsto (n \bmod 43,\ n \bmod 47)$ is a bijection from $\{1,\dots,2021\}$ to $(\mathbb{Z}/43) \times (\mathbb{Z}/47)$, and it carries A exactly onto $(\mathbb{Z}/43)^\times \times (\mathbb{Z}/47)^\times$, because $\gcd(n, 2021) = 1$ says precisely that n is a nonzero residue modulo each prime. In particular $|A| = 42 \cdot 46 = 1932$.

Now fix a residue $a \not\equiv 0 \pmod{43}$. The elements of A lying above a are indexed by the 46 choices of the second coordinate, so each such a is hit exactly 46 times. Therefore

$$S(j) \equiv 46 \sum_{a=1}^{42} a^j \pmod{43}, \qquad S(j) \equiv 42 \sum_{b=1}^{46} b^j \pmod{47}.$$

Conclusion. Since $46 \equiv 3 \pmod{43}$ is invertible mod 43, and 42 is invertible mod 47, these multipliers cannot create or destroy divisibility. By the Lemma,

$$43 \mid S(j) \iff 42 \nmid j, \qquad 47 \mid S(j) \iff 46 \nmid j.$$

(When $42 \mid j$ the first sum is $\equiv 46 \cdot (-1) \equiv -3 \not\equiv 0 \pmod{43}$; when $46 \mid j$ the second is $\equiv 42 \cdot (-1) \equiv -42 \not\equiv 0 \pmod{47}$.)

Hence $2021 \mid S(j)$ if and only if j is divisible by neither 42 nor 46.

Sanity check at $j = 0$: $S(0) = |A| = 1932$, and $1932 = 43 \cdot 44 + 40$, so $S(0) \equiv 40 \pmod{43}$ — not divisible, as predicted, since $42 \mid 0$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths agree at $2021 = 43 \times 47$ and at "split the divisibility by CRT". They split at (†): "now handle $S(j)$ modulo 43." At (†) the natural attempt *evaluates the sum over the integer range* $1 \le n \le 2021$, by inclusion–exclusion and power-sum formulae. The model *pushes the set A through the reduction map and counts fibres*, replacing a sum over 1932 integers by 46 copies of a sum over the 42 nonzero residues. One works in $\mathbb{Z}$ and reduces at the end; the other reduces first and works in $\mathbb{F}_{43}$.

> *The set A is not a list of integers, it is the unit group $(\mathbb{Z}/2021)^\times$ wearing a disguise. And j is an exponent on a group. So this problem is about the order of elements, not about summation formulae — and the only exponents that will matter are the ones divisible by the group's exponent.*

> *Reduction mod 43 is a group homomorphism from $(\mathbb{Z}/2021)^\times$ onto $(\mathbb{Z}/43)^\times$, and every fibre of a group homomorphism has the same size. So the sum over A is a constant multiple of the sum over $\{1,\dots,42\}$, and that constant is the kernel size 46 — which I can throw away, because it is a unit mod 43.*

> *To decide whether $\sum_n n^j$ vanishes mod p I do not need to evaluate it. I only need one element that reshuffles the sum: if m generates the group, then multiplying by m permutes the terms, so $T = m^j T$, and either $m^j = 1$ or $T = 0$. Symmetry kills the sum without computing the sum.*

Layer 3 THE WHY

Why must the primitive-root thought be thought? Drive it to the definition.

A sum of the form $\sum_{g\in G}\chi(g)$, where G is a finite group and χ a homomorphism to a field, is not a computation. It is a symmetry question, for a reason that is essentially the reason a circle has a centre. Look at what the sum *is*: the total of the values of a function on a set that a group acts on transitively by translation. Translating the set does not change the total. So the total is fixed by a whole group of maps; and if any of those maps acts on the total by multiplication by something other than 1, the total is forced to be 0. That is the entire content of the Lemma, and it is true in vastly more generality than mod p: it is orthogonality of characters, it is why $\sum_{k=0}^{n-1}\omega^k = 0$ for ω a nontrivial nth root of unity, and it is why a nonconstant balanced function averages to zero.

Written out as the sentence a solver actually thinks: *a sum over a group of a function that is multiplicative in the group variable is either trivially large (the function is constant 1) or exactly zero. There is no third option.* Once you believe that dichotomy, there is nothing left to compute — you only have to decide which case you are in, and that is the condition $(p-1)\mid j$, i.e. whether $n\mapsto n^j$ is the trivial character.

Now the second "of course". Why 42 and 46 rather than, say, 1932?

Because $(\mathbb{Z}/2021)^\times \cong (\mathbb{Z}/43)^\times \times (\mathbb{Z}/47)^\times \cong \mathbb{Z}/42\times\mathbb{Z}/46$ is *not cyclic* — $\gcd(42,46)=2$, so its exponent is $\operatorname{lcm}(42,46)=966$, strictly less than its order 1932. The whole problem is the failure of 2021 to have a primitive root. If 2021 were prime, the answer would be the single clean condition "$2020\nmid j$", and the problem would be the Lemma and nothing else. Because 2021 splits, the two prime factors each impose their own condition independently, and the answer is a *union* of two arithmetic progressions of failures rather than one. The phrase "multiple of 2021" is doing real work: it is asking for two conditions at once, and the conditions do not merge.

That also explains the shape of the final answer and why it is a union rather than an intersection. $S(j)$ fails to be divisible by 2021 as soon as *either* prime fails. Failure is easy; success requires both. So the excluded set is $42\mathbb{Z}_{\geq 0}\cup 46\mathbb{Z}_{\geq 0}$.

Kedlaya's solution opens by stating the congruence $S(j)\equiv 46\sum_{n=1}^{42} n^j \pmod{43}$ with no derivation at all. Here is where that comes from. The Chinese Remainder bijection $A \leftrightarrow (\mathbb{Z}/43)^\times\times(\mathbb{Z}/47)^\times$ is not an observation about this problem; it is the definition of what $\gcd(n,2021)=1$ means once you know $2021 = 43\cdot 47$. And the multiplier 46 is the size of $(\mathbb{Z}/47)^\times$, the kernel of the projection — so it is the number of ways to complete a residue mod 43 to an element of A. The step is a fibre count, and fibre counts over a group homomorphism are always constant. Nothing is being noticed; something is being read off.

Finally, the reason the answer must be checked at $j=0$: the problem says *nonnegative* integer j, and 0 is divisible by everything. $S(0)=|A|=1932$, which is not a multiple of 2021 —

reassuringly consistent, but only because the phrase "divisible by 42" was taken to include $j = 0$. A solution that writes "$j \not\equiv 0 \pmod{42}$ and $j \not\equiv 0 \pmod{46}$" gets this right automatically; a solution that writes "j is not a positive multiple of 42 or 46" gets it wrong.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the number 2021 paired with $\gcd(n, 2021) = 1$. Write down two things immediately: $2021 = 43 \times 47$, and *A is the unit group* $(\mathbb{Z}/2021)^\times$. The second is the one that does the work; the first is only how you split it.

2. "Sum of jth powers over the units mod m" suggests: split by CRT into prime moduli, then use the primitive-root/character-orthogonality dichotomy on each prime separately. The method is *reduce, then apply a group symmetry to the sum* — never evaluate.

3. Hesitation occurs at the start, at the sight of $\sum_{n \in A} n^j$ with j unspecified. The instinct is that an unspecified exponent means there is nothing to compute, followed by an attempt to compute it anyway for small j. What reinforces it away is the realisation that the answer is not a value but a *criterion*, and criteria for a sum to vanish come from symmetry, not from formulae. A second hesitation: after proving the mod-43 and mod-47 conditions, wondering whether they interact. They do not — CRT makes them logically independent — and saying so explicitly is worth a line.

4. Three traps. (i) Treating 2021 as though it had a primitive root, giving the single false condition "$1932 \nmid j$" or "$966 \nmid j$". (ii) Forgetting $j = 0$, which is divisible by both 42 and 46 and must be excluded; "nonnegative" in the statement is deliberate. (iii) Dropping the multipliers 46 and 42 without checking they are units — here they are, but if the modulus had been p^2 the corresponding multiplier would be divisible by p and the argument would change completely.

5. Recognisable faster next time by the flag: *a sum indexed by all elements of a finite group*. That is never a computation. Write T for the sum, find a group element m whose translation permutes the index set, deduce $T = \chi(m)T$, and conclude $T = 0$ unless χ is trivial. Recognise the same shape in $\sum_k \omega^k = 0$ for nontrivial roots of unity, in $\sum_\chi \chi(g) = 0$ for $g \neq e$, and in "the average of a nonconstant homomorphism is zero".

6. The reusable decision rule: **before summing over a set, ask what group acts on that set, and whether the summand transforms simply under the action. If it does, the sum is determined by whether the transformation is trivial — a yes/no question — and you never evaluate anything.** Corollary for composite moduli: a condition modulo $m = p_1 \cdots p_k$ with distinct primes splits into k genuinely independent conditions, and the failure set is their *union*; do not expect them to combine into a single congruence unless $(\mathbb{Z}/m)^\times$ happens to be cyclic, which for m with two odd prime factors it never is.

▷ **Cue:** the problem's modulus is a specific composite number such as 2021, 2022, 1001 $\longrightarrow$ **Move:** factor it on sight and split every divisibility condition by the Chinese Remainder Theorem before doing anything else; $2021 = 43 \times 47$

▷ **Cue:** a set is defined as $\{n : 1 \leq n \leq m,\ \gcd(n, m) = 1\}$ $\longrightarrow$ **Move:** stop reading it as a range of integers and read it as the unit group $(\mathbb{Z}/m)^\times$; the problem is then about element orders, not about counting

▷ **Cue:** you must decide whether $\sum_{g \in G} \chi(g)$ vanishes, with χ multiplicative $\longrightarrow$ **Move:** multiply the index by a fixed $m \in G$: the sum is permuted, so $T = \chi(m)T$, hence $T = 0$ unless $\chi(m) = 1$ for all m, i.e. unless χ is trivial

▷ **Cue:** $\sum_{n=1}^{p-1} n^j \bmod p$ appears $\longrightarrow$ **Move:** quote the dichotomy: it is -1 when $(p-1) \mid j$ and 0 otherwise. Memorise this; it is the mod-p analogue of $\sum \omega^k$

▷ **Cue:** a sum over the units mod m must be reduced modulo a prime factor p of m $\longrightarrow$ **Move:** use that reduction is a surjective group homomorphism with constant fibre size $|(\mathbb{Z}/m)^\times|/(p-1)$; the sum becomes that constant times the sum over $(\mathbb{Z}/p)^\times$, and the constant can be discarded exactly when it is coprime to p

▷ **Cue:** the answer to "determine all j" is a divisibility condition and j is allowed to be 0 ⟶ **Move:** check $j = 0$ separately; 0 is divisible by every positive integer, so it lands in the excluded class, and the sum is just the cardinality of the index set

2021 A6

2021 A6. Let $P(x)$ be a polynomial whose coefficients are all either 0 or 1. Suppose that $P(x)$ can be written as a product of two nonconstant polynomials with integer coefficients. Does it follow that $P(2)$ is a composite integer?

Fields: algebra and complex analysis: root location for polynomials with restricted coefficients, factorisation over $\mathbb{Z}$, positivity of coefficients after a shift, intermediate value theorem **Position:** A6 — the hardest problem on the paper, and it deserves the label. The statement is one line, the answer is one word, and the proof needs a quantitative bound on the real parts of the roots. Expect this to take an hour **Answer:** Yes. If P factors as QR with Q, R nonconstant in $\mathbb{Z}[x]$, then $Q(2) \geq 2$ and $R(2) \geq 2$, so $P(2) = Q(2)R(2)$ is composite. (Recall 1 is neither prime nor composite; the statement would be false without the "nonconstant" hypothesis.)

Layer 1 THE WHAT — surface analysis

The natural attempt. The statement reads as a question about integers, so the natural attempt treats it as one. $P(2)$ is the integer whose binary digits are the coefficients of P; the hypothesis is that the *polynomial* factors; the conclusion is that the *number* factors. So the obvious move: if $P = QR$ then $P(2) = Q(2)R(2)$, and we are done as soon as neither factor is ± 1.

That is the right first line, and it is also where the natural attempt stalls, because $Q(2) = \pm 1$ is not obviously impossible. Q has integer coefficients but they need not be 0 or 1 — for instance $x^4 + x^3 + x + 1 = (x+1)^2(x^2 - x + 1)$, where the factor $x^2 - x + 1$ has a negative coefficient and evaluates to 3 at $x = 2$. Nothing so far forbids a factor like $x^3 - x^2 - x + 1$ hmm, or more to the point, a factor evaluating to 1: $Q(2) = 1$ would need the alternating-looking cancellation $\sum q_i 2^i = 1$, which is easy to arrange for an arbitrary integer polynomial ($Q = x^2 - 2x + 1$ gives $Q(2) = 1$). So the burden is real: we must use something about P to rule it out.

The natural attempt now tries small cases. Every $0/1$ polynomial of low degree that factors does give a composite value at 2; you check ten of them and believe the answer is yes. Then you try to prove it by congruences — "$Q(2)$ and $R(2)$ multiply to $P(2)$, and $P(2)$ has such-and-such binary form" — and nothing comes out, because the binary digits of $P(2)$ carry no information that survives multiplication. You try $P(1)$: $P(1)$ is the number of ones, $Q(1)R(1) = P(1)$, and again $Q(1) = 1$ is not excluded by anything you have.

The wall is this: *you have a hypothesis about the coefficients of P and you are trying to deduce something about the values of its factors, with no mechanism connecting the two.* Coefficients and values at a point are linked by exactly one thing — the roots — and until you go to the roots you are manipulating symbols with no leverage. This problem punishes symbol-shuffling more completely than almost any other on the paper.

The model chain. Write $P(x) = a_0 + a_1 x + \cdots + a_n x^n$ with each $a_i \in \{0, 1\}$ and $a_n = 1$ (the leading coefficient is nonzero, hence 1), so P is monic.

Step 0: P has no root in $[0, \infty)$. For real $t > 0$, $P(t) \geq t^n > 0$ since all coefficients are nonnegative and $a_n = 1$; and $P(0) = a_0 \in \{0, 1\}$, so if $P(0) = 0$ the root is 0 itself, not positive. In particular P has no root with $t \geq 1$.

Step 1: if $\alpha \neq 0$ is a root with $\mathrm{Re}(\alpha) \geq 0$, then $|\alpha| \leq \frac{1+\sqrt{5}}{2}$. Divide $P(\alpha) = 0$ by α^n:

$$1 + a_{n-1}\alpha^{-1} + a_{n-2}\alpha^{-2} + \cdots + a_0\alpha^{-n} = 0,$$

so

$$\left|1 + a_{n-1}\alpha^{-1}\right| = \left|a_{n-2}\alpha^{-2} + \cdots + a_0\alpha^{-n}\right| \le \sum_{k=2}^{n} |\alpha|^{-k}.$$

If $\mathrm{Re}(\alpha) \ge 0$ then $\mathrm{Re}(\alpha^{-1}) = \mathrm{Re}(\overline{\alpha})/|\alpha|^2 \ge 0$, so $\mathrm{Re}(1 + a_{n-1}\alpha^{-1}) \ge 1$ and hence the left side is at least 1. Write $t = 1/|\alpha|$. If $t \ge 1$ there is nothing to prove ($|\alpha| \le 1 < \varphi$). If $t < 1$,

$$1 \le \sum_{k=2}^{n} t^k < \frac{t^2}{1-t} \quad \Longrightarrow \quad t^2 + t - 1 > 0 \quad \Longrightarrow \quad t > \frac{\sqrt{5}-1}{2},$$

so $|\alpha| < \frac{2}{\sqrt{5}-1} = \frac{1+\sqrt{5}}{2} =: \varphi$.

Step 2: every root α satisfies $\mathrm{Re}(\alpha) < \frac{3}{2}$. Three cases.

- $\mathrm{Re}(\alpha) \le 0$: immediate.
- $\arg\alpha \in [-\pi/4, \pi/4]$ (so $\alpha \ne 0$): then $\arg(\alpha^{-1}) \in [-\pi/4, \pi/4]$ and $\arg(\alpha^{-2}) \in [-\pi/2, \pi/2]$, so $\mathrm{Re}(\alpha^{-1}) \ge 0$ and $\mathrm{Re}(\alpha^{-2}) \ge 0$. Since $a_{n-1}, a_{n-2} \ge 0$,

  $$\left|1 + a_{n-1}\alpha^{-1} + a_{n-2}\alpha^{-2}\right| \;\ge\; \mathrm{Re}\Big(1 + a_{n-1}\alpha^{-1} + a_{n-2}\alpha^{-2}\Big) \;\ge\; 1,$$

  while the same normalised equation gives

  $$\left|1 + a_{n-1}\alpha^{-1} + a_{n-2}\alpha^{-2}\right| = \left|a_{n-3}\alpha^{-3} + \cdots + a_0\alpha^{-n}\right| \le \sum_{k=3}^{n} t^k$$

  with $t = 1/|\alpha|$. If $t \ge 1$ then $\mathrm{Re}(\alpha) \le |\alpha| \le 1 < 3/2$. Otherwise $1 < \dfrac{t^3}{1-t}$, i.e. $t^3 + t - 1 > 0$. The polynomial $x^3 + x - 1$ has a unique positive root τ, and $\tau > 2/3$ because $(2/3)^3 + 2/3 - 1 = -1/27 < 0$. Hence $t > 2/3$ and $|\alpha| < 3/2$, so $\mathrm{Re}(\alpha) \le |\alpha| < 3/2$.
- Otherwise $\mathrm{Re}(\alpha) > 0$ and $\arg\alpha \in (-\pi/2, -\pi/4) \cup (\pi/4, \pi/2)$, so $\cos(\arg\alpha) < 1/\sqrt{2}$. Step 1 applies, giving

  $$\mathrm{Re}(\alpha) = |\alpha|\cos(\arg\alpha) < \frac{\varphi}{\sqrt{2}} = \frac{1+\sqrt{5}}{2\sqrt{2}} \approx 1.1441 < \frac{3}{2}.$$

Step 3: the shift. Suppose $P = QR$ with $Q, R \in \mathbb{Z}[x]$ nonconstant. Since P is monic, the leading coefficients of Q and R multiply to 1, so both are $+1$ or both are -1; replacing (Q, R) by $(-Q, -R)$ if necessary, take both monic.

Let $\widetilde{Q}(x) := Q(x + 3/2)$. Over $\mathbb{C}$, $\widetilde{Q}$ is a product of factors of two kinds:

$$x - \big(\alpha - \tfrac{3}{2}\big) = x + \big(\tfrac{3}{2} - \alpha\big) \quad (\alpha \text{ a real root of } Q),$$

$$\big(x + \tfrac{3}{2} - \alpha\big)\big(x + \tfrac{3}{2} - \overline{\alpha}\big) = x^2 + 2\,\mathrm{Re}\big(\tfrac{3}{2} - \alpha\big)\,x + \big|\tfrac{3}{2} - \alpha\big|^2 \quad (\alpha, \overline{\alpha} \text{ a nonreal pair}),$$

using that Q has real coefficients so nonreal roots pair with their conjugates. Every root α of Q is a root of P, so $\mathrm{Re}(\alpha) < 3/2$ by Step 2; hence $\frac{3}{2} - \alpha > 0$ in the first kind and both $2\,\mathrm{Re}(\frac{3}{2} - \alpha) > 0$ and $|\frac{3}{2} - \alpha|^2 > 0$ in the second. A product of polynomials all of whose coefficients are positive has all coefficients positive. So $\widetilde{Q}(x) = \sum_{i=0}^{d} c_i x^i$ with every $c_i > 0$ and $d = \deg Q \ge 1$.

Step 4: $Q(2) > Q(1)$. Note $Q(2) = \widetilde{Q}(1/2)$ and $Q(1) = \widetilde{Q}(-1/2)$: the shift $3/2$ is chosen because 1 and 2 are symmetric about it. Then

$$\widetilde{Q}(\tfrac{1}{2}) - \widetilde{Q}(-\tfrac{1}{2}) = 2\sum_{i \text{ odd}} c_i 2^{-i} \;\ge\; 2c_1 2^{-1} = c_1 > 0,$$

since $d \geq 1$ guarantees c_1 exists and is positive. So $Q(2) > Q(1)$.

Step 5: $Q(1) \geq 1$**.** $Q(1)$ is an integer. Q is monic, so $Q(t) \to +\infty$ as $t \to +\infty$. If $Q(1) \leq 0$, the intermediate value theorem would give a real root of Q in $[1, \infty)$, hence a root of P there — impossible by Step 0. So $Q(1) \geq 1$, and therefore the integer $Q(2)$ satisfies $Q(2) > Q(1) \geq 1$, i.e. $Q(2) \geq 2$.

Step 6. The same argument applies to R, so $R(2) \geq 2$. Hence $P(2) = Q(2)R(2)$ is a product of two integers each at least 2: it is composite. ■

Caveat. The constant $3/2$ is proved sufficient above, but it is not claimed to be sharp. A numerical scan of all $0/1$ polynomials of degree at most 14 gives a largest real part of a root of about 1.1323, attained near degree 14, which sits just below the bound $\varphi/\sqrt{2} \approx 1.1441$ from the third case of Step 2. So the true supremum of $\mathrm{Re}(\alpha)$ is very likely close to $\varphi/\sqrt{2}$ and certainly well below $3/2$; the argument does not need this and no claim about the exact supremum is made here.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths agree on the first line: $P = QR \Rightarrow P(2) = Q(2)R(2)$, so it suffices to show neither factor is ± 1 at $x = 2$. They split at (†): "show $Q(2) \neq \pm 1$." At (†) the natural attempt looks for an *arithmetic* reason — congruences, binary digits, size of $P(2)$ — and finds none, because the hypothesis "coefficients in $\{0, 1\}$" is not an arithmetic condition on $P(2)$. The model looks for an *analytic* reason: the coefficient condition constrains where the roots of P can be, the roots of Q are among them, and the location of Q's roots controls the *sign pattern* of Q's coefficients after a shift, which controls $Q(2)$ versus $Q(1)$.

> *The only hypothesis I have is about the coefficients of P, and the only thing I want is a lower bound on $Q(2)$ where Q is some unknown factor. Coefficients of P and values of Q have exactly one thing in common: the roots. P's coefficients constrain where its roots are, and Q's roots are a subset of P's. So the bridge has to be root location, and I should go and find out how far right a root of a $0/1$ polynomial can be.*

> *I want to prove $Q(2) > Q(1)$ and I know nothing about Q's coefficients — except that if every root had negative real part, then Q would have all-positive coefficients, and a polynomial with all-positive coefficients is bigger at $+t$ than at $-t$. The roots are not in the left half-plane, but they are all to the left of some vertical line. So shift that line to the origin.*

> *Where do I put the shift? I want 2 and 1 to become $+t$ and $-t$ for the same t, because that is the comparison a positive-coefficient polynomial wins automatically. So the shift is the midpoint of 1 and 2, namely $3/2$. The whole problem now reduces to: are all the roots to the left of $3/2$?*

> *To bound a root of $\sum a_i x^i$ I should divide by the leading term and look at $1 + a_{n-1}\alpha^{-1} + \cdots$: the tail is a geometric-type series in $1/|\alpha|$, and the head has real part at least 1 whenever α is in the right half-plane, because all the a_i are nonnegative. Nonnegativity of the coefficients is a statement about* directions, *so I should use real parts, not just absolute values.*

Layer 3 THE WHY

Why must the shift-to-$3/2$ thought be thought? Because of what $3/2$ *is*. It is not a tuned constant. It is the midpoint of 1 and 2.

Here is the chain of "of course". You want a lower bound on the integer $Q(2)$. The only integer values of Q you can get hold of cheaply are at small integers, and the only one you can bound from below without knowing Q is $Q(1)$: because Q is monic with no real root in $[1,\infty)$, the intermediate value theorem forces $Q(1) \geq 1$, and integrality upgrades "$Q(2) > Q(1)$" to "$Q(2) \geq 2$". So the target is fixed for you: prove $Q(2) > Q(1)$.

Now, when is a polynomial larger at one point than another, with no knowledge of its coefficients? There is exactly one free lunch: if all coefficients are positive, then $\sum c_i s^i > \sum c_i (-s)^i$ for any $s > 0$, because the even terms tie and the odd terms strictly favour $+s$. That inequality compares a point and its *reflection*. So if you are to use it, 1 and 2 must be reflections of each other — which pins the centre of reflection at $\frac{1+2}{2} = \frac{3}{2}$ and nothing else. *The constant* $3/2$ *is determined by the two points you were always going to compare.*

Having fixed the centre, the remaining question writes itself: when does $Q(x+3/2)$ have all coefficients positive? A monic real polynomial has all-positive coefficients as soon as all its roots lie in the open left half-plane, because each real linear factor $x + c$ and each conjugate-pair quadratic $x^2 + 2\mathrm{Re}(c)x + |c|^2$ then has positive coefficients, and positivity is preserved under multiplication. Translating back: $Q(x+3/2)$ has positive coefficients as soon as every root of Q has real part less than $3/2$. So the whole problem has been converted, with no cleverness at any step, into a single quantitative question: *how far to the right can a root of a* $0/1$ *polynomial be?*

And that question has an "of course" too. The hypothesis $a_i \in \{0,1\}$ is a hypothesis of *nonnegativity plus boundedness*. Boundedness (the a_i are at most 1) gives you $|a_{n-k}\alpha^{-k}| \leq |\alpha|^{-k}$, and a geometric series, and hence a bound on $|\alpha|$: this is the standard Cauchy root bound and it alone gives only $|\alpha| < 2$. Nonnegativity gives you something the standard bound throws away: all the terms $a_{n-k}\alpha^{-k}$ point into the same half-plane when α does. So instead of estimating $|1 + a_{n-1}\alpha^{-1}| \geq 1 - |\alpha|^{-1}$ (which is weak, and useless when $|\alpha|$ is near 1), you estimate $|1 + a_{n-1}\alpha^{-1}| \geq \mathrm{Re}(1 + a_{n-1}\alpha^{-1}) \geq 1$ (which is strong, and is exactly the improvement that brings 2 down past $3/2$). *Keeping one more term on the left is not a trick; it is the systematic way to convert "the coefficients are nonnegative" into "the roots avoid the right half-plane".* Keep two terms and you get the $3/2$; keep three and you would get a little more.

Kedlaya's solution opens by deriving $|\alpha| < (1+\sqrt{5})/2$ and then $\mathrm{Re}(\alpha) \leq 3/2$, and only afterwards introduces $Q(x+3/2)$. Read in that order, the shift by $3/2$ looks like a rabbit. Read in the order above — target $Q(2) > Q(1)$, so reflect about $3/2$, so demand roots left of $3/2$, so go and bound the roots — every constant in the proof is forced by the preceding line. The bound is the servant of the shift, not the other way round.

One more "why" worth having. Why does the problem ask about $x = 2$ specifically, and why must the factors be nonconstant? Because the theorem behind it is Brillhart–Filaseta–Odlyzko (1981): if a prime p is written in base $b \geq 2$ as $\sum a_i b^i$, then $\sum a_i x^i$ is irreducible over $\mathbb{Q}$. This problem is the base-2 contrapositive, where the digits are automatically 0 or 1. The statement "$P(2)$ is composite" really means "$P(2)$ is not 1 and not prime": 1 is excluded because $Q(2), R(2) \geq 2$ forces $P(2) \geq 4$, and the nonconstant hypothesis is what supplies both factors. Drop it and $P = 1 \cdot P$ proves nothing.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that the hypothesis and the conclusion live in different worlds — a coefficient condition on P, a factorisation statement about the integer $P(2)$. Write down at once: $P(2) = Q(2)R(2)$, so the whole problem is *show* $|Q(2)| \geq 2$; and write down the only object that coefficients and values share, *the roots*.

2. "Bound a value of an unknown factor" plus "sign condition on coefficients" suggests: locate the roots of P in the complex plane, shift the variable so that all roots lie in the left half-plane, exploit that the shifted polynomial has positive coefficients, and compare values at a point and its reflection. The auxiliary tools are the Cauchy root bound sharpened by real parts, and the intermediate value theorem plus integrality for the final step.

3. Hesitation occurs at the moment you realise that Q may have negative coefficients, so nothing

about P's $0/1$ structure is inherited by Q. That is the moment most solutions die. What reinforces it away is the observation that Q inherits something better than coefficients: it inherits *roots*. Any property of P expressible as root location passes to every factor for free. A second hesitation comes when the bound $|\alpha| < 2$ from the crude Cauchy estimate is not good enough; what reinforces that away is noticing that the crude estimate uses only $|a_i| \leq 1$ and discards $a_i \geq 0$, so there must be a sharper version.

4. Three traps. (i) Concluding from $P(2) = Q(2)R(2)$ that $P(2)$ is composite without excluding $Q(2) = \pm 1$ — this is the entire problem and a solution that skips it earns nothing. (ii) Assuming the factors Q, R also have $0/1$ coefficients; they need not, as $x^4 + x^3 + x + 1 = (x+1)^2(x^2 - x + 1)$ shows. (iii) Forgetting to normalise Q and R to be monic, or forgetting that $Q(2)$ could a priori be negative — the argument $Q(2) > Q(1) \geq 1$ handles both, but only once monicity is established from $a_n = 1$.

5. Recognisable faster next time by the slogan: *a sign condition on coefficients is a statement about where the roots are, and root conditions are inherited by factors.* Whenever a hypothesis restricts the coefficients (nonnegative, bounded, alternating, $0/1$) and the conclusion concerns an arbitrary divisor, translate the hypothesis into a region of the complex plane containing no roots — that translated form is what the divisor inherits.

6. The reusable decision rule: **to compare a polynomial's values at two points $u < v$ with no information about its coefficients, translate so the two points become $\pm\frac{v-u}{2}$, and then try to prove the translated polynomial has all-positive coefficients; that is equivalent to all roots lying left of the midpoint $\frac{u+v}{2}$.** Two corollaries worth carrying: (a) a real monic polynomial has all coefficients positive as soon as every root has negative real part, so "positive coefficients" and "roots in a half-plane" are interchangeable hypotheses; (b) when bounding roots of a polynomial with *nonnegative* coefficients, estimate Re of the leading block rather than its modulus — keeping k terms on the left and $n-k$ on the right, and increasing k, systematically sharpens the bound.

▷ **Cue:** a hypothesis restricts the coefficients of P but the conclusion concerns an arbitrary factor Q of P $\longrightarrow$ **Move:** translate the hypothesis into a statement about the location of the roots of P; root location is the only property a factor inherits automatically

▷ **Cue:** you must prove a polynomial is larger at v than at u, knowing nothing about its coefficients $\longrightarrow$ **Move:** shift by the midpoint $(u+v)/2$ so the two points become $\pm(v-u)/2$; if the shifted polynomial has positive coefficients the inequality is immediate from the odd-degree terms

▷ **Cue:** you want a real monic polynomial to have all-positive coefficients $\longrightarrow$ **Move:** prove every root has negative real part; each real factor $x + c$ and each conjugate pair $x^2 + 2\mathrm{Re}(c)x + |c|^2$ is then positive, and positivity survives multiplication

▷ **Cue:** bound the roots of $\sum a_i x^i$ with all $a_i \geq 0$ and $a_n = 1$ $\longrightarrow$ **Move:** normalise to $1 + a_{n-1}\alpha^{-1} + \cdots + a_0\alpha^{-n} = 0$, keep the first two or three terms on the left, and bound them below by their *real part* (which is ≥ 1 in the right half-plane) rather than by a modulus estimate; the rest is a geometric series in $1/|\alpha|$

▷ **Cue:** you need an integer value of a monic integer polynomial to be at least 2 $\longrightarrow$ **Move:** show it exceeds its value at a smaller integer point and use the intermediate value theorem to force that smaller value to be at least 1 (no real root to the right); integrality converts the strict inequality into ≥ 2

▷ **Cue:** a polynomial with nonnegative coefficients and nonzero constant or leading term $\longrightarrow$ **Move:** it has no root in $(0, \infty)$ — a one-line fact that is repeatedly the hinge of an intermediate value theorem argument

Putnam 2021 — Paper B

2021 B1

2021 B1. Suppose that the plane is tiled with an infinite checkerboard of unit squares. If another unit square is dropped on the plane at random with position and orientation independent of the checkerboard tiling, what is the probability that it does not cover any of the corners of the squares of the checkerboard?

Fields: geometric probability, integral geometry **Position:** B1, first problem of the paper **Answer:** $2 - \frac{6}{\pi}$

Layer 1 THE WHAT — surface analysis

The natural attempt. Set up coordinates with the checkerboard fixed and axis-aligned: corners at every point of $\mathbb{Z}^2$. By periodicity you can reduce the dropped square's centre to a point in a single unit cell $[0,1)^2$, and by the checkerboard's own symmetry you can reduce the rotation angle θ of the dropped square to $[0, \pi/2)$. Since the dropped square has half-diagonal $\sqrt{2}/2 \approx 0.707 < 1$, only the four corners of the cell it lies in can possibly be covered — that much is a clean and correct first cue. The natural next step is to write down, for fixed θ, the four "forbidden regions" (the centres for which the dropped square would cover each of the four corners) directly in the checkerboard's own axes. Each forbidden region is itself a unit square, but tilted by θ relative to the axes you are working in. You now need the area of the union of four tilted squares, one per corner, and that area depends on exactly how they overlap each other — which pairs overlap, and by how much, changes as θ crosses $\pi/4$. You can push this through for a fixed value of θ (say $\theta = \pi/6$) by intersecting the relevant half-planes by hand, but you are solving a new case-work problem for every value of θ, with no way to write a single formula for the area as a function of θ without first sorting the overlap pattern into sub-cases. This is not a dead end — it is recoverable — but it is expensive, and expensive in a way that does not visibly reduce as you work.

The model chain. Rotate the frame by $-\theta$ so that the *dropped* square is axis-aligned, with centre C' and half-width $1/2$ in both new coordinates. The background cell S (still the unit square with corners $(0,0), (1,0), (0,1), (1,1)$) becomes a tilted unit square with corners

$$V_1 = (0,0),\ V_2 = (\cos\theta, -\sin\theta),\ V_3 = (\sin\theta, \cos\theta),\ V_4 = (\cos\theta + \sin\theta, \cos\theta - \sin\theta).$$

In this frame, "the dropped square covers corner V_i" means $|C'_x - (V_i)_x| \le \frac{1}{2}$ *and* $|C'_y - (V_i)_y| \le \frac{1}{2}$ — an axis-aligned condition, because it is a statement about the dropped square's own sides. So each forbidden region U_i is now an ordinary axis-aligned unit square centred at V_i, and all four share the same axes.

Let S_θ be the axis-aligned bounding box of the tilted square (the V_i span $x \in [0, \sin\theta + \cos\theta]$, $y \in [-\sin\theta, \cos\theta]$), a square of side $\sin\theta + \cos\theta$. Let S'_θ be S_θ inset by $\frac{1}{2}$ on every side, a square of side $\sin\theta + \cos\theta - 1$ (positive for $\theta \in (0, \pi/2)$). *Check the easy direction directly*: take $C' = (x, y)$ with $x \in [\frac{1}{2}, \sin\theta + \cos\theta - \frac{1}{2}]$, $y \in [-\sin\theta + \frac{1}{2}, \cos\theta - \frac{1}{2}]$. Then $|x - 0| \ge \frac{1}{2}$ and

$|x-(\sin\theta+\cos\theta)|\geq\frac{1}{2}$, ruling out U_1 and U_4 on the x-coordinate alone; and $|y-(-\sin\theta)|\geq\frac{1}{2}$, $|y-\cos\theta|\geq\frac{1}{2}$ rule out U_2 and U_3 on the y-coordinate alone. So every point of S'_θ avoids all four forbidden squares. The reverse containment — that every point of $S_\theta\setminus S'_\theta$ *is* caught by some U_i — is proved by dissecting S_θ into S'_θ together with four $\sin\theta\times\cos\theta$ rectangles, one along each side, and checking that each V_i lands as an interior point of one rectangle's side, so that (using $\sin\theta,\cos\theta\leq 1$) the corresponding U_i covers that entire rectangle.

> **Caveat.** I verified the first (easy) containment directly above and spot-checked the second with explicit points near several edges of S_θ; I have not re-derived the full case-by-case dissection independently of Kedlaya–Bhargava–Ng, so I present it as their construction rather than a fully re-proved lemma.

Granting the lemma, the conditional probability that the dropped square covers no corner, given angle θ, is $(\sin\theta+\cos\theta-1)^2$. Since θ is uniform on $[0,\pi/2]$ (density $2/\pi$), the answer is

$$\frac{2}{\pi}\int_0^{\pi/2}(\sin\theta+\cos\theta-1)^2\,d\theta.$$

Expand: $(\sin\theta+\cos\theta-1)^2=(\sin\theta+\cos\theta)^2-2(\sin\theta+\cos\theta)+1=2+\sin 2\theta-2\sin\theta-2\cos\theta$. An antiderivative is $2\theta-\frac{1}{2}\cos 2\theta+2\cos\theta-2\sin\theta$. At $\theta=\pi/2$ this is $\pi+\frac{1}{2}+0-2=\pi-\frac{3}{2}$; at $\theta=0$ it is $-\frac{1}{2}+2-0=\frac{3}{2}$. The definite integral is $\pi-3$, so the probability is $\frac{2}{\pi}(\pi-3)=2-\dfrac{6}{\pi}$. (I carried out this integral myself and it matches the value on record.)

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree on the reduction to one cell, one angle range, and (crucially) that only four corners matter. They split at the sentence: "each forbidden region is a unit square tilted by θ *relative to the axes I am using*." The natural attempt keeps the checkerboard's axes fixed, so every forbidden region is tilted and their pairwise overlaps must be worked out case by case. The model chain rotates so the *dropped* square is axis-aligned instead, and every forbidden region becomes axis-aligned along with it.

> *I don't have to keep the grid's axes fixed just because the grid was drawn first in the problem — the "covers a corner" condition is defined by the dropped square's own two independent side-directions, so if I put my coordinate axes on the dropped square instead of on the background, that condition becomes a plain product of two one-dimensional inequalities, and unions of axis-aligned squares are things I already know how to dissect.*

Layer 3 THE WHY

The condition "the dropped square covers point P" is, by definition, $|C'_x-P'_x|\leq\frac{1}{2}$ and $|C'_y-P'_y|\leq\frac{1}{2}$ in the coordinate system whose axes run along the dropped square's own sides — it is a conjunction of two independent one-dimensional conditions only in that frame. Read in any other frame (in particular the checkerboard's frame), the same set of points is still a square, but it is no longer bounded by lines parallel to your coordinate axes, and the product structure — the very thing that lets you dissect a union of such squares into a bounding box minus an inset box plus edge rectangles — is gone. So the choice of frame is not a cosmetic convenience; it determines whether the covering condition is legible at all. This is the same move as "a circle is the set of points equidistant from its centre, so draw the line of centres": here, a square's defining property is stated in terms of *its own* sides, so you align to the square whose sides define the condition you are trying to characterise, not to whichever object the problem happened to introduce first.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the mismatch between the dropped square's reach ($\sqrt{2}/2 < 1$) and the lattice spacing (1): write down immediately that only finitely many — here exactly four — lattice points can possibly be covered, before attempting any area computation.
2. That cue suggests the standard reduction to a conditional probability given the angle θ, followed by an integral over θ; it does not by itself tell you which frame to compute the area in.
3. Hesitation occurs exactly when you have four forbidden regions and must decide whose axes to use for the case-work; it is reinforced away by asking "which object's own geometry defines the covering condition?" and aligning to that object, even though it is the one described as "dropped" rather than "fixed."
4. The trap is letting the order of the problem's prose (grid described first, dropped square second) dictate which object stays axis-aligned in your working. The two squares play symmetric geometric roles; the prose order carries no mathematical information.
5. Recognise this family faster by asking, whenever a problem defines "covers/overlaps" via an object's own sides or own boundary, whether your current coordinate frame is aligned to that specific object — if not, rotate to it before doing any case analysis.
6. Reusable rule: in a "does a moving axis-defined shape cover a feature of a fixed background" probability problem, choose coordinates aligned to whichever shape's own boundary defines the covering condition (often the moving one, not the background), reducing a union of oddly-tilted forbidden regions to a union of shapes you can dissect via bounding-box-minus-inset.

2021 B2

2021 B2. Determine the maximum value of the sum

$$S = \sum_{n=1}^{\infty} \frac{n}{2^n} (a_1 a_2 \cdots a_n)^{1/n}$$

over all sequences $a_1, a_2, a_3, \cdots$ of nonnegative real numbers satisfying

$$\sum_{k=1}^{\infty} a_k = 1.$$

Fields: analysis (AM-GM optimisation, series manipulation) **Position:** B2, second problem of the paper
Answer: $2/3$, attained at $a_k = 3/4^k$

Layer 1 THE WHAT — surface analysis

The natural attempt. The objective mixes a geometric mean $(a_1 \cdots a_n)^{1/n}$ with a linear constraint $\sum a_k = 1$; the standard reflex is AM-GM, and the standard *default* form of AM-GM is unweighted: $(a_1 \cdots a_n)^{1/n} \le \frac{a_1 + \cdots + a_n}{n}$. Apply it term by term:

$$S \le \sum_{n=1}^{\infty} \frac{n}{2^n} \cdot \frac{a_1 + \cdots + a_n}{n} = \sum_{n=1}^{\infty} \frac{a_1 + \cdots + a_n}{2^n}.$$

Swap the order of summation ($\sum_k a_k \sum_{n \ge k} 2^{-n} = \sum_k a_k \cdot 2^{1-k}$, using $\sum_{n \ge k} 2^{-n} = 2^{1-k}$):

$$S \le \sum_{k=1}^{\infty} a_k \, 2^{1-k}.$$

Since 2^{1-k} is largest at $k = 1$ (where it equals 1) and $\sum a_k = 1$, this gives $S \leq 1$. This bound is valid, and it took a genuinely standard first move to get it. But check whether it is achievable: equality in the unweighted AM-GM step needs $a_1 = a_2 = \cdots = a_n$ for every n *simultaneously*, forcing all a_k equal — impossible for an infinite sum of nonnegative terms with $\sum a_k = 1$ unless they are all 0, a contradiction. So the bound 1 is provably not attained, and worse, this computation gives no candidate sequence at all — try the obvious guess $a_1 = 1$, rest 0: then $S = \frac{1}{2} \cdot 1^{1/1} \cdot 1 = \frac{1}{2}$, nowhere near 1. The unweighted approach is not wrong, but it is loose, and loose in a way that hides the true extremiser rather than pointing toward it.

The model chain. Introduce weights $w_k = 4^k$ inside the AM-GM step instead of leaving it unweighted:

$$2^{n+1}(a_1 \cdots a_n)^{1/n} = \big((4a_1)(4^2 a_2) \cdots (4^n a_n)\big)^{1/n} \leq \frac{1}{n} \sum_{k=1}^{n} 4^k a_k,$$

using $\prod_{k=1}^{n} 4^k = 4^{n(n+1)/2}$ so that $\big(\prod 4^k a_k\big)^{1/n} = 4^{(n+1)/2}(a_1 \cdots a_n)^{1/n} = 2^{n+1}(a_1 \cdots a_n)^{1/n}$. Divide by 2^{n+1}:

$$(a_1 \cdots a_n)^{1/n} \leq \frac{1}{n \cdot 2^{n+1}} \sum_{k=1}^{n} 4^k a_k.$$

Multiply by $n/2^n$ and sum over n:

$$S \leq \sum_{n=1}^{\infty} \frac{1}{2^{2n+1}} \sum_{k=1}^{n} 4^k a_k = \frac{1}{2} \sum_{n=1}^{\infty} \frac{1}{4^n} \sum_{k=1}^{n} 4^k a_k.$$

Swap the order of summation:

$$2S \leq \sum_{k=1}^{\infty} a_k \sum_{n \geq k} 4^{k-n} = \sum_{k=1}^{\infty} a_k \sum_{m \geq 0} 4^{-m} = \sum_{k=1}^{\infty} a_k \cdot \frac{4}{3} = \frac{4}{3},$$

using $\sum a_k = 1$. Hence $S \leq \frac{2}{3}$. Equality requires the weighted AM-GM to be tight for every n simultaneously, i.e. $4a_1 = 4^2 a_2 = \cdots = 4^n a_n$ for all n, i.e. $4^k a_k = \lambda$ constant, i.e. $a_k = \lambda/4^k$. The constraint $\sum a_k = 1$ gives $\lambda \sum (1/4)^k = \lambda \cdot \frac{1/4}{1-1/4} = \lambda/3 = 1$, so $\lambda = 3$ and $a_k = 3/4^k$. This sequence is admissible (nonnegative, sums to 1) and one checks directly that it makes every AM-GM step above an equality, so $S = 2/3$ is attained and the maximum is exactly $2/3$. (I re-derived every algebraic step above myself, including the $\prod 4^k = 2^{n(n+1)}$ identity and both order-of-summation swaps, and they check out.)

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains apply AM-GM to the geometric mean term by term and then swap the order of summation over n and k — that structural skeleton is shared. They split at the line "apply AM-GM": the natural attempt uses the default, unweighted form (nothing tells you to weight it, so you don't); the model chain inserts weights $w_k = 4^k$ before applying it.

> *The unweighted bound I got is loose and doesn't correspond to any sequence I can write down — instead of trusting the default form of AM-GM, let me guess what the extremal a_k* should *look like (a geometric sequence, since the objective and constraint are both built from geometric-flavoured prefactors $n/2^n$ and an infinite sum), and then choose AM-GM's weights specifically so that this guessed sequence is exactly the equality case, solving backward for what the weights must be.*

Layer 3 THE WHY

AM-GM has a free parameter — any positive weights $w_1, \ldots, w_n$ give a valid inequality $(\prod w_k a_k)^{1/n} \leq \frac{1}{n} \sum w_k a_k$ with equality exactly when every $w_k a_k$ is equal. The unweighted

version is not more "natural" than any other choice; it is simply the choice that requires no decision. Once you notice that the equality condition is "$w_k a_k$ constant," two things pin the weights down: first, the equality case should be a geometric sequence $a_k = \lambda r^k$ for some ratio r (a reasonable guess from the shape of the problem — an infinite constraint $\sum a_k = 1$ combined with an infinite sum with a fixed exponential prefactor $n/2^n$ is the standard setting where geometric extremisers occur), which forces $w_k \propto r^{-k}$; second, the resulting double sum $\sum_n (\cdot) \sum_{k \le n} w_k a_k$ must telescope, after swapping order, into a *pure multiple* of $\sum_k a_k$ — and that telescoping only closes into a finite geometric series if w_k and the prefactor 2^{-n} share a common geometric ratio. Solving $w_k = 4^k$ is what is forced once you insist on both of these simultaneously; it is not a guess pulled from nowhere, it is the unique choice consistent with the equality case you conjectured and with the sum actually closing. This is the general lesson: when an inequality tool has a free parameter, use the problem's own structure (the shape of the objective, the shape of the constraint) to conjecture the equality case first, and only then solve backward for the parameter that makes the tool exact there.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the combination of a geometric mean $(a_1 \cdots a_n)^{1/n}$ with a linear constraint $\sum a_k = 1$: write "AM-GM" down immediately, but also write down that AM-GM has a free choice of weights before committing to the unweighted default.
2. That cue suggests weighted AM-GM as the method, with the weights left as unknowns to be solved for rather than fixed in advance.
3. Hesitation occurs right after "apply AM-GM," when it isn't obvious which weights to use; it is reinforced away by conjecturing the equality case first (here, a geometric sequence, suggested by the exponential prefactors already present in the problem) and solving backward for the weights that make that sequence exact.
4. The trap is the unweighted bound: it is a genuinely valid inequality ($S \le 1$), it comes from zero extra thought, and it is loose enough to produce no usable extremal example, which can read as "the problem is just hard" rather than "the wrong weights were chosen."
5. Recognise this family faster by checking, the moment you reach for AM-GM (or any inequality with adjustable weights/multipliers), whether the problem's own structure already hints at a plausible equality case — a geometric sequence, an arithmetic one, a specific symmetric point — before you fix a default, unweighted version.
6. Reusable rule: for a $\sum f(n) \cdot (\text{geometric mean over } n \text{ terms})$ problem, apply weighted AM-GM with weights chosen so that (i) the weighted terms are constant at your conjectured extremiser, and (ii) the resulting double sum, after swapping summation order, telescopes into a pure multiple of the linear constraint — solve for the weights that satisfy both simultaneously rather than defaulting to equal weights.

2021 B3

2021 B3. Let $h(x, y)$ be a real-valued function that is twice continuously differentiable throughout $\mathbb{R}^2$, and define

$$\rho(x, y) = y h_x - x h_y.$$

Prove or disprove: For any positive constants d and r with $d > r$, there is a circle $\mathcal{S}$ of radius r whose center is a distance d away from the origin such that the integral of ρ over the interior of $\mathcal{S}$ is zero.

Fields: vector calculus, integral geometry (existence via averaging + IVT) **Position:** B3, third problem of the paper **Answer:** The statement is true: such a circle always exists

Layer 1 THE WHAT — surface analysis

The natural attempt. Write $x = s\cos\varphi$, $y = s\sin\varphi$ (polar coordinates centred at the origin)

and compute $\partial h/\partial\varphi = -s\sin\varphi\, h_x + s\cos\varphi\, h_y = -(yh_x - xh_y) = -\rho$. So $\rho = -\partial h/\partial\varphi$: a clean, memorable recognition — ρ is (minus) the angular derivative of h about the origin. This immediately gives a true and useful fact for free: on any full circle *centred at the origin*, $\oint \rho\, d\varphi = -[h(s,2\pi) - h(s,0)] = 0$, since $(s,0)$ and $(s,2\pi)$ are the same point and h is single-valued. It is tempting to try to convert this directly into the answer: perhaps test simple functions and look for a pattern. Try $h(x,y) = x^2 + y^2$: then $h_x = 2x, h_y = 2y$, so $\rho = 2xy - 2xy = 0$ identically, and the integral is trivially 0 over *every* circle — a true but useless test case, since $\rho \equiv 0$ hides all the structure of the problem. Try $h(x,y) = x$: then $\rho = y$, and $\iint_S y\, dA$ over a disc of radius r centred at $(d\cos\varphi_0, d\sin\varphi_0)$ is (average value of y on the disc) times (area), i.e. $\pi r^2 \cdot d\sin\varphi_0$ — zero exactly when $\varphi_0 = 0$ or π, i.e. when the circle's centre lies on the x-axis. This is a genuine, fully checkable instance of the claim, and it suggests a natural guess: perhaps the required circle is always the one centred on some axis of symmetry of h. But for a general h with no symmetry at all — say $h(x,y) = x^3$, where $\rho = 3x^2y$ has no visible axis to aim at — there is no candidate circle to write down, and no way to compute $\iint_S \rho$ in closed form for an arbitrary C^2 function h. The natural attempt has correctly identified the operator and found instructive special cases, but it has no route from "this fact is true for circles centred at the origin" to "some specific off-centre circle at distance d also has this property," because a general h offers no symmetry to exploit and no formula to solve.

The model chain. Fix r, d. Every admissible circle is determined by the angular position $\varphi_0 \in [0, 2\pi)$ of its centre $(d\cos\varphi_0, d\sin\varphi_0)$; call it $\mathcal{S}_{\varphi_0}$ and set $J(\varphi_0) = \iint_{\mathcal{S}_{\varphi_0}} \rho\, dA$. The goal is to show $J(\varphi_0) = 0$ for some φ_0.

Compute $\int_0^{2\pi} J(\varphi_0)\, d\varphi_0$. In polar coordinates (s,θ) about the origin, a point of $\mathcal{S}_0$ (the circle at $\varphi_0 = 0$) at radius $s \in [d-r, d+r]$ occupies an angular range $\theta \in [\theta_1(s), \theta_2(s)]$; rotating the whole picture by φ_0 sends this point to angle $\theta + \varphi_0$ at the same radius s. So

$$J(\varphi_0) = \int_{d-r}^{d+r} \int_{\theta_1(s)}^{\theta_2(s)} \rho(s, \theta+\varphi_0)\, s\, d\theta\, ds,$$

and

$$\int_0^{2\pi} J(\varphi_0)\, d\varphi_0 = \int_{d-r}^{d+r} s \left(\int_{\theta_1(s)}^{\theta_2(s)} d\theta \right) \left(\int_0^{2\pi} \rho(s,\psi)\, d\psi \right) ds,$$

after substituting $\psi = \theta + \varphi_0$ (the inner φ_0-integral over a full period does not depend on the fixed shift θ, so it factors out of the θ-integral). But $\int_0^{2\pi} \rho(s,\psi)\, d\psi = -\int_0^{2\pi} \partial h/\partial\psi\,(s,\psi)\, d\psi = -[h(s,2\pi) - h(s,0)] = 0$ for every s, by the same single-valuedness argument as in the natural attempt — this is the origin-centred fact, now applied at each fixed radius s inside the family average rather than to one specific circle. Hence the whole triple integral is 0: $\int_0^{2\pi} J(\varphi_0)\, d\varphi_0 = 0$.

Finally, J is continuous in φ_0: $\rho = yh_x - xh_y$ is continuous (as $h \in C^2$), and the disc $\mathcal{S}_{\varphi_0}$ varies continuously with φ_0, so $J(\varphi_0)$ varies continuously (the symmetric difference of $\mathcal{S}_{\varphi_0}$ and $\mathcal{S}_{\varphi_0+\epsilon}$ has area $\to 0$ as $\epsilon \to 0$, and ρ is bounded on any compact set containing all the relevant discs). A continuous function with average value 0 over $[0, 2\pi]$ cannot be strictly positive everywhere or strictly negative everywhere, so it attains 0 somewhere — by the intermediate value theorem if it changes sign, or trivially if $J \equiv 0$. Hence some φ_0 gives $J(\varphi_0) = 0$, exactly as claimed. (I re-derived the swap of integration order and the vanishing of $\int_0^{2\pi} \rho(s,\psi)\, d\psi$ myself from the definition of ρ, and cross-checked the whole argument against the concrete case $h = x$, where the explicit answer $\varphi_0 \in \{0, \pi\}$ matches $J(\varphi_0) = \pi r^2 d\sin\varphi_0$ having exactly those zeros.)

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains reach the identity $\rho = -\partial h/\partial\varphi$ and the fact that this vanishes on integration around any *full circle centred at the origin*. They split immediately afterward, at the sentence "but $\mathcal{S}$ is not centred at the origin." The natural attempt tries to repair this mismatch by finding extra structure *in* h (a symmetry axis, a special form) that would let the origin-centred

fact reach the one specific circle $\mathcal{S}$ directly. The model chain does not try to fix $\mathcal{S}$; it integrates the same quantity over the entire family of rotated copies of $\mathcal{S}$, restoring origin-centred symmetry not to the circle but to the calculation.

> *I don't need to make this one specific off-centre circle inherit the origin's cancellation property directly — that property belongs to full rotations about the origin, not to any single circle sitting off to the side. So instead of asking "does $\mathcal{S}$ have it," let me integrate my target quantity over the whole family of rotated copies of $\mathcal{S}$, since that family, as a whole, does have the origin's rotational symmetry even though no individual member of it does; only afterward do I ask which member the resulting zero-average belongs to.*

Layer 3 THE WHY

The origin-centred cancellation $\oint \rho \, d\varphi = 0$ costs nothing: it needs no hypothesis on h beyond being single-valued (continuity is enough), because it is really just the fundamental theorem of calculus applied around a closed loop in the angular variable — the loop returns to its starting point, so any total-derivative-in-φ integrates to zero over it, for *every* C^1 function, symmetric or not. That is precisely why it is the wrong kind of fact to go looking for extra symmetry in h to exploit: the fact is not conditional on h's symmetry, it is conditional on the *loop's* symmetry (being centred at the origin, so that a full sweep in φ closes up). A single off-centre circle is not such a loop about the origin, and no property of h can make it one. But the *family* of all rotations of that circle about the origin is, collectively, invariant under rotation about the origin — and when you integrate over the whole family and swap the order of integration, each fixed radius s gets swept through its full 2π range exactly once, reconstructing the origin-centred loop that the individual circle lacked. The zero, in other words, belongs to the family, not to any one member of it; the intermediate value theorem is then the tool that always transfers a zero-average, continuous-in-the-parameter statement down to a zero at some particular parameter value. This is the general strategy: when a quantity has an exact vanishing identity under some group of symmetries (here, rotation about the origin) but the specific object you are given does not itself possess that symmetry, average the quantity over the group's orbit of that object — the average inherits the symmetry even when the object does not — and then descend from "zero on average" to "zero somewhere" via continuity and IVT.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the exact combination $yh_x - xh_y$: recognise it on sight as (minus) the angular derivative of h in polar coordinates about the origin, and write that identity down before doing anything else.

2. That cue suggests the free fact "integrating ρ around any full circle centred at the origin gives zero" (a closed-loop total derivative), which is the seed of an existence argument whenever the actual object of interest (here, an off-centre circle) is only a piece of such a loop.

3. Hesitation occurs right after noticing the target circle is not centred at the origin, when it is tempting to search for special structure in h (symmetry, a nice closed form) to bridge the gap directly; it is reinforced away by remembering that existence claims over a parametrised family are usually attacked by averaging over the family plus the intermediate value theorem, which needs no structure in h at all.

4. The trap is testing the claim on a highly symmetric h (like $h = x^2 + y^2$, giving $\rho \equiv 0$) first: it "confirms" the statement so cheaply that it offers no insight into why it holds for an arbitrary h, and can leave the false impression that some symmetry of h is doing the real work.

5. Recognise this family faster by asking, whenever a problem wants you to prove "some member of a rotationally- or translationally-parametrised family of regions satisfies [integral]$= 0$," whether the integrand has a known zero-average identity over the full symmetry group — if so, integrate over the family, swap the order of integration to expose that identity, and finish with continuity plus the intermediate value theorem.

6. Reusable rule: existence by averaging — to show some member of a continuous family of configurations satisfies an equality that the family only satisfies in aggregate, prove the equality holds on average across the whole family (typically via a symmetry/Fubini swap that reduces to a known vanishing identity), then upgrade "true on average" to "true for some member" using continuity in the parameter and the intermediate value theorem.

2021 B4

2021 B4. Let $F_0, F_1, \ldots$ be the sequence of Fibonacci numbers, with $F_0 = 0$, $F_1 = 1$, and $F_n = F_{n-1} + F_{n-2}$ for $n \geq 2$. For $m > 2$, let R_m be the remainder when the product $\prod_{k=1}^{F_m - 1} k^k$ is divided by F_m. Prove that R_m is also a Fibonacci number.

Fields: number theory (Wilson's theorem, quadratic residues); Fibonacci identities **Position:** B4 — the first genuinely hard problem on the B paper, and the one whose difficulty is concentrated entirely in a single decision about *which* quantity to compute **Answer:** $R_m = 0 = F_0$ when F_m is composite, and $R_m \in \{F_{m-1}, F_{m-2}\}$ when F_m is prime, because $A := \prod_{k=1}^{F_m - 1} k^k$ satisfies $A^2 \equiv -1$ and Cassini's identity says F_{m-1} is already a square root of -1 modulo F_m.

Layer 1 THE WHAT — surface analysis

The natural attempt. You compute. $m = 3$: $F_3 = 2$, the product is $1^1 = 1$, so $R_3 = 1 = F_1$. $m = 4$: $F_4 = 3$, the product is $1 \cdot 4 = 4 \equiv 1$, so $R_4 = 1$. $m = 5$: $F_5 = 5$, the product is $1 \cdot 4 \cdot 27 \cdot 256$, which mod 5 is $1 \cdot 4 \cdot 2 \cdot 1 = 8 \equiv 3 = F_4$. $m = 6$: $F_6 = 8$ and the product contains $2^2 \cdot 4^4$, so $R_6 = 0 = F_0$. $m = 7$: $F_7 = 13$ and the product comes out to $8 = F_6$.

So the composite case is free — if $F_m = ab$ with $1 < a, b < F_m$ then a^a and b^b both sit inside the product, and (taking $a = b$ if F_m is a square, where $a^2 \mid a^a$ since $a \geq 2$) the product is divisible by F_m, giving $R_m = 0 = F_0$. That case takes two lines and you will get it.

The prime case is where the problem lives, and here the default is to try to *evaluate* $A := \prod_{k=1}^{p-1} k^k \pmod p$ where $p = F_m$. The professional tool for a product over all of $(\mathbb{Z}/p)^\times$ is the discrete logarithm. Fix a primitive root g and write each k as $g^{\ell(k)}$. Then

$$A = \prod_{k=1}^{p-1} k^k = g^{\sum_{k=1}^{p-1} k\,\ell(k)},$$

and the exponent must be read modulo $p - 1$. Now try to evaluate $\sum_k k\,\ell(k)$. Reindexing by $i = \ell(k)$, so that k is the integer in $[1, p-1]$ representing g^i, this is $\sum_{i=0}^{p-2} i\,\langle g^i \rangle$ where $\langle \cdot \rangle$ means "the representative in $[1, p-1]$". You would like to hit this with the standard identity for $\sum i x^i$. You cannot. The identity for $\sum i x^i$ lives modulo p, where g^i makes sense as a residue; the exponent sum has to be known modulo $p - 1$, where the integer representatives $\langle g^i \rangle$ are an essentially random-looking sequence with no algebraic structure at all. That is not a hard computation — it is a computation with no handle anywhere on it. Walk far enough down this road and you are trying to evaluate a character sum that genuinely has no closed form.

A second default is worse: try to determine *which* Fibonacci number R_m is. The data refuses to cooperate. $m = 5$ gives F_{m-1}, $m = 7$ gives F_{m-1}, $m = 11$ gives $F_{m-2} = 34$, $m = 13$ gives $F_{m-1} = 144$, $m = 17$ gives $F_{m-1} = 987$. There is no visible rule, and you can burn twenty minutes looking for one — while the problem, read carefully, never asked.

The model chain. Dispose of the composite case as above: $R_m = 0 = F_0$. For $m \geq 5$, if m is even then $F_m = F_{m/2}(F_{m/2+1} + F_{m/2-1})$ with both factors exceeding 1, so F_m is composite and we are done. Check $m = 3, 4$ by hand: $R_3 = R_4 = 1 = F_1$. So assume $m \geq 5$ is odd and $p := F_m$ is prime, and write $\equiv$ for congruence mod p.

Reflect the product. Replacing k by $p-k$ permutes $\{1,\dots,p-1\}$, so

$$A=\prod_{k=1}^{p-1}(p-k)^{p-k}\equiv\prod_{k=1}^{p-1}(-k)^{p-k}=(-1)^{\sum_{k=1}^{p-1}(p-k)}\prod_{k=1}^{p-1}k^{p-k}=(-1)^{p(p-1)/2}\prod_{k=1}^{p-1}k^{p-k}.$$

The two expressions for A have exponents k and $p-k$ on the same base. Multiply them:

$$A^2\equiv(-1)^{p(p-1)/2}\prod_{k=1}^{p-1}k^k k^{p-k}=(-1)^{p(p-1)/2}\prod_{k=1}^{p-1}k^p=(-1)^{p(p-1)/2}\big((p-1)!\big)^p.$$

By Wilson's theorem $(p-1)!\equiv-1$, so $((p-1)!)^p\equiv(-1)^p$ and

$$A^2\equiv(-1)^{p(p-1)/2+p}=(-1)^{p(p+1)/2}.$$

Now pin the sign. Fibonacci modulo 4 is periodic with period 6: $0,1,1,2,3,1,0,1,1,2,3,1,\dots$, so for odd m we get $F_m\equiv 1$ or $2 \pmod 4$. If $p\equiv 1 \pmod 4$ then p is odd and $(p+1)/2$ is odd, so $p(p+1)/2$ is odd. If $p\equiv 2 \pmod 4$ then $p=2$, and $p(p+1)/2=3$ is odd. Either way

$$A^2\equiv -1 \pmod p. \tag{†}$$

Now produce a second square root of -1. Cassini's identity $F_{n-1}F_{n+1}-F_n^2=(-1)^n$ at $n=m-1$ gives

$$F_{m-1}^2=F_{m-2}F_m+(-1)^m\equiv(-1)^m=-1 \pmod p, \tag{‡}$$

using m odd and $F_m\equiv 0$. Subtracting (‡) from (†),

$$0\equiv A^2-F_{m-1}^2=(A-F_{m-1})(A+F_{m-1}),$$

and since $-F_{m-1}\equiv F_m-F_{m-1}=F_{m-2}$, this factors as $(A-F_{m-1})(A-F_{m-2})\equiv 0$. As p is prime, $A\equiv F_{m-1}$ or $A\equiv F_{m-2}$. Both F_{m-1} and F_{m-2} lie strictly between 0 and F_m, so each is already a remainder in the required range. Hence $R_m\in\{F_{m-2},F_{m-1}\}$ and in every case R_m is a Fibonacci number. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths agree up to the line *"I need to understand $A=\prod_{k=1}^{p-1}k^k$ modulo p."* Call it $(*)$.

At $(*)$ the natural attempt tries to **evaluate** A — to produce a formula for it, which forces the discrete logarithm and the character sum that has no closed form. The model chain tries to **constrain** A — to find one polynomial equation that A satisfies. The split is not about number-theoretic technique. It is about reading the word "prove" in the question. You are asked to show R_m lies in a set, not to say which element it is, and a set of two elements is exactly what a quadratic equation delivers.

> *The problem asks me to prove R_m is a Fibonacci number, not* which *one. So I do not need a formula for A; I need an equation satisfied by A whose roots I can name. If I can pin A^2, I get A up to sign — and "up to sign" modulo F_m means the two values F_{m-1} and $F_m-F_{m-1}=F_{m-2}$, which are both Fibonacci. The ambiguity I cannot resolve is exactly the ambiguity the problem is allowing me.*

> *Why A^2 rather than A? Look at the exponents. The product runs over k with exponent k. The involution $k\mapsto p-k$ fixes the index set $\{1,\dots,p-1\}$ and sends the exponent k to $p-k$. So the reflected product has the same bases with complementary exponents, and $k+(p-k)=p$ is constant. Multiplying A by its own reflection therefore collapses every exponent to p and leaves $((p-1)!)^p$, which Wilson's theorem already knows. Squaring is not a trick here; it is the only way*

> *to use the symmetry the index set actually has.*

> *$A^2 \equiv -1$ says A is a square root of -1 modulo p. There are exactly two of those, and they are negatives of each other, so I only need to* find one *and the other comes free. Where would I find a square root of -1 modulo F_m? I should look for a Fibonacci identity that produces $(-1)^n$ — because $(-1)^n$ appearing on the right of an identity is precisely how a square root of -1 gets manufactured. Cassini's identity is the one with $(-1)^n$ in it.*

- ▷ **Cue:** a problem that asks you to prove a quantity *lies in* a small set rather than to compute it $\longrightarrow$ **Move:** stop trying to evaluate; look for a polynomial equation the quantity satisfies whose roots are that set — a set of size two means find the quantity's square
- ▷ **Cue:** a product $\prod_{k=1}^{p-1} k^{e(k)}$ over all nonzero residues mod p $\longrightarrow$ **Move:** apply the involution $k \mapsto p-k$; if $e(k)+e(p-k)$ is constant, multiplying the product by its reflection collapses it to a power of $(p-1)!$ and Wilson's theorem finishes it
- ▷ **Cue:** $x^2 \equiv -1 \pmod{p}$ arising in a problem whose answer is named in some sequence $\longrightarrow$ **Move:** find one square root of -1 in that sequence by an identity with $(-1)^n$ on the right; the second root is its negative, and mod p the negative of F_{m-1} is F_{m-2}
- ▷ **Cue:** a Fibonacci identity is needed and the target expression carries a sign $(-1)^n$ $\longrightarrow$ **Move:** reach for Cassini $F_{n-1}F_{n+1} - F_n^2 = (-1)^n$ first; it is the unique low-index identity that manufactures a sign out of Fibonacci numbers

Layer 3 THE WHY

Why must the first of those thoughts be thought? Because of what the question is, grammatically.

"Prove that R_m is also a Fibonacci number" is an existential statement dressed as a computation. It says: R_m belongs to a certain set. It does not say which element. Every minute you spend trying to decide between F_{m-1} and F_{m-2} is a minute spent proving something stronger than you were asked — and in this case something that appears to be genuinely hard, since the pattern across $m = 5, 7, 11, 13, 17$ shows both cases occurring with no simple rule. The archive's standing advice applies with unusual force: *read what the conclusion actually quantifies over before deciding what to compute.* A conclusion of the form "$x \in \{a, b\}$" is, algebraically, the statement $(x-a)(x-b) = 0$. So the target is not a value; the target is a quadratic. And the only quadratic in x you can hope to evaluate directly is x^2.

Now why does A^2 yield to computation when A does not? Drive it to the structure of the index set.

The product is over the multiplicative group $G = (\mathbb{Z}/p)^\times$, and it has a base and an exponent that are *the same number* — k^k. That is the whole difficulty. As a base, k is an element of G, living mod p. As an exponent, k is an element of $\mathbb{Z}/(p-1)$. The two structures are unrelated, which is exactly why the discrete-logarithm route dies: it demands you compare k mod p with k mod $p-1$, and no theorem does that.

But the *index set* $\{1, \ldots, p-1\}$ has a symmetry that respects both readings at once: $k \mapsto p-k$. As a base, this is $k \mapsto -k$, which the group understands. As an exponent, this is $k \mapsto p-k$, which the additive structure of the exponents understands, and crucially $k + (p-k) = p$ does not depend on k. That constancy is the point. When two exponents attached to the same base sum to a constant, multiplying the two products makes every base carry the *same* exponent, and a product $\prod_k k^c$ over all of G is just $((p-1)!)^c$. So the impossible object $\prod k^k$ and its reflection $\prod k^{p-k}$ are individually intractable but their product is $((p-1)!)^p$ — a Wilson computation. This is the same move as pairing $\sum_k k$ with $\sum_k (n-k)$ to get Gauss's sum: you do not evaluate the sum, you evaluate the sum plus its reflection, because the reflection is what has constant fibres. It is not special to Fibonacci, to Wilson, or to this year's paper.

Kedlaya's solution opens by writing $A = \prod_{k=1}^{p-1}(p-k)^{p-k}$ with no comment. Here is where that comes from: it is the reflection of the index set, and it is written down because the exponents k and $p-k$ are complementary, not because anybody guessed.

Finally, the deepest layer: why is *Cassini's identity* the thing that finishes it, and why should a Fibonacci number be waiting at the end at all?

Once you have (†), the problem has stopped being about products. It says A is one of the two square roots of -1 modulo p. Those two exist (because $p \equiv 1 \bmod 4$, which is why the parity check on F_m mod 4 was needed), they are unique up to sign, and they are a *property of* p, nothing to do with the original product. So the question becomes: what is $\sqrt{-1}$ modulo F_m? And here the answer is an identity that a reader of the Fibonacci sequence should already own. Cassini's identity $F_{n-1}F_{n+1} - F_n^2 = (-1)^n$ is, read modulo F_n, precisely the statement

$$F_{n-1}^2 \equiv (-1)^{n+1} \pmod{F_n}.$$

That is: *the previous Fibonacci number is a square root of* ± 1 *modulo the current one*, with the sign alternating. For odd m the sign is -1, and F_{m-1} *is* $\sqrt{-1} \bmod F_m$. The problem is therefore the collision of two independent constructions of the same object: one analytic-looking, coming from Wilson's theorem applied to a reflected product; one purely recursive, coming from Cassini. Two square roots of -1 modulo a prime must agree up to sign. That is the entire problem, and stated that way there is nothing clever in it.

The final cosmetic step — $-F_{m-1} \equiv F_{m-2}$ — is the recursion itself: $F_{m-1} + F_{m-2} = F_m \equiv 0$. The set of square roots of $-1 \bmod F_m$ is closed under negation, and negation on consecutive Fibonacci numbers is the recursion. That is why the answer had to be Fibonacci on both branches rather than on one.

Caveat. I verified R_m against direct computation of $\prod_{k=1}^{F_m-1} k^k \bmod F_m$ for $m = 3, \ldots, 19$: the composite cases $m = 6, 8, 9, 10, 12, 14, 15, 16, 18, 19$ all give 0, and the prime cases give $R_5 = 3 = F_4$, $R_7 = 8 = F_6$, $R_{11} = 34 = F_9$, $R_{13} = 144 = F_{12}$, $R_{17} = 987 = F_{16}$ — so both branches of the final dichotomy genuinely occur and neither can be eliminated. The parity claim that $F_m \equiv 1$ or $2 \pmod 4$ for odd m I verified from the period-6 cycle $0, 1, 1, 2, 3, 1$ of Fibonacci mod 4, which I checked directly. I have *not* found, and do not claim, any rule determining which of F_{m-1}, F_{m-2} occurs; the proof above does not need one and the data above does not suggest one.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the word "*a*" in "prove that R_m is *a* Fibonacci number" — an existential conclusion, not a value. On seeing it, write down: *target is a set, therefore target is a polynomial equation, therefore compute* A^2 *before computing* A. The second cue, structural, is the exponent in $\prod_{k=1}^{p-1} k^k$ being equal to the base; write down $k + (p-k) = p$ *is constant.*

2. Those cues suggest: reflect the index set by $k \mapsto p-k$, multiply the product by its reflection to collapse all exponents to p, apply Wilson's theorem, and read the result as a quadratic equation for A. Then find a named square root of -1 modulo F_m using a Fibonacci identity carrying $(-1)^n$, namely Cassini.

3. Hesitation occurs twice. First, immediately: the product $\prod k^k$ looks like it has no structure and the instinct is to reach for primitive roots. What reinforces it away is asking what symmetry the *index set* has, before asking what structure the summand has. Second, after $A^2 \equiv -1$: the instinct is that you have lost information by squaring and must now recover the sign. What reinforces it away is re-reading the conclusion — you were never asked for the sign, and the two signs are F_{m-1} and F_{m-2}, both admissible.

4. Two traps. The visible one is evaluation: computing A directly via discrete logarithms is respectable, well-motivated and infinite. This is zombie maths wearing a number theorist's

coat — the most powerful available machine, deployed before asking whether the question needs a value at all. The subtler trap is over-achievement: trying to determine which of F_{m-1}, F_{m-2} occurs. The data actively encourages it (the first two prime cases both give F_{m-1}, so you will believe you have a pattern until $m = 11$), and it is a dead end. A third, minor trap: forgetting that even $m > 4$ makes F_m composite, so that "F_m prime" silently supplies the parity of m that the sign computation needs.

5. Recognisable faster next time by a one-line test on any product or sum over $\{1, \dots, n-1\}$ modulo n: *what does the involution $k \mapsto n - k$ do to my expression?* If it fixes the expression you have a symmetry to exploit; if it sends it to a complementary expression whose product or sum with the original is uniform, then combine the two rather than evaluating either. Gauss's pairing, Wilson's theorem, the reflection formula for the Beta function and this problem are all the same reflex.
6. Reusable decision rule: *match the algebraic operation to the size of the target set.* If the conclusion is "x equals this", compute x. If the conclusion is "x is one of two things", compute x^2 — or more generally, look for the degree-d symmetric function you can evaluate when the target set has d elements. Squaring is not a loss of information when the answer is allowed to be ambiguous by exactly a sign, and recognising that the permitted ambiguity in the *statement* licenses a specific lossy operation in the *proof* is a transferable move. It applies to any "show that x is $\pm$ something", to norm and trace arguments in field theory, and to every problem whose answer is stated as "an element of the sequence".

2021 B5

2021 B5. Say that an n-by-n matrix $A = (a_{ij})_{1 \le i,j \le n}$ with integer entries is *very odd* if, for every nonempty subset S of $\{1, 2, \dots, n\}$, the $|S|$-by-$|S|$ submatrix $(a_{ij})_{i,j \in S}$ has odd determinant. Prove that if A is very odd, then A^k is very odd for every $k \ge 1$.

Fields: linear algebra over $\mathbb{F}_2$; combinatorics (directed graphs, topological order) **Position:** B5 — hard, and hard in a specific way: the obstacle is not a calculation but the fact that the hypothesis and the conclusion are closed under different operations **Answer:** Every very odd matrix becomes unipotent upper triangular after conjugation by a permutation matrix over $\mathbb{F}_2$; unipotent upper triangular matrices are closed under multiplication *and* under principal restriction, so all powers are very odd.

Layer 1 THE WHAT — surface analysis

The natural attempt. Reduce mod 2 first — everything in sight is a parity statement, and det commutes with reduction, so "very odd" means: over $\mathbb{F}_2$, every principal submatrix of A is invertible. That much is free and everybody does it.

Then you attack the conclusion head on. Take $k = 2$ and a subset S. You want $\det\left((A^2)_{S,S}\right) \neq 0$ over $\mathbb{F}_2$. The natural tool is the Cauchy–Binet formula:

$$\det\left((A^2)_{S,S}\right) = \sum_{|T|=|S|} \det(A_{S,T}) \det(A_{T,S}),$$

the sum over all $T \subseteq \{1, \dots, n\}$ with $|T| = |S|$. The term $T = S$ is $\det(A_{S,S})^2 = 1$, which is what you want. So you need every *other* term to cancel in pairs.

And here is the wall, and it is worth walking right up to it. The hypothesis controls $\det(A_{S,S})$ — *principal* minors, rows and columns indexed by the same set. Every term with $T \neq S$ involves $\det(A_{S,T})$, a *non-principal* minor, about which the hypothesis says nothing whatsoever. There is no pairing in sight: T and its "partner" would have to be some other set of the same size, and the sum is over $\binom{n}{|S|}$ sets with no structure imposed on them. For $k = 3$ it is worse — a double sum over two intermediate sets. You can grind through $n = 2$ by hand ($a_{11} = a_{22} = 1$,

and det $= 1 + a_{12}a_{21}$ odd forces $a_{12}a_{21} = 0$, so one of the off-diagonal entries vanishes and the matrix is already triangular). You can do $n = 3$ with effort. The pattern you find there — somebody is always triangular — is the answer, but Cauchy–Binet will not hand it to you, because Cauchy–Binet is a statement about all minors and the hypothesis is a statement about principal ones only.

A second natural attempt: induct on k, hoping A and A^k very odd implies A^{k+1} very odd. This fails for the same reason, one level up — the product of two very odd matrices need not be very odd, so "very odd" is not closed under multiplication and cannot be the induction hypothesis. (Over $\mathbb{F}_2$, take A with rows $(1,1)$ and $(0,1)$, and B with rows $(1,0)$ and $(1,1)$: both are very odd, and AB has rows $(0,1)$ and $(1,1)$, whose $(1,1)$ entry is 0, so $S = \{1\}$ already fails.) So the property to be propagated is not the one you were given.

The model chain. Work over $\mathbb{F}_2$ throughout; A is a 0/1 matrix and "very odd" means every principal submatrix is invertible.

Step 1: the property to propagate. Call M *unipotent* if it is upper triangular with 1s on the diagonal. Suppose P is a permutation matrix with $P^{-1}AP$ unipotent. Conjugating by a permutation matrix relabels indices: $(P^{-1}AP)_{ij} = a_{\sigma(i)\sigma(j)}$ for the corresponding permutation σ, so the principal submatrix of $P^{-1}AP$ on S is the principal submatrix of A on $\sigma(S)$. Hence A is very odd if and only if $P^{-1}AP$ is. Now a principal submatrix of a unipotent matrix is again unipotent (deleting the same rows and columns preserves "0 below the diagonal" and keeps the 1s), so its determinant is 1. Therefore *unipotent implies very odd*, and unipotent matrices are closed under multiplication, so $(P^{-1}AP)^k = P^{-1}A^kP$ is unipotent for every $k \geq 1$, whence A^k is very odd. It remains only to prove the converse: every very odd matrix is permutation-conjugate to a unipotent one.

Step 2: diagonal entries. Take $S = \{i\}$. The 1×1 submatrix (a_{ii}) has odd determinant, so $a_{ii} = 1$ for every i.

Step 3: the digraph is acyclic. Form the loopless directed graph G on $\{1, \dots, n\}$ with an edge $i \to j$ exactly when $i \neq j$ and $a_{ij} = 1$. Suppose G has a directed cycle; choose one of minimal length $m \geq 2$, with vertices $i_1 \to i_2 \to \cdots \to i_m \to i_1$ (necessarily distinct, since a repeated vertex would contain a shorter cycle). I claim that for all j, k,

$$a_{i_j i_k} = \begin{cases} 1 & k - j \equiv 0 \text{ or } 1 \pmod{m}, \\ 0 & \text{otherwise.} \end{cases}$$

The case $k - j \equiv 0$ is Step 2 and $k - j \equiv 1$ is the cycle's own edges. For the rest: if $a_{i_j i_k} = 1$ with $k - j \not\equiv 0, 1 \pmod{m}$, then the edge $i_j \to i_k$ followed by the cycle arc $i_k \to i_{k+1} \to \cdots \to i_j$ is a closed directed walk of length $1 + \big((j-k) \bmod m\big)$. Since $k - j \not\equiv 0$ we have $(j-k) \bmod m \geq 1$, so the length is at least 2; since $k - j \not\equiv 1$ we have $(j-k) \bmod m \leq m - 2$, so the length is at most $m - 1$. That is a directed cycle shorter than m, contradicting minimality.

So the principal submatrix on $S = \{i_1, \dots, i_m\}$ has exactly two 1s in each row. Over $\mathbb{F}_2$ every row therefore sums to 0, so the all-ones column vector lies in its kernel and the submatrix is singular — contradicting very oddness. Hence G has no directed cycle: it is a DAG.

Step 4: triangularise. A finite DAG admits a topological order — list the vertices so that every edge points forwards. (Proof: a DAG has a vertex with no outgoing edge, namely the endpoint of any maximal directed path; place it last and induct on the remaining induced subgraph, which is still a DAG.) Let σ be the permutation realising a topological order and P the corresponding permutation matrix. Then $(P^{-1}AP)_{ij} = a_{\sigma(i)\sigma(j)} = 0$ whenever $i > j$, because an edge from a later vertex to an earlier one is forbidden; and the diagonal is all 1s by Step 2. So $P^{-1}AP$ is unipotent, and Step 1 finishes the proof. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths agree up to the line *"reduce mod 2: every principal submatrix of A is invertible over $\mathbb{F}_2$."* Call it $(*)$.

At $(*)$ the natural attempt asks *how does this determinant hypothesis behave under multiplication?* — and reaches for Cauchy–Binet, the formula that expresses a minor of a product. The model chain asks a different question: *is "very odd" even the right property to be propagating?* — notices that it is not (it is not closed under multiplication), and goes looking for a stronger property that is closed under both multiplication and principal restriction. The split is between manipulating the hypothesis and *replacing* it.

> *The hypothesis is closed under restriction — delete rows and columns of the same index set and you still have a very odd matrix. The conclusion is about multiplication. Those are different operations, and nothing in my toolkit converts one into the other. So the proof cannot be a manipulation of the hypothesis; it has to be a normal form. I need a property $\mathcal{P}$ that (i) implies very odd, (ii) is preserved by multiplication, (iii) is implied by very odd. Then the whole problem is the classification statement (iii).*

> *What property of matrices is closed under multiplication and under deleting matching rows and columns? Upper triangular with unit diagonal — both closures are immediate, because triangularity is about an ordering of the indices and deleting indices preserves the ordering. That is such a tight fit to what I need that I should conjecture "very odd $\Leftrightarrow$ triangular after reordering" and try to prove it, rather than keep looking.*

> *Over $\mathbb{F}_2$ a matrix with* 1*s on the diagonal is the adjacency matrix of a digraph with all loops present. "Upper triangular after reordering the indices" is exactly "the digraph has a topological order", which is exactly "the digraph is acyclic". So my remaining task translates into:* does very oddness forbid directed cycles? *That is now a question about a single explicit submatrix rather than about determinants in general.*

> *To rule out a cycle I only need to exhibit* one *singular principal submatrix. Take the cycle's own vertex set. If I choose the cycle to have minimal length, no chords are allowed, so the submatrix is exactly diagonal-plus-cyclic-shift: two* 1*s per row. Over $\mathbb{F}_2$, two* 1*s per row means every row sums to zero, and the all-ones vector is in the kernel. The minimality is not decoration — it is what makes the row count exactly two.*

- ▷ **Cue:** a hypothesis closed under one operation, a conclusion about a different operation $\longrightarrow$ **Move:** do not manipulate the hypothesis; find a normal form — a property closed under *both*, implied by the hypothesis and implying it
- ▷ **Cue:** "upper triangular after conjugating by a permutation" is wanted $\longrightarrow$ **Move:** translate to a digraph: it is exactly acyclicity, and you prove acyclicity by exhibiting a bad submatrix on a minimal cycle
- ▷ **Cue:** a $0/1$ matrix condition over $\mathbb{F}_2$ $\longrightarrow$ **Move:** read the matrix as an adjacency matrix; determinant over $\mathbb{F}_2$ counts cycle covers mod 2, and linear-algebraic conditions become graph conditions
- ▷ **Cue:** you must show a specific matrix over $\mathbb{F}_2$ is singular $\longrightarrow$ **Move:** look for a constant row sum or column sum; a row sum of 0 means the all-ones vector is in the kernel, and over $\mathbb{F}_2$ "two 1s per row" is a row sum of 0
- ▷ **Cue:** an induction on k for A^k stalls because the property is not multiplicative $\longrightarrow$ **Move:** strengthen the property until it is multiplicative; the strengthened statement is usually easier to prove, not harder, because it has more structure to work with

Layer 3 THE WHY

Why must the first of those thoughts be thought? Because of an asymmetry in the problem that is visible before any mathematics happens.

Look at the two halves of the statement. The hypothesis quantifies over subsets: *for every S, the submatrix on S is invertible*. Subsets are what you get by *restricting*. The conclusion quantifies over exponents: *for every k, A^k is very odd*. Exponents are what you get by *multiplying*. Restriction and multiplication do not commute: the principal submatrix of AB on S is emphatically not the product of the principal submatrices of A and B on S, because the missing indices contribute to the product. There is no identity that fixes this — Cauchy–Binet tells you honestly what the discrepancy is, and the discrepancy is a sum over non-principal minors, which is to say, over precisely the data the hypothesis withholds.

So the problem is not asking you to compute. It is asking you to find the *right* invariant, and once you accept that, the search is highly constrained. You want a property preserved by multiplication and by principal restriction, strong enough to force invertibility of every principal submatrix, and weak enough that every very odd matrix has it. Being upper triangular with unit diagonal, up to relabelling the indices, satisfies the first three at a glance. Multiplication: the product of two matrices supported on the "forward" index pairs is supported there too, and the diagonal of the product is the product of the diagonals. Restriction: triangularity is a statement about an *order* on the index set, and deleting indices leaves an order. Invertibility: the determinant of a triangular matrix is the product of its diagonal, which is 1. Three for three, on properties that are true for the stupidest possible reason. When a candidate invariant satisfies every closure requirement trivially, that is the signal to commit to it and spend the remaining effort proving the classification.

Now the classification, which is the mathematical content: *why should very oddness force triangularity?*

Here the mod-2 reduction stops being a convenience and becomes the mechanism. Over $\mathbb{F}_2$ the only nonzero scalar is 1, so a matrix is not a table of numbers — it is a table of yes/no, which is to say a directed graph. And the determinant loses its cancellation: over $\mathbb{F}_2$,

$$\det M = \sum_{\pi \in S_n} \prod_i m_{i\pi(i)}$$

with no signs, so $\det M$ is the *number of permutations π with every $m_{i\pi(i)} = 1$, counted modulo 2*. In graph language, with all loops present (Step 2), that is the number of *cycle covers* of the digraph, mod 2. The identity permutation is always one of them. So "det odd" means: *the induced subgraph has an even number of non-identity cycle covers.*

Read the argument again in that light and it is no longer a computation. If the induced subgraph on S is acyclic, the identity is its *only* cycle cover, so the determinant is 1 — odd, always. If the graph does contain a cycle, take a shortest one and let S be its vertices; minimality means the induced subgraph consists of the loops and the cycle and nothing else, so it has exactly two cycle covers, the identity and the cycle itself, and the determinant is even. The "all-ones vector is in the kernel" computation in the model chain is the same fact wearing linear-algebra clothes, and it is worth having both readings: the kernel version is shorter to write, the cycle-cover version is the reason. Very oddness *is* acyclicity, exactly, with no slack on either side. That is the "of course" this problem is hiding.

And acyclicity is triangularity. A directed acyclic graph has a topological order — and here too the reason is worth stating rather than cited, because it is what supplies the permutation P. Follow a directed path as far as it goes; it cannot revisit a vertex, so it terminates, and its final vertex has no outgoing edge. Put that vertex last. Delete it and repeat. Nothing in this is a graph-theory theorem you have to have memorised; it is the observation that a finite acyclic process has a last step.

Kedlaya's solution opens with "if there exists a permutation matrix P such that $P^{-1}AP$ is unipotent, then A is very odd" — stated flatly, as a convenience, before any motivation. Here is where that comes from. It is not a convenient sufficient condition that happened to be available. It is the unique shape of statement that can possibly prove this theorem, because the theorem's

hypothesis and conclusion are closed under incompatible operations and only a normal-form argument bridges them. The solution's structure was forced by the grammar of the problem before anyone looked at a matrix.

One last remark on why the *converse* direction is the load-bearing one. The easy half — unipotent implies very odd, and unipotency survives powers — is four lines and proves the theorem *for unipotent matrices*. All the work is in showing that there are no others. That is the general shape of a normal-form proof, and it is worth internalising: the classification is the theorem, and the deduction afterwards is bookkeeping.

Caveat. I verified the statement exhaustively for $n = 4$ over $\mathbb{F}_2$: of the 2^{16} matrices, 543 are very odd, and for each of them A^2, A^3 and A^5 are very odd with no exceptions. I also verified that "very odd" is not closed under multiplication, using the explicit pair in the natural attempt. The chordless-cycle claim in Step 3 I verified by the walk-length computation given, checking both boundary cases $(j - k) \bmod m = 1$ and $= m - 2$ explicitly.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the mismatch between quantifiers: "for every subset S" in the hypothesis, "for every $k \geq 1$" in the conclusion. On seeing it, write down: *hypothesis is closed under restriction, conclusion is about multiplication, these do not commute — look for a normal form, not a manipulation.* The second cue is "odd determinant", which should immediately produce the line *work over $\mathbb{F}_2$; the matrix is now $0/1$, i.e. a digraph.*

2. Those cues suggest: find a property closed under both multiplication and principal restriction that is equivalent to the hypothesis. Unipotent-up-to-permutation is the candidate; in digraph language it is acyclicity; acyclicity is proved by exhibiting a singular principal submatrix on a minimal cycle.

3. Hesitation occurs at exactly the moment you realise "very odd" is not multiplicative, and the feeling is that you have lost — the obvious induction on k is dead. What reinforces it away is recognising that a failed induction is a *diagnosis*, not a defeat: it tells you the induction hypothesis is too weak, and the standard response is to strengthen the statement until it closes. A second hesitation comes at the cycle argument, where you may worry that the induced submatrix could have extra 1s. What reinforces it away is choosing the cycle to be of minimal length and checking that a chord would produce a shorter cycle.

4. The trap is Cauchy–Binet. It is the correct formula for a minor of a product, it is exactly on topic, and it is a complete waste of the paper: it expresses a principal minor of A^2 in terms of non-principal minors of A, and the hypothesis is silent about those. This is zombie maths in its most seductive form — not crude brute force, but the deployment of the single most relevant named theorem without first asking whether it moves information in the direction you need. A secondary trap is trying to prove the result for $k = 2$ first and generalising: $k = 2$ is not easier, because the difficulty is structural rather than combinatorial, and time spent there is time not spent looking for the normal form.

5. Recognisable faster next time by a two-column habit: before starting, write down what operation the *hypothesis* is stable under and what operation the *conclusion* is stable under. If they differ, no amount of algebra on the hypothesis will reach the conclusion, and the proof must pass through a classification or normal form. This is a thirty-second check and it redirects the entire attempt.

6. Reusable decision rule: *when an induction on a property fails because the property is not closed under the relevant operation, do not weaken the goal — strengthen the property.* Look for the smallest strengthening that is closed under the operation, and then prove the strengthened property is actually equivalent to the original. The strengthened statement is typically *easier* to prove, because it carries more structure, which is the counterintuitive part and the reason people do not try it. The same rule produces the "all principal minors positive $\Rightarrow$ Cholesky"

argument, the strengthening of induction hypotheses in inequality problems, and every proof that classifies before it deduces.

2021 B6

2021 B6. Given an ordered list of $3N$ real numbers, we can *trim* it to form a list of N numbers as follows: We divide the list into N groups of 3 consecutive numbers, and within each group, discard the highest and lowest numbers, keeping only the median.

Consider generating a random number X by the following procedure: Start with a list of 3^{2021} numbers, drawn independently and uniformly at random between 0 and 1. Then trim this list as defined above, leaving a list of 3^{2020} numbers. Then trim again repeatedly until just one number remains; let X be this number. Let μ be the expected value of $|X - \frac{1}{2}|$. Show that

$$\mu \geq \frac{1}{4}\left(\frac{2}{3}\right)^{2021}.$$

Fields: probability (order statistics, distribution functions); real analysis (iterated maps, change of variables) **Position:** B6 — the hardest problem on the paper, and one where the correct global structure is visible within two minutes and the correct *local* step takes the rest of the hour **Answer:** Proved. The CDF of X_k is the k-th iterate of $\Phi(u) = 3u^2 - 2u^3$, and $\Phi' = 6u(1-u) \leq \frac{3}{2}$ everywhere, so one trim can contract the spread by at most $\frac{2}{3}$; hence $\mu_k \geq \frac{2}{3}\mu_{k-1}$ and $\mu_{2021} \geq \frac{1}{4}(\frac{2}{3})^{2021}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The set-up is forced and everybody gets it. Let F_k be the cumulative distribution function of the surviving number after k trims, starting from 3^k independent uniforms. At each stage, every survivor is the median of three independent values from the previous stage — independent because the groups of three are disjoint — so by induction all numbers at stage k are independent and identically distributed with CDF F_k. The CDF of the median of three independent variables with CDF F is the probability that at least two of the three fall below x, namely $3F^2 - 2F^3$. So with

$$\Phi(u) = 3u^2 - 2u^3, \qquad F_0(x) = x \text{ on } [0,1],$$

we have $F_k = \Phi(F_{k-1})$, hence $F_k = \Phi^{\circ k}$, the k-fold iterate of Φ. The target is $\mu_k := \mathbb{E}\left|X_k - \frac{1}{2}\right|$ with $k = 2021$.

Two natural continuations, and both are instructive failures.

First: compute F_k. Φ is a cubic and its iterates are polynomials of degree 3^k with no closed form. There is nothing to compute. You can get $F_1 = 3x^2 - 2x^3$, F_2 is degree 9, and after that the coefficients are meaningless.

Second: linearise at the fixed point. This is the good instinct, and it gets you the answer without the proof. By symmetry $\Phi(\frac{1}{2}) = \frac{1}{2}$, and $\Phi'(u) = 6u(1-u)$, so $\Phi'(\frac{1}{2}) = \frac{3}{2}$. So near the centre the map Φ *expands* by a factor $\frac{3}{2}$ per iteration; equivalently, the distribution of X_k concentrates towards $\frac{1}{2}$ at rate $(\frac{2}{3})^k$, and one expects $\mu_k \asymp c\,(\frac{2}{3})^k$. The constant $\frac{1}{4}$ in the target is $\mu_0 = \mathbb{E}|U - \frac{1}{2}|$ for U uniform, so the statement to be proved reads exactly like "μ shrinks by at most the linearisation rate, every step, starting from μ_0".

And here is the wall, and it is a genuine one, not a failure of nerve. *A derivative at a fixed point controls behaviour near that fixed point, and says nothing whatever about behaviour away from it.* The distribution of X_k is spread over all of $[0,1]$, not concentrated in a neighbourhood of $\frac{1}{2}$. For a general map with $\Phi'(\frac{1}{2}) = \frac{3}{2}$ the conclusion $\mu_k \geq \frac{2}{3}\mu_{k-1}$ is simply false — if the derivative were larger somewhere else, mass could be pulled in faster than the local rate. So the linearisation gives you the shape of the answer and none of the argument, and the temptation is to write it

up as though it were a proof, or to try to patch it with an error estimate that will not close. What is needed is a *global* statement, and finding out which global statement is the whole problem.

The model chain. Retain $\Phi(u) = 3u^2 - 2u^3$, $F_k = \Phi^{\circ k}$, and $\mu_k = \mathbb{E}\left|X_k - \frac{1}{2}\right|$.

Symmetry. $\Phi(1-u) = 3(1-u)^2 - 2(1-u)^3 = 1 - 3u^2 + 2u^3 = 1 - \Phi(u)$. Since $F_0(1-x) = 1 - F_0(x)$, induction gives $F_k(1-x) = 1 - F_k(x)$ for all k; in particular $F_k(\frac{1}{2}) = \frac{1}{2}$. Also $\Phi'(u) = 6u(1-u) > 0$ on $(0,1)$, so each F_k is a strictly increasing bijection of $[0,1]$, and F_k maps $[0, \frac{1}{2}]$ onto $[0, \frac{1}{2}]$.

A formula for μ_k with no absolute value in it. Writing $f_k = F_k'$ for the density and using the symmetry,

$$\mu_k = 2\int_0^{1/2} \left(\tfrac{1}{2} - x\right) f_k(x)\,dx = 2\Big[\left(\tfrac{1}{2} - x\right) F_k(x)\Big]_0^{1/2} + 2\int_0^{1/2} F_k(x)\,dx = 2\int_0^{1/2} F_k(x)\,dx,$$

the boundary terms vanishing at both ends ($x = \frac{1}{2}$ kills the first factor, $F_k(0) = 0$ kills the second). So μ_k is twice the area under the CDF on $[0, \frac{1}{2}]$.

The step. Because $F_k = \Phi^{\circ k} = \Phi^{\circ(k-1)} \circ \Phi$, we may put the new iteration on the *inside*:

$$F_k(x) = F_{k-1}\big(F_1(x)\big), \qquad F_1 = \Phi.$$

Substitute $u = F_1(x)$ in the integral. As x runs over $[0, \frac{1}{2}]$, so does u, and $du = F_1'(x)\,dx = 6x(1-x)\,dx$. The global bound is

$$F_1'(x) = 6x(1-x) \le \tfrac{3}{2} \quad \text{for all } x \in [0,1],$$

with equality only at $x = \frac{1}{2}$. Hence $dx = du/F_1'(x) \ge \frac{2}{3}\,du$, and since $F_{k-1} \ge 0$,

$$\mu_k = 2\int_0^{1/2} F_{k-1}\big(F_1(x)\big)\,dx = 2\int_0^{1/2} F_{k-1}(u)\,\frac{du}{F_1'(F_1^{-1}(u))} \;\ge\; 2\cdot\tfrac{2}{3}\int_0^{1/2} F_{k-1}(u)\,du = \tfrac{2}{3}\,\mu_{k-1}.$$

Conclusion. $\mu_0 = \mathbb{E}|U - \frac{1}{2}| = 2\int_0^{1/2}(\frac{1}{2} - x)\,dx = \frac{1}{4}$, so by induction $\mu_k \ge \frac{1}{4}(\frac{2}{3})^k$, and at $k = 2021$,

$$\mu \;\ge\; \tfrac{1}{4}\left(\tfrac{2}{3}\right)^{2021}. \qquad \square$$

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths agree up to the line "$F_k = \Phi^{\circ k}$ *where* $\Phi(u) = 3u^2 - 2u^3$, *and* $\Phi'(\frac{1}{2}) = \frac{3}{2}$ *is where the* $\frac{2}{3}$ *comes from.*" Call it $(*)$.

At $(*)$ the natural attempt treats $\frac{3}{2}$ as a *local* quantity — the derivative at the fixed point — and tries to convert local expansion into a global statement by an error analysis that never closes. The model chain asks a different question: *is* $\frac{3}{2}$ *also the maximum of* Φ' *over the whole interval?* It is: $\Phi'(u) = 6u(1-u)$ peaks at $u = \frac{1}{2}$. So the same number is both the local rate and a global Lipschitz constant, and the entire argument becomes a change of variables with an inequality in it.

There is a second, subtler split at the same line. The probabilistic recursion is $F_k = \Phi(F_{k-1})$ — Φ applied on the *outside*. The proof needs $F_k = F_{k-1} \circ \Phi$ — Φ applied on the *inside*. These are equal only because F_0 is the identity, so F_k is an honest iterate. The natural attempt never notices the distinction and therefore never finds the substitution.

> $\frac{3}{2}$ *is* $\Phi'(\frac{1}{2})$*, but a derivative at one point cannot prove anything about a distribution spread over the whole interval. So before I try to patch the linearisation, let me ask whether* $\frac{3}{2}$ *is a bound on* Φ' everywhere. $\Phi'(u) = 6u(1-u)$*, a downward parabola with its maximum at* $u = \frac{1}{2}$*. It is. The fixed point is also the point of steepest slope, so the local rate is the global rate and there is nothing to patch — I can use* $\Phi' \le \frac{3}{2}$ *as a hypothesis on the whole of* $[0,1]$.

> *A bound on Φ' is a bound on how Φ distorts* lengths along the x-axis. *The quantity I want, μ_k, must therefore be written as an integral $\mathrm{d}x$ over the x-axis with no absolute values and no densities — integrate by parts until it is $\int F_k\,dx$. Then a change of variables can bite.*

> *The recursion I derived puts the new Φ on the outside, $F_k = \Phi(F_{k-1})$, which changes the* values *of the function. To change variables I need it on the inside, $F_k = F_{k-1} \circ \Phi$, which changes the* argument. *Those are equal here only because $F_0 = \mathrm{id}$ makes F_k a pure iterate of one map — and that observation is the whole proof, because it converts a statement about how the distribution transforms into a statement about how the axis is stretched.*

> *The target constant $\frac{1}{4}$ is exactly $\mathbb{E}|U - \frac{1}{2}|$ for U uniform, and $(\frac{2}{3})^{2021}$ has exactly one factor per trim. The inequality is telling me its own proof: it is* 2021 *copies of a single-step inequality $\mu_k \geq \frac{2}{3}\mu_{k-1}$, anchored at $k = 0$. I should stop looking for anything global and find the one-step statement.*

- ▷ **Cue:** a bound of the form $c \cdot r^n$ where n is the number of steps in an iterative process $\longrightarrow$ **Move:** it is n copies of a one-step inequality anchored at a base case; identify c as the $n = 0$ value and r as the per-step factor, then prove only the one-step statement
- ▷ **Cue:** a derivative at a fixed point gives you the right constant but the wrong kind of argument $\longrightarrow$ **Move:** check whether that derivative is also the extremum of the derivative over the whole domain; if the fixed point is where the slope is steepest, the local rate is a global Lipschitz constant and the linearisation upgrades to a proof
- ▷ **Cue:** repeatedly taking the median of m samples $\longrightarrow$ **Move:** the CDF transforms by the fixed polynomial $\Phi(u) = \sum_{j>m/2} \binom{m}{j} u^j (1-u)^{m-j}$; the whole process is iteration of one map on $[0,1]$, and you should study that map, not the distributions
- ▷ **Cue:** an expectation $\mathbb{E}|X - c|$ with c the symmetry centre $\longrightarrow$ **Move:** rewrite it as $2\int_0^c F(x)\,dx$ by integration by parts; the absolute value and the density both disappear and the quantity becomes an area under the CDF, which is what change of variables can act on
- ▷ **Cue:** a recursion $F_k = \Phi(F_{k-1})$ with F_0 the identity $\longrightarrow$ **Move:** F_k is the k-th iterate of Φ, so Φ may equally be applied on the inside: $F_k = F_{k-1} \circ \Phi$; use the inside version whenever you want to change variables

Layer 3 THE WHY

Why must the first of those thoughts be thought? Start from what the inequality itself is saying.

$\frac{1}{4}(\frac{2}{3})^{2021}$ has two pieces and each names a stage of the process. The $\frac{1}{4}$ is $\mathbb{E}|U - \frac{1}{2}|$ for a single uniform, i.e. the value of μ before any trimming happens at all. The 2021 is the number of trims. So the inequality is not a single global estimate — it is a product of 2021 identical factors and a starting value, and the only way to prove such a thing is to prove one factor and multiply. Before any analysis, the *shape* of the target has told you the shape of the proof: find the one-step inequality $\mu_k \geq \frac{2}{3}\mu_{k-1}$. This is worth generalising in exactly the form given in the backtests: whenever a bound reads $c \cdot r^n$ and n counts the steps of an iteration, read off c as the base case and r as the per-step factor and go looking for the induction.

Now the real content: why is the per-step factor $\frac{2}{3}$, and why is it *provable*?

Here the archive has to say something the official solution leaves implicit. Trimming is not an operation on numbers, it is an operation on distributions, and the thing that makes this problem tractable is that in CDF coordinates the operation is a single fixed map. The CDF of the median of three independent samples from F is $3F^2 - 2F^3$ — and note that this expression involves F only through its values, never through x. That is why the process is iteration of one

function. If the CDF of the survivor after k trims is F_k, then $F_k = \Phi^{\circ k}$ and the entire probabilistic apparatus has been retired: what remains is the real-analytic question of how the iterates of a cubic behave.

That cubic has three fixed points: 0, $\frac{1}{2}$, 1. The two ends are attracting ($\Phi'(0) = \Phi'(1) = 0$), the middle one is repelling ($\Phi'(\frac{1}{2}) = \frac{3}{2} > 1$). So the iterates F_k converge pointwise to a step function — mass piles up at $\frac{1}{2}$ — and the question "how fast does X_k concentrate at $\frac{1}{2}$?" is the question "how fast do the iterates of Φ steepen at its repelling fixed point?". Near $\frac{1}{2}$ the answer is $\frac{3}{2}$ per step, which is the linearisation. The problem is that the concentration of X_k is not a local phenomenon: at every finite k the distribution has support on all of $[0,1]$, so a purely local rate proves nothing.

And now the fact that makes this problem work, which is the "of course" the whole analysis was hunting for. $\Phi'(u) = 6u(1-u)$. This is not an arbitrary cubic derivative. It is the probability density of the median of three uniforms — $\Phi' = f_1$ — and that density is symmetric about $\frac{1}{2}$ and unimodal. Its maximum is therefore at $\frac{1}{2}$. But $\frac{1}{2}$ is also, by the same symmetry, the fixed point of Φ. Two different consequences of one symmetry land on the same point: *the fixed point of the map is the mode of the corresponding density, hence the point where the map is steepest*. So the local expansion rate $\Phi'(\frac{1}{2})$ is simultaneously the global supremum $\sup_{[0,1]} \Phi'$, and the linearisation heuristic is not a heuristic — it is a global Lipschitz bound in disguise. This is why the problem has the answer it has, and it is the reason the constant $\frac{2}{3}$ can be attained as a per-step factor rather than merely approached asymptotically.

Once you have $\Phi' \le \frac{3}{2}$ globally, the mechanics follow from what a derivative bound controls. $\Phi' \le \frac{3}{2}$ is a statement about the stretching of the x-axis: the map Φ carries an interval of length ℓ to one of length at most $\frac{3}{2}\ell$, and therefore its inverse carries length ℓ to length at least $\frac{2}{3}\ell$. So the quantity you feed it must be a length — or an integral dx — along the x-axis. That is the entire reason for the integration by parts that turns $\mu_k = 2\int_0^{1/2}(\frac{1}{2} - x) f_k(x)\,dx$ into $\mu_k = 2\int_0^{1/2} F_k(x)\,dx$. In the first form the object is an expectation against a density and the derivative bound has nothing to grip. In the second it is an area over the x-axis, and a substitution $u = \Phi(x)$ with $dx \ge \frac{2}{3}du$ finishes in one line.

The last piece is the composition order, and it deserves naming because it is the step most likely to be skipped and it is not automatic. Probability gives you $F_k = \Phi(F_{k-1})$: the new trim acts on the *values* of the old CDF. The proof needs $F_k = F_{k-1} \circ \Phi$: the new trim acts on the *argument*. These say different things in general, and they agree here for exactly one reason — $F_0(x) = x$ is the identity, so F_k is a pure iterate of a single map and iterates commute with themselves. The moment you write $F_k = \Phi^{\circ k}$ you have earned the right to peel a Φ off either end, and peeling it off the inside is what makes the change of variables legal. That the uniform distribution is the identity in CDF coordinates is not a coincidence of this problem; it is the reason the uniform distribution is the reference distribution for every order-statistic computation.

Caveat. The presentation of the key step above is not Kedlaya's. The official first solution transposes the axes to write $\mu_k = 2\int_0^{1/2}(\frac{1}{2} - F_k^{-1}(y))\,dy$ and then applies the bound $\frac{1}{2} - F_1(x) \le \frac{3}{2}(\frac{1}{2} - x)$ at $x = F_k^{-1}(y)$; the official second solution compares f_k against an explicit step function and invokes a rearrangement argument. Both are correct. The change-of-variables route given here uses the same inequality $\Phi' \le \frac{3}{2}$ and avoids inverse functions entirely, and I have checked it line by line: the substitution $u = F_1(x)$ is legitimate because F_1 is a strictly increasing bijection of $[0, \frac{1}{2}]$ onto itself, and the inequality $dx \ge \frac{2}{3}\,du$ requires only $0 < F_1' \le \frac{3}{2}$, which holds on $(0, \frac{1}{2}]$ with the single vanishing point $x = 0$ an endpoint of measure zero (the integrand $F_{k-1}(u)$ vanishes there, and the integral is convergent since $F_{k-1}(u)/F_1'(F_1^{-1}(u))$ is bounded near 0).

I verified μ_k numerically by iterating the CDF on a grid of 2×10^6 points: $\mu_0, \ldots, \mu_9 = 0.2500, 0.1875, 0.1336, 0.0921, 0.0624, 0.0419, 0.0280, 0.0187, 0.0125, 0.0083$, and the ratio $\mu_k / \left(\frac{1}{4}(\frac{2}{3})^k\right)$ increases monotonically to about 1.2806 — so the bound holds with room and $(\frac{2}{3})^k$ is the correct decay rate, not a lossy one. The exact value $\mu_1 = \frac{3}{16}$ I confirmed by hand against $2\int_0^{1/2}(3x^2 - 2x^3)\,dx$.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the pair of constants in the target: $\frac{1}{4}$ and $(\frac{2}{3})^{2021}$, with 2021 equal to the number of trims. On seeing it, write down $\frac{1}{4} = \mathbb{E}|U - \frac{1}{2}|$ *is the base case;* $\frac{2}{3}$ *is the one-step factor; prove* $\mu_k \geq \frac{2}{3}\mu_{k-1}$. The second cue is "median of three", which should produce *CDF transforms by* $u \mapsto 3u^2 - 2u^3$ immediately, and then $F_0 = \mathrm{id}$, *so* F_k *is the k-th iterate.*

2. Those cues suggest studying the single map $\Phi(u) = 3u^2 - 2u^3$ rather than the distributions: locate its fixed points, compute $\Phi' = 6u(1-u)$, observe the maximum $\frac{3}{2}$ is attained at the repelling fixed point $\frac{1}{2}$, rewrite μ_k as $2\int_0^{1/2} F_k$, and change variables by $u = \Phi(x)$.

3. Hesitation occurs at the moment the linearisation gives the right constant. You now have the answer and no proof, and the instinct is either to write up the heuristic with hedging language, or to attempt a local-to-global error estimate. What reinforces it away is a single question that takes ten seconds: *is this derivative the largest one?* Plot $6u(1-u)$. It is. A second hesitation occurs at $\mu_k = 2\int_0^{1/2}(\frac{1}{2} - x)f_k(x)dx$, where the presence of both $(\frac{1}{2} - x)$ and f_k makes the integral look intractable; what reinforces it away is asking what *form* an integral must have for a derivative bound to act on it — a plain $\int F\,dx$ — and integrating by parts until it has that form.

4. Two traps. The first is attempting to compute F_{2021}, or any F_k for $k \geq 2$: these are polynomials of degree 3^k with no structure, and the appearance of a concrete finite number like 3^{2021} invites exactly that. Zombie maths — the instinct that a specified quantity should be evaluated. The second, and the one that will actually cost a strong solver the problem, is stopping at the linearisation. It is seductive because it produces the exact constant in the target, which feels like confirmation; but a derivative at a point is not an estimate on an interval, and a write-up resting on it is a confident wrong proof. The tell is that at no stage does the argument use any property of Φ away from $\frac{1}{2}$ — and any such argument would prove the same bound for maps where it is false.

5. Recognisable faster next time by treating iteration problems in the right coordinates from the start: when a random procedure is repeated and each round's output feeds the next, look for the coordinate in which one round is a fixed map with no memory. For order statistics that coordinate is the CDF; for sums it is the characteristic function; for products it is the logarithm. Having found it, stop thinking about probability and study the map — fixed points, derivatives at fixed points, and the supremum of the derivative.

6. Reusable decision rule: *a linearisation at a fixed point becomes a proof exactly when the fixed point is an extremum of the derivative.* When you obtain the right constant from $\Phi'(p)$ at a fixed point p, do not try to bound the error — check instead whether $\Phi'(p) = \sup \Phi'$ or $\inf \Phi'$ on the relevant domain. If it is, the local rate is a global Lipschitz bound and the heuristic converts into a one-line proof by change of variables; if it is not, the bound you guessed is probably false and you should look for the true extremum instead. This applies to every iterated-map problem, to convergence rates of Newton's method, and to any estimate where a per-step factor has been read off a derivative.