DAWN PUTNAM SESSION

Putnam 2014 — Papers A and B

Putnam 120 Archive

Contents

Putnam 2014 — Paper A **5**
2014 A1 5
2014 A2 7
2014 A3 9
2014 A4 12
2014 A5 15
2014 A6 18

Putnam 2014 — Paper B **23**
2014 B1 23
2014 B2 25
2014 B3 27
2014 B4 30
2014 B5 33
2014 B6 37

Putnam 2014 — Paper A

2014 A1

2014 A1. Prove that every nonzero coefficient of the Taylor series of

$$(1 - x + x^2)e^x$$

about $x = 0$ is a rational number whose numerator (in lowest terms) is either 1 or a prime number.

Fields: number theory (prime factorisation in factorials, Legendre-type counting) **Position:** A1 – the opener, but the fiddly kind: the algebra collapses in three lines and the entire problem is then a single case-split that has one genuine trap in it **Answer:** The coefficient of x^n ($n \geq 2$) is $(n-1)^2/n!$; its numerator in lowest terms is 1 for every n except $n = 5$, where it is the prime 2.

Layer 1 THE WHAT — surface analysis

The natural attempt. Multiplying out $(1 - x + x^2)\sum_{n\geq 0} x^n/n!$ and reading off the coefficient of x^n is not the hard part – everyone reaches

$$c_n = \frac{1}{n!} - \frac{1}{(n-1)!} + \frac{1}{(n-2)!} = \frac{(n-1)^2}{n!}$$

for $n \geq 2$ within a minute. The trouble starts once this is rewritten as $c_n = (n-1)/(n(n-2)!)$ and the student tries to argue about the reduced numerator directly. The instinct is: "if $n-1$ is prime, it survives; if $n-1$ is composite, it factors and gets absorbed into $(n-2)!$" – and then to declare the problem finished, because every composite number factors as a product of two smaller numbers. This is true, but it is stated too fast. A student who checks $n = 4, 6, 7, \ldots$ by hand and sees the numerator collapse to 1 every time gains false confidence that "composite $\Rightarrow$ numerator 1" is simply correct, and stops looking for a counterexample. The gap is invisible unless you ask, specifically, whether *every* composite number factors as a product of two *distinct* smaller numbers – and most students never ask that question, because in ordinary factoring practice ($12 = 3 \times 4$, $15 = 3 \times 5$, $20 = 4 \times 5$) it never comes up as an issue.

The model chain. Write $c_n = (n-1)^2/n! = (n-1)/(n\,(n-2)!)$ for $n \geq 2$. Since $\gcd(n-1, n) = 1$ always, the only cancellation available is between $n-1$ and $(n-2)! = 1 \cdot 2 \cdots (n-2)$.

Case 1: $n-1$ is prime. Call it $p = n-1$. Every factor of $(n-2)! = (p-1)!$ is strictly less than p, so p divides none of them and $\gcd(p, (n-2)!) = 1$. The numerator survives whole: it is the prime p.

Case 2: $n-1 = ab$ with $1 < a < b < n-1$ (so $n-1$ is composite and not a prime square). Both a and b are at most $n-2$ and distinct, so both occur as separate factors in the product $1 \cdot 2 \cdots (n-2)$. Hence $ab = n-1$ divides $(n-2)!$, the numerator cancels completely, and the reduced numerator is 1.

Case 3: $n-1 = p^2$ for a prime p. This is the case Case 2 does not cover, because the only factorisation of p^2 into two factors less than p^2 is $p \times p$ – the same number twice, not two distinct

terms of $1 \cdots (n-2)$. Whether $p^2 \mid (n-2)! = (p^2-1)!$ now depends on whether $(p^2-1)!$ contains *two* multiples of p, namely p and $2p$. This needs $2p \leq p^2 - 1$, i.e. $p^2 - 2p - 1 \geq 0$, true for every prime $p \geq 3$. So for $p \geq 3$, $p^2 \mid (p^2-1)!$ and the numerator is again 1. The inequality fails only at $p = 2$: $2p = 4$ but $p^2 - 1 = 3$, so $(p^2-1)! = 3! = 6$ contains only one factor of 2. Here $n - 1 = 4$, $n = 5$, and $c_5 = (5-1)^2/5! = 16/120 = 2/15$ – numerator 2, a prime, and the unique exception.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths agree up to the line "if $n-1$ is composite, it factors and is absorbed into $(n-2)!$." The natural attempt treats this as a single closed case; the model chain asks whether the factorisation can always be split into two *distinct* factors, and finds that it cannot – precisely when $n-1$ is the square of a prime.

> *Composite doesn't mean "factors into two smaller numbers" in the strong sense I need – it means $n-1 = ab$ for* some $1 < a \leq b < n-1$, *and if $a = b$ that's only one distinct value appearing in $1 \cdot 2 \cdots (n-2)$, not two. I need to isolate exactly which composite numbers have no factorisation into two* different *factors, and that set is exactly $\{p^2 : p \text{ prime}\}$.*

> *Once I'm inside the p^2 case, the question "does p^2 divide $(p^2-1)!$" is really "are there two multiples of p at or below p^2-1" – that's a size comparison, $2p$ versus p^2-1, and I should check it as an inequality in p rather than trust the pattern from case 2 to continue automatically. The inequality is tight at small p, which is exactly where I should expect an exception, not at large p where everything grows apart.*

Layer 3 THE WHY

The whole problem reduces to one fact about factorials: $(n-2)!$ contains a given integer $m < n-1$ as a factor if and only if m can be exhibited as a product of *distinct* terms already present in $1 \cdot 2 \cdots (n-2)$ – divisibility by a product is not automatic just because the product's value is small enough, it requires the individual multiples to occur as separate entries in the factorial. A prime $p = n-1$ fails this outright, since nothing below p can supply it. A composite $n-1 = ab$ with $a \neq b$ succeeds trivially, since a and b are simply two different entries in the list $1, \ldots, n-2$. The case $n - 1 = p^2$ is the honest edge of the argument: the "two factors" the composite case relies on have collapsed into one, so the argument must be rebuilt from Legendre's principle directly – counting multiples of p up to $p^2 - 1$ – rather than inherited from case 2. That $p = 2$ is the sole failure is not an accident of arithmetic but of scale: $2p$ grows linearly in p while $p^2 - 1$ grows quadratically, so the two multiples-of-p needed are guaranteed to fit for every prime past the smallest one, and the smallest prime is exactly where a linear and a quadratic function are still close enough to cross.

Caveat. I re-derived $c_n = (n-1)^2/n!$ independently from the product $(1 - x + x^2)e^x$, re-derived both the "prime" and "composite, distinct factors" cases from scratch, and located the $p = 2$ exception by direct inequality ($2p \leq p^2 - 1 \iff p \geq 3$) rather than by citing it – then checked $c_5 = 16/120 = 2/15$ by hand. This matches Kedlaya and Ng's stated solution exactly, including their same $p = 2$ exception, so the chain is fully verified, not paraphrased.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that the coefficient collapses to a single fraction with a squared numerator, $(n-1)^2/n!$ – a square in the numerator of a "numerator is 1 or prime" problem should

immediately raise the question of what happens when that square is a prime squared, since squares behave differently from products of two distinct primes under cancellation.

2. That cue suggests a three-way case split on the arithmetic nature of $n-1$: prime, composite with two distinct factors, and prime-squared – because these are exactly the three ways an integer can (or cannot) be written as a product of two integers less than itself.

3. Hesitation occurs at "if $n-1$ is composite, it's absorbed into $(n-2)!$" – it feels finished. The reinforcement is to ask, explicitly, whether the composite number's factorisation uses two *different* values, since a factorial is a product of distinct consecutive integers, not a product of a number with itself.

4. The trap is treating "composite" as a single case instead of splitting off perfect squares of primes, and then further assuming the perfect-square case behaves uniformly across all primes without checking the small-prime boundary explicitly.

5. Recognisable faster next time by treating any "does m divide $N!$" question as "can m be written as a product of distinct integers $\leq N$", and flagging $m = p^2$ as the standing exception to that being automatic.

6. Reusable rule: when a numerator's reduction to lowest terms depends on absorbing an integer m into a factorial in the denominator, do not stop at "composite m factors and cancels" – check separately whether m is a perfect square (or higher prime power), since those are exactly the values that cannot be split into two distinct factors below themselves, and verify the resulting divisibility by an explicit inequality rather than by extrapolating the generic composite case.

▷ **Cue:** a fraction's numerator, after full cancellation, is claimed to be "1 or prime" and the raw numerator before cancellation is a perfect square m^2 $\longrightarrow$ **Move:** split into three cases – m itself prime, m composite with two distinct factors both below the cancelling bound, and m prime – then separately check whether m^2 (not just m) divides the denominator, since m occurring twice needs two distinct multiples of m below the bound

▷ **Cue:** "does integer m divide $N!$" for $m < N$ $\longrightarrow$ **Move:** ask whether m can be written as a product of distinct integers $\leq N$; if $m = p^2$ for a prime p, check the inequality $2p \leq N$ explicitly rather than assuming it from small cases

2014 A2

2014 A2. Let A be the $n \times n$ matrix whose entry in the i-th row and j-th column is

$$\frac{1}{\min(i,j)}$$

for $1 \leq i, j \leq n$. Compute $\det(A)$.

Fields: linear algebra (structured determinants via non-standard row reduction) **Position:** A2 – routine slot on paper, but the determinant of a full $n \times n$ matrix cannot be brute-forced, so the whole problem is one structural observation **Answer:** $\det(A) = \dfrac{(-1)^{n-1}}{(n-1)!\,n!}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The entries look familiar – everyone recognises a "min" matrix from covariance-of-Brownian-motion contexts, but that recognition doesn't itself hand you a determinant. The default move for an unfamiliar $n \times n$ determinant is standard Gaussian elimination: pick the top row (all entries 1, since $\min(1, j) = 1$ for every j) and subtract multiples of it from every row below to clear the first column. This does clear column 1, but the entries that remain are no longer of the clean form $1/\min(i, j)$ – row i becomes $\left(0, \frac{1}{\min(i,2)} - 1, \ldots\right)$, and

further elimination against row 2 produces yet messier expressions. By row three or four the pattern in the entries has been destroyed by the accumulated subtractions, and the student is looking at a matrix of increasingly arbitrary rational numbers with no visible structure, trying to spot a pattern in the resulting 3×3 or 4×4 determinants computed by hand and hoping to guess the general formula from $n = 1, 2, 3$. This can eventually succeed – the guessed pattern $(-1)^{n-1}/((n-1)!n!)$ is not hard to conjecture from three data points – but "guess from small cases, then prove by induction" is a second, separate task bolted onto a computation that never needed to be messy in the first place.

The model chain. Look at the entries of two *adjacent* rows, not row 1 against everything. For fixed j, compare $\min(k-1, j)$ and $\min(k, j)$: for $j \leq k-1$ these are equal (both equal j), and they differ only for $j \geq k$, where $\min(k-1, j) = k-1$ and $\min(k, j) = k$. So row k and row $k-1$ agree in their first $k-1$ entries and diverge only from column k onward.

Replace row k by (row k) $-$ (row $k-1$), working from $k = n$ down to $k = 2$ so that each subtraction uses an original, unmodified row $k-1$. The new row k has zero entries in columns $1, \ldots, k-1$, and constant value $\frac{1}{k} - \frac{1}{k-1}$ in every column $j \geq k$. Row 1 is untouched: $(1, 1, \ldots, 1)$.

The resulting matrix is upper triangular, with diagonal entries 1 (row 1) and $\frac{1}{k} - \frac{1}{k-1}$ (row k, $k = 2, \ldots, n$). The determinant of an upper triangular matrix is the product of its diagonal:

$$\det(A) = \prod_{k=2}^{n} \left(\frac{1}{k} - \frac{1}{k-1} \right) = \prod_{k=2}^{n} \frac{-1}{k(k-1)} = \frac{(-1)^{n-1}}{\prod_{k=2}^{n} k(k-1)}.$$

Since $\prod_{k=2}^{n} k = n!$ and $\prod_{k=2}^{n} (k-1) = (n-1)!$, this is $\det(A) = (-1)^{n-1}/(n!(n-1)!)$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths start from the same untouched row $1 = (1, \ldots, 1)$. They split at the choice of *which* row to eliminate against: the natural attempt fixes row 1 as the pivot for every subsequent row; the model chain eliminates each row against its immediate predecessor instead.

> *Row* 1 *is special only in that it's constant – but every row of this matrix is a step function of* j *that is constant beyond some threshold. Two rows whose thresholds are one apart will agree everywhere except past the later threshold, so subtracting* adjacent *rows should cancel almost the whole row, not just clear one column the way subtracting row* 1 *does.*

> *I don't need the elimination to happen in the usual order (top row first, working down). If I eliminate from the bottom row upward – row* n *minus row* $n-1$*, then row* $n-1$ *minus row* $n-2$*, and so on – each subtraction still uses a row that hasn't itself been modified yet, so the cancellation I predicted for adjacent rows actually happens cleanly, rather than being corrupted by an earlier elimination step.*

Layer 3 THE WHY

The entries $A_{ij} = 1/\min(i, j)$ are, for fixed row i, a step function of j: constant at $1/i$ from column i onward, and equal to $1/j$ before that. That is the entire content of min as an operation – it is symmetric, and it is piecewise constant in exactly this way. Two rows i and $i-1$ are the *same* step function up to column $i-1$, because $\min(i, j) = \min(i-1, j) = j$ whenever $j < i$; they diverge only once j reaches the later row's own index, where one function has already flattened out and the other hasn't yet. So subtracting adjacent rows is not a trick chosen to make the algebra nice – it is the operation that matches the object's own definition: differencing a step function against a copy of itself shifted by one threshold necessarily produces zero everywhere except past that threshold. Standard Gaussian elimination against a fixed pivot row ignores this structure entirely and elimination against row 1 treats every row as equally related to row

1, when in fact each row is far more closely related to its immediate neighbour. The reusable content is not "eliminate bottom-up" as a trick, but "identify what makes each row a small perturbation of the next, and difference in that direction."

Caveat. I re-derived the adjacent-row-difference computation independently: verified that $\min(k, j) = \min(k-1, j)$ for $j < k$ and computed the surviving entry $1/k - 1/(k-1)$ for $j \geq k$ directly from the definition, then multiplied out the resulting upper-triangular determinant by hand for $n = 2, 3, 4$ against the closed form to confirm $(-1)^{n-1}/(n!(n-1)!)$. This matches Kedlaya and Ng's stated answer exactly, and their remark that this order of elimination (rather than the standard top-down order) is what keeps the computation clean – so the chain is fully verified, not paraphrased.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that $A_{ij} = 1/\min(i, j)$ is, row by row, a step function of the column index – constant from the diagonal onward, varying before it. That should be written down before any elimination is attempted, since it says exactly how any two rows relate to each other.

2. That cue suggests row-reducing by differencing rows against their *neighbours* rather than against a fixed pivot row, since neighbouring step functions (thresholds one apart) cancel almost entirely against each other.

3. Hesitation occurs right after eliminating column 1 against row 1 and finding the remaining entries no longer resemble $1/\min(i, j)$; the reinforcement is to notice that the mess is caused by the choice of pivot, not by the problem, and to restart the elimination using adjacent rows instead.

4. The trap is defaulting to the textbook elimination order (always subtract a multiple of the pivot row currently at the top) without asking whether the matrix's own structure suggests a better direction to eliminate in.

5. Recognisable faster next time by treating any matrix built from a symmetric, monotone bivariate function like $\min(i, j)$, $\max(i, j)$, or $\gcd(i, j)$ as a candidate for adjacent-row or adjacent-column differencing before reaching for general elimination.

6. Reusable rule: for a determinant of a matrix whose entries are a known function of (i, j), before eliminating, ask which pair of rows (or columns) are *closest* to being equal as functions of the other index – usually neighbours under whatever ordering the defining function respects – and difference those, since the resulting cancellation reflects the function's own structure rather than an arbitrary pivot choice.

▷ **Cue:** a determinant whose entries are $f(\min(i, j))$, $f(\max(i, j))$, or another symmetric monotone function of the indices $\longrightarrow$ **Move:** difference adjacent rows (or columns), working from the far end inward so each step uses an unmodified neighbour, since such entries are step functions of one index and adjacent rows/columns share almost the entire step function

▷ **Cue:** Gaussian elimination against a fixed pivot row destroys visible structure in the remaining entries $\longrightarrow$ **Move:** re-choose the elimination direction to match the matrix's own defining structure (e.g. adjacent-index differencing) rather than continuing with the standard top-down pivot

2014 A3

2014 A3. Let $a_0 = 5/2$ and $a_k = a_{k-1}^2 - 2$ for $k \geq 1$. Compute

$$\prod_{k=0}^{\infty} \left(1 - \frac{1}{a_k}\right)$$

in closed form.

Fields: algebra / analysis (recurrence recognition, telescoping products) **Position:** A3 – the recurrence is the whole problem; once the right substitution is found the computation is short, but without it there is no visible path to a closed form **Answer:** $\displaystyle\prod_{k=0}^{\infty} \left(1 - \frac{1}{a_k}\right) = \frac{3}{7}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The recursion $a_k = a_{k-1}^2 - 2$ is concrete, so the honest first move is to compute: $a_0 = 2.5$, $a_1 = 2.5^2 - 2 = 4.25$, $a_2 = 4.25^2 - 2 = 16.0625$, and the sequence is clearly exploding. A student now tries to get a handle on the *product* directly – writing out partial products $\prod_{k=0}^{N}(1 - 1/a_k)$ numerically, hoping the decimal value stabilises fast enough to recognise a nice fraction (it does: $0.6 \times 0.7647 \times 0.9377 \times \cdots \to 0.4286\ldots = 3/7$), but recognising a limit numerically is not a proof, and there is no way to bound the tail of the product without first understanding how a_k grows. The other natural move is to attack $1 - 1/a_k = (a_k - 1)/a_k$ algebraically, hoping the product telescopes against the recursion $a_k = a_{k-1}^2 - 2$ directly – but $a_{k-1}^2 - 2 - 1 = a_{k-1}^2 - 3$ does not factor against $a_{k-1}^2 - 2$ in any way that cancels with the next or previous term, and the student is left manipulating messy quadratic expressions in a_{k-1} with no telescoping structure in sight. Both routes stall for the same underlying reason: the recursion $a_k = a_{k-1}^2 - 2$, looked at as raw algebra on the numbers a_k themselves, does not reveal why it telescopes – that requires seeing the numbers as something else.

The model chain. Notice that $a_0 = \frac{5}{2} = 2 + \frac{1}{2}$ is of the form $t + \frac{1}{t}$ with $t = 2$. Set $x_k = 2^{2^k}$, so $x_0 = 2$, and define the claim $a_k = x_k + x_k^{-1}$. This is true for $k = 0$; assuming it for $k - 1$,

$$a_k = a_{k-1}^2 - 2 = \left(x_{k-1} + x_{k-1}^{-1}\right)^2 - 2 = x_{k-1}^2 + x_{k-1}^{-2} = x_k + x_k^{-1},$$

using $x_{k-1}^2 = x_k$. So $a_k = x_k + x_k^{-1}$ for all k, and in particular $a_k > 2$ for every k, so the product converges absolutely.

Now write $1 - \frac{1}{a_k} = \dfrac{a_k - 1}{a_k} = \dfrac{x_k - 1 + x_k^{-1}}{x_k + x_k^{-1}}$. Introduce $P_k = x_k + 1 + x_k^{-1}$. The algebraic identity $(x^2 + 1 + x^{-2}) = (x - 1 + x^{-1})(x + 1 + x^{-1})$ (direct expansion of $(x + x^{-1})^2 - 1$) gives, with $x = x_k$,

$$x_k - 1 + x_k^{-1} = \frac{x_k^2 + 1 + x_k^{-2}}{x_k + 1 + x_k^{-1}} = \frac{P_{k+1}}{P_k},$$

since $x_k^2 = x_{k+1}$. So $1 - \frac{1}{a_k} = \dfrac{P_{k+1}}{P_k\, a_k}$, and the partial product telescopes in P:

$$\prod_{k=0}^{N} \left(1 - \frac{1}{a_k}\right) = \frac{P_{N+1}}{P_0} \cdot \frac{1}{\prod_{k=0}^{N} a_k}.$$

For the remaining product, set $D_k = x_k - x_k^{-1}$; then $D_k\, a_k = (x_k - x_k^{-1})(x_k + x_k^{-1}) = x_k^2 - x_k^{-2} = D_{k+1}$, so $a_k = D_{k+1}/D_k$ and $\prod_{k=0}^{N} a_k = D_{N+1}/D_0$ telescopes too. With $P_0 = x_0 + 1 + x_0^{-1} = \frac{7}{2}$ and $D_0 = x_0 - x_0^{-1} = \frac{3}{2}$,

$$\prod_{k=0}^{N} \left(1 - \frac{1}{a_k}\right) = \frac{P_{N+1}}{7/2} \cdot \frac{3/2}{D_{N+1}} = \frac{3}{7} \cdot \frac{P_{N+1}}{D_{N+1}}.$$

As $N \to \infty$, $x_{N+1} = 2^{2^{N+1}} \to \infty$, so $P_{N+1} = x_{N+1} + 1 + x_{N+1}^{-1}$ and $D_{N+1} = x_{N+1} - x_{N+1}^{-1}$ are asymptotically equal, and $P_{N+1}/D_{N+1} \to 1$. Hence the product converges to $\frac{3}{7}$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths reach $a_0 = 2.5$ and the raw recursion $a_k = a_{k-1}^2 - 2$ honestly. They split at the very next step: the natural attempt treats a_k as a number to be manipulated on its own terms; the model chain asks what *kind* of number produces a "square minus 2" recursion, and rewrites a_0 as $t + 1/t$ before doing any further algebra.

> *$2.5 = 2 + 0.5$ is $t + 1/t$ for $t = 2$ – and $(t + 1/t)^2 - 2 = t^2 + 1/t^2$ exactly, with no error term, because the cross terms in $(t + 1/t)^2$ are $2 \cdot t \cdot \frac{1}{t} = 2$, precisely what's being subtracted. That means the whole recursion $a_k = a_{k-1}^2 - 2$ is not really a statement about a_k; it's the doubling law for $t + 1/t$ under $t \mapsto t^2$, and I should stop treating a_k as primitive and instead track $x_k = 2^{2^k}$.*

> *Once $a_k = x_k + x_k^{-1}$, a product like $1 - 1/a_k$ should be attacked by looking for a difference-of-squares or product identity in x_k alone, because x_k is the object that actually doubles its exponent cleanly ($x_k^2 = x_{k+1}$) – I need two auxiliary telescoping quantities, one built from $x - 1 + x^{-1}$ dividing into $x^2 + 1 + x^{-2}$, and one from $x - x^{-1}$ multiplying $x + x^{-1}$ to give $x^2 - x^{-2}$, since those are exactly the two algebraic identities that turn "squaring x_k" into "telescoping a product."*

Layer 3 THE WHY

The recursion $a_k = a_{k-1}^2 - 2$ is, at the axiom level, the algebraic shadow of the identity $\cosh(2\theta) = 2\cosh^2\theta - 1$, or equivalently, of the fact that $t^{2^k} + t^{-2^k}$ satisfies exactly this doubling recursion for any t: this is not a coincidence to be verified case by case, it follows immediately from squaring $t^n + t^{-n}$ and using $t^n \cdot t^{-n} = 1$. Once a_0 is recognised as $t + 1/t$ for a rational t, every subsequent a_k is forced to be $t^{2^k} + t^{-2^k}$ by the same one-line algebra, and the problem stops being about an abstract recursively-defined sequence and becomes about explicit powers of a single number $t = 2$. This is why the product can telescope at all: telescoping requires consecutive terms of a product to share a common factor that cancels, and the only reason $x_k^2 = x_{k+1}$ produces that shared structure is that x_k was chosen to be an actual power tower, not merely "some quantity satisfying the recursion." The two auxiliary sequences $P_k = x_k + 1 + x_k^{-1}$ and $D_k = x_k - x_k^{-1}$ are not guesswork either: they are exactly the factors that appear when $x^3 + 1$ and $x^2 - 1$ are factored, and $x^3 + 1 = (x+1)(x^2 - x + 1)$, $x^2 - 1 = (x-1)(x+1)$ are the only two factorisations of low-degree polynomials in x that could plausibly turn a rational function of x_k into a ratio of consecutive terms of a new sequence.

> **Caveat.** I derived $a_k = x_k + x_k^{-1}$ by induction independently, verified the identity $(x - 1 + x^{-1})(x + 1 + x^{-1}) = x^2 + 1 + x^{-2}$ by direct expansion, verified $D_k a_k = D_{k+1}$ by direct expansion of $(x_k - x_k^{-1})(x_k + x_k^{-1})$, computed $P_0 = 7/2$, $D_0 = 3/2$ by hand, and confirmed $P_{N+1}/D_{N+1} \to 1$ from $x_{N+1} \to \infty$. This gives $3/7$ independently of Kedlaya and Ng's own telescoping argument (which proceeds via a single combined fraction rather than two separately telescoping auxiliary sequences P_k, D_k), and the two derivations agree on the final answer, so the chain is fully verified.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is $a_0 = 5/2 = 2 + 1/2$ – any time a recursion of the form "square minus a constant" is paired with a starting value of the form $t + 1/t$ (or $2\cos\theta$, $2\cosh\theta$), that constant should be recognised as coming from the cross term of $(t + 1/t)^2$, i.e. the constant is 2 precisely because $2 \cdot t \cdot (1/t) = 2$.

2. That cue suggests substituting $a_k = x_k + x_k^{-1}$ with $x_k = t^{2^k}$, converting a recursively-defined sequence into explicit powers of a fixed rational number, which is the substitution that makes both growth rate and algebraic manipulation tractable.

3. Hesitation occurs right after writing $1 - 1/a_k = (a_k - 1)/a_k$ and finding no obvious cancellation between consecutive terms in a_k alone; the reinforcement is to translate immediately into x_k and hunt for a difference-of-squares or sum-of-cubes factorisation, since those are the only algebraic identities that turn a squaring recursion into a telescoping ratio.

4. The trap is trying to telescope the product directly in terms of a_k, or trying to prove convergence to $3/7$ by numerical stabilisation of partial products rather than by an exact telescoping argument.

5. Recognisable faster next time by treating any recurrence of the shape $a_k = a_{k-1}^2 - c$ with c a small integer as a candidate for the substitution $a_k = t^{2^k} + t^{-2^k} \cdot c^{1/2}$-type doubling identities, starting from whichever constant c matches $2 \cdot (\text{something}) \cdot (\text{its reciprocal})$.

6. Reusable rule: when a sequence is defined by $a_k = a_{k-1}^2 - 2$, first check whether $a_0 = t + 1/t$ for a recognisable t; if so, replace the sequence entirely by the explicit power sequence $x_k = t^{2^k}$, and look for telescoping in any product or sum involving a_k by factoring the numerator and denominator as $x^n \pm x^{-n}$ expressions, which are the only quantities guaranteed to relate cleanly across the doubling step $x_k \mapsto x_k^2$.

▷ **Cue:** a sequence recursion of the form $a_k = a_{k-1}^2 - 2$ ⟶ **Move:** check whether $a_0 = t + 1/t$ for a rational t; if so, replace a_k by the explicit doubling sequence $x_k = t^{2^k}$ via $a_k = x_k + x_k^{-1}$, since $(t + 1/t)^2 - 2 = t^2 + 1/t^2$ exactly

▷ **Cue:** a product or sum built from a doubling sequence $x_k = t^{2^k}$ fails to telescope in its original form ⟶ **Move:** introduce auxiliary sequences from low-degree factorisations of $x^n \pm 1$ (e.g. $x^2 - 1 = (x-1)(x+1)$, $x^3 + 1 = (x+1)(x^2 - x + 1)$) built from x_k, since these are exactly the expressions whose ratio across the doubling step $x_k \to x_k^2$ is itself a clean telescoping quantity

2014 A4

2014 A4. Suppose X is a random variable that takes on only nonnegative integer values, with $E[X] = 1$, $E[X^2] = 2$, and $E[X^3] = 5$. (Here $E[y]$ denotes the expectation of the random variable Y.) Determine the smallest possible value of the probability of the event $X = 0$.

Fields: probability (moment problems); linear programming duality in disguise; polynomial inequalities on the integers **Position:** A4 — mid-paper. Short once the right polynomial is on the page; the whole difficulty is producing that polynomial rather than guessing it. **Answer:** The minimum of $P(X = 0)$ is $\frac{1}{3}$, attained uniquely by $P(X = 0) = \frac{1}{3}$, $P(X = 1) = \frac{1}{2}$, $P(X = 3) = \frac{1}{6}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You write $a_n = P(X = n)$ and translate every condition into a linear equation in the unknowns $a_0, a_1, a_2, \ldots$:

$$\sum_n a_n = 1, \qquad \sum_n n a_n = 1, \qquad \sum_n n^2 a_n = 2, \qquad \sum_n n^3 a_n = 5, \qquad a_n \geq 0.$$

The objective is to make a_0 small. The instinct is to try small supports. Four equations, so try four unknowns. Support $\{0,1,2,3\}$: subtracting the equations pairwise gives $a_1 + a_2 + a_3 = 1 - a_0$, $a_1 + 2a_2 + 3a_3 = 1$, $a_1 + 4a_2 + 9a_3 = 2$, $a_1 + 8a_2 + 27a_3 = 5$. The last three determine a_1, a_2, a_3 outright: $a_3 = \frac{1}{6}$, $a_2 = 0$, $a_1 = \frac{1}{2}$, and then $a_0 = \frac{1}{3}$. A strong student now has a candidate.

Then the doubt. Why should the support be $\{0,1,2,3\}$? Try $\{0,1,4\}$: $a_1 + 4a_4 = 1$, $a_1 + 16a_4 = 2$ gives $a_4 = \frac{1}{12}$, $a_1 = \frac{2}{3}$, and then $a_1 + 64a_4 = \frac{2}{3} + \frac{16}{3} = 6 \neq 5$. Fails. Try $\{0,1,2,4\}$,

$\{0,2,3\}$, $\{0,1,5\}$, ... — each one a small linear system, some infeasible, some giving $a_0 > \frac{1}{3}$. After twenty minutes you have a table of supports and a strong suspicion that $\frac{1}{3}$ is the minimum, but the support can be any subset of an infinite set, and checking cases never closes. This is the wall: a feasible example proves min $\leq \frac{1}{3}$, and nothing in this approach can prove min $\geq \frac{1}{3}$.

The model chain. *Step 1: look for a single inequality.* I want to show $a_0 \geq \frac{1}{3}$, i.e. $\sum_{n\geq 1} a_n \leq \frac{2}{3}$. The only information available is the four linear sums $E[1], E[X], E[X^2], E[X^3]$, so any conclusion must come from $E[g(X)]$ for some cubic $g(n) = c_0 + c_1 n + c_2 n^2 + c_3 n^3$, whose expectation is known:

$$E[g(X)] = c_0 + c_1 + 2c_2 + 5c_3.$$

If $g(0) = 0$ and $g(n) \geq 1$ for every integer $n \geq 1$, then

$$\sum_{n\geq 1} a_n \;\leq\; \sum_{n\geq 1} g(n)a_n \;=\; E[g(X)],$$

and I am done if $E[g(X)] = \frac{2}{3}$.

Step 2: where must g touch 1? If the bound is to be sharp at the candidate distribution (mass on $0, 1, 3$), every inequality $g(n)a_n \geq a_n$ must be an equality there, so $g(1) = g(3) = 1$. Take the cubic that equals 1 at $n = 1, 2, 3$ and 0 at $n = 0$:

$$g(n) = 1 + \tfrac{1}{6}(n-1)(n-2)(n-3) = \tfrac{1}{6}\left(n^3 - 6n^2 + 11n\right).$$

Check $g(0) = 1 + \frac{1}{6}(-6) = 0$.

Step 3: $g \geq 1$ on positive integers. For $n \in \{1,2,3\}$, $(n-1)(n-2)(n-3) = 0$. For $n \geq 4$ all three factors are positive. So $g(n) \geq 1$ for every integer $n \geq 1$, with equality exactly at $n = 1, 2, 3$. (For real n it fails: $g(2.5) = 1 - \frac{1}{16} < 1$. The integrality of X is used here and nowhere else.)

Step 4: compute.

$$E[g(X)] = \tfrac{1}{6}\big(E[X^3] - 6E[X^2] + 11E[X]\big) = \tfrac{1}{6}(5 - 12 + 11) = \tfrac{2}{3}.$$

Hence $1 - a_0 = \sum_{n\geq 1} a_n \leq \frac{2}{3}$, so $a_0 \geq \frac{1}{3}$.

Step 5: sharpness. $a_0 = \frac{1}{3}$, $a_1 = \frac{1}{2}$, $a_3 = \frac{1}{6}$ gives total mass 1, $E[X] = \frac{1}{2} + \frac{3}{6} = 1$, $E[X^2] = \frac{1}{2} + \frac{9}{6} = 2$, $E[X^3] = \frac{1}{2} + \frac{27}{6} = 5$. So the minimum is $\frac{1}{3}$.

Equality in Step 4 forces $a_n = 0$ for $n \geq 4$, and the linear system on $\{0,1,2,3\}$ solved above then forces $a_2 = 0$: the extremal distribution is unique.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains translate the problem into the unknowns a_n and both find the candidate $(\frac{1}{3}, \frac{1}{2}, 0, \frac{1}{6})$. They split at (†): "I have a candidate; now how do I rule out everything else?" The natural attempt keeps generating candidates (other supports). The model stops generating candidates and asks for one inequality valid for *every* distribution: a polynomial g with $g(0) = 0$ and $g \geq 1$ on the positive integers.

> *I know exactly four numbers about X: the expectations of $1, X, X^2, X^3$. So whatever I prove has to be the expectation of some cubic in X. I want to bound $P(X \geq 1)$, so I want a cubic that sits above the indicator of $\{X \geq 1\}$ and is 0 at 0.*

> *My candidate puts mass on 1 and 3. For the bound to be tight there, the cubic has to equal 1 at 1 and 3. With $g(0) = 0$ as well, that's three conditions on four coefficients — and making it also touch at 2 gives $1 + \frac{1}{6}(n-1)(n-2)(n-3)$, which is visibly ≥ 1 on integers.*

Layer 3 THE WHY

Why is "find a polynomial" the natural move rather than a trick? Go back to what the hypotheses are. Each is a statement $E[h(X)] = c$ for $h \in \{1, n, n^2, n^3\}$, and expectation is linear. So the set of admissible distributions is cut out by linear equations in the a_n, and the objective a_0 is linear too. That is a linear programme. In any linear programme, the *only* way to certify a lower bound on the objective is to take a linear combination of the constraints — here, a cubic g — that dominates the objective term by term. There is no other source of inequalities: positivity $a_n \geq 0$ plus the four equations is everything. So once you recognise the problem as "minimise a linear quantity subject to linear constraints and positivity", the polynomial is not an inspiration; it is the form every proof must take.

Why touch at $1, 2, 3$? Because of complementary slackness, which you can reconstruct without the name: if a distribution attains the bound, every term dropped in $\sum_{n\geq 1} a_n \leq \sum_{n\geq 1} g(n)a_n$ must be zero, so $g(n) = 1$ wherever that distribution has mass. The candidate you found by small-support search tells you the touching points. That is the rôle of the natural attempt: it was not wasted, it located where the certificate has to be tight. The search and the certificate are the two halves of one argument.

Kedlaya's solution writes down $f(n) = 11n - 6n^2 + n^3$ in its first line and notes $f(1) = f(2) = f(3) = 6$. Here is where that comes from: $f - 6 = (n-1)(n-2)(n-3)$, the cubic vanishing at the three positive integers where the dual bound should be tight, shifted so that it vanishes at 0. Written as $f(n) = 6 + (n-1)(n-2)(n-3)$, the claim $f(n) \geq 6$ on positive integers needs no calculation at all.

There is a second route to the same polynomial, and it is worth seeing because it explains the numbers $1, 2, 5$. The factorial moments are $E[X] = 1$, $E[X(X-1)] = 2 - 1 = 1$, $E[X(X-1)(X-2)] = 5 - 3 \cdot 2 + 2 \cdot 1 = 1$. Every factorial moment up to order three equals 1 — exactly as for a Poisson(1) variable. For a probability generating function $G(z) = \sum a_n z^n$ that means $G(1) = G'(1) = G''(1) = G'''(1) = 1$, so Taylor's theorem at $z = 1$ gives

$$a_0 = G(0) = 1 - 1 + \tfrac{1}{2} - \tfrac{1}{6} + \tfrac{G''''(c)}{24} = \tfrac{1}{3} + \tfrac{G''''(c)}{24}, \qquad c \in (0, 1),$$

and $G''''(c) \geq 0$ because G has nonnegative coefficients. The $\frac{1}{3}$ is the cubic Taylor polynomial of e^{z-1} about $z = 1$, evaluated at $z = 0$. (Taylor at the endpoint is legitimate: $E[X^3] < \infty$ makes G''' continuous on $[0, 1]$, and G'''' exists on $[0, 1)$.) Both routes are the same fact: the third-order Taylor remainder is a fourth derivative, and fourth derivatives of a PGF on $[0, 1]$ are nonnegative.

Caveat. Verified independently: a linear-programming solver over distributions on $\{0, 1, \ldots, 60\}$ returns minimum $a_0 = 0.33333\ldots$ with support $\{0, 1, 3\}$ and weights $(\frac{1}{3}, \frac{1}{2}, \frac{1}{6})$, matching the proof. Note the dual certificate is not unique: $g(n) = 1 + (n-1)(n-3)(an - \frac{1}{3})$ works for every $a \in [\frac{1}{12}, \frac{1}{6}]$, since $E[X(X-1)(X-3)] = 5 - 8 + 3 = 0$; Kedlaya's choice is $a = \frac{1}{6}$.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "smallest possible value" of a probability, with only moments known. Write down $a_n = P(X = n)$ and the four linear equations, and then write the target inequality in the form $\sum_{n\geq 1} a_n \leq E[g(X)]$ with g a cubic to be found.
2. Linear programming duality, done by hand: find the extremal candidate by a small-support search, then find a polynomial certificate that is tight exactly on that support.
3. Hesitation occurs after the candidate $\frac{1}{3}$ is found, when case-checking other supports feels endless. Reinforce with the principle that a candidate only proves one direction; the other direction needs a statement about *all* distributions, and with linear data the only such statements are expectations of polynomials.

4. The trap is enumeration: supports are subsets of an infinite set, so no amount of checking closes the argument. A subtler trap is proving $g \geq 1$ for real $n \geq 1$: it is false ($g(2.5) = 1 - \frac{1}{16}$), and the student who tries it will think the method has failed. The inequality holds on integers only, and that is where the hypothesis "nonnegative integer values" is spent.

5. Slogan: *moments known, probability wanted: find the polynomial that sits above the indicator and touches it where the extremal distribution lives.* This is the discrete Chebyshev–Markov moment method, and it covers Markov's inequality ($g(n) = n/t$), Chebyshev's inequality ($g(n) = (n-\mu)^2/t^2$), and this problem alike.

6. Decision rule: **when every hypothesis and the objective are linear in unknown nonnegative weights, the lower bound is a weighted combination of the hypotheses; find the candidate optimum first, then build the combination to be tight exactly where the candidate puts weight.** Companion rule: moments $1, 2, 5$ (Bell numbers) mean all factorial moments are 1; translate raw moments to factorial moments before doing anything else.

▷ **Cue:** bound a probability $P(X = 0)$ or $P(X \geq t)$ given only the first few moments of X $\longrightarrow$ **Move:** find a polynomial g of degree at most the number of known moments with $g \geq$ the indicator on the support of X; then $P \leq E[g(X)]$

▷ **Cue:** an extremal candidate distribution has been found by solving small linear systems $\longrightarrow$ **Move:** use its support as the list of points where the certificate polynomial must equal the indicator (complementary slackness); build g from the product $\prod(n - s)$ over those points

▷ **Cue:** moments of an integer-valued variable are $1, 2, 5, \ldots$ or otherwise look structured $\longrightarrow$ **Move:** convert to factorial moments $E[X(X-1)\cdots(X-k+1)]$; these are the derivatives of the generating function at 1

▷ **Cue:** a probability generating function G with known derivatives at $z = 1$ and a question about $G(0)$ $\longrightarrow$ **Move:** Taylor expand at 1 with Lagrange remainder; the remainder has a sign because G has nonnegative coefficients

2014 A5

2014 A5. Let

$$P_n(x) = 1 + 2x + 3x^2 + \cdots + nx^{n-1}.$$

Prove that the polynomials $P_j(x)$ and $P_k(x)$ are relatively prime for all positive integers j and k with $j \neq k$.

Fields: algebra (polynomials, roots in $\mathbb{C}$); real analysis (Rolle's theorem, counting zeros of exponential-polynomial functions) **Position:** A5 — hard. Nothing on the page suggests analysis, and the natural algebraic attacks stall completely; the proof needs one reformulation and one counting argument. **Answer:** A common root z of P_j and P_k would make $w = 1/z$ satisfy $w^m = m(w-1) + 1$ for four distinct integers m; taking $|\cdot|^2$ turns this into a real function $|w|^{2t} - q(t)$ (q quadratic) with four zeros, which Rolle forbids.

Layer 1 THE WHAT — surface analysis

The natural attempt. "Relatively prime" means no common factor, and over $\mathbb{C}$ that means no common root, so you suppose $P_j(z) = P_k(z) = 0$ with $j < k$ and look for a contradiction. First you tidy P_n. It is the derivative of $1 + x + \cdots + x^n = \frac{x^{n+1}-1}{x-1}$, so

$$P_n(x) = \frac{nx^{n+1} - (n+1)x^n + 1}{(x-1)^2}.$$

Now the attack forks, and every branch stalls.

Branch 1: Euclid. Divide P_k by P_j. $P_k - P_j = (j+1)x^j + \cdots + kx^{k-1}$, which is x^j times a polynomial of the same shape with shifted coefficients. The remainders never simplify to a

pattern; for $k = j+1$ the argument is immediate ($P_{j+1} - P_j = (j+1)x^j$, and 0 is not a root), but for $k = j+2$ you already have $(j+1)x^j + (j+2)x^{j+1}$, whose root $-\frac{j+1}{j+2}$ you now have to test in P_j, and for general k you are lost.

Branch 2: irreducibility. If each P_n were irreducible over $\mathbb{Q}$ and they had different degrees, you would be done. You try Eisenstein, reduction mod p, anything. None of it works — and Kedlaya and Ng remark that irreducibility of P_n "appears difficult to prove". This branch is a wall you cannot see from the inside.

Branch 3: the two root equations. From the closed form, a common root satisfies $jz^{j+1} - (j+1)z^j + 1 = 0$ and $kz^{k+1} - (k+1)z^k + 1 = 0$. You try to eliminate z: solve the first for z^j, substitute into the second, and get one more polynomial equation in z with no visible reason to be impossible. Forty minutes, three branches, no contradiction.

The common feature of all three is that they stay inside algebra, where the exponents j, k are fixed integers and there is no tool that compares the j-equation with the k-equation.

The model chain. Step 1: the root equation, inverted. Suppose $P_j(z) = P_k(z) = 0$ with $1 \le j < k$. Then $z \neq 0$ (since $P_n(0) = 1$) and $z \neq 1$ (since $P_n(1) = \frac{n(n+1)}{2} > 0$). For $n \in \{j, k\}$ the closed form gives $nz^{n+1} - (n+1)z^n + 1 = 0$. Put $w = 1/z$ (so $w \neq 0, 1$) and multiply by w^{n+1}:

$$n - (n+1)w + w^{n+1} = 0, \qquad \text{i.e.} \qquad w^m = m(w-1) + 1 \quad \text{with } m = n+1.$$

So $w^m = m(w-1)+1$ holds for $m = j+1$ and $m = k+1$. It also holds, for any w, at $m = 0$ (both sides are 1) and $m = 1$ (both sides are w). Since $2 \le j+1 < k+1$, these are four distinct nonnegative integers $0 < 1 < j+1 < k+1$.

Step 2: make it real. Take the squared modulus of both sides. Writing $r = |w| > 0$,

$$r^{2m} = |m(w-1)+1|^2 = m^2|w-1|^2 + 2m\,\mathrm{Re}(w-1) + 1.$$

Define, for real $t \ge 0$,

$$f(t) = r^{2t} - \big(t^2|w-1|^2 + 2t\,\mathrm{Re}(w-1) + 1\big).$$

Then f is smooth and $f(0) = f(1) = f(j+1) = f(k+1) = 0$.

Step 3: count zeros. The subtracted part is a polynomial of degree at most 2 in t, so its third derivative vanishes and

$$f'''(t) = (2\ln r)^3\, r^{2t}.$$

Case $r \neq 1$. Then $f'''(t) \neq 0$ for all t. By Rolle's theorem, four distinct zeros of f give three distinct zeros of f' (one strictly between each consecutive pair), hence two of f'', hence one of f'''. Contradiction.

Case $r = 1$. Then $r^{2t} = 1$ and $f(t) = -\big(t^2|w-1|^2 + 2t\,\mathrm{Re}(w-1)\big)$, a polynomial of degree at most 2 with four zeros. So it is the zero polynomial, and in particular $|w-1|^2 = 0$, i.e. $w = 1$. Contradiction with $w \neq 1$.

Either way no common root exists. Since P_j and P_k have no common root in $\mathbb{C}$, their greatest common divisor (over $\mathbb{Q}$, $\mathbb{R}$ or $\mathbb{C}$ — the gcd over a field is unchanged by extending the field) has no root, so it is a nonzero constant.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains reach the same pair of equations: $nz^{n+1} - (n+1)z^n + 1 = 0$ for $n = j$ and $n = k$. At (†) they split. The natural attempt treats j and k as two fixed exponents and tries to eliminate z. The model treats the exponent as the *variable*: it rewrites the equation as $w^m = m(w-1)+1$ and reads it as "the exponential $m \mapsto w^m$ agrees with the linear function $m \mapsto m(w-1)+1$ at $m = j+1$ and $m = k+1$".

> *Two equations that are the same equation with different values of n — I can't eliminate between them, but I can ask how many n could possibly satisfy one fixed equation. If I invert, $w^{n+1} = (n+1)w - n$: an exponential in $n+1$ equals a linear function of $n+1$. An exponential and a line don't meet often.*

> *It's already true at $m = 0$ and $m = 1$ for free, so a common root gives four meeting points. For real $w > 0$ a convex exponential meets a line at most twice, but w is complex — so take $|\cdot|^2$: that turns the line into a quadratic in m and the exponential into $|w|^{2m}$, and a real exponential minus a quadratic has at most three zeros. Four beats three.*

Layer 3 THE WHY

Why should anyone think of varying the exponent? Because of the shape of the claim. "P_j and P_k are coprime for all $j \neq k$" is a statement that the zero sets $Z_1, Z_2, Z_3, \ldots$ of the P_n are pairwise disjoint. Turn it around: fix a complex number z and ask *for how many* n is $z \in Z_n$? The claim is that the answer is always at most one. That is a counting question in the variable n, with z held fixed — and counting solutions of an equation in one real variable is exactly what Rolle's theorem does. The pairwise-disjointness phrasing is the cue; "relatively prime" is its algebraic disguise.

Why invert z to $w = 1/z$? Because the equation $nz^{n+1} - (n+1)z^n + 1 = 0$ has the variable n tangled into both the exponents and the coefficients. Dividing by z^{n+1} separates them: $n - (n+1)w + w^{n+1} = 0$ puts all the n-dependence of the exponential into one term, w^{n+1}, and leaves the rest *linear* in n. That separation — exponential on one side, polynomial on the other — is what makes the zero-counting possible.

Why do $m = 0$ and $m = 1$ come free, and why does that matter? The line $m \mapsto m(w-1)+1$ is by construction the line through $(0, w^0) = (0, 1)$ and $(1, w^1) = (1, w)$: it is the chord of the exponential $m \mapsto w^m$ between $m = 0$ and $m = 1$. So the equation $w^m = m(w-1)+1$ says "the point (m, w^m) lies on the chord from 0 to 1". For a real base $w > 0$, $w \neq 1$, the function w^t is strictly convex and meets any line in at most two points — the two chord endpoints are already used, so no integer $m \geq 2$ can satisfy the equation, and the problem is over. The only obstacle is that w is complex. Taking $|\cdot|^2$ is the cheapest way to get a real equation; it costs one degree (the line becomes a quadratic), which allows three zeros instead of two. A common root supplies four. The margin is exactly one, and it is supplied by the two free points. That is why the proof works, and it is why the Remark in the official solution generalises to four distinct integers $a < b < c < d$: any four points on a line, against an exponential, after squaring.

Kedlaya's first solution defines $f(t) = |w|^{2t} - t^2|w|^2 + 2t(t-1)\mathrm{Re}(w) - (t-1)^2$ without comment. Expand $|t(w-1)+1|^2 = t^2|w|^2 - 2t(t-1)\mathrm{Re}(w) + (t-1)^2$ and you recover exactly the f above; it is the squared modulus of the chord. It also asserts $f'''(t) > 0$, which needs $|w| > 1$; that is true (every root of P_n has $|z| < 1$, by the Eneström–Kakeya argument of the second solution) but not proved in the first solution. The two-case version above avoids needing it.

Caveat. Verified by computer: $\gcd(P_j, P_k) = 1$ over $\mathbb{Q}$ for all $1 \leq j < k \leq 30$, and numerically every root of P_n ($2 \leq n \leq 39$) has modulus below 1, consistent with the remark that $f''' > 0$ in the official solution. The proof above does not depend on either check.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the family structure: a claim about P_j and P_k for *all* pairs $j \neq k$. Write down the closed form $P_n(x) = \frac{nx^{n+1} - (n+1)x^n + 1}{(x-1)^2}$ and then rewrite the root condition with n as the unknown: $w^{n+1} = (n+1)w - n$.

2. Zero counting via Rolle's theorem for exponential-minus-polynomial functions: $e^{ct} - q(t)$ with $\deg q = d$ has at most $d + 1$ real zeros when $c \neq 0$.

3. Hesitation occurs when w turns out to be complex and the convexity argument does not apply. Reinforce with the standard move: to get a real equation from a complex one, take $|\cdot|^2$; it doubles the degree of polynomial parts and turns w^m into $|w|^{2m}$, which is still exponential. Then recount. A second hesitation: the case $|w| = 1$, where the exponential degenerates to a constant — handle it separately rather than hoping it cannot occur.

4. The trap is irreducibility. It looks like the obvious route to coprimality, and it is an open-looking problem. Close behind is Euclid's algorithm, which works for $k = j + 1$ and gives false hope. Both treat j, k as fixed; the problem is only tractable when the index is allowed to vary.

5. Slogan: *"the zero sets are pairwise disjoint" is a counting statement in the index; fix the root and count the indices.* More broadly: whenever the index n appears as an exponent, promote it to a real variable t and use calculus on t.

6. Decision rule: **when a family of equations indexed by n must have at most one solution n for each fixed parameter, isolate the exponential dependence on n from the polynomial dependence, extend n to a real variable, and count zeros with Rolle.** Companion rule: an exponential meets a polynomial of degree d in at most $d + 1$ points, so identify the "free" zeros ($n = 0, 1$, trivial cases) and check whether the given ones push the count past $d + 1$.

▷ **Cue:** a statement that polynomials P_j, P_k in a family share no root, for all $j \neq k$ $\longrightarrow$ **Move:** fix a putative common root and count how many indices n it can serve; aim to show at most one

▷ **Cue:** a root equation in which the index n appears both as an exponent and as a coefficient, e.g. $nz^{n+1} - (n+1)z^n + 1 = 0$ $\longrightarrow$ **Move:** divide through by the top power ($w = 1/z$) to isolate a single exponential term w^{n+1} against a polynomial in n

▷ **Cue:** w^m equals a linear function of m at several integers m, with w complex $\longrightarrow$ **Move:** take $|\cdot|^2$ to get $|w|^{2t} = q(t)$ with q quadratic; count zeros of $|w|^{2t} - q(t)$ by three applications of Rolle; treat $|w| = 1$ separately

▷ **Cue:** a sum $\sum_{k=1}^{n} kx^{k-1}$ $\longrightarrow$ **Move:** recognise it as $\frac{d}{dx}\frac{x^{n+1}-1}{x-1}$ and write the closed form $\frac{nx^{n+1}-(n+1)x^n+1}{(x-1)^2}$

2014 A6

2014 A6. Let n be a positive integer. What is the largest k for which there exist $n \times n$ matrices $M_1, \ldots, M_k$ and $N_1, \ldots, N_k$ with real entries such that for all i and j, the matrix product $M_i N_j$ has a zero entry somewhere on its diagonal if and only if $i \neq j$?

Fields: linear algebra (dimension counting, multilinear forms / tensor products); combinatorics (extremal construction) **Position:** A6 — the hardest slot on paper A. The construction is findable; the upper bound needs a vector space nobody names in the statement. **Answer:** $k = n^n$. Permutation-like 0/1 matrices indexed by functions $\{1, \ldots, n\} \to \{1, \ldots, n\}$ achieve it; the upper bound comes from the linear independence of the k multilinear forms $y \mapsto \prod_r \langle \mathrm{row}_r(M_i), y_r \rangle$ in a space of dimension n^n.

Layer 1 THE WHAT — surface analysis

The natural attempt. You start small. For $n = 1$ the matrices are numbers, $m_i n_j = 0$ iff $i \neq j$. Two pairs would need $m_1 n_2 = 0$ with $m_1 \neq 0$, so $n_2 = 0$, contradicting $m_2 n_2 \neq 0$. So $k = 1$.

For $n = 2$ you try to build examples. Diagonal matrices: $M_i = \mathrm{diag}(a_i, b_i)$, $N_j = \mathrm{diag}(c_j, d_j)$, and $M_i N_j$ has diagonal $(a_i c_j, b_i d_j)$. You want, for each $i \neq j$, one of $a_i c_j$, $b_i d_j$ to be zero, while $a_i c_i b_i d_i \neq 0$. Since a_i, c_i etc. are all nonzero on the diagonal pairs, $a_i c_j = 0$ is impossible — diagonal matrices give only $k = 1$. So you allow off-diagonal entries and try the matrices with a

single 1 in each row. Row r of M picks a column $\sigma(r)$; column r of N picks a row $\tau(r)$. Then $(MN)_{rr} = [\sigma(r) = \tau(r)]$. With $\sigma = \tau$ the diagonal is all 1s; with $\sigma \neq \tau$ some entry is 0. There are n^n functions σ. So $k \geq n^n$, and for $n = 2$ that is 4.

Now the upper bound. You try to argue about ranks: if $M_i N_j$ has a zero diagonal entry… but a zero diagonal entry says nothing about rank or determinant. You try pigeonhole on the zero patterns: there are $2^n - 1$ nonempty sets of zero positions, but the condition is a whole $k \times k$ array of such patterns, and you cannot see what forbids $k = n^n + 1$. You try to perturb: can a fifth pair be added for $n = 2$? Every attempt fails, but you cannot say why. The wall is that "has a zero somewhere on the diagonal" is an *or* of n conditions, and no linear-algebra tool you know handles an *or*.

The model chain. *Step 1: the construction, $k \geq n^n$.* Let $e_1, \ldots, e_n$ be the standard basis of $\mathbb{R}^n$ (as row vectors). For each function $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$ let M_σ be the matrix whose r-th row is $e_{\sigma(r)}$, and $N_\sigma = M_\sigma^T$, whose r-th column is $e_{\sigma(r)}^T$. Then

$$(M_\sigma N_\tau)_{rr} = e_{\sigma(r)} \cdot e_{\tau(r)} = \begin{cases} 1 & \sigma(r) = \tau(r) \\ 0 & \text{otherwise.} \end{cases}$$

So $M_\sigma N_\tau$ has a zero on the diagonal iff $\sigma(r) \neq \tau(r)$ for some r, i.e. iff $\sigma \neq \tau$. There are n^n functions σ.

Step 2: turn "or" into "product". The diagonal of $M_i N_j$ contains a zero iff the product of its diagonal entries is zero:

$$D(i,j) := \prod_{r=1}^{n} (M_i N_j)_{rr} = \prod_{r=1}^{n} \langle \mu_r^{(i)}, \nu_r^{(j)} \rangle,$$

where $\mu_r^{(i)}$ is the r-th row of M_i, $\nu_r^{(j)}$ the r-th column of N_j, and $\langle \cdot, \cdot \rangle$ the dot product in $\mathbb{R}^n$. The hypothesis is: $D(i,j) = 0$ for $i \neq j$ and $D(i,i) \neq 0$.

Step 3: the space of forms. Consider real-valued functions of n vectors $y_1, \ldots, y_n \in \mathbb{R}^n$ (so n^2 real variables $y_{r,s}$, the s-th coordinate of y_r). Let V be the span of the monomials

$$y_{1,s_1}\, y_{2,s_2} \cdots y_{n,s_n}, \qquad s_1, \ldots, s_n \in \{1, \ldots, n\},$$

one factor from each vector. There are n^n such monomials, so $\dim V \leq n^n$. For each i define

$$F_i(y_1, \ldots, y_n) = \prod_{r=1}^{n} \langle \mu_r^{(i)}, y_r \rangle = \prod_{r=1}^{n} \Big(\sum_{s=1}^{n} \mu_{r,s}^{(i)}\, y_{r,s} \Big).$$

Expanding the product, each term takes one $y_{r,s}$ from each factor r, so $F_i \in V$. By Step 2, $F_i(\nu_1^{(j)}, \ldots, \nu_n^{(j)}) = D(i,j)$.

Step 4: independence. Suppose $\sum_i c_i F_i = 0$ as a function. Evaluate at $(\nu_1^{(j)}, \ldots, \nu_n^{(j)})$:

$$0 = \sum_i c_i D(i,j) = c_j D(j,j),$$

and $D(j,j) \neq 0$, so $c_j = 0$ for every j. Hence $F_1, \ldots, F_k$ are linearly independent in V, and $k \leq \dim V \leq n^n$.

Combining, the largest k is n^n.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains build the same construction and both face the upper bound with the same hypothesis in hand. At (†) — "what can I do with *some diagonal entry is zero*?" — they split. The natural attempt keeps the condition as an "or" of n events and looks for rank or counting arguments. The model replaces the "or" by the vanishing of a product, $\prod_r (M_i N_j)_{rr} = 0$, and then notices that this product is linear in each row of M_i separately.

> *"Some diagonal entry is zero" is the same as "the product of the diagonal entries is zero". And I have a $k \times k$ array $D(i,j)$ that is zero off the diagonal and nonzero on it — that's an invertible diagonal matrix. If I can write $D(i,j)$ as a dot product $\langle m_i, n_j \rangle$ in some space of dimension d, then $k \leq d$.*

> *The product $\prod_r \langle row_r(M_i), col_r(N_j) \rangle$ is linear in each row of M_i and in each column of N_j separately. Multilinear in n vectors of length n: expand it and count the monomials. n choices in each of n factors — n^n. That matches the construction, so this is the bound.*

Layer 3 THE WHY

Why is the "diagonal matrix of dot products" the thing to look for? Go back to what the hypothesis is: a $k \times k$ table of numbers $D(i,j)$, indexed by an M and an N, that is zero exactly off the diagonal. The only general principle that bounds the size of such a table is rank: a diagonal matrix with nonzero diagonal has rank k, and a matrix of the form $D(i,j) = \langle m_i, n_j \rangle$ with $m_i, n_j \in \mathbb{R}^d$ is a product of a $k \times d$ and a $d \times k$ matrix, so has rank at most d. Every upper bound of the form "k objects pairwise orthogonal-but-not-self-orthogonal" is this argument. So the question the solver must ask is not "what trick bounds this?" but "in what vector space is $D(i,j)$ a dot product?" — and, because the construction shows $k = n^n$ is attainable, any space that gives the sharp bound must have dimension exactly n^n. The target dimension is known before the space is found. That is a strong hint: the space must have a basis indexed by functions $\{1,\dots,n\} \to \{1,\dots,n\}$, the same objects that index the construction.

Why a product and not some other combination of the diagonal entries? Because the argument needs a single number that is zero iff *at least one* entry is zero. Sums fail (entries can cancel); the product is the canonical "or-detector" for zeros. The price is that the product is not linear in M_i — but it is linear in each row separately, because each diagonal entry $(M_i N_j)_{rr}$ involves only row r of M_i. Multilinearity is the replacement for linearity, and the standard way to linearise a multilinear expression is to expand it into monomials, one coordinate chosen from each argument. Counting those monomials gives n^n.

Kedlaya's solution introduces the tensor product $(\mathbb{R}^n)^{\otimes n}$ and "the tensor product of the rows of M_i". Here is where that comes from: the tensor product is *defined* as the space in which multilinear expressions become linear. The vector $m_i = \mu_1^{(i)} \otimes \cdots \otimes \mu_n^{(i)}$ has coordinates $\mu_{1,s_1}^{(i)} \cdots \mu_{n,s_n}^{(i)}$ — exactly the coefficients of the monomials in the expansion of F_i. The polynomial space V in the model chain is the same construction with the vocabulary removed, and it is how to find the argument without knowing the word "tensor".

The problem is phrased with $M_i N_j$ rather than a single family precisely because it is a rank question in disguise: two families, a pairing between them, and a diagonal pattern. Seen that way, it belongs with the Oddtown/Eventown theorems and the Frankl–Wilson method, where a combinatorial "$\neq$" is converted into a vanishing polynomial and dimension does the rest.

Caveat. Construction and bound both verified by hand; for $n = 1$ the argument in the natural attempt independently gives $k = 1 = 1^1$. The inequality $\dim V \leq n^n$ is all that is needed (in fact $\dim V = n^n$, the monomials being distinct and hence independent).

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the pattern "$M_i N_j$ has property $\mathcal{P}$ if and only if $i \neq j$": two families, a pairing, a diagonal pattern. Write down the $k \times k$ table $D(i,j)$ and the goal: express $D(i,j)$ as a dot product in a space of known dimension.
2. The linear-algebra (dimension) method: produce k vectors that are provably independent because their pairings with a second family form an invertible (here diagonal) matrix.

3. Hesitation occurs at "some entry is zero", which looks non-algebraic. Reinforce with the translation: an *or* of vanishing conditions is the vanishing of a product. Then check what the product is linear in.

4. The trap is working with the matrices as wholes: rank, determinant, trace of M_iN_j. None of them sees the diagonal entries individually, and the condition is entirely about individual diagonal entries. A second trap: believing the construction is not optimal and hunting for a better one, instead of trusting n^n and looking for the space of that dimension.

5. Slogan: *a table that is zero exactly off the diagonal bounds its size by the dimension of any space in which it is a table of dot products.* Find the space by expanding: products become tensor products, multilinear becomes linear.

6. Decision rule: **when two families $\{A_i\}, \{B_j\}$ interact so that some quantity vanishes iff $i \neq j$, write that quantity as $\langle a_i, b_j \rangle$ in a vector space and conclude $k \leq$ its dimension; if the quantity is a product of n bilinear pairings, the space is the n-fold tensor product, of dimension the product of the dimensions.** Companion rule: when the construction gives a clean number like n^n, look for a space whose basis is indexed by the same objects as the construction.

▷ **Cue:** two indexed families with a pairing that is zero iff $i \neq j$ (a "diagonal" incidence pattern) $\longrightarrow$ **Move:** realise the pairing as a dot product $\langle a_i, b_j \rangle$ in a space of dimension d; linear independence of the a_i gives $k \leq d$

▷ **Cue:** a condition "at least one of the quantities $q_1, \ldots, q_n$ is zero" $\longrightarrow$ **Move:** replace it by $\prod_r q_r = 0$; check whether the product is multilinear in natural blocks of variables

▷ **Cue:** a product of n expressions, each linear in its own vector of length n $\longrightarrow$ **Move:** expand into monomials choosing one coordinate from each factor; they span a space of dimension n^n (the tensor product $(\mathbb{R}^n)^{\otimes n}$)

▷ **Cue:** an extremal problem whose construction is indexed by functions $\{1, \ldots, n\} \to \{1, \ldots, n\}$ $\longrightarrow$ **Move:** expect the upper bound to come from a vector space with a basis indexed by the same functions

Putnam 2014 — Paper B

2014 B1

2014 B1. A *base* 10 *over-expansion* of a positive integer N is an expression of the form

$$N = d_k 10^k + d_{k-1} 10^{k-1} + \cdots + d_0 10^0$$

with $d_k \neq 0$ and $d_i \in \{0, 1, 2, \ldots, 10\}$ for all i. For instance, the integer $N = 10$ has two base 10 over-expansions: $10 = 10 \cdot 10^0$ and the usual base 10 expansion $10 = 1 \cdot 10^1 + 0 \cdot 10^0$. Which positive integers have a unique base 10 over-expansion?

Fields: number theory, combinatorics (positional representations) **Position:** B1 – opener, but the converse direction carries the real content **Answer:** Exactly the positive integers whose usual base-10 expansion has no digit equal to 0.

Layer 1 THE WHAT — surface analysis

The natural attempt. The worked example is generous: $N = 10$ branches into two over-expansions precisely because its usual expansion has a 0 digit, and the second expansion is obtained by "borrowing" – replace $d_{i+1} \cdot 10^{i+1} + 0 \cdot 10^i$ by $(d_{i+1} - 1) \cdot 10^{i+1} + 10 \cdot 10^i$. Generalising this forward direction is immediate: any 0 digit anywhere in the usual expansion (with a nonzero digit above it) gives a second over-expansion by the same borrowing move, so a number with a 0 digit is never unique. The natural attempt then tries to prove the converse – no 0 digits implies uniqueness – by induction on the number of digits, and this is where it goes soft. The instinct is: "the last digit d_0 of any over-expansion satisfies $d_0 \equiv N \pmod{10}$, and since N's digits run over $\{0, \ldots, 9\}$ in the usual sense, this pins d_0 down uniquely." That sentence is true for the *usual* expansion, where d_0 is confined to $\{0, \ldots, 9\}$ – ten symbols for ten residues, a clean bijection. But an over-expansion allows $d_0 \in \{0, \ldots, 10\}$, eleven symbols, and the induction as stated never asks whether that extra symbol reopens the very ambiguity it is trying to rule out. Carried through uncritically, the "proof" reaches the correct final answer (it has been pattern-matched from the example) but has a silent gap exactly at the step that decides the problem.

The model chain. Forward direction: if some digit d_i of the usual expansion of N is 0 and $d_{i+1} \neq 0$ (such an i exists whenever any digit is 0, since $d_k \neq 0$), replace $d_{i+1} 10^{i+1} + 0 \cdot 10^i$ by $(d_{i+1} - 1) 10^{i+1} + 10 \cdot 10^i$ to produce a second, distinct over-expansion. So any N with a 0 digit has more than one over-expansion.

Converse, by induction on the number of digits of N: if $1 \leq N \leq 9$ the claim is immediate (the only over-expansion is $N = N \cdot 10^0$, since a two-term expression $d_1 \cdot 10 + d_0$ with $d_1 \geq 1$ already exceeds 9). For $N \geq 10$ with no 0 digit in its usual expansion, take any over-expansion $N = d_k 10^k + \cdots + d_0$. Reducing mod 10 forces $d_0 \equiv N \pmod{10}$ with $d_0 \in \{0, \ldots, 10\}$. Among the eleven values $\{0, 1, \ldots, 10\}$, every residue class mod 10 has exactly one representative *except* residue 0, which has two: 0 and 10. Since N's usual last digit is nonzero by hypothesis, $N \not\equiv 0 \pmod{10}$, so d_0 is forced to equal that nonzero usual last digit – uniquely. Then $(N - d_0)/10$

is a positive integer whose usual digits are exactly the remaining (higher) digits of N, still all nonzero, and $d_k 10^{k-1} + \cdots + d_1 10^0$ is an over-expansion of it; by the induction hypothesis this is unique, so the original over-expansion of N is unique too.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths write down "$d_0 \equiv N \pmod{10}$." They split on what happens next. The natural attempt treats this as already finishing the job, importing the habit that base-10 digits run over ten values. The model stops and counts: d_0 ranges over *eleven* values, $\{0, \ldots, 10\}$, and asks how many of those eleven collide in each residue class mod 10 before declaring d_0 unique.

> *The digit alphabet here is $\{0, \ldots, 10\}$, eleven symbols, not the usual ten – so "determined mod 10" is not automatically "determined." I need to count, for each residue class mod 10, how many elements of $\{0, \ldots, 10\}$ land in it, before I'm allowed to call d_0 unique.*

> *Residue class 0 mod 10 contains two elements of $\{0, \ldots, 10\}$ – namely 0 and 10 – and every other residue class contains exactly one. So the* only *place an over-expansion can branch at the last digit is where $N \equiv 0 \pmod{10}$, i.e. where the ordinary last digit is 0. That is precisely the hypothesis I am being asked to characterise, not a side issue to wave past.*

Layer 3 THE WHY

The usual base-10 expansion is the special case of a positional representation where the digit alphabet $\{0, \ldots, 9\}$ has exactly as many symbols as the base has residues: ten symbols, ten classes mod 10, a bijection, hence a unique representative in each class, hence a unique expansion overall (this is why ordinary decimal notation is unique in the first place – a fact usually taken for granted rather than derived). An over-expansion enlarges the alphabet by exactly one symbol, to eleven. Enlarging a set of representatives for ten classes by one element forces, by the pigeonhole principle, exactly one class to end up with two representatives and no class to end up with zero (there are still eleven items, one modulus of size ten). Which class gets the extra representative is not free – it is determined by where the new symbol 10 was inserted, and $10 \equiv 0 \pmod{10}$, so the doubled class is residue 0. This is the whole content of the problem: extending a positional alphabet by one symbol beyond the base creates the minimum possible amount of redundancy, and that redundancy is localised, digit-position by digit-position, exactly at the residue class matching the new symbol. A number's over-expansion is unique if and only if none of its digit positions ever needs to resolve that ambiguous class, i.e. if and only if none of its usual digits is 0. Kedlaya's solution states the borrowing move and the induction without discussing why the alphabet size of eleven is the source of non-uniqueness; that structural reason – one extra symbol, one collided residue class, located at 0 – is what makes the "no zero digits" answer inevitable rather than merely verified case by case.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the worked example itself: $N = 10$ gets a second over-expansion exactly where its usual expansion has a 0. On seeing it, write down immediately: this is evidence of a general counting fact about the digit alphabet, not just a one-off curiosity to be pattern-matched.
2. That cue suggests the method: count how many elements of the extended digit range $\{0, \ldots, 10\}$ fall into each residue class mod 10 – a direct pigeonhole count – rather than trying to verify uniqueness by checking examples up to some bound and guessing the pattern holds.
3. Hesitation occurs right after writing "$d_0 \equiv N \pmod{10}$, hence d_0 is determined" – the sentence that feels finished but is not. Reinforce it away by asking explicitly: determined within which set, and how many elements of that set share this residue?

4. The trap is the habit of thinking of base-10 digits as automatically confined to $\{0,\dots,9\}$; the brain autocompletes back to the familiar range even after the problem has explicitly redefined it to $\{0,\dots,10\}$, and that autocomplete is exactly what hides the one case that matters.

5. Recognisable faster next time by treating any "digit range extended beyond the usual $[0, \text{base}-1]$" setup as an automatic cue to count collisions mod the base in the new range before asserting anything is unique.

6. Reusable rule: uniqueness of a positional representation is a counting fact about digit-alphabet size versus the base's residue classes. Enlarging the alphabet by k symbols beyond the base creates exactly k residue classes with more than one representative (by pigeonhole, generically one extra representative per class, distributed according to where the new symbols sit); non-uniqueness at a given digit position occurs if and only if the true digit for that position lies in one of those ambiguous classes.

▷ **Cue:** a positional-representation uniqueness question where the digit range strictly contains the usual $[0, \text{base}-1]$ $\longrightarrow$ **Move:** count, for each residue class mod the base, how many elements of the enlarged digit range represent it; a digit position branches into multiple valid values exactly where its class has more than one representative

▷ **Cue:** a problem statement supplies one small worked example of non-uniqueness (here $N = 10$) $\longrightarrow$ **Move:** do not pattern-match the answer from the example alone; extract the counting principle the example is evidence for, and check it accounts for every case, including proving the converse rigorously

2014 B2

2014 B2. Suppose that f is a function on the interval $[1,3]$ such that $-1 \le f(x) \le 1$ for all x and $\int_1^3 f(x)\,dx = 0$. How large can $\int_1^3 \frac{f(x)}{x}\,dx$ be?

Fields: analysis (extremal problems under a pointwise bound and a linear constraint) **Position:** B2
Answer: $\log\frac{4}{3}$, attained by $f = 1$ on $[1,2)$ and $f = -1$ on $(2,3]$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The integrand is a product of $f(x)$ and the weight $1/x$, and the reflex on seeing "bound this integral of a product" is Cauchy–Schwarz:

$$\left|\int_1^3 \frac{f(x)}{x}\,dx\right| \le \left(\int_1^3 f(x)^2\,dx\right)^{1/2}\left(\int_1^3 \frac{1}{x^2}\,dx\right)^{1/2}.$$

Since $-1 \le f \le 1$, $\int_1^3 f^2 \le 2$; and $\int_1^3 x^{-2}\,dx = 1 - \frac{1}{3} = \frac{2}{3}$. So the bound comes out to $\sqrt{2}\cdot\sqrt{2/3} = \sqrt{4/3} = 2/\sqrt{3} \approx 1.1547$. This is a completely correct inequality, and it is also nearly four times the true answer $\log\frac{4}{3} \approx 0.2877$: it is true but useless. The reason it is useless is visible on inspection – the bound never used $\int_1^3 f = 0$ at all, and equality in Cauchy–Schwarz would require $f(x)$ proportional to $1/x$, which is impossible under the pointwise constraint $|f| \le 1$ except in a degenerate limiting sense. Carrying this further – trying to patch it by subtracting off a multiple of the mean, or squeezing the f^2 bound tighter – does not close the gap, because Cauchy–Schwarz is structurally blind to the sign pattern of f; it only sees the size $\int f^2$, and the size alone cannot know that f should be $+1$ where $1/x$ is large and -1 where $1/x$ is small.

The model chain. Guess the extremal shape first: since $1/x$ is decreasing on $[1,3]$ and f is pointwise-bounded with a fixed (zero) mean, the best use of the "budget" is to push f to $+1$ where the weight $1/x$ is largest and to -1 where it is smallest, splitting at the point that keeps

$\int f = 0$. By symmetry of length (not weighted length) on $[1,3]$, that split is the midpoint $x = 2$: let

$$g(x) = \begin{cases} 1 & 1 \le x < 2 \\ -1 & 2 < x \le 3. \end{cases}$$

Then $\int_1^3 g\,dx = 1 - 1 = 0$ and

$$\int_1^3 \frac{g(x)}{x}\,dx = \int_1^2 \frac{1}{x}\,dx - \int_2^3 \frac{1}{x}\,dx = \log 2 - (\log 3 - \log 2) = 2\log 2 - \log 3 = \log \frac{4}{3}.$$

Now prove no admissible f beats g. Let $h = g - f$; then $\int_1^3 h\,dx = 0 - 0 = 0$, and $h(x) = g(x) - f(x) \ge 1 - 1 = 0$ on $[1,2)$ (since $g = 1 \ge f$), while $h(x) = g(x) - f(x) \le -1 - (-1) = 0$ on $(2,3]$ (since $g = -1 \le f$). Split the (zero) integral of h against the constant weight $\frac{1}{2}$:

$$0 = \tfrac{1}{2}\int_1^3 h\,dx = \tfrac{1}{2}\int_1^2 h\,dx + \tfrac{1}{2}\int_2^3 h\,dx.$$

On $[1,2)$, $1/x \ge 1/2$ and $h \ge 0$, so $h(x)/x \ge h(x)/2$; on $(2,3]$, $1/x \le 1/2$ and $h \le 0$, so $h(x)/x \ge h(x)/2$ as well (multiplying a nonpositive number by a smaller positive weight makes it less negative). Hence

$$\int_1^3 \frac{h(x)}{x}\,dx \;\ge\; \tfrac{1}{2}\int_1^3 h(x)\,dx = 0,$$

so $\int_1^3 f/x\,dx = \int_1^3 g/x\,dx - \int_1^3 h/x\,dx \le \log\frac{4}{3} - 0 = \log\frac{4}{3}$. Since g itself attains this value, the supremum is a maximum, equal to $\log\frac{4}{3}$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths face the same integrand $f(x)/x$. Cauchy–Schwarz treats it as an abstract inner product and bounds it by the sizes $\|f\|_2$, $\|1/x\|_2$ alone. The model instead never asks "how big can this integral be in the abstract" – it asks "what specific competitor function should f be compared against," constructs that competitor g directly from the shape of the constraints, and proves optimality by a pointwise sign argument on $h = g - f$ rather than a norm inequality.

> *The constraint on f is pointwise – $|f(x)| \le 1$ for every x – not a bound on $\|f\|_2$. Cauchy–Schwarz answers a question about L^2 size, and that is not the question this problem is asking; I should instead ask what is the best possible sign pattern for f subject to the zero-mean constraint, and prove that pattern is optimal by direct comparison, not by an inequality that discards the sign information entirely.*

> *Because $\int_1^3 f = 0$ is fixed, I am free to compare f against any other function with the same mean by looking at their difference: $h = g - f$ has $\int h = 0$ automatically for any competitor g I choose with $\int g = 0$. That turns a global integral inequality into a pointwise argument, provided I choose g so that $h = g - f$ has a controlled, single sign on each piece of the domain where the weight $1/x$ itself is one-signed relative to its value at the split point.*

Layer 3 THE WHY

The reason Cauchy–Schwarz cannot reach the sharp constant is structural: $\{f : |f(x)| \le 1 \text{ for all } x\}$ is (the ball of) L^∞, not L^2, and $f \mapsto \int f \cdot (1/x)\,dx$ is a linear functional. Maximising a linear functional over a convex set is achieved at an extreme point of that set, and the extreme points of the L^∞ unit ball are exactly the ± 1-valued ("bang-bang") functions – any f that is

strictly between -1 and 1 on a set of positive measure can be perturbed to increase the functional while staying feasible, so it cannot be optimal. This is why the correct first move is to guess a ± 1-valued candidate rather than to bound a norm: the extremal object is forced by the geometry of the constraint set, not merely permitted by it. Which ± 1-valued function is optimal is then decided by the monotonicity of the weight $1/x$: reallocating a unit of "$+1$-measure" from a point where $1/x$ is small to a point where $1/x$ is large, while removing the same measure of "-1" to preserve the zero mean, strictly increases $\int f/x$ whenever the swap moves mass toward higher weight – so the optimal f must be $+1$ exactly on the region of highest weight and -1 on the region of lowest weight, i.e. an initial segment $[1, c)$ and a final segment $(c, 3]$ for some threshold c; the zero-mean constraint with equal lengths on each side (since the weight's only role once the shape is fixed is monotonicity, and $[1, 3]$ has the same amount of "space" on either side of 2) forces $c = 2$. The comparison argument with $h = g - f$ is then this optimality reasoning made fully rigorous and elementary, without invoking the general extreme-point theorem by name: it is that theorem's proof, carried out by hand for this specific weight and constraint.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the combination of a pointwise bound $|f| \leq 1$ (not an L^2 bound) together with a monotone weight $1/x$ inside the integral. On seeing it, write down immediately: this is a bang-bang extremal question, not a norm-inequality question.

2. That cue suggests the method: guess a ± 1-valued candidate whose sign matches the ranking of the weight (high weight $\to f = +1$, low weight $\to f = -1$), with the split point fixed by the zero-mean constraint, then prove optimality by direct comparison rather than by an abstract inequality.

3. Hesitation occurs right at the moment the integrand is recognised as a product $f(x) \cdot (1/x)$ and the hand reaches for Cauchy–Schwarz out of habit. Reinforce it away by actually computing the Cauchy–Schwarz bound and checking whether it uses the zero-mean constraint at all – if a hypothesis of the problem never enters the bound, the bound is not using all the information and is very unlikely to be sharp.

4. The trap is that Cauchy–Schwarz produces a completely correct inequality that feels like progress – it is progress toward a bound, but not toward the sharp value, and the gap between $2/\sqrt{3}$ and $\log \frac{4}{3}$ is large enough that no amount of tightening the same method closes it.

5. Recognisable faster next time by treating "pointwise (L^∞) bound on the unknown function, plus a monotone weight, plus a linear (integral) side constraint" as a fixed pattern that calls for a bang-bang guess and a comparison proof.

6. Reusable rule: to maximise $\int f\,w\,dx$ subject to $|f| \leq 1$ pointwise and $\int f = c$ fixed, with w monotone, guess the extremal as a ± 1-valued function with sign matching the order of w, split so as to satisfy the mean constraint; then prove optimality by setting $h =$(extremal)$-f$ and showing $\int h\,w\,dx \geq 0$ using only the single sign of h on each region where w is comparable to its value at the split point.

▷ **Cue:** maximise $\int f\,w\,dx$ subject to a pointwise bound $|f| \leq 1$ and a linear constraint on f, with w monotone $\longrightarrow$ **Move:** guess the bang-bang extremal $f = \pm 1$ with sign matching the order of w, split point fixed by the linear constraint; prove optimality via $h = (\text{extremal}) - f$ and a pointwise sign argument, not a norm inequality

▷ **Cue:** an integral of a product $f(x)w(x)$ appears and the reflex reaches for Cauchy–Schwarz $\longrightarrow$ **Move:** first check whether the actual constraint on f is pointwise (L^∞) rather than a size bound (L^2); if it is pointwise, Cauchy–Schwarz discards the binding constraint and a comparison/exchange argument is needed instead

2014 B3

2014 B3. Let A be an $m \times n$ matrix with rational entries. Suppose that there are at least $m + n$ distinct prime numbers among the absolute values of the entries of A. Show that the rank of A is at least 2.

Fields: linear algebra + number theory (unique factorization), via a combinatorial (graph-cycle) reformulation **Position:** B3 **Answer:** Rank$(A) \geq 2$ always; proved by contradiction from assuming rank$(A) \leq 1$.

Layer 1 THE WHAT — surface analysis

The natural attempt. Rank$(A) \leq 1$ is a concrete algebraic statement: every 2×2 minor vanishes, $A_{11}A_{ij} = A_{i1}A_{1j}$ for all i, j (after arranging, without loss of generality, that $A_{11} \neq 0$). The natural instinct is to work with this condition directly: fix the corner entry A_{11}, and try to show that having $m + n$ distinct primes scattered through the grid forces some minor to fail. This runs into a genuine obstacle almost immediately – the prime entries are not conveniently placed at row 1 or column 1; they can sit anywhere in the $m \times n$ grid, in any pattern consistent with there being at least $m + n$ of them. Trying to reason about "entry A_{ij} is prime" and "entry A_{kl} is prime" for arbitrary, unrelated (i, j) and (k, l) produces a large, unstructured case split: sometimes two prime entries share a row, sometimes a column, sometimes neither, and the minor condition connecting any two of them requires a third and fourth entry (to complete the 2×2 submatrix) that need not themselves be prime or even distinctive in any way. Counting arguments are tried too – with $m + n$ primes distributed among m rows, pigeonhole guarantees some row (or column) contains more than one prime entry – but knowing one row is "crowded" does not by itself contradict rank 1, since a rank-1 matrix can perfectly well have several prime entries in a single row (they only need to be a fixed row-vector $b_1, \ldots, b_n$ scaled by that row's a_i). The natural attempt has the right algebraic tool (the minor condition) and the right general instinct (counting), but no organising structure for which primes to chain together, and stalls in casework.

The model chain. Assume, for contradiction, rank$(A) \leq 1$; discard any all-zero rows or all-zero columns first (this can only reduce m, n, so at least $m + n$ distinct primes remain among the surviving entries, an even stronger hypothesis relative to the smaller dimensions). Build a bipartite graph G: one vertex for each remaining row, one vertex for each remaining column – $m + n$ vertices in total – and an edge joining row i to column j whenever $|A_{ij}|$ is one of the distinct primes. By hypothesis G has at least $m + n$ edges. A graph with no cycles (a forest) on V vertices has at most $V - 1$ edges; since G has $\geq V$ edges on $V = m + n$ vertices, G is not a forest and must contain a cycle. Walk that cycle: row i_1 – col j_1 – row i_2 – col j_2 – $\cdots$ – row i_k – col j_k – back to row i_1, for some $k \geq 2$ (bipartite graphs have no odd cycles and no repeated edges, so the shortest cycle has even length $2k$ with $k \geq 2$). Let $p_1, \ldots, p_{2k}$ be the (distinct) primes at the entries $A_{i_1j_1}, A_{i_2j_1}, A_{i_2j_2}, A_{i_3j_2}, \ldots, A_{i_kj_k}, A_{i_1j_k}$ around the cycle. Since rank$(A) \leq 1$, every 2×2 minor vanishes; applied to consecutive entries sharing a row or column around the cycle, this gives $|A_{i_1j_1}||A_{i_2j_2}| = |A_{i_2j_1}||A_{i_1j_2}|$-type relations that chain together (via the shared auxiliary factors from rank 1) to the single identity

$$p_1 p_2 \cdots p_{2k} = \big(\text{a product of the "in-between" magnitudes}\big)^2,$$

a perfect square. (Checked in the smallest case $k = 2$: a 2×2 submatrix with all four entries prime, $|A_{i_1j_1}| = p_1$, $|A_{i_2j_1}| = p_2$, $|A_{i_2j_2}| = p_3$, $|A_{i_1j_2}| = p_4$, rank 1 forces $p_1p_3 = p_2p_4$; multiplying both sides by p_2p_4 gives $p_1p_2p_3p_4 = (p_2p_4)^2$, exactly this shape.) But $p_1, \ldots, p_{2k}$ are distinct primes, so $p_1 \cdots p_{2k}$ is squarefree with every prime exponent equal to 1 in its factorisation; a nonzero perfect square has every prime exponent even. These two factorisations of the same integer disagree, which is impossible by unique factorisation. Contradiction, so rank$(A) \geq 2$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths start from the same algebraic fact: $\text{rank}(A) \leq 1$ means every 2×2 minor vanishes. They split at what to do with the phrase "at least $m + n$ distinct primes." The natural attempt reads it as a bare count to feed into pigeonhole on rows or columns separately. The model reads m and n together as $m + n$ – the exact size of the vertex set of the natural bipartite graph on rows and columns – and reads the primes as edges of that graph, turning a scattered algebraic hypothesis into a single counting fact about a graph (edges $\geq$ vertices $\Rightarrow$ a cycle exists).

> *The numbers m rows, n columns, and "$m + n$ primes" are begging to be read together as one number: $m + n$ vertices' worth of edges in a graph built from rows and columns. A graph on V vertices with at least V edges cannot be a forest, so it must contain a cycle – and that cycle, not any individual prime entry in isolation, is the object I should be analysing.*

> *Once the cycle is found, every step around it – row to column, column to row – is exactly a shared-row or shared-column pair of entries, which is exactly the data a 2×2 minor needs. Chaining the vanishing-minor conditions all the way around the cycle telescopes the non-prime auxiliary factors away entirely (each one appears once as a "numerator" contribution at one step and once as a "denominator" contribution at the next), leaving a bare equation between the primes themselves – something unique factorisation can refute directly.*

Layer 3 THE WHY

$\text{Rank}(A) \leq 1$ is not really a local, entry-by-entry fact even though it is checked entry by entry: it says $A_{ij} = a_i b_j$ for fixed sequences $a_1, \ldots, a_m$ and $b_1, \ldots, b_n$, i.e. every ratio A_{ij}/A_{i1} (for fixed i, varying j) equals the row-independent ratio b_j/b_1, and likewise for columns. This is a global consistency condition on ratios, and consistency conditions on ratios are exactly the kind of statement that composes around a cycle: walking row $\to$ column $\to$ row $\to \cdots$ and multiplying the local ratio-equalities at each step returns, at the end of a closed loop, a single equation purely in terms of the values visited, because every auxiliary a_i or b_j that is not itself one of the primes gets introduced at one step and cancelled at the next. This is the same principle as a discrete "holonomy": a locally-consistent structure (rank 1) produces a global obstruction (the squareness of the cyclic product) whenever the underlying incidence structure has a cycle, and produces no obstruction at all along a tree, where every ratio can be chosen independently with no loop to force a relation. That is exactly why the forest bound (edges $\leq$ vertices -1) is the right threshold: below it, the prime-entries graph could in principle be a tree, and a rank-1 matrix can absolutely place primes along a tree's worth of positions without contradiction (each new prime just fixes one new ratio, no accumulated constraint). The hypothesis "$m + n$ distinct primes" is titled at exactly one more than that threshold, which is the minimum amount of data that forces a cycle to exist, which is the minimum amount of data capable of producing the squareness obstruction at all. Kedlaya's second solution states the graph-and-cycle argument directly; the reason $m + n$ is the threshold – rather than some other function of m, n – is this forest bound, and it is worth deriving rather than taking on faith.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the phrase "at least $m + n$ distinct primes" attached to an $m \times n$ matrix. On seeing a hypothesis that counts special entries against exactly (row count) + (column count), write down immediately: this number matches the size of the natural row-column vertex set, which is the signal for a bipartite-graph argument.
2. That cue suggests the method: build the bipartite graph with rows and columns as vertices and prime entries as edges, then use the elementary forest bound (a graph with no cycle on V vertices has at most $V - 1$ edges) to force a cycle.

3. Hesitation occurs right after assuming $\text{rank}(A) \leq 1$ and reaching for the 2×2 minor condition applied to arbitrary, unrelated prime entries – the case split explodes because nothing organises which pairs of entries to compare. Reinforce it away by asking: is there a structure (a cycle through shared rows and columns) that tells me which minors to chain together, rather than checking minors ad hoc?

4. The trap is treating $\text{rank}(A) \leq 1$ as a fact to verify locally, one pair of entries at a time, when it is really a global consistency condition on ratios that only produces a usable contradiction once it is composed all the way around a closed loop.

5. Recognisable faster next time by treating any hypothesis of the form "at least (number of rows) + (number of columns) special entries in a matrix" as an automatic cue for a row-column bipartite graph and a tree/cycle-counting argument.

6. Reusable rule: to contradict $\text{rank}(A) \leq 1$ using specially-marked entries (primes, or any set with a strong multiplicative or arithmetic invariant), form the bipartite graph on rows and columns with edges at the marked entries; if the edge count reaches the vertex count, a cycle is guaranteed, and chaining the 2×2-minor ratio identities around that cycle produces one algebraic equation involving only the marked entries, which the chosen invariant (here: unique factorisation, since a squarefree product of primes can never equal a nonzero perfect square) can then refute directly.

▷ **Cue:** a hypothesis counts at least (row count) + (column count) specially-marked entries in a matrix, and a bound on rank is wanted $\longrightarrow$ **Move:** form the bipartite graph on rows and columns with edges at the marked entries; edge count $\geq$ vertex count forces a cycle (forest bound: at most $V - 1$ edges without one)

▷ **Cue:** $\text{rank}(A) \leq 1$ is assumed for contradiction, with special (e.g. prime) entries scattered through the grid $\longrightarrow$ **Move:** do not check the 2×2 minor condition on arbitrary unrelated pairs; instead locate a cycle in the row-column incidence graph of the special entries and chain the minor conditions all the way around it, so the auxiliary (non-special) factors telescope away and only the special entries remain in the final identity

2014 B4

2014 B4. Show that for each positive integer n, all the roots of the polynomial

$$\sum_{k=0}^{n} 2^{k(n-k)} x^k$$

are real numbers.

Fields: algebra / real-rootedness; sign changes and the intermediate value theorem; log-concave coefficient sequences **Position:** B4 — mid-paper, but the only honest route needs you to find $n+1$ test points, and the whole problem is where to put them **Answer:** For $n \neq 2$ the values at $x = -2^{2j-n}$, $j = 0, \ldots, n$, have sign $(-1)^j$, so there are n distinct negative roots; for $n = 2$ the polynomial is $(1+x)^2$.

Layer 1 THE WHAT — surface analysis

The natural attempt. Write $f_n(x) = \sum_{k=0}^{n} 2^{k(n-k)} x^k$. A strong student's reflexes for "all roots are real" are, in order: factor it; find a recursion in n that interlaces roots; use Descartes' rule of signs; look for a self-adjoint matrix whose characteristic polynomial this is.

Factoring works for $n = 1, 2$: $f_1 = 1 + x$, $f_2 = 1 + 2x + x^2 = (1+x)^2$. Then $f_3 = 1 + 4x + 4x^2 + x^3 = (1+x)(1+3x+x^2)$, which is encouraging, but $f_4 = 1 + 8x + 16x^2 + 8x^3 + x^4$ has no rational root ($f_4(-1) = -6$), so there is no product formula waiting.

A recursion? The exponent $k(n-k)$ changes by k when n increases by 1, so $f_{n+1}(x)$ is $\sum 2^{k(n-k)}(2^k x)^k$-ish — the coefficients get rescaled by a factor depending on k itself, not a clean $f_n(2x)$ or xf_n. No interlacing recursion falls out in the time you have.

Descartes' rule says there are no positive roots (all coefficients positive) and at most n negative roots, counting multiplicity — which is an upper bound, and we need a lower bound. The reflex that remains is to exhibit n sign changes of f_n on the negative axis. So the student tries $x = -1$: $f_n(-1) = \sum(-1)^k 2^{k(n-k)}$, which is dominated by the middle term, sign $(-1)^{\lfloor n/2 \rfloor}$. Then $x = -2$, $x = -\frac{1}{2}$, $x = -4$, …, on a hunch that powers of 2 are right. For $n = 4$: $f_4(-1) = -6$, $f_4(-2) = 1 - 16 + 64 - 64 + 16 = 1$, $f_4(-\frac{1}{2}) = 1 - 4 + 4 - 1 + \frac{1}{16} = \frac{1}{16}$. The signs look like they might alternate, but some values are tiny ($\frac{1}{16}$ out of terms of size 4 and 64), so each one is a fresh delicate calculation, and the student has no rule for which power of 2 to test next or why the sign should come out as it does. This is where the attempt stalls: the method (IVT between test points) is right; the test points are being guessed.

The model chain. **Step 1: choose the test points from the coefficients.** Let $a_k = 2^{k(n-k)}$. The ratio of consecutive coefficients is

$$\frac{a_k}{a_{k+1}} = 2^{k(n-k)-(k+1)(n-k-1)} = 2^{2k+1-n}.$$

At $|x| = t$ the terms $a_k t^k$ and $a_{k+1} t^{k+1}$ are equal exactly when $t = 2^{2k+1-n}$. So the term $a_j t^j$ is the largest on the stretch $2^{2j-1-n} \le t \le 2^{2j+1-n}$, whose geometric midpoint is $t_j = 2^{2j-n}$. Test at $x_j = -t_j = -2^{2j-n}$, $j = 0, 1, \ldots, n$.

Step 2: evaluate exactly. The k-th term at x_j is

$$(-1)^k 2^{k(n-k)+k(2j-n)} = (-1)^k 2^{2jk-k^2} = (-1)^k 2^{j^2-(k-j)^2}.$$

Hence, putting $m = k - j$,

$$f_n(x_j) = (-1)^j 2^{j^2} S_j, \qquad S_j = \sum_{m=-j}^{n-j} (-1)^m 2^{-m^2}.$$

So it is enough to show $S_j > 0$ for every j when $n \neq 2$.

Step 3: the sign of S_j. Split the $m = 0$ term $1 = \frac{1}{2} + \frac{1}{2}$ and write $S_j = L_j + R_j$ with

$$R_j = \tfrac{1}{2} + \sum_{m=1}^{n-j} (-1)^m 2^{-m^2}, \qquad L_j = \tfrac{1}{2} + \sum_{m=1}^{j} (-1)^m 2^{-m^2}.$$

Consider R_j (the argument for L_j is identical with $n - j$ replaced by j, because 2^{-m^2} is even in m). Its terms are $\frac{1}{2}, -\frac{1}{2}, +2^{-4}, -2^{-9}, +2^{-16}, \ldots$, alternating in sign with magnitudes $\frac{1}{2} = \frac{1}{2} > 2^{-4} > 2^{-9} > \cdots$. Group them as $(\frac{1}{2} - \frac{1}{2}) + (2^{-4} - 2^{-9}) + (2^{-16} - 2^{-25}) + \cdots$, with possibly one unpaired positive term at the end. Every group is ≥ 0, so $R_j \ge 0$; and:

- if $n - j = 0$, $R_j = \frac{1}{2} > 0$;
- if $n - j = 1$, $R_j = \frac{1}{2} - \frac{1}{2} = 0$;
- if $n - j \ge 2$, $R_j \ge 2^{-4} - 2^{-9} > 0$ (the first group is 0, the second is positive, the rest are ≥ 0).

So $R_j = 0$ iff $n - j = 1$, and likewise $L_j = 0$ iff $j = 1$. Therefore $S_j = 0$ only if $j = 1$ and $n - j = 1$, i.e. $n = 2$. For $n \neq 2$, $S_j > 0$ for all j.

Step 4: count. For $n \neq 2$, $f_n(x_j)$ has sign $(-1)^j$ at the $n + 1$ points $x_0 > x_1 > \cdots > x_n$. By the intermediate value theorem f_n has a root in each open interval (x_{j+1}, x_j): n distinct real roots. A degree-n polynomial has at most n roots, so these are all of them. For $n = 2$, $f_2 = (1 + x)^2$ has the real double root -1. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains reach the same plan: "show n sign changes on the negative axis and use the IVT." They split at the choice of test points.

(N) "Try $x = -1, -2, -\frac{1}{2}, -4, \ldots$ and see." The test points are chosen from the answer one hopes to see.

(M) "The test point for the j-th sign is where the j-th term is biggest." The test points are read off the coefficient ratios $a_k/a_{k+1} = 2^{2k+1-n}$.

On (N) every evaluation is a new computation with no structure. On (M) every evaluation is the *same* computation — a two-sided Gaussian sum $\sum(-1)^m 2^{-m^2}$ truncated differently — and one sign argument handles all $n+1$ points.

> *To get a sign change I need a point where one term decides the sign. A term $a_j x^j$ decides the sign where it is the biggest, and it is the biggest between the two places where it ties with its neighbours. Those ties are at $|x| = a_{j-1}/a_j$ and $|x| = a_j/a_{j+1}$, and here those ratios are powers of 4 apart, so I'll test at the geometric midpoint 2^{2j-n}.*

> *At that midpoint the middle term is exactly the sum of its two neighbours: $2^{j^2} = 2^{j^2-1} + 2^{j^2-1}$. So the middle term doesn't win outright — it ties with its neighbours, and the sign is decided by the next layer out. Rather than fight that, I'll split the middle term in half and give each half to one side; each side is then an alternating series with decreasing terms, whose sign I know.*

- ▷ **Cue:** a polynomial with positive coefficients and a claim that all roots are real $\longrightarrow$ **Move:** look for $n+1$ points on the negative axis where the sign alternates; place the j-th point where the j-th term dominates, i.e. between a_{j-1}/a_j and a_j/a_{j+1}
- ▷ **Cue:** coefficients of the form $c^{k(n-k)}$ or c^{-k^2} $\longrightarrow$ **Move:** substitute $x = -c^{2j-n}$; each term becomes $(-1)^k c^{j^2-(k-j)^2}$, a Gaussian centred at $k = j$
- ▷ **Cue:** an alternating sum whose central term exactly equals the sum of its two neighbours $\longrightarrow$ **Move:** split the central term in half between the two sides; each side becomes a standard alternating series with nonincreasing terms
- ▷ **Cue:** a "show for all n" statement that fails the uniform argument at one small n $\longrightarrow$ **Move:** check whether that case is genuinely exceptional (here a double root) and dispose of it by hand

Layer 3 THE WHY

Why must the test points be -2^{2j-n}, and why was that not a guess?

Start from what the intermediate value theorem needs: a point where you *know* the sign of f_n. For a sum of terms of different signs you know the sign only when one term, or one tidy group, outweighs the rest. For $x < 0$ the terms $a_k x^k$ alternate in sign, so the question at each x is "which term is largest?" — and that has a mechanical answer. $a_k t^k$ versus $a_{k+1} t^{k+1}$ is a comparison of t with the ratio a_k/a_{k+1}. If the ratios a_k/a_{k+1} increase with k, then as t runs from 0 to ∞ the largest term moves from $k = 0$ to $k = n$ one index at a time, and the sign of $f_n(-t)$ ought to follow $(-1)^{\text{index of the largest term}}$, which alternates n times. That is n roots. The ratios increasing is precisely *log-concavity*: $a_k^2 \geq a_{k-1}a_{k+1}$. Here $\log_2 a_k = k(n-k)$ is a downward parabola, the most log-concave thing there is. So the definition of the coefficients already says: the dominant term marches across, and you should sample once per index, at the centre of its reign.

Now the reason the problem is B4 and not B1. Compute how strongly log-concave the sequence is:

$$\frac{a_k^2}{a_{k-1}a_{k+1}} = 2^{2k(n-k)-(k-1)(n-k+1)-(k+1)(n-k-1)} = 2^2 = 4.$$

Exactly 4, for every k and n. There is a classical sufficient condition (Kurtz, *Amer. Math. Monthly* 99 (1992)): if all $a_k > 0$ and $a_k^2 > 4a_{k-1}a_{k+1}$ for $1 \le k \le n-1$, then all roots are real and distinct. The proof of that theorem is exactly the model chain: at the midpoint the middle term beats its two neighbours combined because the ratio exceeds 4. This problem sits *on* the threshold, where the middle term only ties with its neighbours — $2^{j^2} = 2 \cdot 2^{j^2-1}$ — and $n = 2$, $(1+x)^2$, shows that on the threshold a double root genuinely can occur. So the problem is Kurtz's theorem at its boundary, and the extra work (splitting the middle term, looking at the second neighbours 2^{j^2-4}) is the price of being on the boundary. That is also why $n \ge 3$ is needed in the official solution: you need at least one term beyond the tying trinomial to break the tie.

Kedlaya's solution groups the terms into "a trinomial that evaluates to 0" with binomials on either side "in which one term dominates". Here is where that comes from: the trinomial is the tie at the threshold ratio 4, and the binomials are the pairs $(2^{-m^2}, 2^{-(m+1)^2})$ of the alternating tail. Mahlburg's remark in the official text — split the central term into two halves — is the same computation written as two alternating series, which is the version given above because the sign of an alternating series with decreasing terms is a fact every A-level student already owns.

Finally, why the $(-1)^j 2^{j^2}$ factors out so neatly: $k(n-k) + k(2j-n) = 2jk - k^2 = j^2 - (k-j)^2$ is completing the square. The substitution $x = -2^{2j-n}$ is the one that recentres the parabola $k(n-k)$ at $k = j$. Completing the square is not a trick either; it is what "put the vertex of the parabola at j" means algebraically.

Caveat. Verified by exact rational arithmetic: for every $n \le 40$ the values $f_n(-2^{2j-n})$, $j = 0, \dots, n$, have sign $(-1)^j$ except at $n = 2$ (signs $+, 0, +$, the double root), and numerically all roots are real for $n \le 14$. The ratio identity $a_k^2 = 4a_{k-1}a_{k+1}$ was checked exactly for $n \le 40$. The Kurtz citation is given for context only; the proof above does not depend on it.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the exponent $k(n-k)$: a quadratic in k. On seeing it, write the ratio $a_k/a_{k+1} = 2^{2k+1-n}$ — the whole problem's geometry is in that line.

2. The cue suggests the dominant-term method: sample f_n at $-t_j$ with t_j inside the range where the j-th term is largest, and apply the intermediate value theorem n times, then count against the degree.

3. Hesitation occurs at the discovery that the middle term exactly ties with its two neighbours, which looks as if the method has failed. It is reinforced away by noticing that the tie is *exact* and symmetric, so it cancels cleanly, and the sign is then decided by the next pair of terms, whose sign is known.

4. The trap is testing at convenient points $(-1, -2)$ chosen for arithmetic rather than for dominance; a secondary trap is to forget that for $n = 2$ the uniform argument gives a zero rather than a sign, and so to claim distinct roots when the root is double.

5. Recognise it faster by asking, whenever a sign change or a root count is wanted for a sum of many terms: *at each point, which term is biggest, and where does the lead change hands?* The hand-over points are ratios of consecutive terms.

6. Reusable decision rule: *to locate roots of a sum with widely varying terms, tabulate where each term is largest (ratios of consecutive coefficients) and sample once per reign; the sign there is the sign of the leader, provided the leader beats the rest.* If the leader only ties, cancel the tie exactly and let the next layer decide. The same rule counts real zeros of lacunary polynomials, estimates theta-type sums, and underlies Newton-polygon arguments over the p-adics.

2014 B5

2014 B5. In the 75th annual Putnam Games, participants compete at mathematical games. Patniss and Keeta play a game in which they take turns choosing an element from the group of invertible $n \times n$ matrices with entries in the field $\mathbb{Z}/p\mathbb{Z}$ of integers modulo p, where n is a fixed positive integer and p is a fixed prime number. The rules of the game are:

(1) A player cannot choose an element that has been chosen by either player on any previous turn.

(2) A player can only choose an element that commutes with all previously chosen elements.

(3) A player who cannot choose an element on his/her turn loses the game.

Patniss takes the first turn. Which player has a winning strategy? (Your answer may depend on n and p.)

Fields: group theory (centralisers, Cauchy's theorem); combinatorial games; linear algebra over finite fields **Position:** B5 — a hard problem that becomes a parity count once the right invariant is named; the $p = 2$ half needs one genuine construction **Answer:** Keeta wins if p is odd; Patniss wins if $p = 2$ (for every n).

Layer 1 THE WHAT — surface analysis

The natural attempt. A game problem, so the student's first reflex is a pairing (mirror) strategy for one of the players, and the second is to play small cases.

For p odd the mirror appears at once. The scalar matrix $-I$ commutes with everything and $-I \neq I$. If Patniss plays g, Keeta plays $-g$: it commutes with everything g commutes with, it commutes with g, and it is new, because after each of Keeta's moves the chosen set is closed under $g \mapsto -g$ and $-g \neq g$. So Keeta always has a reply and never loses. That half is done in five minutes, and the student rightly feels good.

For $p = 2$ the mirror is gone: $-I = I$. The student plays $n = 1$: $GL_1(\mathbb{F}_2) = \{1\}$, Patniss takes it and wins. Then $n = 2$: $GL_2(\mathbb{F}_2)$ has order 6 (it is S_3). If Patniss opens with I, every element is still available and the game is a mess of cases. The student tries to find a mirror for Patniss — "Patniss opens with something, then answers each h with some partner h'" — and cannot find a fixed-point-free involution that respects commuting. Then the student tries to analyse the game tree for $n = 2$ directly: after Patniss plays an element g of order 3, the only elements commuting with g are I, g, g^2, so exactly two more moves are made and Patniss makes the last; after Patniss plays an element of order 2, its centraliser is $\{I, g\}$, one more move, Keeta wins. So opening with order 3 wins. Fine for S_3, but for $n = 3$ ($|G| = 168$) and general n the tree is hopeless, and the student has no idea which opening generalises. The attempt stalls on a search for the right first move with no principle for choosing it.

The model chain. **Step 1: the game in any finite group.** Let G be a finite group and, for $S \subseteq G$, let $C(S) = \{g \in G : gs = sg \ \forall s \in S\}$ be its centraliser, a subgroup. Let S be the set chosen so far. Rule (2) makes chosen elements commute pairwise, so $S \subseteq C(S)$, and the legal moves are exactly $C(S) \setminus S$. The game therefore ends precisely when $S = C(S)$, and at that moment the number of moves made is

$$|S| = |C(S)|,$$

the order of a subgroup. Patniss made the last move, and wins, iff this number is odd.

Step 2: parity is decided by the first move. As S grows, $C(S)$ shrinks, each one a subgroup of the last.

- If $|C(S)|$ is odd at some moment, every later $C(S')$ is a subgroup of it, so has odd order by Lagrange, and the final count is odd: Patniss wins however either side plays.

- If some chosen element h has even order, then at every later moment $h \in S' \subseteq C(S')$, so $|C(S')|$ is divisible by the order of h, which is even: the final count is even and Keeta wins.

Suppose Patniss opens with g. If $|C(\{g\})|$ is odd, Patniss has won. If $|C(\{g\})|$ is even, Keeta replies with an element $h \in C(\{g\})$ of even order, $h \neq g$: if g has odd order, Cauchy's theorem gives h of order 2 in $C(\{g\})$; if g itself has even order, the second bullet already applies and Keeta makes any legal move (one exists, since $|C(\{g\})| \geq 2$). So:

$$\text{Patniss wins} \iff \text{some } g \in G \text{ has } |C(\{g\})| \text{ odd.}$$

Step 3: p **odd.** $-I$ is central of order 2, so every $C(\{g\})$ contains it and has even order. Keeta wins. (The mirror $g \mapsto -g$ is an explicit strategy.)

Step 4: $p = 2$**.** Take a monic irreducible polynomial P of degree n over $\mathbb{F}_2$ (one exists: the multiplicative group of the field $\mathbb{F}_{2^n}$ is cyclic, and the minimal polynomial over $\mathbb{F}_2$ of a generator has degree n, since the generator's powers fill $\mathbb{F}_{2^n}^\times$). Let g be its companion matrix, so that $e_1, ge_1, \ldots, g^{n-1}e_1$ are the standard basis vectors $e_1, \ldots, e_n$ and $P(g) = 0$. $\det g = \pm P(0) \neq 0$ since P is irreducible of degree $n \geq 1$ and not x (for $n = 1$ take $P = x + 1$).

Claim: every matrix A commuting with g is a polynomial in g. Write $Ae_1 = q(g)e_1$ for a polynomial q (possible because $e_1, ge_1, \ldots, g^{n-1}e_1$ is a basis). Then for each k,

$$A\, g^k e_1 = g^k A e_1 = g^k q(g) e_1 = q(g)\, g^k e_1,$$

so A and $q(g)$ agree on a basis: $A = q(g)$.

Now $\mathbb{F}_2[g] \cong \mathbb{F}_2[x]/(P)$ is a field with 2^n elements, since P is irreducible and is the minimal polynomial of g. Hence $C(\{g\}) = \mathbb{F}_2[g] \setminus \{0\}$ has order $2^n - 1$, which is odd. Patniss opens with g and wins. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains dispose of p odd with the mirror $g \mapsto -g$. They split at $p = 2$.

(N) "Find Patniss's first move by exploring the game tree / find a pairing for her."

(M) "The game ends when the chosen set equals its own centraliser; so the total number of moves is the order of a group, and only its parity matters."

On (N) the question "who wins?" is a question about strategies. On (M) it becomes a question about a single number — the size of the final centraliser — and the strategies drop out.

> *Who wins depends only on whether the total number of moves is odd or even. When does the game stop? When every element commuting with everything chosen has already been chosen: $S = C(S)$. So the length of the game is the order of a subgroup. I don't need to understand the play; I need the parity of that order.*

> *The centraliser only shrinks, and a subgroup of an odd-order group is odd. So if Patniss can make $|C(g)|$ odd with her first move, the parity is locked for the rest of the game. For p odd, $-I$ is in every centraliser and locks it even. For $p = 2$ I need one matrix whose centraliser has odd order — something like a field's multiplicative group, $2^n - 1$.*

▷ **Cue:** a game where the set of legal moves is "everything compatible with all moves so far" and the loser is whoever cannot move ⟶ **Move:** ask what the chosen set looks like when the game stops; if it is forced to equal some canonical object (a subgroup, a maximal clique), the winner is decided by the parity of that object's size

- ▷ **Cue:** you need the parity of the order of a subgroup $\longrightarrow$ **Move:** odd order is inherited by subgroups (Lagrange); even order is certified by one element of even order, and exists iff there is an element of order 2 (Cauchy)
- ▷ **Cue:** a matrix group over $\mathbb{F}_p$ with p odd $\longrightarrow$ **Move:** $-I$ is a central involution: it gives a mirror strategy and makes every centraliser even
- ▷ **Cue:** you want a matrix whose centraliser is as small as possible $\longrightarrow$ **Move:** take the companion matrix of an irreducible polynomial; its centraliser is $\mathbb{F}_{p^n}^{\times}$, of order $p^n - 1$

Layer 3 THE WHY

Why is the parity invariant the thing to look for, and why is it natural rather than inspired?

Start from rule (3). The winner is whoever makes the last move, so the winner is determined by one bit: the parity of the length of the game. Every impartial "last player to move wins" game invites the question: *is the length of the game forced?* Usually it is not, and one needs Sprague–Grundy theory. But it costs one line to ask what the position looks like at the end, and here rules (1) and (2) answer it by definition. Rule (2) says the legal set is $C(S)$; rule (1) subtracts S; the game ends when $C(S) \setminus S$ is empty. So the end state is "S is its own centraliser", and the length of the game is $|C(S)|$. That is the circle-and-centre moment: the rules define legality through the centraliser, so the centraliser is the object, and asking for its size is not a trick.

Once the length is the order of a group, parity questions about group orders have exactly two tools in the A-level-plus toolbox: Lagrange (orders of subgroups divide the order of the group, so odd stays odd) and Cauchy (even order forces an element of order 2). The proof uses both, once each. Nothing else is available, and nothing else is needed.

Now $p = 2$. What does "$|C(g)|$ odd" ask for? That no involution commutes with g. The smaller the centraliser, the better the chance, so look for the g with the smallest centraliser. Every matrix commutes with its own polynomials, so $C(g) \supseteq \mathbb{F}_2[g] \setminus \{0\}$ (the invertible ones), and the best one can hope for is equality. When is the centraliser no bigger than the polynomials in g? When g has a cyclic vector — the claim proved above, which is the whole content of the step. And when is $\mathbb{F}_2[g] \setminus \{0\}$ a group of odd order? When $\mathbb{F}_2[g]$ is a field, i.e. the minimal polynomial is irreducible, of degree n so that g still has a cyclic vector. That is exactly a companion matrix of an irreducible polynomial. Its centraliser is $\mathbb{F}_{2^n}^{\times}$, of order $2^n - 1$, odd because 2^n is even. So the construction is forced by the requirement "smallest possible centraliser", not pulled from nowhere.

Kedlaya's solution proves $|C(g)|$ odd by passing to the algebraic closure: g has distinct eigenvalues there, so anything commuting with it is diagonal in the same eigenbasis, hence diagonalisable, hence of odd order (eigenvalues lie in finite fields of characteristic 2, whose multiplicative groups have odd order). Here is where that comes from, and why the version above is cheaper. "Distinct eigenvalues" is the same as "has a cyclic vector" for a matrix whose characteristic polynomial is squarefree, and the cyclic-vector argument stays inside $\mathbb{F}_2$ and identifies the centraliser exactly, rather than bounding the order of each element. The official remark then says the same thing in the language of $\mathbb{F}$ as a vector space: $C(g)$ is the multiplicative group of the field.

Finally, why must $p = 2$ and p odd differ? Every centraliser contains the centre of G, which is the scalar matrices $\mathbb{F}_p^{\times}$, of order $p - 1$. That is even exactly when p is odd. For $p = 2$ the centre is trivial and there is no obstruction. The answer's dependence on p alone, and not on n, is the order of the centre.

Caveat. The answer was confirmed by exhaustive game-tree search for $(n, p) = (1, 2), (2, 2), (3, 2)$ (Patniss wins) and $(1, 3), (1, 5), (2, 3)$ (Keeta wins). The existence of an irreducible polynomial of every degree over $\mathbb{F}_2$ is used as a standard fact about finite fields (existence of $\mathbb{F}_{2^n}$ and cyclicity of its multiplicative group); it is not reproved here.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is rule (3): the last mover wins. On seeing it, write "winner = parity of the game's length" and then "the game ends when $C(S) = S$".

2. The cue suggests an invariant/parity method: reduce the game to the parity of the order of the final centraliser, and control that parity with Lagrange and Cauchy.

3. Hesitation occurs at $p = 2$, after the mirror strategy for p odd has been found, because the student keeps looking for a second mirror. It is reinforced away by noticing that the mirror was only a symptom: it works because $-I$ makes every centraliser even. The question to ask for $p = 2$ is the same parity question, now looking for an odd centraliser.

4. The trap is the game-tree: small cases ($GL_2(\mathbb{F}_2) \cong S_3$) can be done by hand, which lures the solver into trying to generalise a strategy rather than an invariant. A second trap is to claim that Keeta's reply of even order is always available without handling the case where Patniss's own move already has even order.

5. Recognise it faster by asking, in any game where legality means "compatible with every previous move": *what is the terminal position, and is its size forced to be the size of some structured object?* If yes, the game is a parity problem about that object.

6. Reusable decision rule: *in a last-mover-wins game, compute the end state before thinking about strategy; if the length is determined by an invariant of the end state, the first player wins iff she can fix that invariant's parity on move one.* For the construction half: *to make a centraliser as small as possible, use a matrix with a cyclic vector, whose centraliser is the polynomials in it; with an irreducible characteristic polynomial, that is a finite field.*

2014 B6

2014 B6. Let $f : [0,1] \to \mathbb{R}$ be a function for which there exists a constant $K > 0$ such that $|f(x) - f(y)| \leq K|x - y|$ for all $x, y \in [0,1]$. Suppose also that for each rational number $r \in [0,1]$, there exist integers a and b such that $f(r) = a + br$. Prove that there exist finitely many intervals $I_1, \ldots, I_n$ such that f is a linear function on each I_i and $[0,1] = \bigcup_{i=1}^{n} I_i$.

Fields: real analysis (Lipschitz functions); elementary number theory (Farey fractions, $ps' - p's = \pm 1$)
Position: B6 — the last problem of the paper, hard; but every step is elementary once the right pairs of rationals are chosen **Answer:** True: f agrees with its piecewise-linear interpolant on the Farey fractions of denominator at most $2K$.

Layer 1 THE WHAT — surface analysis

The natural attempt. A student with some analysis sees "Lipschitz" and reaches for derivatives. A Lipschitz function is differentiable almost everywhere (Rademacher, or Lebesgue in one dimension) with $|f'| \leq K$. At a rational $r = p/q$, $f(r) = a + bp/q$ says $qf(r) \in \mathbb{Z}$. So the plan is: show f' is an integer wherever it exists, conclude f' takes finitely many values (integers in $[-K, K]$), and hence f is piecewise linear.

The first half can be pushed through: difference quotients between nearby rationals p/q, p'/q' are

$$\frac{f(p'/q') - f(p/q)}{p'/q' - p/q} = \frac{q\,[q'f(p'/q')] - q'\,[qf(p/q)]}{p'q - pq'},$$

an integer divided by $p'q - pq'$. The student hopes this forces integer slopes. But the second half is false as stated: "$f' \in \mathbb{Z}$ almost everywhere and $|f'| \leq K$" does *not* make f piecewise linear with finitely many pieces. The integral of the indicator function of a fat Cantor set has derivative in $\{0,1\}$ almost everywhere and is linear on no neighbourhood of any point of the Cantor set. So the derivative route discards exactly the information that matters — the *intercepts*

are integers too — and it hits a wall that is real, not a matter of effort. The student who has spent twenty minutes on measure theory has nowhere to go.

The model chain. **Reformulation.** For a rational $r = p/q$ in lowest terms ($q \geq 1$), the condition "$f(r) = a + br$ with $a, b \in \mathbb{Z}$" says $qf(r) = aq + bp \in \mathbb{Z}$; conversely any integer is $aq + bp$ for some integers a, b because $\gcd(p, q) = 1$. So the hypothesis is exactly: $q\, f(p/q) \in \mathbb{Z}$ for every reduced $p/q \in [0, 1]$. Call a line $x \mapsto a + bx$ *integral* if $a, b \in \mathbb{Z}$.

Farey facts. Let F_N be the fractions in $[0, 1]$ with denominator at most N, in increasing order. (i) Consecutive $\frac{r}{s} < \frac{r'}{s'}$ in F_N satisfy $r's - rs' = 1$. (ii) If $\frac{r}{s} < \frac{r'}{s'}$ satisfy $r's - rs' = 1$ and $\frac{x}{y}$ lies strictly between them, then $xs - ry \geq 1$ and $r'y - xs' \geq 1$, so

$$y = y(r's - rs') = s'(xs - ry) + s(r'y - xs') \geq s + s',$$

with equality only for $\frac{x}{y} = \frac{r+r'}{s+s'}$. Now suppose (i) holds for F_{N-1} and let $\frac{r}{s} < \frac{r'}{s'}$ be consecutive there. The mediant $\frac{r+r'}{s+s'}$ lies strictly between them, so it is not in F_{N-1}: $s + s' \geq N$. By (ii), any fraction of denominator N strictly between them forces $N \geq s + s'$, hence $N = s + s'$ and the fraction is the mediant. Hence the fractions entering at stage N are exactly the mediants $\frac{r+r'}{s+s'}$ of consecutive pairs in F_{N-1} with $s + s' = N$, and (i) follows by induction from $F_1 = \{\frac{0}{1}, \frac{1}{1}\}$, since $(r + r')s - r(s + s') = r's - rs' = 1$ and likewise on the other side.

Step 1: chords between Farey neighbours are integral. Let $\frac{r}{s} < \frac{r'}{s'}$ with $r's - rs' = 1$, and let $a + bx$ be the chord of f joining them. Since $\frac{r'}{s'} - \frac{r}{s} = \frac{1}{ss'}$,

$$b = ss'\Big(f(\tfrac{r'}{s'}) - f(\tfrac{r}{s})\Big) = s \cdot \big[s' f(\tfrac{r'}{s'})\big] - s' \cdot \big[s f(\tfrac{r}{s})\big] \in \mathbb{Z},$$

and $|b| \leq K$ by the Lipschitz condition. Also $as = sf(\frac{r}{s}) - br \in \mathbb{Z}$ and $as' = s'f(\frac{r'}{s'}) - br' \in \mathbb{Z}$; since $\gcd(s, s') = 1$ (a consequence of $r's - rs' = 1$), $a \in \mathbb{Z}$.

Step 2: no kink at a mediant with large denominator. Keep $\frac{r}{s}, \frac{r'}{s'}$ and their integral chord $a_0 + b_0 x$, and let $\frac{m}{N} = \frac{r+r'}{s+s'}$, $N = s + s'$. Put

$$t = Nf(\tfrac{m}{N}) - a_0 N - b_0 m \in \mathbb{Z}, \qquad \text{so} \qquad f(\tfrac{m}{N}) = a_0 + b_0 \tfrac{m}{N} + \tfrac{t}{N}.$$

Using $\frac{m}{N} - \frac{r}{s} = \frac{1}{sN}$ and $\frac{r'}{s'} - \frac{m}{N} = \frac{1}{s'N}$, the slopes of the two chords of f through $\frac{m}{N}$ are

$$\frac{f(\frac{m}{N}) - f(\frac{r}{s})}{1/(sN)} = b_0 + st, \qquad \frac{f(\frac{r'}{s'}) - f(\frac{m}{N})}{1/(s'N)} = b_0 - s't.$$

Both are difference quotients of f, so both lie in $[-K, K]$. Subtracting, $(s + s')|t| \leq 2K$, i.e. $N|t| \leq 2K$. If $N > 2K$ then $t = 0$: $f(\frac{m}{N})$ lies on the chord.

Step 3: stabilisation. Let f_N be the function that agrees with f on F_N and is linear between consecutive elements. Fix $N_0 = \max(1, \lfloor 2K \rfloor)$. For $N > N_0$ we have $N > 2K$, and each point of $F_N \setminus F_{N-1}$ is a mediant $\frac{r+r'}{s+s'}$ of consecutive elements of F_{N-1} with $s + s' = N$; by Step 2 the value of f there lies on the chord of f_{N-1}, so $f_N = f_{N-1}$. By induction $f_N = f_{N_0}$ for all $N \geq N_0$.

Step 4: conclusion. Every rational in $[0, 1]$ lies in some F_N, so $f = f_{N_0}$ on $\mathbb{Q} \cap [0, 1]$. Both are continuous (f is Lipschitz), and the rationals are dense, so $f = f_{N_0}$ on $[0, 1]$. Taking $I_1, \ldots, I_n$ to be the closed intervals between consecutive elements of F_{N_0}, f is linear (indeed integral-linear, by Step 1) on each. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains translate the hypothesis into "$qf(p/q) \in \mathbb{Z}$" and both look at difference quotients between two rationals. They split at the choice of the pair.

(N) "Take any two nearby rationals and let them tend together" — aiming at f'. The difference quotient then has denominator $p'q - pq'$, uncontrolled, and the limit throws away the intercepts.

(M) "Take two rationals with $p'q - pq' = 1$" — Farey neighbours. Then the difference quotient is an integer outright, the intercept is an integer too, and no limit is taken.

The second split is at the mediant: (M) asks what integrality says about the *change* of slope at a rational of denominator N, and finds it is a multiple of N.

> *The slope between p/q and p'/q' is an integer divided by $p'q - pq'$. I can't control that denominator for arbitrary pairs — but I can choose the pairs. If I only ever compare rationals with $p'q - pq' = 1$, every chord slope is an integer. That's the Farey sequence.*

> *If f bends at m/N between two integral lines, those lines meet at m/N, so $(a_1 - a_2)N = (b_2 - b_1)m$ and N divides the change of slope. But Lipschitz caps the change of slope at $2K$. So f can only bend at rationals with denominator at most $2K$ — finitely many points. Now I need to make that rigorous without assuming f is already piecewise linear: compare with the chord one Farey level up.*

- ▷ **Cue:** a function whose values at p/q lie in $\frac{1}{q}\mathbb{Z}$ (or any "integral at rationals" condition) $\longrightarrow$ **Move:** compare only pairs of rationals with $p'q - pq' = \pm 1$ (Farey neighbours); differences and chords then have integer data
- ▷ **Cue:** two integral lines $a_1 + b_1 x$, $a_2 + b_2 x$ meeting at a rational m/N in lowest terms $\longrightarrow$ **Move:** N divides $b_2 - b_1$: a kink at a point of denominator N costs a slope change of at least N
- ▷ **Cue:** a Lipschitz (bounded-slope) function and a discreteness condition $\longrightarrow$ **Move:** bound the slope *change* by $2K$ and play it against the discreteness to rule out all but finitely many kinks
- ▷ **Cue:** a property to be shown on all of $[0,1]$ for a continuous function $\longrightarrow$ **Move:** prove it on a dense set (the rationals, level by level) and pass to the closure

Layer 3 THE WHY

Why Farey neighbours, and why is the bound $2K$ the heart of the problem?

Start with the hypothesis read literally. It says: at each rational, the value of f is the value of *some* integral line. Integral lines form a lattice $\mathbb{Z}^2$ of (a, b), and each rational p/q gives a linear map $(a, b) \mapsto a + bp/q$. A single rational pins down very little. Two rationals p/q and p'/q' together pin down a line through the two values — and that line is integral exactly when the 2×2 system

$$\begin{pmatrix} q & p \\ q' & p' \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} qf(p/q) \\ q'f(p'/q') \end{pmatrix}$$

has an integer solution for every integer right-hand side, i.e. exactly when the determinant $qp' - pq'$ is ± 1. That is the definition of Farey neighbours. So the Farey sequence is not a piece of number-theoretic folklore imported from outside; it is the answer to "when does the integrality of two values force the integrality of the chord?" — the unimodular pairs. That is the circle-and-centre level of "of course": the hypothesis is about a lattice, and unimodular pairs are the bases of that lattice.

Next, why the Lipschitz condition is there. Without it nothing stops integral pieces of steeper and steeper slope from accumulating at a point, and no finite list of intervals could be expected. What Lipschitz provides is a cap on the *change* of slope, $2K$. What integrality provides is a floor: at a kink between two integral lines meeting at m/N, the slopes differ by a multiple of N (from $(a_1 - a_2)N = (b_2 - b_1)m$ and $\gcd(m, N) = 1$). Cap meets floor: kinks can occur only at

denominators $N \leq 2K$. The whole problem is that one inequality. Step 2 of the model chain is the rigorous version: at the mediant, the two new chords have slopes $b_0 + st$ and $b_0 - s't$, whose difference is $t(s + s') = tN$ — the "multiple of N" — and Lipschitz caps it at $2K$.

Kedlaya's solution starts "For each positive integer n, define the n-th Farey sequence". Here is where that comes from: the determinant of the system above. And the choice "$n > 2K$" is the cap-meets-floor inequality, stated before its motivation. (A small correction for the reader of the official text: it says the slope from $\frac{m}{n}$ to $\frac{r'}{s'}$ "is at most $b_0 - s't$"; the computation gives exactly $b_0 - s't$.)

Why the derivative route fails, precisely: differentiation keeps only slope information, and slope integrality alone permits a fat-Cantor mixture of slopes 0 and 1. The intercept integrality is what forbids kinks at irrationals (two integral lines meet only at a rational point) and forces kinks at rationals to be expensive. The model chain never forgets the intercept: Step 1 proves $a \in \mathbb{Z}$ as carefully as $b \in \mathbb{Z}$, and Step 2's integer t is built from both.

Finally, why the proof can avoid measure theory and limits altogether: the Farey refinement is a finite process at each level, and Step 2 shows it stops changing after level $2K$. The only analytic input is at the very end — continuity, to pass from rationals to reals.

Caveat. The argument was checked line by line; the Farey facts are proved in the model chain rather than cited. The fat-Cantor example in the natural attempt is standard (the integral of the indicator of a closed nowhere-dense set of positive measure is Lipschitz with derivative in $\{0, 1\}$ almost everywhere and is not piecewise linear with finitely many pieces); it is offered to show where the derivative route breaks, not as part of the proof.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "$f(r) = a + br$ with integers a, b". On seeing it, write the equivalent form $qf(p/q) \in \mathbb{Z}$, and then write the chord slope between two rationals with its denominator $p'q - pq'$ visible.

2. The method is Farey refinement: interpolate f on F_N, show the interpolant is integral, and show it stops changing once $N > 2K$, then conclude by continuity and density.

3. Hesitation occurs at the moment of choosing which pairs of rationals to compare — arbitrary nearby pairs feel more natural to an analyst. It is reinforced away by looking at the denominator $p'q - pq'$ and asking the one question that makes it harmless: can I choose pairs with this equal to 1?

4. The trap is Rademacher: differentiate, get integer slopes a.e., and believe you are done. It is a trap because slope integrality alone does not give finitely many pieces; the intercepts carry the essential information. A secondary trap is to prove local linearity near each point and forget to use compactness (or an explicit bound like $2K$) to get *finitely* many intervals.

5. Recognise it faster by pairing two sources of constraint and asking what each says about the same quantity: here Lipschitz says the slope change is at most $2K$, integrality says a slope change at denominator N is a multiple of N. Whenever a problem gives an analytic bound and an arithmetic condition, look for the quantity both constrain.

6. Reusable decision rule: *when a hypothesis says values at rationals lie in a lattice, work with unimodular pairs (Farey neighbours, consecutive convergents); they are exactly the pairs for which integrality transfers from values to chords.* And: *an arithmetic "floor" that grows with the denominator, set against an analytic "cap" that does not, confines all the interesting behaviour to finitely many denominators.*