DAWN PUTNAM SESSION

Putnam 2011 — Papers A and B

Putnam 120 Archive

Contents

Putnam 2011 — Paper A **5**
2011 A1 5
2011 A2 7
2011 A3 10
2011 A4 13
2011 A5 16
2011 A6 19

Putnam 2011 — Paper B **23**
2011 B1 23
2011 B2 25
2011 B3 28
2011 B4 31
2011 B5 34
2011 B6 37

Putnam 2011 — Paper A

2011 A1

2011 A1. Define a *growing spiral* in the plane to be a sequence of points with integer coordinates $P_0 = (0,0), P_1, \ldots, P_n$ such that $n \geq 2$ and: the directed line segments $P_0P_1, P_1P_2, \ldots, P_{n-1}P_n$ are in the successive coordinate directions east (for P_0P_1), north, west, south, east, etc.; the lengths of these line segments are positive and strictly increasing. How many of the points (x, y) with integer coordinates $0 \leq x \leq 2011$, $0 \leq y \leq 2011$ *cannot* be the last point, P_n, of any growing spiral?

Fields: combinatorics / elementary number theory (alternating sums of monotone integer sequences) **Position:** A1 – an opener in position only; the exclusion argument is a genuine two-sided inequality proof, not arithmetic **Answer:** 10053.

Layer 1 THE WHAT — surface analysis

The natural attempt. The problem gives you a picture and a grid, and the honest first move is to draw. Plot $P_1, P_2, P_3, \ldots$ for a few small spirals – lengths $1, 2, 3$; lengths $1, 2, 4$; lengths $2, 3, 5$ – and read off coordinates, hoping a pattern in which (x, y) show up announces itself. This finds real information fast: P_1 is always on the positive x-axis, P_2 always has $y < x$ is false (in fact $y_1 > x_1$ forces $P_2 = (x_1, y_1)$ with $y_1 > x_1$), and no early point ever has $x = 0$. But turning "no early point has $x = 0$" into "no point of *any* spiral, at *any* stage, ever has $x = 0$" by staring at examples is not a proof; it is a conjecture with a growing sample size. The natural attempt keeps generating spirals, keeps failing to find a counterexample, and eventually has to assert the general claim on the strength of never having seen it fail. That is exactly the gap a Putnam grader is paid to notice: the coordinate of P_n is a genuinely variable alternating sum ($x_1 - x_2 + x_3 - \cdots$), and "it looks like it's always positive" is not the same statement as "it is always positive for every strictly increasing sequence of positive integers," which is what must actually be proved.

The model chain. Relabel the segment lengths, in the order they occur, as $x_1 < y_1 < x_2 < y_2 < x_3 < y_3 < \cdots$ – so $x_1, x_2, \ldots$ are the lengths of the east/west segments and $y_1, y_2, \ldots$ the lengths of the north/south segments, and the single increasing chain records exactly the order the segments are drawn in. Then

$$P_1 = (x_1, 0),\ P_2 = (x_1, y_1),\ P_3 = (x_1 - x_2, y_1),\ P_4 = (x_1 - x_2, y_1 - y_2),\ \ldots$$

and in general, after n east/west segments have been used, the x-coordinate of the current point is $x_1 - x_2 + x_3 - \cdots \pm x_n$, and symmetrically for y.

No point beyond P_0 has $x = 0$. If n is odd, regroup the alternating sum from the front:

$$x_1 - x_2 + \cdots + (-1)^{n-1}x_n = x_1 + (x_3 - x_2) + (x_5 - x_4) + \cdots + (x_n - x_{n-1}) > 0,$$

since $x_1 > 0$ and every bracket $x_{2k+1} - x_{2k}$ is a difference of consecutive terms of a strictly increasing sequence, hence ≥ 1. If n is even,

$$x_1 - x_2 + \cdots - x_n = (x_1 - x_2) + (x_3 - x_4) + \cdots + (x_{n-1} - x_n) < 0,$$

by the same regrouping with the sign reversed. So the x-coordinate is never 0 beyond P_0: all 2012 points $(0, y)$, $0 \le y \le 2011$, are unreachable.

No point beyond P_3 has $y \in \{0,1,2,3\}$. The y-coordinate after $n \ge 2$ north/south segments is $y_1 - y_2 + \cdots \pm y_n$. For n even the same bracketing gives $(y_1 - y_2) + (y_3 - y_4) + \cdots < 0$. For $n \ge 3$ odd,

$$y_1 + (y_3 - y_2) + (y_5 - y_4) + \cdots + (y_n - y_{n-1}).$$

Here $y_1 \ge x_1 + 1 \ge 2$, and $y_3 > x_2 > y_1$ forces $y_3 \ge y_1 + 2$, so the first bracket alone is ≥ 2; the sum is $\ge y_1 + 2 \ge 4$. So every point from P_4 onward has $y < 0$ or $y \ge 4$; only P_0, P_1, P_2, P_3 can land in the strip $0 \le y \le 3$, and P_0, P_1 have $y = 0$, so only P_2, P_3 (with $y = y_1$) can give $y \in \{1,2,3\}$, and $P_2 = (x_1, y_1)$, $P_3 = (x_1 - x_2, y_1)$ with $x_1 - x_2 < 0$, so only P_2 can give a nonnegative x with $y_1 \le 3$, and then $x_1 < y_1 \le 3$. This rules out $(x,0), (x,1)$ for $x \ge 1$, $(x,2)$ for $x \ge 2$, $(x,3)$ for $x \ge 3$ – exactly matching what is *not* achievable as $P_2 = (x_1, y_1)$ with $x_1 < y_1 \le 3$.

Everything else is achievable. If $0 < x < y$, take $P_2 = (x, y)$ directly (lengths $x < y$). If $x \ge y \ge 4$, take P_6 with lengths $1, 2, 3, x+1, x+2, x+y-1$ (strictly increasing since $x \ge 4$, $y \ge 4$): $x_1 - x_2 + x_3 = 1 - 3 + (x+2) = x$ and $y_1 - y_2 + y_3 = 2 - (x+1) + (x+y-1) = y$.

Counting the excluded points: $2012 + 2011 + 2011 + 2010 + 2009 = 10053$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths notice the same phenomenon on small examples: x never seems to hit 0, and small y only shows up early. They split at how that observation gets turned into a proof. The natural attempt stays at the level of individual coordinates and tries to track the running sum $x_1 - x_2 + x_3 - \cdots$ term by term, where the sign of the next partial sum is not obviously determined. The model chain regroups the same sum into pairs *from the correct end* – $(x_1) + (x_3 - x_2) + (x_5 - x_4) + \cdots$ for odd n, $(x_1 - x_2) + (x_3 - x_4) + \cdots$ for even n – so that every bracket has a fixed, visible sign before any arithmetic is done.

> *I should not evaluate $x_1 - x_2 + x_3 - \cdots$ term by term and watch the running total wobble; I should decide, before computing anything, which end of the alternating sum to bracket from, because pairing from the wrong end leaves a leftover unpaired term with the wrong sign and pairing from the right end makes every bracket's sign a one-line consequence of "the sequence is increasing."*

> *The parity of n (how many east/west or north/south segments have been used) is not incidental bookkeeping – it is the switch that decides which bracketing is available, so the two cases (odd n, even n) must be handled as genuinely separate arguments, not unified into one formula.*

Layer 3 THE WHY

The underlying fact is about alternating sums of any strictly increasing sequence, not about spirals: if $u_1 < u_2 < u_3 < \cdots$ are integers, then $u_1 - u_2 + u_3 - \cdots \pm u_n$ has a sign forced entirely by the parity of n, because grouping from the leading term that is *not* cancelled leaves every remaining pair a positive difference of consecutive terms. This is the integer-valued cousin of the fact that a partial sum of an alternating series with monotonically increasing terms is trapped between consecutive partial sums – except here, since the terms grow without any convergence requirement, the trapping argument collapses to a clean strict inequality rather than a limiting bound. Once the x- and y-coordinates of P_n are recognised as exactly this object (an alternating sum of a strictly increasing integer sequence, sliced at the east/west and north/south positions respectively), the sign is not something to search for empirically; it is read off the parity of how many terms have been used, which is visible directly from which P_i is being asked about. The reason P_2 and P_3 are the only points that can sit at $y \in \{1,2,3\}$ is the same fact applied at its smallest possible n: $n = 1$ is one raw term with no bracket available at all, so it is the only case where the sum is not pushed away from small values by a forced positive or negative bracket.

Caveat. I re-derived the alternating-sum sign argument, the P_2/P_3 boundary case, and the two achievability constructions independently and they match Kedlaya–Ng's stated forbidden set and count of 10053 exactly; the arithmetic $2012+2011+2011+2010+2009=10053$ was checked by direct addition. The full case analysis in the official solution (which I have paraphrased and expanded rather than reproduced verbatim) is not repeated here in every sub-case; I verified the regrouping argument in general and against several concrete numerical spirals rather than exhausting all n.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that a coordinate of P_n is an alternating sum $u_1-u_2+u_3-\cdots$ of a sequence that is given to be strictly increasing; on seeing that structure, write down the two regroupings (from the front, paired by consecutive terms) for odd and even numbers of terms before doing anything else.

2. That cue suggests proving a sign statement by algebraic regrouping rather than by exhausting small cases – the method is "bracket an alternating sum of a monotone sequence from the unpaired end," which converts an inequality about n unknown numbers into a sum of at most one leftover term plus manifestly-signed pairs.

3. Hesitation occurs exactly at deciding which end to bracket from, since bracketing $x_1-x_2+x_3-\cdots-x_n$ (even n) as $x_1-(x_2-x_3)-\cdots$ leaves a sign that is not immediately visible; the reinforcement is to always bracket starting from the first term with no partner, so every remaining pair is a "later minus earlier" difference with a known sign.

4. The trap is treating the problem as a search over small spirals and inferring the forbidden set from pattern-matching a handful of examples instead of proving the sign of the alternating sum in general; a plausible-looking pattern from five spirals is not a proof for a 2012×2012 grid.

5. Recognisable faster next time by tagging any coordinate defined as a running alternating sum of a strictly monotone sequence as "regroup from the unpaired end, split by parity of the term count" on sight, rather than tracking partial sums numerically.

6. Reusable rule: for an alternating sum $u_1-u_2+u_3-\cdots\pm u_n$ of a strictly monotone sequence, its sign is forced by the parity of n – bracket the terms into consecutive pairs starting from whichever end leaves no partner, and each bracket's sign is then read directly off the monotonicity, with no case-by-case numerical checking required.

▷ **Cue:** a coordinate (or quantity) defined as a running alternating sum of a strictly increasing sequence $\longrightarrow$ **Move:** regroup into consecutive pairs from the unpaired end and read the sign of each pair off the monotonicity, splitting by parity of the number of terms

▷ **Cue:** a counting problem over a grid phrased as "how many points cannot occur" $\longrightarrow$ **Move:** characterise the achievable set with an explicit two-sided argument (a general impossibility proof plus an explicit construction for everything else), not a pattern extrapolated from small cases

2011 A2

2011 A2. Let $a_1,a_2,\ldots$ and $b_1,b_2,\ldots$ be sequences of positive real numbers such that $a_1=b_1=1$ and $b_n=b_{n-1}a_n-2$ for $n=2,3,\ldots$. Assume that the sequence (b_j) is bounded. Prove that

$$S=\sum_{n=1}^{\infty}\frac{1}{a_1\cdots a_n}$$

converges, and evaluate S.

Fields: analysis (telescoping series, sequences and bounds) **Position:** A2 – routine in difficulty position but the telescoping identity has to be manufactured, not found **Answer:** $S = 3/2$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The recursion $b_n = b_{n-1}a_n - 2$ rearranges immediately to $a_n = (b_n + 2)/b_{n-1}$ for $n \geq 2$, so a strong student writes down

$$\frac{1}{a_1 \cdots a_n} = \frac{1}{a_1} \cdot \prod_{j=2}^{n} \frac{b_{j-1}}{b_j + 2} = \frac{b_1 b_2 \cdots b_{n-1}}{(b_2+2)(b_3+2)\cdots(b_n+2)},$$

since $a_1 = 1$. The natural next move is to try to bound this term directly: since (b_j) is bounded, say $0 < b_j < B$, each factor $b_{j-1}/(b_j+2)$ is less than 1, so the product shrinks – comparison with a geometric series looks promising for *convergence*. But the problem does not ask only for convergence; it asks for the exact value S, and a geometric-comparison bound proves $\sum 1/(a_1 \cdots a_n) < \infty$ without ever producing a number. At this point the natural attempt is stuck choosing between two unattractive options: try to solve the recursion for a_n or b_n in closed form (there is no reason to expect one – (a_n), (b_n) are barely constrained), or attempt to sum the series termwise using only the bound just derived, which throws away the exact algebraic structure of the recursion and can only ever produce an inequality, not an equality.

The model chain. Do not try to solve the recursion. Instead, notice that the very relation $a_n = (b_n + 2)/b_{n-1}$ that gives the general term of the sum also gives a candidate *difference* of a simpler sequence. Define

$$S_m = \frac{3}{2}\left(1 - \frac{b_1 \cdots b_m}{(b_1+2)\cdots(b_m+2)}\right).$$

This is guessed, not derived, by working backward from the desired shape: a telescoping sum $\sum_{n=1}^{m} \frac{1}{a_1 \cdots a_n}$ should equal (partial sum of m terms) = (something at m) − (something at $m-1$'s absence, i.e. a constant) once telescoped, and the natural "something" built from the same $b_j/(b_j+2)$ factors already in play is $\frac{b_1 \cdots b_m}{(b_1+2)\cdots(b_m+2)}$ times a constant to be fixed by the base case.

Check the claim by direct computation. First, $S_1 = \frac{3}{2}\left(1 - \frac{b_1}{b_1+2}\right) = \frac{3}{2}\left(1 - \frac{1}{3}\right) = 1 = \frac{1}{a_1}$, using $b_1 = 1$. Next, for $m \geq 2$,

$$S_m - S_{m-1} = \frac{3}{2}\left(\frac{b_1 \cdots b_{m-1}}{(b_1+2)\cdots(b_{m-1}+2)} - \frac{b_1 \cdots b_m}{(b_1+2)\cdots(b_m+2)}\right) = \frac{3}{2} \cdot \frac{b_1 \cdots b_{m-1}}{(b_1+2)\cdots(b_{m-1}+2)} \cdot \left(1 - \frac{b_m}{b_m+2}\right)$$

Since $1 - \frac{b_m}{b_m+2} = \frac{2}{b_m+2}$, and since $b_1 = 1$ gives $b_1 + 2 = 3$, this simplifies to

$$S_m - S_{m-1} = \frac{b_1 \cdots b_{m-1}}{(b_2+2)\cdots(b_m+2)} = \frac{1}{a_1 \cdots a_m},$$

which is exactly the identity derived from the recursion above. By induction, $S_m = \sum_{n=1}^{m} \frac{1}{a_1 \cdots a_n}$ for every $m \geq 1$.

Now use boundedness. Since every b_j is a positive real with $b_j < B$ for some constant B, the function $t \mapsto t/(t+2)$ is increasing on $t > 0$, so $b_j/(b_j+2) \leq B/(B+2) =: \rho < 1$ for every j. Hence

$$0 < \frac{b_1 \cdots b_m}{(b_1+2)\cdots(b_m+2)} = \prod_{j=1}^{m} \frac{b_j}{b_j+2} \leq \rho^m \to 0$$

as $m \to \infty$. Since every term of the original series is positive, (S_m) is increasing, and it is bounded above by $\frac{3}{2}$ (the subtracted quantity is always positive), so (S_m) converges; combined with $\rho^m \to 0$, $S_m \to \frac{3}{2}(1-0) = \frac{3}{2}$. Thus $S = \frac{3}{2}$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths reach the same rewritten general term $\dfrac{1}{a_1 \cdots a_n} = \dfrac{b_1 \cdots b_{n-1}}{(b_2+2)\cdots(b_n+2)}$ from the recursion. They split immediately after that, at the question: what do I do with this expression? The natural attempt treats it as something to be *bounded* (compare to a geometric series, get convergence, stop). The model treats it as something to be *recognised as a difference* – it goes looking for a sequence S_m with $S_m - S_{m-1}$ equal to exactly this term, before ever invoking the boundedness hypothesis.

> *The hypothesis that (b_j) is bounded is there to control a limit, not to control the individual terms of the sum – so I should not reach for it until I have already turned the sum into something with a limit to control, i.e. I should telescope first and use boundedness only to kill the leftover term at the end.*

> *The general term $b_1 \cdots b_{n-1}/[(b_2+2)\cdots(b_n+2)]$ is missing exactly the factor (b_1+2) that would make it look like a full product $\prod b_j/(b_j+2)$ – so multiplying and dividing by $(b_1+2)=3$ turns it into (constant)×(a clean product that decreases each step), which is precisely the form a telescoping difference needs.*

Layer 3 THE WHY

The recursion $a_n = (b_n+2)/b_{n-1}$ says that each reciprocal factor $1/a_n$ is a ratio of consecutive terms of the auxiliary sequence $c_n := b_n + 2$ against b_{n-1}; a product of such ratios is the classic signature of a telescoping product, because $\prod_{j=2}^{n} \frac{b_{j-1}}{b_j+2}$ has every "inner" value of b appearing once in a numerator-like role and once in a denominator-like role. Whenever a sum's general term arises from a first-order recursion of this multiplicative shape, the sum itself is a telescoping sum in disguise, and the entire task reduces to finding the right additive constant so that the boundary term at $n = 1$ matches. That constant is not free to guess blindly: it is fixed the moment you require S_1 to equal the first term $1/a_1$, which is why the $3/2$ in the model chain is not an inspired guess but the unique value forced by plugging in $b_1 = 1$. Once the telescoping form is secured, "the sequence (b_j) is bounded" stops being a vague side condition and becomes exactly the ingredient needed to show the geometric-type factor $\prod b_j/(b_j+2)$ tends to 0 – since $t/(t+2)$ is a fixed increasing function of t on the positives, boundedness of b_j is precisely what caps this ratio uniformly away from 1, which is the only way a product of such factors can be forced to vanish rather than merely stay bounded.

Caveat. I derived the telescoping identity $S_m - S_{m-1} = 1/(a_1 \cdots a_m)$ independently from the recursion (not by copying the official solution's stated formula for S_m and checking it), confirmed $S_1 = 1/a_1$ directly from $b_1 = 1$, and verified the boundedness argument gives $\rho = B/(B+2) < 1$. This matches Kedlaya–Ng's answer $S = 3/2$ exactly; I consider this fully verified.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that the recursion $b_n = b_{n-1}a_n - 2$ solves instantly for a_n in terms of b_n, b_{n-1}; on seeing a recursion mixing two sequences where one is being summed as a product-reciprocal, immediately substitute to express the summand purely in terms of the other sequence.
2. That substitution, once it produces a product of ratios of consecutive terms of an auxiliary sequence, should suggest telescoping – specifically, hunting for S_m with $S_m - S_{m-1}$ equal to the general term, rather than bounding the term directly.
3. Hesitation occurs right after the substitution, when the boundedness hypothesis is sitting unused and looks like it should be applied immediately to bound the term; the reinforcement

is to ask "does this hypothesis control a term, or a limit?" – boundedness controls a limit, so it belongs at the end of a telescoping argument, not the start of a comparison test.

4. The trap is proving convergence by geometric comparison and stopping, since the problem explicitly demands the value S, not just convergence; a proof that only bounds the tail cannot answer "evaluate S" and is graded as an incomplete solution.

5. Recognisable faster next time by treating "the summand is a product of ratios coming from a first-order recursion" as an immediate telescoping cue, and by fixing the telescoping sequence's additive constant from the very first term of the sum, not by trial and error.

6. Reusable rule: when a series' general term is a product $\prod$(ratio of consecutive terms of an auxiliary recursive look for an explicit partial-sum formula $S_m = C\,(1 - \prod(\text{same ratios}))$ with C fixed by matching S_1 to the first term; use any boundedness hypothesis only afterward, to show the leftover product tends to 0.

▷ **Cue:** a series whose n-th term is a product of reciprocals defined by a first-order recursion mixing two sequences $\longrightarrow$ **Move:** substitute to write the term as a product of ratios of consecutive values of one sequence, then guess a telescoping partial-sum formula and fix its constant from the first term

▷ **Cue:** a boundedness hypothesis on an auxiliary sequence appearing after the sum has already been telescoped $\longrightarrow$ **Move:** use it to show the leftover product/ratio tends to 0, not to bound the general term directly

2011 A3

2011 A3. Find a real number c and a positive number L for which

$$\lim_{r\to\infty} \frac{r^c \int_0^{\pi/2} x^r \sin x\, dx}{\int_0^{\pi/2} x^r \cos x\, dx} = L.$$

Fields: analysis (asymptotics of integrals, Laplace-type concentration, integration by parts) **Position:** A3 – middling in position, but requires two-sided squeeze estimates, not just a heuristic **Answer:** $c = -1$, $L = 2/\pi$.

Layer 1 THE WHAT — surface analysis

The natural attempt. With a large parameter r inside x^r on a fixed interval, the reflex is Laplace's method: x^r is overwhelmingly concentrated near the right endpoint $x = \pi/2$ as $r \to \infty$, since x^r is maximised there. So a natural attempt approximates both integrals by their value at $x = \pi/2$: $\sin(\pi/2) = 1$ and $\cos(\pi/2) = 0$. This immediately signals danger rather than giving an answer – the denominator's integrand vanishes exactly where the mass concentrates, so "replace $\cos x$ by its endpoint value" gives 0, and the ratio blows up, which is at least consistent with needing a compensating power r^c with $c < 0$ to tame it. But it does not, by itself, tell you *which* power, nor the constant L: $\cos x$ near $\pi/2$ behaves like $\pi/2 - x$ to first order, so the denominator integral should be of a smaller order than the numerator, and matching the orders requires expanding $\cos x$ to first order near the endpoint and doing a genuine asymptotic integral (a Laplace/Watson's-lemma-style computation), not merely plugging in the endpoint. A student who stops at "the denominator is smaller because cosine vanishes there" has correctly diagnosed the problem but produced no numbers, and attempting the full asymptotic expansion of both integrals from scratch, while it can be made to work, is a long road: substituting $x = \pi/2 - t/r$, expanding sin and cos in t, and taking $r \to \infty$ inside the integral requires justifying the interchange of limit and integral (dominated convergence or a comparison argument), which is more machinery than the problem needs.

The model chain. Route around asymptotic expansion entirely by relating the two integrals to *each other* via integration by parts, so that only one integral ever needs a genuine estimate.

Let $f(r) = \int_0^{\pi/2} x^r \sin x\, dx$. Integrating by parts (with $u = \sin x$, $dv = x^r\, dx$, so $du = \cos x\, dx$, $v = x^{r+1}/(r+1)$):

$$\int_0^{\pi/2} x^r \cos x\, dx = \left[\frac{x^{r+1}}{r+1}\sin x\right]_0^{\pi/2} - \int_0^{\pi/2} \frac{x^{r+1}}{r+1}\cos x\, dx.$$

That is the wrong integration by parts to try first – it reproduces a cosine integral on the right. Integrate by parts the other way instead: with $u = \cos x$, $dv = x^r\, dx$,

$$\int_0^{\pi/2} x^r \cos x\, dx = \left[\frac{x^{r+1}}{r+1}\cos x\right]_0^{\pi/2} + \int_0^{\pi/2} \frac{x^{r+1}}{r+1}\sin x\, dx = 0 + \frac{f(r+1)}{r+1},$$

since $\cos(\pi/2) = 0$ and the boundary term at $x = 0$ vanishes (as $x^{r+1} \to 0$). So the denominator is exactly $f(r+1)/(r+1)$: an *exact* identity, no approximation needed. The original expression, with $c = -1$, becomes

$$\frac{r^{-1}f(r)}{f(r+1)/(r+1)} = \frac{r+1}{r}\cdot\frac{f(r)}{f(r+1)}.$$

Since $(r+1)/r \to 1$, only $\lim_{r\to\infty} f(r)/f(r+1)$ is needed.

Bound $f(r)$ with a genuine squeeze, using only elementary estimates on $[0, \pi/2]$. Since $\sin x \le 1$,

$$f(r) < \int_0^{\pi/2} x^r\, dx = \frac{(\pi/2)^{r+1}}{r+1}.$$

Since sin is concave on $[0, \pi/2]$, it lies above its chord from $(0,0)$ to $(\pi/2, 1)$, i.e. $\sin x \ge \frac{2x}{\pi}$ there, so

$$f(r) > \int_0^{\pi/2} \frac{2x^{r+1}}{\pi}\, dx = \frac{2}{\pi}\cdot\frac{(\pi/2)^{r+2}}{r+2} = \frac{(\pi/2)^{r+1}}{r+2}.$$

So $\frac{(\pi/2)^{r+1}}{r+2} < f(r) < \frac{(\pi/2)^{r+1}}{r+1}$, i.e. $\frac{r}{r+2} < r(2/\pi)^{r+1} f(r) < \frac{r}{r+1}$, and both bounds $\to 1$, so

$$g(r) := r\left(\frac{2}{\pi}\right)^{r+1} f(r) \longrightarrow 1 \quad \text{as } r \to \infty.$$

Now express $f(r)/f(r+1)$ in terms of g:

$$f(r) = \frac{g(r)}{r}\left(\frac{\pi}{2}\right)^{r+1}, \qquad f(r+1) = \frac{g(r+1)}{r+1}\left(\frac{\pi}{2}\right)^{r+2},$$

so

$$\frac{f(r)}{f(r+1)} = \frac{g(r)}{g(r+1)}\cdot\frac{r+1}{r}\cdot\frac{2}{\pi} \longrightarrow 1\cdot 1\cdot\frac{2}{\pi} = \frac{2}{\pi}$$

as $r \to \infty$, since $g(r), g(r+1) \to 1$. Therefore

$$\lim_{r\to\infty} \frac{r+1}{r}\cdot\frac{f(r)}{f(r+1)} = 1\cdot\frac{2}{\pi} = \frac{2}{\pi}.$$

So $c = -1$ and $L = 2/\pi$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths correctly diagnose that the denominator is asymptotically smaller than the numerator because $\cos x$ vanishes exactly where $x^r \sin x$ concentrates. They split at the line "so I need to estimate both integrals near $x = \pi/2$": the natural attempt tries to estimate *both* integrals directly by a local expansion near the endpoint (Laplace's method from scratch, on two separate integrals). The model instead asks whether the cosine integral can be traded for a *sine* integral by integration by parts first, so that only one integral – $f(r)$ – ever needs an asymptotic estimate, and the relationship between numerator and denominator becomes an exact algebraic identity plus a single ratio $f(r)/f(r+1)$.

> *Before doing any asymptotics, I should ask whether integration by parts can convert the cosine integral into another copy of f at a shifted parameter – because the problem is symmetric enough (both integrands are x^r times a trig function, and $\sin' = \cos$, $\cos' = -\sin$) that one integral is very likely a disguised instance of the other, and if so I only ever need to understand the asymptotics of* one *family of integrals, not two.*

> *When I integrate by parts, I must choose $u = \cos x$ (not $u = \sin x$) specifically because $\cos(\pi/2) = 0$ kills the boundary term at the upper limit – the choice of which factor to differentiate is dictated by which choice makes the boundary term vanish, not by which looks more standard.*

> *To pin down $\lim f(r)/f(r+1)$ rigorously I need a genuine two-sided bound on $f(r)$, not just "$\sin x \approx 1$ near $\pi/2$" – and the concavity of $\sin$ on $[0, \pi/2]$ hands me a clean linear lower bound $\sin x \geq 2x/\pi$ for free, which is exactly matched in form to the crude upper bound $\sin x \leq 1$, so both bounds integrate to the same power of r and squeeze the ratio to a limit.*

Layer 3 THE WHY

The move to integrate by parts is not a trick reached for because it "often helps" – it is forced by recognising that $\frac{d}{dx}\sin x = \cos x$ and $\frac{d}{dx}\cos x = -\sin x$ means the numerator and denominator integrands are literally derivatives of one another (up to sign), so any relation between $\int x^r \sin x$ and $\int x^r \cos x$ that exists at all is almost certainly found by moving the derivative from the trig factor onto the x^r factor via parts, converting one trig function into the other. The specific choice of which factor to differentiate is then not aesthetic but is dictated by the boundary values already given in the problem: $\sin(\pi/2) = 1 \neq 0$ would leave a nonzero boundary term if sin were the part left undifferentiated, whereas $\cos(\pi/2) = 0$ makes that same boundary term vanish – so the correct choice of u is read directly off which trig value is zero at the interval's endpoint, not guessed. Finally, the squeeze on $f(r)$ is not an arbitrary pair of bounds – concavity of sin on $[0, \pi/2]$ is the exact structural fact that supplies a linear *lower* bound matching the trivial constant *upper* bound $\sin x \leq 1$ in the same power of x, so that both bounds, once integrated against x^r, differ only in the polynomial factor $r+1$ versus $r+2$ and squeeze to the same limit; this is why the concentration of $x^r\,dx$ near $\pi/2$ can be captured without ever explicitly computing a limiting Gaussian-type integral.

Caveat. I re-derived the integration-by-parts identity $\int_0^{\pi/2} x^r \cos x\,dx = f(r+1)/(r+1)$, the two-sided bound on $f(r)$ from $\sin x \leq 1$ and the concavity bound $\sin x \geq 2x/\pi$, and the resulting limit $f(r)/f(r+1) \to 2/\pi$ from scratch, independently of the official write-up's presentation order; the final answer $c = -1$, $L = 2/\pi$ matches Kedlaya–Ng exactly. I consider this fully verified.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that sin and cos are derivatives of each other and the interval's endpoint gives $\cos(\pi/2) = 0$, $\sin(\pi/2) = 1$; on seeing an integral pairing $\int x^r \sin x$ against $\int x^r \cos x$ on $[0, \pi/2]$, write down "integrate by parts to relate them" before attempting any asymptotic expansion.
2. That cue suggests choosing, in the integration by parts, whichever trig function is undifferentiated to be the one that vanishes at the relevant endpoint, so the boundary term disappears and the two integrals become related by an exact identity rather than an approximation.
3. Hesitation occurs at picking the direction of integration by parts (differentiate sin or differentiate cos?); the reinforcement is to check both boundary terms explicitly and choose the

direction that kills the boundary term, rather than picking the "usual" direction (differentiating the transcendental factor down in complexity) out of habit.

4. The trap is attempting Laplace's method on both integrals separately from scratch – substituting $x = \pi/2 - t/r$ and expanding each integrand – which works eventually but requires justifying interchange of limits twice and produces no simplification from the fact that the two integrands are related by differentiation.

5. Recognisable faster next time by treating any pair of integrals of the form $\int g(x)h(x)^r\,dx$ where the "extra" factor h's derivative is (up to sign) the other integral's factor as a signal to integrate by parts first, converting the pair into a single family evaluated at two nearby parameters.

6. Reusable rule: to estimate $\lim r^c\, I(r)/J(r)$ for two integrals of x^r against related trig or exponential factors, first look for an exact algebraic identity (integration by parts, a substitution, a recursion) linking I and J before attempting a from-scratch asymptotic expansion of each; reduce to a single family's self-ratio $I(r)/I(r+1)$ wherever possible, and bound that ratio with matching-order elementary inequalities (e.g. a trivial upper bound and a concavity-based linear lower bound).

▷ **Cue:** a ratio of two integrals $\int x^r f(x)\,dx$ and $\int x^r g(x)\,dx$ where $f' = \pm g$ or $g' = \pm f$ on the interval $\longrightarrow$ **Move:** integrate by parts, choosing the factor left undifferentiated to be the one vanishing at the endpoint, to get an exact relation between the two integrals rather than approximating each separately

▷ **Cue:** needing $\lim_{r\to\infty} h(r)/h(r+1)$ for $h(r) = \int_0^a x^r\phi(x)\,dx$ with ϕ concave, positive, and $\phi(a) > 0$ $\longrightarrow$ **Move:** squeeze $h(r)$ between $\int x^r\phi(a)\,dx$ (using $\phi \le \phi(a)$ near the max, or a trivial bound) and $\int x^r \cdot$ (the chord through 0 and $\phi(a)$)$\,dx$ (using concavity), both of which give the same power of r

2011 A4

2011 A4. For which positive integers n is there an $n \times n$ matrix with integer entries such that every dot product of a row with itself is even, while every dot product of two different rows is odd?

Fields: linear algebra over $\mathbb{F}_2$; parity arguments; Gram matrices and determinants **Position:** A4 — mid-paper. The construction is one line; the impossibility proof is a dimension count once the problem is read modulo 2. **Answer:** Exactly the odd n. For odd n take $J - I$ (all ones, zeros on the diagonal); for even n no such matrix exists.

Layer 1 THE WHAT — surface analysis

The natural attempt. You start with small cases, which is right. $n = 1$: the matrix (0) works (its one row has self-dot 0, and there are no pairs). $n = 2$: rows $r = (r_1, r_2)$ and $s = (s_1, s_2)$. You want $r_1^2 + r_2^2$ even, $s_1^2 + s_2^2$ even, $r_1s_1 + r_2s_2$ odd. You try $r = (1,1)$, $s = (1,1)$: the cross product is 2, even. You try $r = (1,1)$, $s = (1,3)$: cross product 4. Everything you try gives an even cross product. With some care you see why: $r_1^2 + r_2^2$ even forces $r_1 \equiv r_2 \pmod 2$, and likewise $s_1 \equiv s_2$, so $r_1s_1 + r_2s_2 \equiv 2r_1s_1 \equiv 0$. So $n = 2$ fails.

$n = 3$: you look for three rows with self-dots even and cross-dots odd. The first thing to hand is $(0,1,1), (1,0,1), (1,1,0)$: self-dots 2, cross-dots 1. That is $J - I$, and you check it for general n: self-dot $n - 1$, cross-dot $n - 2$. Both parities are right exactly when n is odd.

So you have the conjecture "odd n only" and a proof for $n = 2$. The natural next step is to push the $n = 2$ argument to $n = 4$. Rows $v_1, \ldots, v_4$, each with an even number of odd entries. You start listing the eight even-weight patterns mod 2 in $\{0,1\}^4$, and try to pick four with pairwise odd overlaps. 1100 and 1010 overlap in one place: good. Add 1001: overlaps with both in one place: good. Now a fourth, overlapping each of those three in an odd number of

places, with even weight... 0111 has odd weight. 1111 overlaps each in two places. You can grind through all eight candidates and verify there is no fourth row, but the case check does not tell you what happens at $n = 6$ or $n = 100$. The wall: the $n = 2$ argument was a coincidence of two coordinates, and the hand search does not scale.

The model chain. *Step 0: construction for odd n.* Let J be the $n \times n$ all-ones matrix and $M = J - I$. Row i has $n - 1$ ones. Its self-dot is $n - 1$, even. Two different rows share a 1 in every column except the two columns i and j, so their dot product is $n - 2$, odd.

Step 1: read everything mod 2. Suppose n is even and M exists, with rows $v_1, \ldots, v_n$. Reduce all entries mod 2 and work in the vector space $\mathbb{F}_2^n$. Since $x^2 \equiv x \pmod 2$ for every integer x,

$$v_i \cdot v_i \;\equiv\; \sum_t (v_i)_t \;=\; v_i \cdot \mathbf{1} \pmod 2,$$

where $\mathbf{1} = (1, \ldots, 1)$. So the hypothesis "self-dot even" says exactly that every row is orthogonal (mod 2) to $\mathbf{1}$:

$$v_1, \ldots, v_n \in W := \{w \in \mathbb{F}_2^n : w \cdot \mathbf{1} = 0\}, \qquad \dim W = n - 1.$$

Step 2: the rows are independent. The Gram matrix mod 2 is $G = (v_i \cdot v_j) = J - I = J + I$. Suppose $\sum_i c_i v_i = 0$ with $c \in \mathbb{F}_2^n$. Dotting with v_j gives $\sum_i c_i (v_i \cdot v_j) = 0$ for every j, i.e. $Gc = 0$. Now

$$G^2 = (J + I)^2 = J^2 + 2J + I = nJ + I \equiv I \pmod 2 \quad (n \text{ even}),$$

since $J^2 = nJ$. So G is invertible, $c = G^2 c = G(Gc) = 0$, and $v_1, \ldots, v_n$ are linearly independent over $\mathbb{F}_2$.

Step 3: contradiction. n independent vectors cannot live in the $(n - 1)$-dimensional space W. So no such matrix exists for even n.

The answer is: all odd n.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains find $J - I$ and both find the parity fact at $n = 2$. They split at (†): "$r_1^2 + r_2^2$ even forces $r_1 \equiv r_2$." The natural attempt treats this as a fact about two coordinates. The model reads the same fact as a fact about the vector $\mathbf{1}$: self-dot even $\iff$ row $\perp \mathbf{1}$ mod 2. Once the rows are seen to live in a hyperplane, the problem becomes a dimension count.

> *Every condition here is a parity. So I should stop thinking about integers and do the whole problem in $\mathbb{F}_2$, where $x^2 = x$. Then "$v \cdot v$ even" is "the entries of v sum to 0", which is "$v \cdot \mathbf{1} = 0$". All my rows are stuck in a hyperplane.*

> *If the rows are in a space of dimension $n - 1$, there are n of them, so they are dependent. The cross-dot condition fixes the Gram matrix completely: it is $J + I$. If that Gram matrix is invertible, the rows cannot be dependent. So the whole question is whether $J + I$ is invertible mod 2, and that depends on the parity of n.*

Layer 3 THE WHY

Why is "go to $\mathbb{F}_2$" forced rather than inspired? Look at the phrasing: *even* and *odd*. The problem does not care about the size of any entry, only its parity, and dot products respect reduction mod 2 because reduction mod 2 is a ring homomorphism $\mathbb{Z} \to \mathbb{F}_2$. So nothing is lost by reducing, and something is gained: $\mathbb{F}_2$ is a field, and over a field you have linear algebra — dimension, independence, invertibility. Over $\mathbb{Z}$ with parity conditions you have none of that in usable form.

Why does the self-dot turn into orthogonality to $\mathbf{1}$? Because the Frobenius identity $x^2 = x$ in $\mathbb{F}_2$ makes the quadratic form $v \cdot v$ *linear*. This is the fact that does all the work, and the

$n = 2$ computation "$r_1^2 + r_2^2$ even $\Rightarrow$ $r_1 \equiv r_2$" was that fact in disguise: in two coordinates, orthogonality to $(1,1)$ says the two coordinates agree. You had the idea at $n = 2$ and did not name it. Naming it is the step.

Why does the Gram matrix decide dependence? A general principle: if vectors $v_1, \ldots, v_n$ have Gram matrix G, any linear relation $\sum c_i v_i = 0$ gives $Gc = 0$. So an invertible Gram matrix certifies independence. The problem hands you the entire Gram matrix mod 2 — that is what "every dot product of a row with itself … every dot product of two different rows" specifies. So a question about the rows is really a question about $J + I$.

Why does parity of n enter? J satisfies $J^2 = nJ$, because each entry of J^2 counts n products $1 \cdot 1$. So J has eigenvalues n (once) and 0, and $J + I$ has eigenvalues $n + 1$ and 1. Mod 2, $n + 1 \equiv 0$ exactly when n is odd. So for odd n the Gram matrix is singular and nothing stops the rows from living in a hyperplane — indeed $J - I$ does. For even n it is invertible and the hyperplane is too small. The dichotomy in the answer is the dichotomy in the eigenvalue $n + 1$.

Kedlaya's solution says "$Mv = 0$ where v is the vector $(1, 1, \ldots, 1)$, and so M is singular", then computes $(\det M)^2 = \det(A - I) = 1$. Here is where that comes from: it is Step 1 and Step 2 in determinant language. $Mv \equiv 0$ is "every row is $\perp$ **1**"; $\det(MM^T) \equiv 1$ is "the Gram matrix is invertible". Singular M and invertible MM^T cannot coexist. My version and Kedlaya's are the same argument; the dimension-count phrasing shows *why* you would think of multiplying M by **1**.

Caveat. Checked by exhaustive search over rows in $\{0,1\}^n$ for $n \leq 7$: suitable rows exist for $n = 1, 3, 5, 7$ and not for $n = 2, 4, 6$, consistent with the proof. (Parity is all that matters, so $\{0,1\}$ entries suffice for the search.)

Layer $+\alpha$ BACKTESTING

1. The words "even" and "odd" applied to dot products. Write down: "reduce mod 2; $v \cdot v \equiv v \cdot \mathbf{1}$".

2. Linear algebra over $\mathbb{F}_2$: identify a subspace the vectors must lie in, and compare its dimension with the rank forced by the Gram matrix.

3. Hesitation after $n = 2$ and $n = 4$ are ruled out by hand: the case check does not generalise. Reinforce with: "a parity fact that works for two coordinates is usually an orthogonality statement over $\mathbb{F}_2$; find the vector it is orthogonal to."

4. The trap is working over $\mathbb{Z}$ or $\mathbb{R}$, where $J - I$ is always invertible (its determinant is $(-1)^{n-1}(n-1)$) and nothing distinguishes odd from even n except the parity of $n - 1$ — which is a hint, but not a proof. A second trap: forgetting that the self-dot condition is linear mod 2 and trying to handle it as a quadratic constraint.

5. Slogan: *parity conditions on a quadratic form become linear conditions over* $\mathbb{F}_2$*, because* $x^2 = x$. Any time a sum of squares is required to be even, you have a hyperplane.

6. Decision rule: **when a problem prescribes all pairwise dot products of n vectors, compute the determinant (or invertibility) of the prescribed Gram matrix over the relevant field; if it is invertible, the vectors are independent and any subspace containing them has dimension at least n.** Pair it with the eigenvalue fact $J^2 = nJ$ for any matrix of the form $aI + bJ$.

▷ **Cue:** a condition says a dot product, sum of squares, or quadratic expression is even or odd $\longrightarrow$ **Move:** reduce to $\mathbb{F}_2$ and use $x^2 = x$: $v \cdot v \equiv v \cdot \mathbf{1}$, so "self-dot even" means "orthogonal to **1**"

▷ **Cue:** all pairwise dot products of a family of vectors are prescribed $\longrightarrow$ **Move:** the Gram matrix is known; if it is invertible the vectors are linearly independent, so count dimensions

▷ **Cue:** a matrix of the form $aI + bJ$ appears $\longrightarrow$ **Move:** use $J^2 = nJ$; the eigenvalues are $a + bn$ (once) and a ($n - 1$ times), and invertibility depends on whether $a + bn$ vanishes in the relevant field

▷ **Cue:** an answer depends on the parity of n and the construction is $J - I \longrightarrow$ **Move:** look for the place where n appears as a scalar (here $J^2 = nJ$) and reduce it modulo the relevant prime

2011 A5

2011 A5. Let $F : \mathbb{R}^2 \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ be twice continuously differentiable functions with the following properties:

- $F(u, u) = 0$ for every $u \in \mathbb{R}$;
- for every $x \in \mathbb{R}$, $g(x) > 0$ and $x^2 g(x) \leq 1$;
- for every $(u, v) \in \mathbb{R}^2$, the vector $\nabla F(u, v)$ is either $\mathbf{0}$ or parallel to the vector $\langle g(u), -g(v) \rangle$.

Prove that there exists a constant C such that for every $n \geq 2$ and any $x_1, \ldots, x_{n+1} \in \mathbb{R}$, we have

$$\min_{i \neq j} |F(x_i, x_j)| \leq \frac{C}{n}.$$

Fields: multivariable calculus (first-order PDE, change of variables); real analysis (pigeonhole on a bounded interval, mean value theorem) **Position:** A5 — hard end of the paper. The hypotheses look like three unrelated technical conditions; the problem is to see that together they say "F is a function of one bounded variable". **Answer:** With $G(x) = \int_0^x g$, the image of G is a bounded interval (A, B) and $F(u, v) = h(G(u) - G(v))$ for a C^1 function h with $h(0) = 0$; pigeonhole gives two $G(x_i)$ within $(B - A)/n$, and the mean value theorem gives $C = (B - A) \max |h'|$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The conclusion looks like pigeonhole: $n + 1$ points, something of size $1/n$. So you try pigeonhole on the x_i directly. It fails at once: the x_i can be $0, 10^6, 2 \cdot 10^6, \ldots$, and then every pair is far apart. Whatever C/n bounds, it is not $|x_i - x_j|$.

So you try to understand F. You write the gradient condition as $\nabla F(u, v) = \lambda(u, v) \langle g(u), -g(v) \rangle$ and try to bound $|F(x_i, x_j)|$ by integrating from the diagonal, where $F = 0$:

$$F(x_i, x_j) = F(x_i, x_j) - F(x_j, x_j) = \int_{x_j}^{x_i} F_u(t, x_j)\, dt = \int_{x_j}^{x_i} \lambda(t, x_j)\, g(t)\, dt.$$

The factor $g(t)$ is integrable over all of $\mathbb{R}$, which feels promising. But λ is not controlled: nothing in the hypotheses bounds it, and on an unbounded region λ could be as large as it likes. You try the example $g(x) = 1/(1 + x^2)$ and $F = \arctan u - \arctan v$, where $\lambda \equiv 1$ and the bound works; but then you realise $F = (\arctan u - \arctan v)^3$ or $\sin(\arctan u - \arctan v)$ satisfy the hypotheses too, so F is not determined by g. The wall: you are estimating F along paths without knowing what F is, and there is no uniform bound on λ over $\mathbb{R}^2$.

The model chain. *Step 1: the natural coordinate.* Let $G(x) = \int_0^x g(t)\, dt$. Since $g > 0$ is continuous, G is C^1 (indeed C^3) and strictly increasing, with $G' = g$. It is bounded: for $x > 1$,

$$0 < G(x) - G(1) = \int_1^x g(t)\, dt \leq \int_1^x \frac{dt}{t^2} < 1,$$

and symmetrically $-1 < G(x) - G(-1) < 0$ for $x < -1$. So G maps $\mathbb{R}$ onto an open interval (A, B) with $B - A < \infty$ (open because G is strictly increasing and continuous, so it attains no maximum or minimum). By the inverse function theorem ($G' = g > 0$), $G^{-1} : (A, B) \to \mathbb{R}$ is C^1.

Step 2: in the new coordinates F depends only on the difference. Define on the open square $Q = (A, B)^2$

$$H(s, t) = F\big(G^{-1}(s), G^{-1}(t)\big).$$

H is C^1. Write $u = G^{-1}(s)$, $v = G^{-1}(t)$; then $(G^{-1})'(s) = 1/g(u)$ and by the chain rule

$$H_s + H_t = \frac{F_u(u,v)}{g(u)} + \frac{F_v(u,v)}{g(v)}.$$

By hypothesis $(F_u, F_v) = \lambda\,(g(u), -g(v))$ for some real λ (with $\lambda = 0$ when $\nabla F = \mathbf{0}$), so $H_s + H_t = \lambda - \lambda = 0$ on Q.

Fix $d \in (-(B-A),\, B-A)$. The set $\{(s,t) \in Q : s - t = d\}$ is a line segment (the intersection of a line with a convex set), and along it, parametrised as $\tau \mapsto (\tau + d, \tau)$, the derivative of H is $H_s + H_t = 0$. So H is constant on it. Hence $H(s,t) = h(s-t)$ where

$$h(d) := H\Big(\tfrac{A+B+d}{2},\, \tfrac{A+B-d}{2}\Big), \qquad |d| < B - A,$$

which is C^1 as a composition of C^1 maps. Translating back, $F(u,v) = h(G(u) - G(v))$, and $F(u,u) = 0$ gives $h(0) = 0$.

Step 3: pigeonhole in the bounded coordinate. Let $L = B - A$ and $M = \max\{|h'(d)| : |d| \le L/2\}$, finite because h' is continuous on this closed bounded interval, which lies inside the open domain $(-L, L)$. Given $x_1, \ldots, x_{n+1}$, the $n+1$ numbers $G(x_i)$ lie in (A, B). Sort them; the n consecutive gaps sum to less than L, so some gap is at most L/n: there are $i \neq j$ with $|G(x_i) - G(x_j)| \le L/n \le L/2$ (here $n \ge 2$ is used).

Step 4: mean value theorem. With $d = G(x_i) - G(x_j)$,

$$|F(x_i, x_j)| = |h(d) - h(0)| \le M|d| \le \frac{ML}{n}.$$

So $C = ML = (B-A)\max_{|d| \le (B-A)/2} |h'(d)|$ works, and it depends only on F and g.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains see pigeonhole in the C/n. They split at (†): "pigeonhole on what?" The natural attempt pigeonholes on x_i (fails: $\mathbb{R}$ is unbounded) and then tries to estimate F by path integrals without a formula for it. The model asks which bounded quantity the hypotheses provide, finds $G = \int g$, and changes coordinates so that the gradient condition becomes $H_s + H_t = 0$.

> *Pigeonhole needs the points in a bounded interval. The x_i are not. What in the hypotheses is bounded? $x^2 g(x) \le 1$ says $g \le 1/x^2$, which is integrable — so $\int_{-\infty}^{\infty} g$ is finite. The bounded ruler is $G(x) = \int_0^x g$. I should measure the x_i with G, not with the identity.*

> *∇F is parallel to $(g(u), -g(v))$. That is the gradient of $G(u) - G(v)$. A function whose gradient is everywhere parallel to the gradient of ϕ has the same level sets as ϕ, so F should be a function of $G(u) - G(v)$. To prove it, change variables to $s = G(u)$, $t = G(v)$ and check that the $(1,1)$ directional derivative vanishes.*

> *Once $F = h(G(u) - G(v))$ with $h(0) = 0$, the estimate is the mean value theorem: $|h(d)| \le M|d|$. I need M finite, so I need d in a compact subinterval of h's domain — that's what $n \ge 2$ buys: $|d| \le (B-A)/2$.*

Layer 3 THE WHY

Why is G the object of interest, and not a trick? Read each hypothesis for what it says about the geometry.

The third hypothesis constrains only the *direction* of ∇F. A gradient's direction is the normal to the level curve through that point, so the hypothesis is a statement about the level curves of F: at every point, they are perpendicular to $(g(u), -g(v))$. The vector $(g(u), -g(v))$ is itself a gradient — of $\phi(u, v) = G(u) - G(v)$, where $G' = g$. So the level curves of F run along the level curves of ϕ. Two functions whose gradients are everywhere parallel share their level sets, and a function constant on the level sets of ϕ is a function of ϕ. That is the circle-and-centre moment here: once you ask "what function has gradient $(g(u), -g(v))$?", G appears because g is a derivative waiting to be integrated, not because someone pulled it from a hat. The change of variables $s = G(u)$, $t = G(v)$ is then the standard way to prove that "F is a function of ϕ": in those coordinates $\phi = s - t$, and "constant along the level lines of $s - t$" is "zero derivative in direction $(1, 1)$".

The second hypothesis, $x^2 g(x) \leq 1$, looks like a technical growth bound, and that is what it is for: it says g is dominated by $1/x^2$ at infinity, which is precisely integrability of g on $\mathbb{R}$ (the condition gives nothing near 0, but there g is continuous, hence bounded on $[-1, 1]$). Integrability of g means G is bounded, and that is the only reason pigeonhole can work: the $n + 1$ values $G(x_i)$ live in an interval of finite length $B - A$. The positivity $g > 0$ makes G strictly increasing, so G^{-1} exists and the coordinate change is legal.

The first hypothesis, $F(u, u) = 0$, becomes $h(0) = 0$: the reference point for the mean value theorem. And $n \geq 2$ in the conclusion is not decoration. It guarantees $|d| \leq (B - A)/2$, keeping d inside a closed interval strictly within h's open domain $(-(B - A), B - A)$. Near the ends of that domain h' could blow up, since G^{-1} has unbounded derivative near A and B. So every hypothesis maps to exactly one step of the proof, and the problem is less mysterious than its statement: *F is a C^1 function of one bounded variable, vanishing at 0.*

Abhinav Kumar's solution (reproduced by Kedlaya and Ng) opens with "Define $G(x) = \int_0^x g(t)\,dt$." Here is where that comes from: $(g(u), -g(v)) = \nabla(G(u) - G(v))$, and the problem asks for pigeonhole, which needs a bounded ruler, which $\int g < \infty$ supplies. The two reasons point at the same function.

One more remark on why the natural attempt's path integral stalled: in that integral, $\lambda(t, x_j) g(t) = F_u$, and in the new coordinates $\lambda = h'(G(u) - G(v))$. The factor you could not bound, λ, is h' in disguise. It is bounded only on the region $|G(u) - G(v)| \leq (B - A)/2$, which is not a region you would ever describe in the (u, v) plane without G.

Caveat. Only C^1 regularity of F and continuity of g are used; the problem's C^2 hypotheses are more than enough. The domain of h is exactly $(-(B - A), B - A)$, and $|h'|$ may be unbounded near its endpoints, which is why the proof restricts to $|d| \leq (B - A)/2$.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is C/n with $n + 1$ points: write "pigeonhole on a bounded interval of length L gives two points within L/n". The second is $x^2 g(x) \leq 1$: write "$\int_{\mathbb{R}} g < \infty$, so $G = \int g$ is bounded".

2. A change of variables that turns a gradient-direction condition into a translation-invariance condition, followed by pigeonhole and the mean value theorem.

3. Hesitation occurs when pigeonhole on the x_i fails and the student concludes pigeonhole is the wrong tool. Reinforce with: "if pigeonhole fails in one coordinate, find a bounded monotone reparametrisation and pigeonhole there."

4. The trap is treating the three hypotheses separately. A second trap: bounding $\max |h'|$ over the whole domain of h, which may be infinite; the restriction to $|d| \leq (B - A)/2$ is where $n \geq 2$ is spent. A third: forgetting that λ may be 0 (the "either **0**" clause), which the argument handles because $H_s + H_t = \lambda - \lambda$ in every case.

5. Slogan: *a gradient known only up to direction pins down the level sets; find a function whose gradient has that direction and make it a coordinate.* And: *an integrable density on $\mathbb{R}$ is a bounded ruler for*

the whole line.

6. Decision rule: **when a hypothesis fixes the direction of ∇F as $\nabla\phi$ for some computable ϕ, conclude $F = h \circ \phi$ by changing to coordinates in which ϕ is linear, and transfer every question about F to a one-variable question about h.** Companion rule: when a pigeonhole bound is wanted on an unbounded set, look for a bounded increasing function supplied by the hypotheses.

- ▷ **Cue:** ∇F is parallel to a given vector field that is itself a gradient $\nabla\phi$ $\longrightarrow$ **Move:** F is constant on level sets of ϕ; change variables so ϕ becomes linear and show $F = h(\phi)$
- ▷ **Cue:** a vector $\langle g(u), -g(v)\rangle$ with the same function in both slots, opposite signs $\longrightarrow$ **Move:** it is $\nabla(G(u) - G(v))$ with $G' = g$; the relevant variable is $G(u) - G(v)$
- ▷ **Cue:** a bound like $x^2 g(x) \le 1$ or $g(x) = O(1/x^2)$ on a positive function $\longrightarrow$ **Move:** g is integrable on $\mathbb{R}$; $G = \int g$ is a bounded, strictly increasing reparametrisation of the line
- ▷ **Cue:** $n+1$ arbitrary points and a conclusion of size C/n $\longrightarrow$ **Move:** pigeonhole: find a bounded interval containing some transform of the points; two of them are within length $/n$
- ▷ **Cue:** $|h(d)|$ to be bounded with $h(0) = 0$ and h only C^1 on an open interval $\longrightarrow$ **Move:** mean value theorem on a closed subinterval where $\max|h'|$ is finite; use a spare hypothesis (here $n \ge 2$) to stay in it

2011 A6

2011 A6. Let G be an abelian group with n elements, and let

$$\{g_1 = e, g_2, \ldots, g_k\} \subsetneqq G$$

be a (not necessarily minimal) set of distinct generators of G. A special die, which randomly selects one of the elements $g_1, g_2, ..., g_k$ with equal probability, is rolled m times and the selected elements are multiplied to produce an element $g \in G$. Prove that there exists a real number $b \in (0,1)$ such that

$$\lim_{m\to\infty} \frac{1}{b^{2m}} \sum_{x\in G} \left(\text{Prob}(g = x) - \frac{1}{n}\right)^2$$

is positive and finite.

Fields: probability (random walks on groups); Fourier analysis on finite abelian groups (characters, Plancherel); linear algebra (diagonalisation) **Position:** A6 — the last problem of the morning paper, and hard for anyone who has not met characters. With them it is a direct computation; the difficulty is knowing to reach for them. **Answer:** Let $\lambda_\chi = \frac{1}{k}\sum_i \chi(g_i)$ for each character χ of G and $b = \max_{\chi\neq 1}|\lambda_\chi|$. Then $0 < b < 1$ and the limit equals $\frac{1}{n}\#\{\chi \neq 1 : |\lambda_\chi| = b\}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You set it up as a Markov chain: $p_m(x) = \text{Prob}(g = x)$ after m rolls, with $p_0 = \delta_e$ and

$$p_{m+1}(x) = \frac{1}{k}\sum_{i=1}^{k} p_m(xg_i^{-1}).$$

So $p_m = T^m p_0$ for a linear operator T on functions $G \to \mathbb{R}$, and the uniform distribution $u \equiv 1/n$ is fixed by T. The quantity in the problem is $\|p_m - u\|^2$. You expect geometric decay governed by the second-largest eigenvalue of T, and you guess b should be that eigenvalue's modulus.

To make this precise you need the eigenvalues. You try a small case: $G = \mathbb{Z}/3$, $S = \{0,1\}$. T is the 3×3 matrix with $\frac{1}{2}$ on the diagonal and on one off-diagonal cycle. You find its eigenvalues

by the characteristic polynomial: 1 and $\frac{1}{2}(1+\omega)$, $\frac{1}{2}(1+\omega^2)$, where $\omega = e^{2\pi i/3}$; both have modulus $\frac{1}{2}$. Good. But for a general abelian G you do not have the characteristic polynomial, T need not be symmetric (so real eigenvectors, orthogonality, and "$\|T^m v\| \approx |\lambda|^m \|v\|$" are all in doubt), and you can prove neither that the limit exists nor that it is positive: a general non-normal T could have Jordan blocks, which contribute $m\,|\lambda|^m$ and wreck the limit. The wall: Perron–Frobenius tells you $p_m \to u$, but not with the exact rate and exact constant the problem demands.

The model chain. Notation. A *character* of G is a homomorphism $\chi : G \to \mathbb{C}^\times$; since $x^n = e$, $\chi(x)$ is an n-th root of unity, so $|\chi(x)| = 1$ and $\overline{\chi(x)} = \chi(x^{-1})$. For a finite abelian group there are exactly n characters, forming a group $\widehat{G}$ with trivial character $\mathbf{1}$, and they satisfy the orthogonality relations

$$\sum_{x\in G} \chi(x)\overline{\psi(x)} = n\,[\chi = \psi], \qquad \sum_{\chi\in\widehat{G}} \chi(x) = n\,[x = e].$$

For $f : G \to \mathbb{C}$ define $\hat{f}(\chi) = \sum_x f(x)\overline{\chi(x)}$. Then Plancherel holds: $\sum_x |f(x)|^2 = \frac{1}{n}\sum_\chi |\hat{f}(\chi)|^2$ (expand the right side and apply the first relation).

Step 1: the transform of p_m. Let μ be the die's distribution, $\mu(g_i) = \frac{1}{k}$. The product of independent rolls has distribution the convolution $p_m = \mu * \cdots * \mu$, and since χ is multiplicative, $\hat{p}_m(\chi) = \mathbb{E}\,\overline{\chi(g)} = \prod_{\text{rolls}} \mathbb{E}\,\overline{\chi(\text{roll})} = \overline{\lambda_\chi}^{\,m}$, where

$$\lambda_\chi = \frac{1}{k}\sum_{i=1}^{k} \chi(g_i).$$

For the uniform $u \equiv \frac{1}{n}$: $\hat{u}(\mathbf{1}) = 1$ and $\hat{u}(\chi) = 0$ for $\chi \neq \mathbf{1}$. Also $\lambda_{\mathbf{1}} = 1$. So $\widehat{p_m - u}(\mathbf{1}) = 0$ and $\widehat{p_m - u}(\chi) = \overline{\lambda_\chi}^{\,m}$ otherwise.

Step 2: Plancherel.

$$D_m := \sum_{x\in G}\Big(p_m(x) - \frac{1}{n}\Big)^2 = \frac{1}{n}\sum_{\chi\neq\mathbf{1}} |\lambda_\chi|^{2m}.$$

Step 3: $b < 1$. Put $b = \max_{\chi\neq\mathbf{1}} |\lambda_\chi|$ (a maximum over a finite nonempty set; $n \geq 2$ since $e \in S \subsetneq G$). By the triangle inequality $|\lambda_\chi| \leq 1$. If $|\lambda_\chi| = 1$, the k unit complex numbers $\chi(g_i)$ have average of modulus 1, which forces them all equal; one of them is $\chi(g_1) = \chi(e) = 1$, so $\chi(g_i) = 1$ for every i. The g_i generate G and χ is a homomorphism, so $\chi \equiv 1$. Hence $|\lambda_\chi| < 1$ for every $\chi \neq \mathbf{1}$, and $b < 1$.

Step 4: $b > 0$. If $b = 0$, then $\hat{\mu}(\chi) = \overline{\lambda_\chi} = 0$ for all $\chi \neq \mathbf{1}$ and $\hat{\mu}(\mathbf{1}) = 1$, which is the transform of u. The transform is injective (Fourier inversion $f(x) = \frac{1}{n}\sum_\chi \hat{f}(\chi)\chi(x)$, from the second orthogonality relation), so $\mu = u$: every element of G has probability $\frac{1}{n} > 0$ under the die. Then $\{g_1, \ldots, g_k\} = G$, contradicting $\subsetneq$. So $b > 0$.

Step 5: the limit. Let $N = \#\{\chi \neq \mathbf{1} : |\lambda_\chi| = b\} \geq 1$. Then

$$\frac{D_m}{b^{2m}} = \frac{N}{n} + \frac{1}{n}\sum_{|\lambda_\chi|<b}\Big(\frac{|\lambda_\chi|}{b}\Big)^{2m} \longrightarrow \frac{N}{n} \in \Big[\tfrac{1}{n}, 1\Big),$$

since each ratio in the finite sum is in $[0, 1)$. This is positive and finite.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains write $p_m = T^m p_0$ and both expect the answer to be the second-largest eigenvalue. They split at (†): "I need the eigenvalues and eigenvectors of T." The natural attempt tries to compute them from the matrix. The model uses the fact that T is a convolution on an abelian group, so its eigenvectors are known in advance — they are the characters — and the eigenvalues are the averages λ_χ.

> *The walk is "multiply by a random element". On an abelian group, the functions that multiplication turns into scalar multiples of themselves are the characters: $\chi(xg) = \chi(x)\chi(g)$. So the characters diagonalise every step of this walk at once, and the eigenvalue for χ is the average of χ over the die's faces.*

> *The quantity is a sum of squares over G — an L^2 norm. Characters are orthogonal, so the norm splits by Plancherel into $\sum |\lambda_\chi|^{2m}$ with no cross terms. After that, the limit is about the largest $|\lambda_\chi|$ with $\chi \neq 1$, and I have to show it is strictly between 0 and 1 using the three hypotheses: e on the die, generators, proper subset.*

Layer 3 THE WHY

Why characters, and why is this natural rather than arcane? Start from what the die does: it multiplies. The problem wants to follow a distribution through repeated multiplication by random elements. The central question for any repeated linear operation is: *which vectors does it only rescale?* For multiplication in a group, those are functions satisfying $\chi(xg) = c_g\,\chi(x)$ for all x — and setting $x = e$ shows $c_g = \chi(g)$, so χ is a homomorphism to $\mathbb{C}^\times$. That is the definition of a character; it is not a borrowed piece of machinery but the answer to "what are the eigenvectors of translation?" If $G = \mathbb{Z}/n$, these are the functions $x \mapsto e^{2\pi i j x/n}$, and the whole argument is the discrete Fourier transform. The $\mathbb{Z}/3$ computation in the natural attempt found eigenvalues $\frac{1}{2}(1+\omega)$: that is λ_χ for $\chi(x) = \omega^x$, and you had the character in your hand without naming it.

Why does this also solve the Jordan-block worry? Because G is abelian, all translation operators commute, and the characters form an orthogonal basis of simultaneous eigenvectors. So T is normal — diagonalisable by an orthogonal basis — and $\|T^m v\|^2$ is exactly $\sum |\lambda|^{2m}|\text{coefficient}|^2$. That is what makes a *limit* (not only a lim sup) exist. Kedlaya and Ng remark that for a general Markov chain only the lim sup statement survives; abelianness is precisely what buys the exact limit.

Why does each hypothesis appear exactly where it does? Take each away.

- Drop $g_1 = e$: on $\mathbb{Z}/2$ with the die $\{1\}$, the walk alternates between 0 and 1, the sign character has $\lambda = -1$, and $D_m = \frac{1}{2}$ for every m — no $b < 1$ works. Having e on the die is what pins one of the $\chi(g_i)$ to 1 in Step 3, killing any rotation.
- Drop "generators": if the g_i lie in a proper subgroup H, a nontrivial character trivial on H has $\lambda_\chi = 1$ and the walk never leaves H.
- Drop $\subsetneq$: if the die is all of G, then p_1 is already uniform, $D_m = 0$ for all $m \geq 1$, and the limit is 0 for every b — not positive. This is Step 4.

Every hypothesis is load-bearing, and Step 3 and Step 4 are where each is spent.

Kedlaya and Ng's solution begins by building the $n \times n$ transition matrix M and asserting that $v_\chi = (\chi(h_i))$ is an eigenvector. Here is where that comes from: v_χ is an eigenvector because multiplication by g scales χ by $\chi(g)$, and M is an average of such multiplications. The rest of their solution is Plancherel written with explicit vectors.

Caveat. The final line of the official solution states the limit as $\#\{\chi : |\lambda_\chi| = b\}$. With their normalisation ($v = \frac{1}{n}\sum_\chi v_\chi$, $\|v_\chi\|^2 = n$) the squared norm is $\frac{1}{n^2} \cdot n \sum_{\chi \neq \mathbf{1}} |\lambda_\chi|^{2m}$, so the limit is $\frac{1}{n}\#\{\chi \neq \mathbf{1} : |\lambda_\chi| = b\}$; the official line omits the factor $\frac{1}{n}$. This does not affect the conclusion (positive and finite). Checked numerically: $\mathbb{Z}/6$ with die $\{0,1\}$ gives limit $0.3333 = 2/6$; $\mathbb{Z}/5$ with die $\{0,1\}$ gives $0.4000 = 2/5$; $\mathbb{Z}/7$ with die $\{0,1,3\}$ gives $0.8571 = 6/7$ — each matching N/n, not N.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "abelian group" together with "rolled m times and multiplied": a random walk on an abelian group. Write down: "characters diagonalise it; $\hat{p}_m(\chi) = \overline{\lambda_\chi}^m$". The second cue is "$\sum_x (\cdot)^2$": write "Plancherel".

2. Fourier analysis on a finite abelian group: transform, observe that convolution becomes multiplication, apply Plancherel, and extract the dominant eigenvalue.

3. Hesitation occurs at "I need the eigenvalues of T" when no formula for the characteristic polynomial is available. Reinforce with: "for a translation-invariant operator on an abelian group, the eigenvectors are the characters; compute eigenvalues by applying T to χ."

4. The trap is proving only $p_m \to u$ (Perron–Frobenius, ergodicity) and not the exact asymptotic. The problem wants a limit that is *positive and finite*, which needs both an upper bound and a lower bound at the same rate, and existence of the limit, which needs normality. A second trap: forgetting to prove $b > 0$ — it is the only place $\subsetneq$ is used, and without it the statement is false.

5. Slogan: *repeated random multiplication on an abelian group is diagonalised by characters; the L^2 distance to uniform is $\frac{1}{n}\sum_{\chi \neq 1} |\lambda_\chi|^{2m}$.* Recognise it faster by asking of any linear iteration: "what does this commute with, and what are the joint eigenvectors of those symmetries?"

6. Decision rule: **when an iteration commutes with a group of symmetries, find the functions those symmetries only rescale; they diagonalise the iteration, and quantities measured in L^2 split without cross terms.** Companion rule: when a problem lists several hypotheses and asks for a strict two-sided bound, test each hypothesis by deleting it — each deletion shows which inequality it guards.

▷ **Cue:** a random walk "multiply by a random element of S" on a finite abelian group $\longrightarrow$ **Move:** use characters: $\hat{p}_m(\chi) = \overline{\lambda_\chi}^m$ with $\lambda_\chi = \frac{1}{|S|}\sum_{s \in S} \chi(s)$

▷ **Cue:** a sum of squares $\sum_{x \in G} (f(x))^2$ over a finite abelian group $\longrightarrow$ **Move:** Plancherel: $\sum_x |f(x)|^2 = \frac{1}{n}\sum_\chi |\hat{f}(\chi)|^2$

▷ **Cue:** an average of unit complex numbers has modulus 1 $\longrightarrow$ **Move:** equality in the triangle inequality: all the numbers are equal

▷ **Cue:** need exact geometric rate b^m and a positive limiting constant $\longrightarrow$ **Move:** isolate the eigenvalues of maximal modulus below 1; the constant is their count (times the normalisation); requires the operator to be normal

▷ **Cue:** a list of hypotheses guarding a strict two-sided conclusion $0 < b < 1$ $\longrightarrow$ **Move:** delete each hypothesis in turn to find which inequality it protects, then use it exactly there

Putnam 2011 — Paper B

2011 B1

2011 B1. Let h and k be positive integers. Prove that for every $\epsilon > 0$, there are positive integers m and n such that

$$\epsilon < |h\sqrt{m} - k\sqrt{n}| < 2\epsilon.$$

Fields: Diophantine approximation, density of $\mathbb{Q}$ in $\mathbb{R}$, elementary estimates **Position:** B1 — the opener
Answer: Take $m = k^2(a^2 + b)$, $n = h^2a^2$ for positive integers a, b with $3a^2 > b$ and b/a a rational number lying inside $\left(\frac{3\epsilon}{hk}, \frac{4\epsilon}{hk}\right)$; then $\epsilon < h\sqrt{m} - k\sqrt{n} < 2\epsilon$.

Layer 1 THE WHAT — surface analysis

The natural attempt. Fix $n = 1$ and look at the sequence $x_m = h\sqrt{m} - k$ for $m = 1, 2, 3, \ldots$. This increases without bound, and its consecutive gaps

$$x_{m+1} - x_m = h(\sqrt{m+1} - \sqrt{m}) = \frac{h}{\sqrt{m+1} + \sqrt{m}}$$

shrink to 0 as $m \to \infty$. So choose M with $h/(2\sqrt{M}) < \epsilon$, i.e. $M > h^2/(4\epsilon^2)$; every gap past index M is then smaller than ϵ. The sequence $(x_m)_{m \geq M}$ climbs from some finite value to $+\infty$ in steps smaller than ϵ, and the target window $(\epsilon, 2\epsilon)$ has length ϵ – a step smaller than the window's length cannot jump clean over it. So there is a first index $m_0 \geq M$ with $x_{m_0} > \epsilon$; because the previous term was $\leq \epsilon$ and the jump to x_{m_0} is under ϵ, $x_{m_0} < 2\epsilon$ (with a one-line patch – move to $m_0 + 1$ – if $x_{m_0} = \epsilon$ exactly, since the next gap is again under ϵ). Taking $n = 1$, $m = m_0$ solves the problem.

This works, and it is honest mathematics: gaps shrinking to 0 plus a first-crossing argument is a standard way to show a set of achievable values must enter any small window far enough out. But notice the cost. A threshold $M(\epsilon)$ has to be produced and justified before anything else happens; the boundary case $x_{m_0} = \epsilon$ needs a patch; and the whole argument only ever uses $n = 1$, which never explains why the problem hands over *two* free integers m and n rather than one.

The model chain. Use both free integers at once. If $m = k^2M$ and $n = h^2N$ for positive integers M, N, then

$$h\sqrt{m} - k\sqrt{n} = hk\sqrt{M} - hk\sqrt{N} = hk\left(\sqrt{M} - \sqrt{N}\right):$$

the coefficients h, k have been pulled entirely outside a single square-root difference. Now set $M = a^2 + b$, $N = a^2$ for positive integers a, b, so that

$$h\sqrt{m} - k\sqrt{n} = hk\left(\sqrt{a^2 + b} - a\right).$$

This quantity is controlled directly. Rationalising,

$$\sqrt{a^2 + b} - a = \frac{b}{\sqrt{a^2 + b} + a}.$$

If also $3a^2 > b$ (so $a^2 + b < 4a^2$, i.e. $\sqrt{a^2+b} < 2a$), the denominator lies strictly between $2a$ and $3a$, giving the squeeze

$$\frac{b}{3a} < \sqrt{a^2+b} - a < \frac{b}{2a}.$$

So the whole problem reduces to choosing a rational number b/a, with $3a^2 > b$, lying in the window

$$\frac{3\epsilon}{hk} < \frac{b}{a} < \frac{4\epsilon}{hk},$$

since then $b/(3a) > \epsilon/hk$ and $b/(2a) < 2\epsilon/hk$, and multiplying through by hk gives exactly $\epsilon < hk(\sqrt{a^2+b} - a) < 2\epsilon$. Such a rational b/a exists because $\mathbb{Q}$ is dense in $\mathbb{R}$ and the target window is a nonempty open interval. Having found positive integers a, b realising that ratio, scale both up by a common positive integer factor t if needed to force $3a^2 > b$: the ratio b/a is unchanged by this scaling, while $3a^2$ grows like t^2 against b's t, so the inequality holds for t large. Set $m = k^2(a^2+b)$, $n = h^2a^2$; these are positive integers, and $\epsilon < h\sqrt{m} - k\sqrt{n} < 2\epsilon$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains want the same thing: make $h\sqrt{m} - k\sqrt{n}$ small and land it in the window $(\epsilon, 2\epsilon)$ by choosing m, n. They split at the very first decision – how to use the two free integers. The natural attempt treats them asymmetrically, freezing $n = 1$ and hunting for m along a single increasing sequence. The model chain treats m, n as a matched pair from the outset, writing $m = k^2M,\ n = h^2N$ so the coefficients h, k cancel out of the roots.

> *The problem gives me two free integers, not one – so before hunting along a sequence in a single variable, I should ask what choosing m and n together buys me. Setting $m = k^2M$, $n = h^2N$ makes $h\sqrt{m} = hk\sqrt{M}$ and $k\sqrt{n} = hk\sqrt{N}$: the coefficients h, k vanish from inside the roots, and the target collapses to hk times a single square-root difference $\sqrt{M} - \sqrt{N}$, which I get to name.*

> *$\sqrt{a^2+b} - a$, for b small compared to a^2, is a controllable ratio of two integers I choose, not a term buried inside an unbounded increasing sequence. Since $\mathbb{Q}$ is dense in $\mathbb{R}$, I can place b/a in any target window directly, with no threshold on a to compute in advance – "large enough" is only needed afterwards, for a crude bound, and it's free because scaling a, b together doesn't move b/a.*

Layer 3 THE WHY

Why does writing $m = k^2M$, $n = h^2N$ occur to the model solver first? Because $h\sqrt{k^2M} = hk\sqrt{M}$ is the most basic fact about square roots – a perfect-square factor comes straight out – and the goal is a single tunable real number instead of two sequences scaled by different constants h and k. Once the coefficients are unified into one prefactor hk, the entire problem is "make $\sqrt{M} - \sqrt{N}$ land in $(\epsilon/hk, 2\epsilon/hk)$," which is a cleaner question than the original.

The second fact driving the model chain is that $\sqrt{a^2+b} - a$, rationalised to $b/(\sqrt{a^2+b} + a)$, behaves for $b \ll a^2$ almost like $b/(2a)$ – a genuinely linear, order-preserving function of the ratio b/a. That is exactly what makes density of $\mathbb{Q}$ the right tool: once the window for $hk(\sqrt{a^2+b} - a)$ is translated (via the squeeze between $b/3a$ and $b/2a$, valid once $b < 3a^2$) into a window for the single real number b/a, placing a rational number inside an open interval is not an estimate at all, it is the definition of density. The natural attempt's crossing argument is, at bottom, doing the same job – covering a target interval with a discrete but sufficiently fine set of achievable values – but it has to rediscover that fineness by hand, via an explicit threshold $M(\epsilon)$ and a case check at the boundary, instead of invoking a fact (density of $\mathbb{Q}$ in $\mathbb{R}$) that is already available and requires no threshold at all.

The role of $3a^2 > b$ is purely to license the crude bound $\sqrt{a^2+b} < 2a$; it is not a separate creative step, only bookkeeping that the linear approximation to $\sqrt{a^2+b} - a$ is trustworthy,

and it can always be arranged after the ratio b/a is fixed, since scaling a, b by a common factor never changes that ratio.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that two integers m, n are free to choose, and they sit under square roots multiplied by fixed coefficients h, k whose squares are directly available. On seeing it, write down: try $m = k^2 \cdot (\text{something})$, $n = h^2 \cdot (\text{something})$ to cancel h, k out of the roots.

2. That cue suggests the problem reduces to controlling a bare square-root difference $\sqrt{M} - \sqrt{N}$ (times hk), which in turn suggests rationalising it into a fraction that can be squeezed between simple bounds.

3. Hesitation sets in exactly at the choice between the sequential/threshold route, which feels safe because it obviously terminates eventually, and a direct rational-density construction, which feels like it needs a new idea before any progress is visible. What reinforces the switch is noticing the sequential route still needs an explicit threshold $M(\epsilon)$ justified from scratch and an edge case patched, doubling the bookkeeping for a result the direct squeeze gives in one line.

4. The trap is treating "land in a small window $(\epsilon, 2\epsilon)$" as an automatic cue for a limiting or asymptotic argument. The presence of two free integers, not one, tied to fixed coefficients whose squares are available, is a signal that a direct two-parameter algebraic construction is both available and cheaper.

5. Recognise it faster: whenever the target is "a coefficient times $\sqrt{\text{integer}}$ lies in an interval," and the coefficient multiplies a root of something you may choose, check first whether absorbing that coefficient as a perfect-square factor turns the whole expression into a single tunable rational-controlled quantity.

6. Reusable rule: to force $c\sqrt{X}$ into a prescribed window for X a positive integer of your choosing, write $X = a^2 + b$, rationalise $c(\sqrt{a^2+b} - a) = cb/(\sqrt{a^2+b} + a)$, squeeze the denominator between $2a$ and $3a$ (valid once $b < 3a^2$), and place the rational b/a inside the window obtained by dividing the target window by c – density of $\mathbb{Q}$ in $\mathbb{R}$ answers that placement immediately.

▷ **Cue:** a target quantity is a fixed coefficient times a difference (or sum) of square roots of integers you may choose $\longrightarrow$ **Move:** try to absorb the coefficient as a perfect-square factor inside the root so it cancels, reducing to a bare square-root expression

▷ **Cue:** need $\sqrt{a^2+b} - a$ to land in a prescribed window, for integers a, b you may choose $\longrightarrow$ **Move:** rationalise to $b/(\sqrt{a^2+b} + a)$, squeeze the denominator between $2a$ and $3a$ (valid once $b < 3a^2$), and choose the rational b/a in the corresponding window using density of $\mathbb{Q}$ in $\mathbb{R}$

▷ **Cue:** a problem hands you two free integer parameters and a single asymmetric inequality to satisfy $\longrightarrow$ **Move:** look for a substitution that uses both parameters jointly (e.g. tied to squares of given coefficients) before defaulting to fixing one and searching over the other

2011 B2

2011 B2. Let S be the set of all ordered triples (p, q, r) of prime numbers for which at least one rational number x satisfies $px^2 + qx + r = 0$. Which primes appear in seven or more elements of S?

Fields: Number theory, discriminants, elementary modular bounding **Position:** B2 — the workhorse
Answer: Exactly the primes 2 and 5. Every other prime ℓ appears in at most six triples of S.

Layer 1 THE WHAT — surface analysis

The natural attempt. $px^2 + qx + r = 0$ has a rational root exactly when its discriminant $q^2 - 4pr$ is a perfect square, say a^2 for some integer $a \geq 0$: the quadratic formula gives $x = (-q \pm a)/(2p)$,

and this is rational precisely under that condition (a standard fact – an integer-coefficient quadratic's roots are rational iff the discriminant is a perfect square). So S is the set of prime triples (p, q, r) with $q^2 - 4pr$ a perfect square.

Faced with "which primes occur seven or more times," the direct instinct is to fix a candidate prime ℓ in the p-slot (say), then search: try $r = 2, 3, 5, 7, \ldots$ in turn, and for each, look for primes q making $q^2 - 4\ell r$ a perfect square, by trial. This does find triples – for $\ell = 2$, $r = 5$: $q^2 - 40 = a^2$ gives $q = 7, a = 3$ working ($(2, 7, 5) \in S$); for $\ell = 5$, $r = 2$: $q^2 - 40 = a^2$ again, same pair. Enough patient search this way, for small ℓ, does turn up seven or more hits for $\ell = 2$ and $\ell = 5$.

But this approach cannot prove the other half of the problem – that every prime $\ell \neq 2, 5$ (and $\ell = 3$) appears in at most six triples. Trial search over $r = 2, 3, 5, 7, \ldots$ and q never terminates in a proof; it only ever produces more examples or more non-examples, never a certificate that no further q, r work. To close the problem this way would require, for every candidate ℓ, an argument that no larger r or q can work – and nothing about brute casework on parities of p, q, r separately (odd/odd/odd, odd/odd/even, and so on, six cases) pins down cleanly which prime is forced to be small; it produces a sprawling case split with no single number falling out as "the" constrained one.

The model chain. From $q^2 - 4pr = a^2$: since $p, r \geq 2$, $4pr \geq 16$, so $q^2 > 16$, forcing $q \geq 5$; hence q is odd (all primes ≥ 5 are), and $a^2 = q^2 - 4pr$ is odd, so a is odd too. Every odd square is $\equiv 1 \pmod 8$ (write the odd number as $2j + 1$: $(2j + 1)^2 = 4j(j + 1) + 1$, and $j(j + 1)$ is always even, so the square is $\equiv 1 \pmod 8$). Hence $q^2 - a^2 \equiv 0 \pmod 8$, i.e. $4pr \equiv 0 \pmod 8$, i.e. pr is even. Since p, r are primes, one of them must equal 2.

Say $r = 2$. Then $q^2 - a^2 = 8p$, i.e. $(q - a)(q + a) = 8p$. Both factors are even (as q, a are odd) and positive (their sum is $2q > 0$ and they have the same sign), so writing $q - a = 2s$, $q + a = 2t$ with $0 < s \leq t$, $st = 2p$. Since p is prime, the only factorisations with $s \leq t$ are $(s, t) = (1, 2p)$ or $(2, p)$ (the latter only sensible when $p \neq 2$; when $p = 2$ it collapses into the same case), giving $q = s + t \in \{2p + 1,\ p + 2\}$. By the identical computation with $p = 2$ instead, $q \in \{2r + 1,\ r + 2\}$.

Now fix a target prime ℓ and ask in how many ordered triples of S it can appear, in any of the three slots. If ℓ occupies the p-slot with $r = 2$: need $q \in \{2\ell + 1, \ell + 2\}$ prime. If ℓ occupies the r-slot with $p = 2$: by the same formula (symmetric in p, r), again $q \in \{2\ell + 1, \ell + 2\}$ prime. If ℓ occupies the q-slot: either $p = 2$ and $\ell = 2r + 1$ or $\ell = r + 2$, i.e. $r = (\ell - 1)/2$ or $r = \ell - 2$ is prime; or $r = 2$ and symmetrically $p = (\ell - 1)/2$ or $p = \ell - 2$ is prime. So altogether there are exactly four candidate primality conditions governing every occurrence of ℓ:

$$2\ell + 1, \quad \ell + 2, \quad \tfrac{\ell - 1}{2}, \quad \ell - 2 \quad \text{prime.}$$

Each true condition produces exactly one triple with $p = 2$ and one with $r = 2$ – since $(p, q, r) \in S \iff (r, q, p) \in S$, discriminant being symmetric in p, r – so it contributes exactly 2 to ℓ's count. Hence ℓ appears in at most 8 triples in general, and *exactly* $2 \times$ (number of true conditions among the four).

For $\ell = 3$: $\ell - 2 = 1$ is not prime, killing one of the four conditions outright, so at most three conditions survive, giving at most 6 occurrences.

For $\ell \geq 7$ prime: the three numbers $\ell - 2, \ell, \ell + 2$ are spaced 2 apart, and since $\gcd(2, 3) = 1$, they cover all three residues mod 3 – exactly one of them is divisible by 3. It isn't ℓ itself ($\ell \geq 7$ is prime, so $\ell \neq 3$ means $3 \nmid \ell$). So it's $\ell - 2$ or $\ell + 2$, and since $\ell \geq 7$ makes that number $\geq 5 > 3$, divisibility by 3 makes it composite – killing one of the two conditions "$\ell + 2$ prime," "$\ell - 2$ prime." At most three of the four conditions survive, so at most 6 occurrences.

For $\ell = 2$ and $\ell = 5$ the bound simply is not tight, and the cleanest proof is to exhibit triples directly. $(2, 5, 2)$: $25 - 16 = 9 = 3^2$, valid. $(2, 5, 3)$: $25 - 24 = 1 = 1^2$, valid. $(2, 7, 5)$: $49 - 40 = 9 = 3^2$, valid. $(2, 11, 5)$: $121 - 40 = 81 = 9^2$, valid. Together with their reversals $(2, 5, 2)$ [self-reversed], $(3, 5, 2)$, $(5, 7, 2)$, $(5, 11, 2)$, this gives 2 appearing in (at least) $(2, 5, 2), (2, 5, 3), (2, 7, 5), (2, 11, 5), (3, 5, 2), (5, 7, 2), (5, 11, 2)$ – seven triples – and 5 appearing in

$(2,5,2)$ [twice over, as both slots], $(2,7,5)$, $(2,11,5)$, $(5,7,2)$, $(5,11,2)$ and more from the q-slot conditions – also seven or more. So 2 and 5 both clear the bound that every other prime is capped under.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start from the same discriminant condition $q^2 - 4pr = a^2$. They split at the question "now what?" The natural attempt turns immediately to numerical search over r and q for a fixed candidate ℓ. The model chain instead asks what constraint parity places on $q^2 - a^2 = 4pr$, and pushes that constraint to the modulus (mod 8) that actually forces one of p, r to be the specific prime 2 – collapsing three free primes to one before any search begins.

> *q and a are both forced odd here, and $q^2 - a^2$ is a difference of two odd squares – before searching for triples numerically, I should check what divisibility that forces on $4pr$, because a difference of two odd squares is always divisible by 8, and that is more than parity alone; it should pin down which of p, r has to be the smallest prime, not just that one of them is even.*

> *ℓ is not only allowed to sit in the p- or r-slot of a triple – it can equally be the q that some other pair $(2, \ell, r)$ or $(p, \ell, 2)$ produces. So the primality conditions I need to check for ℓ aren't just $2\ell + 1$ and $\ell + 2$; I also need to run the formula backwards, solving $\ell = 2r + 1$ or $\ell = r + 2$ for r, which brings in $(\ell - 1)/2$ and $\ell - 2$ as two more candidates.*

Layer 3 THE WHY

Why mod 8 and not mod 4 or mod 2? Because the target fact is not merely "$q^2 - a^2$ is even" (mod 2 gives that for free from parity) but a statement strong enough to force *one specific prime factor*, 2, out of pr entirely. Odd numbers squared always land on 1 mod 8 – this is not a coincidence to be checked case by case, it is the content of $(2j+1)^2 = 4j(j+1) + 1$ together with $j(j+1)$ always being a product of consecutive integers, hence even. Mod 4 would only give $q^2 - a^2 \equiv 0 \pmod 4$, i.e. $pr \equiv 0 \pmod 2$ up to a possible factor of 2 ambiguity that mod 8 resolves cleanly; mod 8 is the smallest modulus at which "odd squared" becomes a genuine constant, and that is exactly the level of information needed to conclude pr even without qualification, hence (both being prime) one of p, r equal to 2.

Why does the factorisation $(q - a)(q + a) = 8p$ then hand over a two-candidate formula rather than an open-ended search? Because p is prime, and the divisor pairs of $8p$ splitting into two same-parity factors are exhausted by a short, finite list – primality of p is what makes the factor pairs countable in the first place, converting "solve a quadratic Diophantine equation" into "enumerate the divisors of a number with exactly four prime-power-shaped factorisations."

Why do the extra two candidates $(\ell - 1)/2, \ell - 2$ exist at all, beyond the two from ℓ as p or r? Because the discriminant $q^2 - 4pr$ is symmetric in p and r – the equation $px^2 + qx + r$ and its reversal $rx^2 + qx + p$ have exactly the same discriminant – and this symmetry is precisely what makes "reversal" a free extra triple for every valid one found, and simultaneously what makes the same derived formula ($q \in \{2p + 1, p + 2\}$ when $r = 2$) do double duty: read forwards, it bounds occurrences of ℓ as p or r; read backwards (solving for the other unknown given $q = \ell$), it bounds occurrences of ℓ as q. Missing this second reading is the single most likely way to under-count and arrive at a wrong bound of 4 instead of 8 (before the mod-3 argument trims it to 6).

Finally, why does mod 3 finish the job? Because $\ell - 2, \ell, \ell + 2$ advance by steps of 2, and $\gcd(2, 3) = 1$ means three such steps visit every residue class mod 3 exactly once – a fact about arithmetic progressions with common difference coprime to the modulus, not a coincidence special to this problem. Exactly one of the three numbers is a multiple of 3; ruling out that it is ℓ itself (since $\ell \neq 3$ in this case) forces it to be $\ell - 2$ or $\ell + 2$, and since that number exceeds 3,

being a multiple of 3 makes it composite. One condition among the four dies for every prime $\ell \geq 7$, and separately for $\ell = 3$ the elementary fact that 1 is not prime kills the $\ell - 2$ condition directly.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "$px^2 + qx + r = 0$ has a rational root." On seeing it, write down immediately: this holds iff the discriminant $q^2 - 4pr$ is a perfect square – convert the root condition to a Diophantine condition before doing anything else.
2. That cue suggests treating $a^2 = q^2 - 4pr$ as an equation in a new integer unknown a, and specifically checking what parity and small-modulus constraints it places on p, q, r, since p, q, r are all primes and primes interact well with modular arithmetic.
3. Hesitation sets in when deciding which modulus to reduce mod – mod 2 (parity) is the reflexive first move and only shows q, a share parity, which is not enough to force any prime to be 2. What reinforces the push to mod 8 specifically is computing $(2j+1)^2$ mod 8 once, seeing it is always 1, and recognising that is exactly the extra strength needed to pin pr's parity down completely.
4. The trap is two-fold: first, believing numerical search can substitute for a bound – it can produce examples for 2 and 5 but can never certify that no other prime clears the bar; second, forgetting that the target prime ℓ can sit in the q-slot as well as the p- or r-slot, which silently halves the candidate list and produces a wrong (too generous) bound.
5. Recognise it faster: whenever a Diophantine condition reduces to "a difference of two squares equals a small multiple of a product of primes," go straight for the factorisation $(X - Y)(X + Y) = \ldots$ and enumerate the (necessarily few, since one factor is prime) divisor pairs, rather than searching numerically.
6. Reusable rule: to bound how often a target object (here a prime ℓ) can appear across all valid configurations, first find every distinct *role* it could play (here: p, r, or q) and derive the governing condition separately for each role by applying the same structural fact (the $r = 2$ formula) both forwards and backwards – an undercount almost always comes from missing a role, not from a wrong formula.

▷ **Cue:** quadratic $px^2 + qx + r = 0$ with integer coefficients has a rational root $\longrightarrow$ **Move:** set the discriminant $q^2 - 4pr$ equal to a perfect square a^2 and work with the resulting Diophantine equation

▷ **Cue:** an equation of the form (odd)^2 − (odd)^2 = (something) $\longrightarrow$ **Move:** reduce mod 8, since every odd square is $\equiv 1 \pmod 8$, forcing the difference to be $\equiv 0 \pmod 8$

▷ **Cue:** a difference of two squares equals a small multiple of a product involving a prime $\longrightarrow$ **Move:** factor as $(X - Y)(X + Y) = \ldots$ and enumerate the necessarily short list of divisor pairs directly

▷ **Cue:** counting occurrences of a target value across several structurally symmetric roles (here p, q, r in a symmetric condition) $\longrightarrow$ **Move:** derive the governing condition once, then apply it in every role the target could occupy, not only the first one noticed

2011 B3

2011 B3. Let f and g be (real-valued) functions defined on an open interval containing 0, with g nonzero and continuous at 0. If fg and f/g are differentiable at 0, must f be differentiable at 0?

Fields: Single-variable analysis, differentiability at a point, algebraic identities **Position:** B3 — the analysis problem **Answer:** Yes. Writing $u = fg$, $v = f/g$, the identity $f^2 = uv$ (valid wherever $g \neq 0$) forces f differentiable at 0: via the chain rule on $\sqrt{\cdot}$ when $f(0) \neq 0$, and via $f'(0) = v'(0)g(0)$, computed directly from the definition, when $f(0) = 0$.

Layer 1 THE WHAT — surface analysis

The natural attempt. Since g is nonzero at 0, fg and f/g multiply to f^2 exactly wherever $g \neq 0$ – an identity, true pointwise, not something that needs differentiability to justify: $(fg) \cdot (f/g) = f^2 \cdot (g/g) = f^2$. Write $u = fg$, $v = f/g$, both given differentiable at 0. The natural next move is to try to recover f itself by taking a square root: $f = \pm\sqrt{uv}$, hoping the chain rule for $\sqrt{\cdot}$ closes the problem in one step.

This works beautifully when $f(0) \neq 0$. There, $u(0)v(0) = f(0)^2 > 0$, uv is differentiable at 0 (a product of two functions each differentiable at that single point), and $\sqrt{\cdot}$ is differentiable at any positive argument, so the chain rule applies directly to $\pm\sqrt{uv}$ and gives f differentiable at 0 with essentially no further work.

But the same idea runs into a wall exactly where the problem needs it most: at $f(0) = 0$. There, $u(0)v(0) = 0$, and $\sqrt{\cdot}$ is not differentiable at 0 – its derivative $1/(2\sqrt{t})$ blows up as $t \to 0^+$. The formula $f = \pm\sqrt{uv}$ is still true as an identity, but differentiating it via the chain rule is no longer legitimate at the one point that matters. Having solved the generic case cleanly, the natural attempt stalls precisely at the boundary case, with no second idea yet in hand.

The model chain. First establish that f itself is continuous at 0, without assuming it. Since g is continuous and nonzero at 0, there is a neighbourhood of 0 on which $g \neq 0$ (continuity plus $g(0) \neq 0$ forces g to keep one sign, hence stay away from 0, nearby); on that neighbourhood $f = u/g$ identically, since $u = fg$ and $g \neq 0$ there. Now u is continuous at 0 (differentiable there, hence continuous there) and g is continuous and nonzero at 0, so $f = u/g$ is continuous at 0 as a quotient of two functions continuous at 0 with nonzero denominator.

Case $f(0) \neq 0$. By continuity of f at 0 and $f(0) \neq 0$, f keeps one sign, say $\varepsilon \in \{\pm 1\}$, on some neighbourhood of 0; there, $\varepsilon f = \sqrt{f^2} = \sqrt{uv}$. Now uv is a product of two functions each differentiable at the single point 0 (need not be differentiable anywhere else), so uv is differentiable at 0 by the product rule applied directly from the definition of the derivative at a point:

$$\frac{(uv)(x) - (uv)(0)}{x} = \frac{u(x) - u(0)}{x} v(x) + u(0) \frac{v(x) - v(0)}{x} \to u'(0)v(0) + u(0)v'(0)$$

as $x \to 0$, using $v(x) \to v(0)$ (differentiability of v at 0 gives continuity of v at 0). Since $(uv)(0) = f(0)^2 > 0$ and $\sqrt{\cdot}$ is differentiable at any positive point, the chain rule applies to $\varepsilon f = \sqrt{uv}$ at 0, giving f differentiable at 0.

Case $f(0) = 0$. Then $v(0) = f(0)/g(0) = 0$, so by the definition of the derivative,

$$v'(0) = \lim_{x \to 0} \frac{v(x) - v(0)}{x} = \lim_{x \to 0} \frac{f(x)}{xg(x)}$$

exists, by hypothesis. Since g is continuous at 0 with $g(0) \neq 0$, $\lim_{x \to 0} g(x) = g(0) \neq 0$ also exists. Multiplying the two existing limits,

$$\lim_{x \to 0} \frac{f(x)}{x} = \lim_{x \to 0} \frac{f(x)}{xg(x)} \cdot g(x) = v'(0)\, g(0),$$

which exists. Since $f(0) = 0$, this is exactly $f'(0)$, so f is differentiable at 0 with $f'(0) = v'(0)g(0)$.

Either way, f is differentiable at 0.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains open identically: form $u = fg$, $v = f/g$, note $uv = f^2$ pointwise, and reach for $f = \pm\sqrt{uv}$. They agree completely on the case $f(0) \neq 0$, where that formula differentiates cleanly. They split exactly at $f(0) = 0$: the natural momentum is to keep pushing the same square-root formula through the chain rule; the model chain instead recognises that $t \mapsto \sqrt{t}$ is not differentiable at $t = 0$, abandons the chain rule at that single point, and returns to the raw definition of the derivative, working with $v(x)/x$ directly.

> *$f^2 = uv$ is an identity, true at every point near 0 where $g \neq 0$, not merely a numerical fact at 0 – and since u, v are each differentiable at the single point 0, their product is differentiable at 0 too, by the product rule applied directly from the definition, even though I have no idea whether u, v are differentiable anywhere else.*

> *I can't just assume f keeps one sign near 0 – but I don't need to assume it: g is continuous and nonzero at 0, so on some interval $g \neq 0$, and there $f = u/g$ identically; since u is continuous at 0 (being differentiable there) and g is continuous and nonzero at 0, their quotient f is continuous at 0, and continuity plus $f(0) \neq 0$ is what forces f to keep one sign nearby – the sign-consistency falls out of algebra I already have, not an extra hypothesis I need to invent.*

> *When $f(0) = 0$, $\sqrt{\cdot}$ is exactly the wrong outer function to differentiate at 0 – so drop the square root and go back to the definition: $v'(0) = \lim_{x\to 0} f(x)/(xg(x))$ already exists by hypothesis, and $g(x) \to g(0) \neq 0$, so multiplying these two existing limits together hands me $\lim_{x\to 0} f(x)/x$ for free.*

Layer 3 THE WHY

Why is $f^2 = uv$ the load-bearing identity, rather than some other combination of u and v? Because the problem gives exactly two differentiable quantities, $u = fg$ and $v = f/g$, built from f and g with no other stated regularity on f itself; any usable relation has to be an algebraic consequence that holds regardless of whether f is well-behaved, and g cancelling out of $uv = (fg)(f/g) = f^2$ is exactly such a relation – it needs nothing beyond $g \neq 0$, which is already given at 0 and hence nearby.

Why does a product of two functions, each merely differentiable at one point, turn out to be differentiable at that point, with no assumption of differentiability nearby? Because the definition of the derivative at a point is itself a statement about a single limit, and the algebraic identity

$$\frac{u(x)v(x) - u(0)v(0)}{x} = \frac{u(x) - u(0)}{x} v(x) + u(0)\frac{v(x) - v(0)}{x}$$

only ever uses u, v at the two points x and 0, plus the fact that v is continuous at 0 (itself implied by differentiability at 0, not an extra hypothesis). No behaviour of u, v away from those two points is invoked. This is the general fact that differentiability, unlike smoothness in a neighbourhood, is a purely local (in fact pointwise) notion once the definition is unpacked, and the product rule survives being applied at a single isolated point of differentiability for exactly that reason.

Why does the square root break at $t = 0$, and why does that matter here rather than being a technicality to route around quickly? Because $\sqrt{\cdot}$ has derivative $1/(2\sqrt{t}) \to \infty$ as $t \to 0^+$ – this is not a gap in the argument to be patched, it is a genuine failure of the chain rule's hypothesis (the outer function must be differentiable at the inner function's value), and the problem's phrasing – allowing $f(0) = 0$ at all – is precisely constructed to force a solver through this failure. The fact that the fallback (case $f(0) = 0$) succeeds by a completely different route, computing $\lim f(x)/x$ as a product of two limits that already exist rather than via any composition, is what makes the two-case structure necessary rather than an avoidable inelegance: there is no single formula that differentiates through $t = 0$, so no unified one-line proof exists, and the case split is not optional.

Finally, why is $f = u/g$ (rather than $f = vg$, which would work symmetrically) the right identity to extract continuity and sign-consistency of f from? Either works in principle, but u/g uses only u (differentiable, hence continuous, at 0) and g (given continuous and nonzero at 0) – the two ingredients the problem hands over with the least processing – whereas deriving the

same conclusion from v alone would additionally require knowing g is nonzero throughout a neighbourhood, which is true (continuity plus $g(0) \neq 0$ gives it) but is one more fact to invoke rather than one already sitting in the hypothesis.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that fg and f/g are given together, and g cancels between them. On seeing it, write down immediately: $(fg) \cdot (f/g) = f^2$, an identity requiring nothing but $g \neq 0$.

2. That cue suggests recovering f from f^2 by a square root, and differentiating $\sqrt{uv}$ via the chain rule, since u, v are both given differentiable at 0.

3. Hesitation sets in exactly at $f(0) = 0$, where the natural continuation (differentiate $\sqrt{uv}$ by the chain rule) silently needs $uv(0) \neq 0$ and nobody notices the omission until they ask directly "is $\sqrt{\cdot}$ differentiable at 0?" What reinforces the switch away from the chain rule is answering that question explicitly and seeing it is false, which forces a return to the bare definition of the derivative rather than a patched version of the same formula.

4. The trap is treating $\sqrt{f^2} = |f|$ as a formula that differentiates painlessly everywhere the identity holds, when in fact it is exactly as differentiable as $\sqrt{\cdot}$ is at the value $f(0)^2$ – which fails precisely at $f(0) = 0$, the case the problem is built to test.

5. Recognise it faster: whenever a derivative is being recovered by taking a square root of a product identity, check the value of that product at the point in question *before* reaching for the chain rule; a zero value there is a standing signal to fall back on the raw definition of the derivative instead of trying to patch the composition.

6. Reusable rule: to prove a target function h is differentiable at a point using only that certain algebraic combinations of h are differentiable there, first find a pointwise algebraic identity expressing something clean (here h^2) in terms of the given combinations, with no unproven hypothesis on h itself; differentiate that identity directly from the definition at the single point; then invert it via the chain rule where the inversion is smooth (nonzero value), and via a direct limit computation using one of the original given quantities where it is not.

▷ **Cue:** given that fg and f/g are differentiable at a point, with g continuous and nonzero there $\longrightarrow$ **Move:** form the identity $f^2 = (fg)(f/g)$, valid wherever $g \neq 0$, and note it is a product of two functions each differentiable at that single point

▷ **Cue:** about to differentiate $\sqrt{\text{something}}$ via the chain rule at a specific point $\longrightarrow$ **Move:** check the something's value at that point is nonzero first – $\sqrt{\cdot}$ is not differentiable at 0

▷ **Cue:** need f continuous or sign-consistent near a point, but only fg (or f/g) and g are known continuous there $\longrightarrow$ **Move:** write $f = (fg)/g$ (or $f = (f/g) \cdot g$), valid where $g \neq 0$, to get f's continuity or sign for free as a quotient (or product) of known-continuous pieces

2011 B4

2011 B4. In a tournament, 2011 players meet 2011 times to play a multiplayer game. Every game is played by all 2011 players together and ends with each of the players either winning or losing. The standings are kept in two 2011×2011 matrices, $T = (T_{hk})$ and $W = (W_{hk})$. Initially, $T = W = 0$. After every game, for every (h, k) (including for $h = k$), if players h and k tied (that is, both won or both lost), the entry T_{hk} is increased by 1, while if player h won and player k lost, the entry W_{hk} is increased by 1 and W_{kh} is decreased by 1.

Prove that at the end of the tournament, $\det(T + iW)$ is a non-negative integer divisible by 2^{2010}.

Fields: linear algebra (Hermitian and Gram matrices, determinants); Gaussian integers **Position:** B4 — mid-paper; one factorisation does all the work, and the naive congruence argument gets exactly half the required power of 2 **Answer:** $T + iW = \overline{A}^T A$ for a 2011×2011 matrix A with entries in $\{1, i\}$, so $\det(T + iW) = |\det A|^2 = 2^{2010}(a^2 + b^2)$ with $a, b \in \mathbb{Z}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. Write $n = 2011$. A strong student reads off the structure of the matrix first. T is symmetric ("h and k tied" is symmetric), W is antisymmetric (every update to W_{hk} is mirrored with the opposite sign in W_{kh}), and $T_{hh} = n$ because a player always ties with himself. So $M = T + iW$ satisfies $\overline{M}^T = T^T - iW^T = T + iW = M$: it is Hermitian. That gives "$\det M$ is real" at once, since the eigenvalues of a Hermitian matrix are real.

For divisibility the student works modulo 2. In each game the pair (h, k) either ties or does not, so if d_{hk} is the number of games in which exactly one of h, k won, then $T_{hk} = n - d_{hk}$ and $W_{hk} \equiv d_{hk} \pmod 2$. With n odd,

$$T_{hk} + iW_{hk} \equiv 1 - d_{hk} + i\,d_{hk} = 1 - (1-i)d_{hk} \pmod 2.$$

So modulo the Gaussian prime $1 - i$ (note $2 = i(1-i)^2$), every entry of M is $\equiv 1$: M is the all-ones matrix modulo $1 - i$. Subtract the first row from every other row; each of the remaining $n - 1$ rows becomes divisible by $1 - i$, so $(1-i)^{n-1} \mid \det M$ in $\mathbb{Z}[i]$. Taking norms, $2^{n-1} \mid (\det M)^2$, and since $\det M$ is a rational integer this gives only

$$2^{(n-1)/2} = 2^{1005} \mid \det M.$$

That is the wall: the congruence argument sees one factor of $(1 - i)$ per row, and needs two. Pushing it (working modulo 2 instead of modulo $1 - i$) fails because M is not of rank 1 modulo 2. Non-negativity is also still missing: Hermitian gives real, not non-negative.

The model chain. Number the games $1, \ldots, n$ with $n = 2011$. Define the $n \times n$ matrix $A = (a_{jk})$ by

$$a_{jk} = \begin{cases} 1 & \text{player } k \text{ wins game } j, \\ i & \text{player } k \text{ loses game } j. \end{cases}$$

Step 1: $M = T + iW = \overline{A}^T A$**.** The (h, k) entry of $\overline{A}^T A$ is $\sum_j \overline{a_{jh}}\, a_{jk}$. Check the four cases for one game j: both win: $1 \cdot 1 = 1$; both lose: $(-i)(i) = 1$; h wins, k loses: $1 \cdot i = i$; h loses, k wins: $(-i) \cdot 1 = -i$. These are exactly the increments of $T_{hk} + iW_{hk}$ in game j: $+1$ for a tie, $+i$ when h beats k, $-i$ when k beats h (because then W_{kh} goes up by 1 and W_{hk} down by 1). Summing over j gives the identity, including $h = k$.

Step 2: non-negative integer. $\det M = \det \overline{A}^T \det A = \overline{\det A} \det A = |\det A|^2$. Since the entries of A are Gaussian integers, $\det A = a + bi$ with $a, b \in \mathbb{Z}$, so $\det M = a^2 + b^2$ is a non-negative integer.

Step 3: the power of 2. Subtract row 1 of A from each row $j \geq 2$; the determinant is unchanged. Each new entry $a_{jk} - a_{1k}$ lies in $\{0,\ 1 - i,\ i - 1\}$, so each of rows $2, \ldots, n$ equals $(1 - i)$ times a row of Gaussian integers. Pulling out these factors,

$$\det A = (1-i)^{n-1}(c + di), \qquad c, d \in \mathbb{Z}.$$

Hence $\det M = |1 - i|^{2(n-1)}(c^2 + d^2) = 2^{2010}(c^2 + d^2)$. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains observe that M is Hermitian. They split at the next line.

(N) "M is Hermitian, so $\det M$ is real; now reduce M modulo 2."

(M) "M is Hermitian and is a sum over games; is each game's contribution of the form $\overline{v}^T v$? Then $M = \overline{A}^T A$."

On (N) the determinant is attacked directly and every factor of $1 - i$ is counted once. On (M) the determinant is a squared modulus, so every factor of $1 - i$ found in A is counted twice, and non-negativity comes free.

> *The target is "non-negative" and "divisible by 2^{2010}". Non-negative determinants come from squares:* $\det(\overline{A}^T A) = |\det A|^2$. *And* $2^{2010} = |1-i|^{2\cdot 2010}$ *— that is* 2010 *factors of* $(1-i)$ *in something, squared. So I should stop looking at M and look for the A that M is the Gram matrix of.*

> *M is a sum over games, and each game adds a Hermitian matrix with entries* 1 *and* $\pm i$ *of rank one if I am lucky. I need numbers* a_k *with* $\overline{a_h} a_k = 1$ *on ties and* $= i$ *when h wins and k loses. Try* $a = 1$ *for a win and* $a = i$ *for a loss.*

- ▷ **Cue:** a matrix built as a sum of contributions from independent events (games, observations), and the question asks for a non-negative determinant $\longrightarrow$ **Move:** write each contribution as $\overline{v}^T v$ and the matrix as a Gram matrix $\overline{A}^T A$; then $\det = |\det A|^2$
- ▷ **Cue:** a required power of 2 that is twice what a congruence argument gives $\longrightarrow$ **Move:** the quantity is a norm $|z|^2$; find z and put the factors of $1-i$ (or 2) in z
- ▷ **Cue:** a matrix whose rows agree modulo some element π except for the first $\longrightarrow$ **Move:** subtract the first row from the others: π^{n-1} divides the determinant
- ▷ **Cue:** a symmetric matrix T and antisymmetric W appearing as $T + iW$ $\longrightarrow$ **Move:** recognise a Hermitian matrix; ask whether it is a Gram matrix

Layer 3 THE WHY

Why is the factorisation the thing to look for, rather than a lucky guess?

Start with what the conclusion says. "Non-negative" is a positivity statement about a determinant. There are very few general reasons a determinant is non-negative, and the one available here is the Gram structure: $\det(\overline{A}^T A) = |\det A|^2$. The student already found that M is Hermitian, which is necessary for $M = \overline{A}^T A$; Hermitian is the shadow of a Gram matrix, and the question "whose Gram matrix?" is the next question to ask, not a trick.

Second, the problem is built additively: T and W start at 0 and each game adds something. A matrix that is a sum over n events of rank-one Hermitian pieces $\overline{v_j}^T v_j$ is, by the definition of matrix multiplication, $\overline{A}^T A$ where A has rows v_j. So the only thing to find is v_j for one game, and the requirement is a two-line table: $\overline{a_h} a_k$ must be 1 when h, k agree and $\pm i$ when they disagree. Take $a = 1$ for a win; then a loss must satisfy $\overline{1} \cdot a = i$, so $a = i$, and the both-lose check $\overline{i}\, i = 1$ confirms it. Nothing is guessed; the table forces it.

Third, the number 2^{2010}. It is 2^{n-1}, and $n-1$ is the number of rows left after one is used as a reference — the fingerprint of "subtract the first row from the others". The natural attempt finds exactly this, but on M, where it produces $(1-i)^{n-1}$ and hence only $2^{(n-1)/2}$. The official solution by Kedlaya and Ng opens with the definition of A and does not say where it comes from. Here is where it comes from: the natural attempt's row-subtraction is the right move applied to the wrong matrix. Applied to A, each row difference is 0 or $\pm(1-i)$, and because $\det M = |\det A|^2$ the factor $(1-i)^{n-1}$ is squared into $|1-i|^{2(n-1)} = 2^{n-1}$. The two conclusions of the problem — non-negative, and a power of 2 of exactly that size — are both consequences of one fact: M is a squared modulus.

Why is $1-i$ the right prime to think about? In $\mathbb{Z}[i]$ the number 2 is not prime: $2 = i(1-i)^2$, and $1-i$ is the unique Gaussian prime above 2. Every Gaussian integer is $\equiv 0$ or 1 modulo $1-i$, which is why 1 and i (the win and lose values) agree modulo $1-i$ and their difference is exactly $\pm(1-i)$. That is the structural reason a factor appears in each row.

Caveat. The identity $\det(T + iW) \in 2^{n-1}\mathbb{Z}_{\geq 0}$ was also checked by exact computation on random tournaments with $n = 3, \ldots, 7$ players (and n games). Note that the argument uses that the number of games equals the number of players only to make A square; with more games than players one would need Cauchy–Binet instead.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "non-negative integer" about a determinant of a matrix whose real part is symmetric and imaginary part antisymmetric. Write down: M is Hermitian; is it $\overline{A}^T A$?
2. The cue suggests factorising M as a Gram matrix over the Gaussian integers, then extracting factors of $1-i$ from $\det A$ by row operations.
3. Hesitation occurs after the congruence argument returns 2^{1005}: the student suspects an arithmetic slip rather than a structural loss. It is reinforced away by recognising the halving as the signature of a norm: whenever the bound you get is the square root of the bound you need, the object is $|z|^2$ and the factors belong in z.
4. The trap is working with T and W separately, or with M directly, where every factor is counted once. A secondary trap is claiming non-negativity from "Hermitian" alone; Hermitian matrices can have negative determinants.
5. Recognise it faster by reading the build-up of the matrix: "initially 0, after every event add ..." means the matrix is a sum of per-event matrices; test whether each one is rank one of the form $\overline{v}^T v$.
6. Reusable decision rule: *when a determinant must be shown non-negative, look for a Gram factorisation; once found, prove divisibility on the factor, where every divisor counts twice.* More generally: *if an argument yields exactly half the exponent required, the quantity is a norm or a square.*

2011 B5

2011 B5. Let $a_1, a_2, \ldots$ be real numbers. Suppose that there is a constant A such that for all n,

$$\int_{-\infty}^{\infty} \left(\sum_{i=1}^{n} \frac{1}{1+(x-a_i)^2} \right)^2 dx \leq An.$$

Prove there is a constant $B > 0$ such that for all n,

$$\sum_{i,j=1}^{n} (1+(a_i-a_j)^2) \geq Bn^3.$$

Fields: real analysis (integrals, convolution of kernels); inequalities (Cauchy–Schwarz) **Position:** B5 — looks like hard analysis, is a three-line inequality once the square is expanded; the difficulty is refusing to reason about the configuration of the a_i **Answer:** Expanding the square turns the hypothesis into $\sum_{i,j} f(a_i-a_j) \leq An$ with $f(y) \geq \frac{1}{4(1+y^2)}$; Cauchy–Schwarz against $\sum_{i,j}(1+(a_i-a_j)^2)$ gives $B = \frac{1}{4A}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. Write $\varphi(x) = \frac{1}{1+x^2}$ and $\Phi_n(x) = \sum_{i=1}^{n} \varphi(x-a_i)$. The student reads the hypothesis geometrically: each $\varphi(x-a_i)$ is a bump of height 1 centred at a_i; if many a_i crowd together, the bumps stack and the squared integral is large. So the hypothesis should say "the a_i are spread out", and the conclusion (large pairwise squared distances) should follow from spreading.

The first quantitative step works. Suppose k of the points $a_1, \ldots, a_n$ lie in a unit cell $[z, z+1)$. For x in that cell each of those k points has $|x-a_i| < 1$, so $\varphi(x-a_i) > \frac{1}{2}$ and $\Phi_n(x) > \frac{k}{2}$. Integrating over the cell only,

$$\frac{k^2}{4} < \int_z^{z+1} \Phi_n^2 \leq An, \qquad \text{so} \qquad k < 2\sqrt{An}.$$

So no unit cell holds more than $K = 2\sqrt{An}$ points, and the points occupy at least $n/K = \frac{1}{2}\sqrt{n/A}$ distinct cells.

Now the student needs a lower bound for $\sum_{i,j}(a_i - a_j)^2 = 2n\sum_i(a_i - \bar{a})^2$ (with $\bar{a}$ the mean) given at most K points per unit cell. Intuitively the minimum is attained by packing the points as tightly as possible around $\bar{a}$: K per cell, occupying $m \approx n/K$ consecutive cells, giving $\sum_i(a_i - \bar{a})^2 \approx K \cdot \frac{m^3}{12} = \frac{nm^2}{12}$ and hence

$$\sum_{i,j}(a_i - a_j)^2 \approx \frac{n^2m^2}{6} \approx \frac{n^3}{24A}.$$

So the exponent 3 does come out. But making "tight packing is the worst case" rigorous needs a rearrangement argument (at most $K(2t+2)$ points within distance t of $\bar{a}$, then summing over dyadic or unit shells), with care over the rounding in m and the cells straddling $\bar{a}$. It is correct in outline and it is a page of bookkeeping — and the first step threw away almost all of the hypothesis by integrating over a single cell.

The model chain. **Step 1: expand the square.** All functions involved are non-negative and integrable, so

$$\int_{-\infty}^{\infty} \Phi_n(x)^2\,dx = \sum_{i,j=1}^{n} \int_{-\infty}^{\infty} \varphi(x-a_i)\varphi(x-a_j)\,dx = \sum_{i,j=1}^{n} f(a_i - a_j),$$

where, substituting $x \mapsto x + a_j$,

$$f(y) = \int_{-\infty}^{\infty} \frac{dx}{(1+x^2)\,(1+(x+y)^2)}.$$

The hypothesis becomes $\sum_{i,j} f(a_i - a_j) \le An$.

Step 2: a lower bound for f. $f(-y) = f(y)$ (substitute $x \mapsto -x$), so take $y \ge 0$. For $x \in [-1,0]$ we have $1 + x^2 \le 2$ and $x + y \in [y-1, y]$, so $(x+y)^2 \le \max\{y^2, (y-1)^2\} \le y^2 + 1$ and $1 + (x+y)^2 \le 2 + y^2 \le 2(1+y^2)$. The integrand is positive everywhere, so keeping only $[-1,0]$,

$$f(y) \ge \int_{-1}^{0} \frac{dx}{4(1+y^2)} = \frac{1}{4(1+y^2)}.$$

Writing $d_{ij} = a_i - a_j$, Step 1 gives

$$\sum_{i,j=1}^{n} \frac{1}{1+d_{ij}^2} \le 4An.$$

Step 3: Cauchy–Schwarz. For the n^2 positive numbers $u_{ij} = 1 + d_{ij}^2$,

$$n^4 = \Big(\sum_{i,j} \sqrt{u_{ij}} \cdot \tfrac{1}{\sqrt{u_{ij}}}\Big)^2 \le \Big(\sum_{i,j} u_{ij}\Big)\Big(\sum_{i,j} \frac{1}{u_{ij}}\Big) \le \Big(\sum_{i,j} u_{ij}\Big) \cdot 4An.$$

(Here $A > 0$: the case $n = 1$ of the hypothesis reads $\int \varphi^2 = \pi/2 \le A$.) Hence

$$\sum_{i,j=1}^{n} \big(1 + (a_i - a_j)^2\big) \ge \frac{1}{4A}\,n^3,$$

and $B = \frac{1}{4A}$ works. □

Remark. Exactly, $f(y) = \frac{2\pi}{4+y^2}$ (the convolution of two Cauchy densities of scale 1 is a Cauchy density of scale 2), which gives $f(y) \ge \frac{\pi}{2(1+y^2)}$ and the better constant $B = \frac{\pi}{2A}$. The proof needs only some bound of the form $f(y) \ge \frac{c}{1+y^2}$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains write down Φ_n. They split at the first line of work on the hypothesis.

(N) "The integrand is large where the a_i cluster; bound the number of a_i in a cell."

(M) "The integrand is the square of a sum; expand it into a double sum over (i, j)."

On (N) the hypothesis is used locally and lossily, and the conclusion must be rebuilt from a configuration. On (M) the hypothesis is already a double sum over the same pairs (i, j) as the conclusion, and the rest is comparing two sums term by term.

> *The conclusion is a sum over all pairs (i, j). The square of a sum over i is a sum over pairs (i, j). So expand the square first and see what each pair contributes — before I try to picture where the points are.*

> *Now I have $\sum g(d_{ij}) \le An$ with g roughly $\frac{1}{1+d^2}$, and I want $\sum(1+d^2)$ large. These are reciprocals of each other, summed over the same n^2 pairs. Cauchy–Schwarz: their product is at least $(n^2)^2$. And $n^4/n = n^3$ — the exponent in the question. That is the proof.*

- ▷ **Cue:** the square of a sum inside an integral (or an expectation) $\longrightarrow$ **Move:** expand into a double sum of pairwise interaction terms; each term depends on one pair only
- ▷ **Cue:** an integral $\int \varphi(x-a)\varphi(x-b)\,dx$ of two translates $\longrightarrow$ **Move:** it is a function $f(a-b)$ of the difference only; bound it below by restricting to an interval where both factors are controlled
- ▷ **Cue:** an upper bound on $\sum_k 1/u_k$ and a required lower bound on $\sum_k u_k$ over the same N terms $\longrightarrow$ **Move:** Cauchy–Schwarz (or AM–HM): $\sum u_k \cdot \sum 1/u_k \ge N^2$
- ▷ **Cue:** a required exponent (here n^3) in an inequality $\longrightarrow$ **Move:** check it against the candidate method before doing any work: $N^2/(\text{hypothesis bound}) = n^4/n$

Layer 3 THE WHY

Why is expanding the square natural, and why is Cauchy–Schwarz the only thing left to do? Look at the objects. The conclusion is $\sum_{i,j=1}^{n}$ of a function of $a_i - a_j$: a pairwise quantity. The hypothesis is $\int(\sum_i \cdot)^2$. The definition of squaring a sum is $\left(\sum_i u_i\right)^2 = \sum_{i,j} u_i u_j$ — a pairwise quantity. When a hypothesis and a conclusion are both secretly sums over the same index set, the first move is to make that visible and compare them term by term. The integral then passes inside the finite sum, and each pair contributes $\int \varphi(x-a_i)\varphi(x-a_j)\,dx$, which by translation invariance of dx depends only on $a_i - a_j$. That is the circle-and-centre moment: the problem is phrased as an integral, but the integral is only packaging for a pairwise kernel.

Next, what does the kernel look like? Two bumps of width 1 at distance $|y|$ overlap by an amount that decays like the bumps themselves, $\sim 1/y^2$. So $f(y) \asymp \frac{1}{1+y^2}$, and to use it one needs only a lower bound of that shape. The cheapest way to bound a positive integral from below is to throw away everything except an interval where you can bound the integrand from below — one unit interval next to the origin suffices. No Fourier transform, no residues. (Kedlaya and Ng use $[-1, 0]$ too, with the slightly weaker estimate $1 + (x+y)^2 \le 1 + (1+y)^2$, which gives 6 instead of 4; the constant does not matter.)

Now the hypothesis says $\sum \frac{1}{1+d_{ij}^2} \lesssim n$ and the conclusion wants $\sum(1+d_{ij}^2) \gtrsim n^3$. Once they are written next to each other, the pairing of u and $1/u$ is on the page. The only general inequality linking $\sum u$ and $\sum 1/u$ is Cauchy–Schwarz (equivalently AM–HM): $\sum u \cdot \sum \frac{1}{u} \ge N^2$ with $N = n^2$ terms. The exponent check $n^4/n = n^3$ matches the problem exactly, which is the confirmation that this is the intended route and that nothing sharper about the configuration is needed.

Finally, why does the natural attempt cost so much? It uses the hypothesis on one unit cell, discarding the interaction between points in different cells, and then has to recover the global spread by a rearrangement argument. The model uses the whole integral at once. The general

lesson: a hypothesis stated as an integral of a square is an L^2 statement, and L^2 statements are about pairs; a counting argument about individual cells is an L^∞-type use of it, and loses.

Caveat. The natural attempt's final step ("tight packing minimises $\sum(a_i-\bar{a})^2$ under at most K points per unit cell") is presented as a heuristic giving the right order n^3; it is not proved here, and it is not needed for the model proof. The exact value $f(y)=\frac{2\pi}{4+y^2}$ in the remark was checked numerically for $y=0$, 0.5, 1, 3, 10; the proof uses only the elementary bound $f(y)\geq\frac{1}{4(1+y^2)}$, which is proved in full. As the official remark notes, the conclusion can be strengthened to Bn^4; that is not pursued here.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the square of a sum inside the integral, sitting next to a conclusion summed over i,j. On seeing it, write $\left(\sum_i u_i\right)^2=\sum_{i,j}u_iu_j$ and push the integral inside.
2. The cue suggests reducing to a pairwise kernel $f(a_i-a_j)$, bounding it below by $\frac{c}{1+y^2}$, and applying Cauchy–Schwarz between the reciprocal sums.
3. Hesitation occurs at the kernel: the student sees $\int\frac{dx}{(1+x^2)(1+(x+y)^2)}$ and thinks it must be evaluated exactly (partial fractions, residues). It is reinforced away by asking what the proof needs: a lower bound of the right shape, which one unit interval supplies.
4. The trap is the geometric picture: "the points must be spread out" is true and leads to a cell-counting argument that works only after a long rearrangement lemma. A second trap is trying to compute f exactly and getting lost in partial fractions.
5. Recognise it faster by matching index sets: when the conclusion sums over pairs and the hypothesis contains a square of a single sum, the problem is already a comparison of two pairwise sums.
6. Reusable decision rule: *expand squares of sums into pairwise sums before reasoning about configurations; when you then hold a bound on $\sum 1/u$ and need one on $\sum u$, apply Cauchy–Schwarz and check the exponent before doing anything else.*

2011 B6

2011 B6. Let p be an odd prime. Show that for at least $(p+1)/2$ values of n in $\{0,1,2,\ldots,p-1\}$,

$$\sum_{k=0}^{p-1}k!n^k \qquad \text{is not divisible by } p.$$

Fields: number theory (Wilson's theorem, polynomials over $\mathbb{F}_p$); algebra (repeated roots via the derivative) **Position:** B6 — genuinely hard; two ideas, each forced once you ask the right question: turn factorials into reciprocal factorials, then count roots of a truncated exponential by making them double roots **Answer:** With $0^0=1$: by Wilson, $f(n)\equiv -g(-1/n)$ for $n\neq 0$, where $g(x)=\sum_{k<p}x^k/k!$; every root of g in $\mathbb{F}_p$ is a double root of the degree-p polynomial $x^p-x+g(x)$, so g (hence f) has at most $(p-1)/2$ roots.

Layer 1 THE WHAT — surface analysis

Throughout, $f(x)=\sum_{k=0}^{p-1}k!\,x^k$ as a polynomial over $\mathbb{F}_p=\mathbb{Z}/p\mathbb{Z}$, with the convention $0^0=1$ so that $f(0)=1$. The claim is: f has at most $(p-1)/2$ roots in $\mathbb{F}_p$.

The natural attempt. The student computes. For $p=3$: $f=1+x+2x^2$, values $1,1,2$ at $x=0,1,2$ — no roots. For $p=5$: $f\equiv 1+x+2x^2+x^3-x^4$, values $1,4,3,3,0$ — one root, $x=4$. For $p=7$ one root, for $p=13$ two, for $p=37$ four (by machine). The bound $(p-1)/2$

is nowhere near tight, which is discouraging: there is no visible structure pairing roots with non-roots.

Next the student tries the general tools for counting roots. f has degree $p-1$ (its leading coefficient $(p-1)! \equiv -1$ is nonzero), so Lagrange's theorem gives at most $p-1$ roots — off by a factor of two. The factor of two suggests quadratic residues, so the student tries to show that $f(n) = 0$ forces n to be, say, a non-residue, or that roots pair up as $n \leftrightarrow$ something with at most one of each pair a root. Neither materialises: in the data the roots are 4 ($p = 5, 7$), 7 ($p = 11$), 5, 12 ($p = 13$), with no residue pattern.

A third attempt: the coefficients $k!$ satisfy $a_{k+1} = (k+1)a_k$, so f ought to satisfy a differential equation. Indeed $\frac{d}{dx}(xf) = \sum (k+1)! x^k$, which is f shifted — $x\,(xf)' = f - 1 + (p)!\, x^p \equiv f - 1$. So $x^2 f' + xf - f + 1 = 0$ in $\mathbb{F}_p[x]$. At a root z, $z^2 f'(z) = -1$, so roots are simple. That is true but is the wrong direction: simple roots give no bound. The student is stuck with a correct differential equation that says roots are *not* repeated.

The model chain. **Step 1: Wilson turns $k!$ into $1/(p-1-k)!$.** For $0 \le k \le p-1$,

$$(p-1)! = k! \prod_{m=k+1}^{p-1} m \equiv k! \prod_{m=k+1}^{p-1} (m-p) = k!\,(-1)^{p-1-k}(p-1-k)! \pmod p,$$

and $(p-1)! \equiv -1$ (Wilson). As $p-1$ is even, $k!\,(p-1-k)! \equiv (-1)^{k+1}$. Hence, for $x \in \mathbb{F}_p^\times$, putting $j = p-1-k$ and using $(-1)^{p-j} = -(-1)^j$,

$$f(x) = \sum_k \frac{(-1)^{k+1} x^k}{(p-1-k)!} = -\sum_{j=0}^{p-1} \frac{(-1)^j x^{p-1-j}}{j!} = -x^{p-1} \sum_{j=0}^{p-1} \frac{(-1/x)^j}{j!} = -g(-1/x),$$

where $g(y) = \sum_{j=0}^{p-1} y^j/j!$ and we used $x^{p-1} = 1$ (Fermat). The map $x \mapsto -1/x$ is a bijection of $\mathbb{F}_p^\times$, and $f(0) = g(0) = 1$. So f and g have the same number of roots in $\mathbb{F}_p$, all nonzero.

Step 2: g is almost its own derivative.

$$g'(y) = \sum_{j=1}^{p-1} \frac{y^{j-1}}{(j-1)!} = g(y) - \frac{y^{p-1}}{(p-1)!} = g(y) + y^{p-1},$$

using $1/(p-1)! \equiv -1$.

Step 3: make every root double. Put $h(y) = y^p - y + g(y) \in \mathbb{F}_p[y]$. Its leading term is y^p (g has degree $p-1$), so h is a nonzero polynomial of degree p, and

$$h'(y) = py^{p-1} - 1 + g'(y) = g(y) + y^{p-1} - 1.$$

Let $z \in \mathbb{F}_p$ with $g(z) = 0$. Then $z \neq 0$, so $z^p = z$ and $z^{p-1} = 1$; hence $h(z) = 0$ and $h'(z) = 0 + 1 - 1 = 0$. So $(y-z)^2 \mid h(y)$ (write $h = (y-z)q$; then $h'(z) = q(z) = 0$).

Step 4: count. If g has r roots in $\mathbb{F}_p$, the distinct factors $(y-z)^2$ are pairwise coprime and all divide h, so $2r \le \deg h = p$, i.e. $r \le (p-1)/2$ since p is odd. So f has at most $(p-1)/2$ roots and at least $(p+1)/2$ non-roots in $\{0, \dots, p-1\}$. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains face "at most half the elements of $\mathbb{F}_p$ are roots". They split at the question each asks of that bound.

(N) "Half: so roots pair with non-roots, or roots are residues."

(M) "Half: so the roots are *double* roots of some polynomial of degree about p."

The second split is inside (M): the natural attempt's differential equation for f proves roots are simple; the model first reflects f by Wilson into a truncated exponential g, for which $g' - g$ is the constant 1 on $\mathbb{F}_p^\times$, and that constant can be cancelled by adding $y^p - y$.

> *Lagrange gives $p-1$ and I need $(p-1)/2$. The standard way to halve a root count is multiplicity: if every root of g is a double root of some h of degree p, then there are at most $p/2$ of them. So I want a polynomial that agrees with g on $\mathbb{F}_p$ and whose derivative also vanishes at the roots. On $\mathbb{F}_p$ I may add any multiple of $y^p - y$ for free.*

> *Factorials in the numerator are awkward; factorials in the denominator are an exponential series, and the exponential is its own derivative. Wilson says $k!$ and $1/(p-1-k)!$ are the same up to sign. Reflect $k \mapsto p-1-k$ and see if f becomes a truncated e^y.*

- ▷ **Cue:** a sum $\sum_{k=0}^{p-1} k!\, x^k$ (factorials up to $p-1$) modulo p $\longrightarrow$ **Move:** use $k!\,(p-1-k)! \equiv (-1)^{k+1}$ (Wilson) to reflect it into $\sum y^j/j!$, a truncated exponential
- ▷ **Cue:** need to show a polynomial has at most half as many roots as its degree bound allows $\longrightarrow$ **Move:** exhibit a polynomial of degree about p for which every root is a double root; count with multiplicity
- ▷ **Cue:** over $\mathbb{F}_p$, a polynomial identity that holds only on $\mathbb{F}_p$ (or on $\mathbb{F}_p^\times$) $\longrightarrow$ **Move:** add a multiple of $y^p - y$ (or use $y^{p-1} = 1$): it changes nothing on $\mathbb{F}_p$ and has derivative -1
- ▷ **Cue:** a truncated exponential $\sum_{j<N} y^j/j!$ $\longrightarrow$ **Move:** its derivative is itself minus the last term: $g' = g - y^{N-1}/(N-1)!$

Layer 3 THE WHY

Why these two moves, and why are they natural rather than lucky?

Why multiplicity. The target is a root count of exactly half. Over a field, the only general upper bound on roots is "at most the degree, counted with multiplicity". To beat the degree bound by a factor of two with that tool, the roots must be counted twice: every root must be a double root of something of degree about p. That is not a trick; it is the only way the one available theorem can produce the number $(p-1)/2$. So the question becomes: which polynomial h vanishes to order two at every root of f? Double roots are detected by the derivative, so one must understand f' at the roots of f.

Why the natural differential equation fails, and the reflected one does not. The student's equation $x^2 f' + xf - f + 1 = 0$ gives $z^2 f'(z) = -1$ at a root: the derivative is controlled but nonzero. To get a double root one needs to cancel that derivative by adding something that vanishes on $\mathbb{F}_p$ — a multiple of $x^p - x$ — whose derivative -1 can absorb a *constant*. The student's $f'(z) = -1/z^2$ is not constant, so no fixed multiple of $x^p - x$ cancels it. The reflected series g has $g'(z) = g(z) + z^{p-1} = 1$ at every nonzero root: a constant, exactly matching the derivative of $y^p - y$ up to sign. That is why the reflection is needed: it converts the non-constant defect of f into the constant defect of g.

Why Wilson produces the exponential. The coefficients $k!$ grow; the exponential's $1/j!$ shrink; modulo p they are mirror images, because $(p-1)! = k! \cdot (k+1) \cdots (p-1)$ and the second block is $\pm(p-1-k)!$ when read as negatives. Anyone who has seen $\binom{p-1}{k} \equiv (-1)^k$ has seen this identity: it says $k!(p-1-k)! \equiv \pm(p-1)!$. And the exponential is the function characterised by $g' = g$; truncating it at degree $p-1$ breaks this only by the single term $y^{p-1}/(p-1)! = -y^{p-1}$, which is -1 on $\mathbb{F}_p^\times$. Every step is dictated by the definitions of g and of $\mathbb{F}_p$.

Kedlaya and Ng's first solution sets $h(x) = x^p - x + g(x)$ without comment. Here is where it comes from: we need a polynomial equal to g on $\mathbb{F}_p$ (so add a multiple $c(x^p - x)$) whose derivative vanishes where g does; $h' = g + x^{p-1} - c$ at a root is $1 - c$, so $c = 1$. Their second solution (Elkies) proves nonvanishing of a Hankel determinant of factorials; it is a different, deeper route through orthogonal polynomials and is not needed.

> **Caveat.** The statement depends on the convention $0^0 = 1$, which the official solution adopts and which is used here ($f(0) = 1$). The official solution remarks that with $0^0 = 0$ the

problem fails for $p = 3$; direct computation here did not reproduce that ($p = 3$, $0^0 = 0$ gives values $0, 1, 2$, so two non-roots, which equals $(p+1)/2$), so that remark is left unconfirmed and plays no role in the proof above. The main result was also checked by computer for all primes $3 \le p < 200$ (the actual number of roots is always far below $(p-1)/2$, e.g. 4 roots for $p = 37$).

Layer $+\alpha$ BACKTESTING

1. The first useful cue is $k!$ for k up to $p-1$, modulo p. On seeing it, write Wilson's reflection $k!\,(p-1-k)! \equiv (-1)^{k+1}$.

2. The cue suggests turning the factorial series into the truncated exponential g, then bounding roots by multiplicity via $h = y^p - y + g$.

3. Hesitation occurs at the factor of two: the student looks for a pairing or a quadratic-residue argument. It is reinforced away by asking which theorem can output "at most half" — only degree counted with multiplicity — and then asking for a polynomial with double roots.

4. The trap is the direct differential equation for f, which is correct and shows the roots of f are simple; it feels like progress and points the wrong way. A second trap is ignoring the $n = 0$ term and the convention $0^0 = 1$.

5. Recognise it faster by looking for any series over $\mathbb{F}_p$ whose derivative equals itself up to a term that is constant on $\mathbb{F}_p^\times$: such a series can be corrected by $y^p - y$ into a polynomial with only double roots.

6. Reusable decision rule: *to prove at most $d/2$ roots, find a degree-d polynomial in which every root is double; over $\mathbb{F}_p$, first use the free modification by $y^p - y$ to cancel the derivative.* And: *factorials modulo p are reciprocal factorials in disguise (Wilson), so look for an exponential.*