DAWN PUTNAM SESSION

Putnam 2008 — Papers A and B

Putnam 120 Archive

Contents

Putnam 2008 — Paper A **5**
2008 A1 5
2008 A2 7
2008 A3 10
2008 A4 13
2008 A5 16
2008 A6 18

Putnam 2008 — Paper B **23**
2008 B1 23
2008 B2 25
2008 B3 27
2008 B4 30
2008 B5 32
2008 B6 35

Putnam 2008 — Paper A

2008 A1

2008 A1. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a function such that $f(x,y) + f(y,z) + f(z,x) = 0$ for all real numbers x, y, and z. Prove that there exists a function $g : \mathbb{R} \to \mathbb{R}$ such that $f(x,y) = g(x) - g(y)$ for all real numbers x and y.

Fields: functional equations, algebra **Position:** A1 — the opener **Answer:** Yes: $g(x) = f(x,0)$ works. Three substitutions into the given relation – $(0,0,0)$, $(x,0,0)$, $(x,y,0)$ – deliver the decomposition directly.

Layer 1 THE WHAT — surface analysis

The natural attempt. The relation is symmetric in a useful way under cycling the three variables, so the first honest moves are to plug in degenerate triples and see what falls out. Setting $x = y = z$ gives $3f(x,x) = 0$, so $f(x,x) = 0$ for every x. Setting $z = x$ gives $f(x,y) + f(y,x) + f(x,x) = 0$, and since $f(x,x) = 0$ this collapses to $f(x,y) = -f(y,x)$: the function is antisymmetric. Both facts are correct and cost nothing.

At this point the shape of the problem looks familiar: a three-term relation among values indexed by pairs of points, antisymmetric, summing to zero around any triangle $x \to y \to z \to x$. That is exactly a *cocycle condition* on the pairs, and the natural next thought is to build g the way one builds a potential function from a conservative field, or a height function from pairwise differences: fix a basepoint, say 0, and define $g(x)$ by accumulating f along a path from 0 to x. Concretely, for any chain $0 = x_0, x_1, \ldots, x_n = x$, set $g(x) := -\sum_{i=1}^{n} f(x_{i-1}, x_i)$, or more simply try $g(x) := f(x_0, x)$ for one intermediate hop.

The trouble is that a definition by chains has an obligation attached: it must not depend on which chain was used, or g is not well-defined. Checking this in general means showing that for any two paths from 0 to x, the two accumulated sums agree – which in turn means showing the three-term relation lets you freely insert or delete an intermediate point in a path without changing the total, for chains of arbitrary length, not just length 2. That is provable (it follows by repeatedly inserting one point at a time and invoking the hypothesis once per insertion), but it is a proof about arbitrary finite chains, carrying its own induction and its own bookkeeping, when the object actually wanted is a two-hop statement about a single triple $(x,y,0)$.

The model chain. Substitute $(x,y,z) = (0,0,0)$: $3f(0,0) = 0$, so $f(0,0) = 0$.

Substitute $(x,y,z) = (x,0,0)$: $f(x,0) + f(0,0) + f(0,x) = 0$, and since $f(0,0) = 0$ this gives

$$f(x,0) + f(0,x) = 0, \qquad \text{i.e.} \qquad f(0,x) = -f(x,0).$$

Substitute $(x,y,z) = (x,y,0)$: $f(x,y) + f(y,0) + f(0,x) = 0$, so

$$f(x,y) = -f(y,0) - f(0,x) = -f(y,0) + f(x,0) = f(x,0) - f(y,0).$$

Define $g(x) := f(x,0)$. Then $f(x,y) = g(x) - g(y)$ for all real x, y, as required.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains reach the same two facts for free – $f(x,x) = 0$ and $f(x,y) = -f(y,x)$ – by substituting equal or cyclically-matched arguments. They split immediately after that, on how to turn the three-term relation into the desired two-term formula. The natural attempt treats the relation as license to build g by chaining, which obligates a path-independence proof over chains of arbitrary length. The model chain substitutes $(x,y,0)$ directly into the original hypothesis – the same hypothesis already used twice – and reads the formula for $g(x) - g(y)$ straight off the result.

> *The hypothesis is not a rule I get to invoke once and then reason abstractly about composing – it is already true for every triple (x,y,z) at once, including the one triple $(x,y,0)$ that contains exactly the two things I want, $f(x,y)$, plus two terms I can name as a single function of one variable. I do not need to build a path from 0 to x to y by hand; the hypothesis already is that path, evaluated once.*

> *Before I trust any formula built from $f(x,0)$, I should find out what $f(0,0)$ and the relationship between $f(x,0)$ and $f(0,x)$ actually are, by substituting the most degenerate values available – all three variables equal, then two of them equal to 0 – because those substitutions are the cheapest possible instances of the hypothesis and they are exactly what the next, less degenerate substitution will need.*

Layer 3 THE WHY

Why does the single substitution $(x,y,0)$ suffice, when the natural attempt's instinct – build g as an accumulated sum along a path – is not wrong, only expensive? Because the hypothesis is universally quantified over *all* triples simultaneously: it is not one equation to be composed with itself along a chosen route, it is an infinite, already-available family of equations, and among them is the one indexed by exactly the triple $(x,y,0)$ that contains $f(x,y)$ and two terms involving the fixed point 0. Choosing the third variable to be a constant is not a trick pulled from outside the problem; it is the definition of "for all x,y,z" read literally – the relation holds in particular when z is fixed at 0, and reading it that way is free.

The reason the degenerate substitutions ($x = y = z$, then $z = x$) come first, and are not a detour, is that they are what license trusting $g(x) := f(x,0)$ at all. Without knowing $f(0,0) = 0$, the substitution $(x,0,0)$ would leave $f(x,0) + f(0,x)$ tangled with an unknown constant $f(0,0)$; without the antisymmetry it implies, there is no reason to expect $f(x,0)$ and $-f(0,x)$ to be interchangeable inside the final formula. Each degenerate substitution removes exactly one unknown quantity ($f(0,0)$, then the relationship between $f(x,0)$ and $f(0,x)$) that the final substitution needs already resolved. The three substitutions are not independent observations; they are a single chain of dependency, each one unlocking the next, and the chain has length three because the problem has three degrees of freedom – x, y, z – and setting them to $(0,0,0)$, then $(x,0,0)$, then $(x,y,0)$ walks down that freedom one variable at a time, always keeping the substitution as cheap as the hypothesis allows.

The natural attempt's chain-construction is not mathematically wrong – it is what a cocycle condition *is*, in general, when the space of points is not equipped with a distinguished "fast" triple containing both endpoints and the basepoint at once. Here it does have one: the triple $(x,y,0)$. Recognising that the target pair (x,y) and the chosen basepoint 0 already sit inside a single instance of the hypothesis is what collapses an inductive path-independence argument into one line.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the shape of the hypothesis itself: a relation holding for *all* triples of a single free variable's worth of points, summing cyclically to zero. On seeing it, the thing to

write down immediately is: this is true for every choice of the three arguments, so I should choose them to contain exactly the quantities in the thing I am trying to prove.

2. That cue suggests substitution, not construction: rather than building g from the hypothesis by an auxiliary procedure (a path, a limit, an integral), look for the single triple (x, y, z) whose substitution turns the hypothesis directly into the target identity.

3. Hesitation sets in right after deriving antisymmetry, when the relation starts to look like a cocycle and the reflex is to reach for the general machinery (path sums, well-definedness) that handles cocycles in unfamiliar settings. What reinforces the cheaper route is asking: does the target identity involve only finitely many named quantities, and is there a single instance of the hypothesis containing all of them at once? Here the answer is yes – $f(x, y)$, $f(x, 0)$, $f(0, y)$ are exactly the three terms of the instance $(x, y, 0)$ – so no chain is needed.

4. The trap is substituting only convenient equal-variable triples ($x = y = z$, or $z = x$) and stopping once antisymmetry is in hand, mistaking "I have found true and pleasant facts" for "I have found the answer." Antisymmetry alone never produces the decomposition $f(x, y) = g(x) - g(y)$; a genuinely new substitution, one that keeps x and y distinct while pinning the third variable, is still required.

5. Recognise this shape faster by treating any hypothesis of the form "a relation holds for all tuples of free variables" as an invitation to ask, before any construction is attempted: which single instance of this relation, under some choice of the free variables, already contains every term appearing in what I am trying to prove?

6. Reusable rule: given a universally-quantified relation among k free variables and a target identity involving only $k - 1$ or fewer named quantities, look first for the one substitution of the remaining variable(s) to constants that turns the hypothesis directly into the target, before reaching for auxiliary constructions (chains, inductions, limits) that solve a more general version of the problem than was asked.

▷ **Cue:** a relation of the form $f(x, y) + f(y, z) + f(z, x) = 0$ (or similar cyclic sum) is asserted for all x, y, z, and the goal is to write $f(x, y)$ as a difference $g(x) - g(y)$ $\longrightarrow$ **Move:** fix the third variable at a constant c (commonly 0) in the general relation to get $f(x, y) = -f(y, c) - f(c, x)$ directly, after first substituting degenerate triples to pin down $f(c, c)$ and the relation between $f(x, c)$ and $f(c, x)$

▷ **Cue:** a universally-quantified relation among several free variables and a target identity naming only a few specific quantities $\longrightarrow$ **Move:** search for the single instance of the relation (a specific choice of the free variables) that already contains every named quantity, instead of building an auxiliary chain or induction that solves a more general problem

2008 A2

2008 A2. Alan and Barbara play a game in which they take turns filling entries of an initially empty 2008×2008 array. Alan plays first. At each turn, a player chooses a real number and places it in a vacant entry. The game ends when all the entries are filled. Alan wins if the determinant of the resulting matrix is nonzero; Barbara wins if it is zero. Which player has a winning strategy?

Fields: linear algebra, combinatorial game theory **Position:** A2 **Answer:** Barbara wins. Pair the entries of row 1 with the entries of row 2, column by column; whenever Alan writes a value in row 1 or 2, Barbara immediately writes the identical value in the paired entry. The two rows end up equal, so the determinant is forced to 0, regardless of everything else that happens on the board.

Layer 1 THE WHAT — surface analysis

The natural attempt. The board has 2008^2 entries, an even number, so Alan and Barbara make equally many moves and Barbara moves last. That fact looks decisive: the determinant is a

linear function of any single entry, by cofactor expansion along that entry's row. Whichever entry is filled last, the determinant equals $\alpha x + \beta$ for some real α (the cofactor) and β (the contribution of everything already placed), where x is the value about to be chosen. If Barbara is the one filling that final cell, she should be able to solve $\alpha x + \beta = 0$ for x and force the determinant to 0 on the spot.

This is genuine leverage, and it is tempting to declare the game won: Barbara simply plays however she likes for $2008^2 - 1$ moves, then, on her last move, computes the cofactor α of the one remaining cell and sets $x = -\beta/\alpha$.

The argument has a hole exactly at the word "computes": it requires $\alpha \neq 0$. But α is the determinant of the 2007×2007 submatrix obtained by deleting the last cell's row and column, and nothing in the setup guarantees that submatrix is invertible. If Alan spends his moves trying to make that particular cofactor vanish – for instance by ensuring two rows of the eventual 2007×2007 submatrix agree – then on Barbara's last move $\alpha = 0$, the determinant is already pinned at β regardless of what she writes in the final cell, and her leverage has evaporated exactly when she needed it. Worse, whether α vanishes depends on the entire history of play, including moves made many turns earlier by both players, so "wait for the last move" is not a strategy that can be verified in advance – it might work, or it might fail, depending on choices neither player is required to make with this in mind.

The model chain. Before the game starts, Barbara fixes a pairing: for each column $j = 1, \ldots, 2008$, pair the cell $(1, j)$ in row 1 with the cell $(2, j)$ in row 2.

Barbara's strategy: whenever Alan writes a value v in a cell of row 1 or row 2, Barbara immediately writes the same value v in that cell's pair-mate (the corresponding cell in the other of rows 1, 2). Whenever Alan writes a value anywhere in rows 3 through 2008, Barbara responds with any legal move anywhere in rows 3 through 2008 (the choice is irrelevant to the outcome).

This is a legal strategy: the pair-mate of the cell Alan just filled is always still empty when Barbara needs it. If it were already filled, it could only have been filled by Barbara in response to an earlier move by Alan on this same pair – but Barbara only ever touches a pair once, immediately after Alan opens it, so a pair is either fully empty or fully filled (by Barbara, in the same round it was opened); Alan can never find a pair half-open with only its mate available to him.

At the end of the game every cell in rows 1 and 2 has been filled in matched pairs with identical values, so row 1 and row 2 of the final matrix are literally the same row, written twice. A matrix with two equal rows has determinant 0 (swapping two equal rows both negates the determinant, by antisymmetry of the determinant under row exchange, and leaves the matrix unchanged, so $\det = -\det$, forcing $\det = 0$). Barbara wins regardless of what Alan does, and regardless of what Barbara herself writes in rows 3 through 2008.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both approaches correctly identify that determinant is linear in a single entry and that parity of 2008^2 gives Barbara the last move – that observation is common ground and not wrong. They split on what kind of leverage to build a strategy around. The natural attempt bets everything on control of one specific, late-arriving numerical coefficient (the cofactor of the last cell), which is a quantity that depends on the entire history of play and can degenerate to zero. The model chain instead forces a purely structural fact – two whole rows made identical – decided cell by cell throughout the game, which guarantees $\det = 0$ with no coefficient to check and no dependence on move order at all.

> *If my strategy's success depends on some cofactor being nonzero at the moment I need it, I have made my win conditional on a numerical accident I do not control – I should instead look for a way to force the determinant to zero that is true by a structural law (antisymmetry, linear dependence) and therefore cannot fail no matter what numbers end up on the board.*

> *A determinant vanishes for free whenever two rows are forced equal, and I get to choose exactly how to respond to every move Alan makes in those two rows – so instead of waiting for one lucky final move, I should spend every single one of my moves in rows* 1 *and* 2 *copying Alan's move there, and let every other cell on the board be irrelevant to the outcome.*

Layer 3 THE WHY

Why does the pairing strategy succeed where the endgame-leverage strategy stalls? Because the determinant is a specific kind of function of the matrix's rows: it is *multilinear and alternating*. Alternating means that swapping two rows negates the determinant; a matrix with two identical rows is fixed under that swap, so its determinant must equal its own negative, forcing it to 0. This is a structural, algebraic identity – it holds for every choice of the other 2006 rows, with no coefficient anywhere that could happen to vanish and spoil the argument. Once a strategy is built to force "row 1 equals row 2," the conclusion $\det = 0$ follows from the definition of what a determinant does under row exchange, not from an arithmetic computation performed under time pressure.

Contrast that with the endgame idea: linearity of det in a single entry is also a true structural fact, but the coefficient α in $\det = \alpha x + \beta$ is itself a determinant – of the deleted submatrix – and determinants of $(n-1) \times (n-1)$ matrices are exactly the kind of quantity that can be forced to zero by an adversary, since it is built from 2006 other rows Alan and Barbara jointly control throughout the game. Leverage that rests on "the cofactor is nonzero" is leverage borrowed against a quantity the opponent has $(2007^2 - 1)$ prior moves' worth of opportunity to sabotage. A robust strategy for a game decided by a single linear-algebraic invariant should target the invariant directly (here: dependence of two rows), not a downstream numerical quantity (a cofactor) that the invariant only affects indirectly and late.

The deeper generalisable fact: any player who wants to force $\det = 0$ in a fill-the-matrix game should look for a way to force linear dependence among the rows (or columns) – two rows equal, one row a fixed multiple of another, or a whole row summing to a target – because linear dependence is a property that, once forced, is permanent and holds regardless of any other entry on the board. That is exactly the kind of win condition a pairing strategy, executed move by move with no memory of the wider board, is suited to force.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the parity of the total number of cells, 2008^2, which is even: whoever moves second gets the last move. On seeing that, the reflex to write down is not yet a strategy, only a question – what can be guaranteed with the last move, and is that guarantee robust or fragile?

2. That cue suggests examining determinant's linearity in a single entry, which is the correct next idea to test – but testing it (does the leverage always work?) is essential before committing to it as the strategy.

3. Hesitation should occur exactly when checking whether the cofactor α of the final cell can be forced nonzero. What reinforces abandoning the endgame plan is realising that α is itself a 2007×2007 determinant, built from rows under joint control – i.e., the plan's success depends on another instance of the very quantity (a determinant) the game is about, which is circular and gives no actual guarantee.

4. The trap is treating "Barbara moves last" as automatically decisive. Moving last only grants leverage over a linear function of the final entry, and that leverage is worthless exactly when the linear coefficient vanishes – which an adversarial Alan (or even an indifferent one) can arrange across 2007^2 jointly-played cells.

5. Recognise this shape faster by asking, for any fill-the-grid game whose win condition is a linear-algebraic property (determinant zero/nonzero, rank, trace, a fixed row sum): can that property be forced by a symmetric pairing of cells, decided move by move, rather than by a single late move whose leverage depends on the rest of the board?

6. Reusable rule: to force a determinant to zero in an adversarial fill-in game, look for a pairing of two rows (or columns) that a fixed strategy can force to become identical (or otherwise linearly dependent) cell by cell, throughout the game – this wins by the alternating property of determinants and needs no numerical computation, whereas strategies based on controlling one final entry's coefficient are fragile because that coefficient is itself a lower-order determinant the opponent can influence.

▷ **Cue:** a game requires forcing a matrix's determinant to zero (or nonzero) through adversarial move-by-move entry choices $\longrightarrow$ **Move:** look first for a pairing strategy that forces two whole rows (or columns) to become equal, or otherwise linearly dependent, cell by cell – this wins by the structural, alternating property of the determinant and needs no numerical computation

▷ **Cue:** a proposed strategy's success depends on a coefficient (a cofactor, a leading term) being nonzero at a specific late move $\longrightarrow$ **Move:** check whether that coefficient is itself built from quantities under joint or adversarial control earlier in the process; if so, treat the strategy as fragile and look for a structural guarantee instead

2008 A3

2008 A3. Start with a finite sequence $a_1, a_2, \ldots, a_n$ of positive integers. If possible, choose two indices $j < k$ such that a_j does not divide a_k, and replace a_j and a_k by $\gcd(a_j, a_k)$ and $\operatorname{lcm}(a_j, a_k)$, respectively. Prove that if this process is repeated, it must eventually stop and the final sequence does not depend on the choices made. (Note: gcd means greatest common divisor and lcm means least common multiple.)

Fields: number theory, invariants, well-ordering **Position:** A3 **Answer:** The process always terminates, and the terminal sequence is uniquely determined by the multiset $\{a_1, \ldots, a_n\}$: for every prime p and exponent $m \geq 1$, the count of terms divisible by p^m is an invariant of every step, and this pins down every term's full prime factorization, hence the whole final sequence, independent of the order of moves.

Layer 1 THE WHAT — surface analysis

The natural attempt. The product $a_1 a_2 \cdots a_n$ is invariant at every step, since $\gcd(a_j, a_k) \cdot \operatorname{lcm}(a_j, a_k) = a_j a_k$ always. That is honest, immediate progress, and it is tempting to look for a second invariant, or a monovariant, in the same spirit – something built from the whole sequence, symmetric, that pins down the answer directly. A natural guess is the sum of squares $\sum a_i^2$: for a fixed product ab, replacing (a, b) by $(\gcd, \operatorname{lcm})$ spreads the two numbers further apart (since $\gcd(a, b) \leq \min(a, b) \leq \max(a, b) \leq \operatorname{lcm}(a, b)$, with at least one inequality strict whenever $a \nmid b$), and spreading two numbers apart while holding their product fixed strictly increases the sum of their squares. So $\sum a_i^2$ is a strictly increasing quantity, bounded above (since the product is fixed, and there are only finitely many multisets of positive integers with a given product decomposed into n parts once the parts are constrained by the fixed product – though making this boundedness fully rigorous takes some care). This is a workable route to termination.

Uniqueness is where this line of attack stalls. Having shown the process must stop, a natural next step is to test small cases by hand – say $n = 3$ with $a = (4, 6, 9)$ – tracing two different orders of moves and checking they land on the same final sequence. They do, and the pattern that emerges is that the final sequence is always totally ordered by divisibility, $a'_1 \mid a'_2 \mid \cdots \mid a'_n$. The temptation is to try to prove uniqueness the way the empirical check was done: an exchange argument, showing that performing one particular pair of operations in the other order produces the same result, then trying to induct on the number of steps or the number of "inversions" (bad pairs $j < k$ with $a_j \nmid a_k$), in the manner of proving that different implementations of bubble sort agree. This can be made to work, but it means tracking

interacting swaps of actual integers, several of which change at once, and checking case by case that reordering two nearby moves does not change the outcome – a combinatorial argument whose bookkeeping grows with n and does not obviously reduce to something checkable in a few lines.

The model chain. **Termination.** The last term a_n of the sequence never decreases: whenever the pair (a_j, a_k) with $j < k$ is operated on, the term at the larger index k becomes $\operatorname{lcm}(a_j, a_k) \geq a_k$. In particular, if $k = n$, then a_n can only increase or stay the same; if $k < n$, then a_n is untouched. So a_n is a nondecreasing sequence of positive integers, bounded above by $a_1 \cdots a_n$ (the fixed product), hence it can only strictly increase finitely many times before it stabilises at some maximum value. Once a_n is at that maximum and constant, index n can never again be the "k" in a move (that would require a further strict increase), so from that point on, no move ever touches a_n again – the process effectively continues on the shorter sequence $a_1, \ldots, a_{n-1}$. By induction on n (base case $n = 1$: no moves are ever possible), the same argument shows a_{n-1} eventually stabilises, then a_{n-2}, and so on down to a_1. After finitely many steps every term is constant, at which point no bad pair $j < k$ with $a_j \nmid a_k$ remains – otherwise a move would still be available and some term would still be able to change. The process terminates.

Uniqueness. Fix a prime p and an integer $m \geq 1$. Claim: the number of terms of the sequence divisible by p^m is unchanged by every single move. Suppose the pair (a_j, a_k) is replaced by $(\gcd(a_j, a_k), \operatorname{lcm}(a_j, a_k))$. If neither a_j nor a_k is divisible by p^m, then $v_p(a_j), v_p(a_k) < m$, so $v_p(\gcd) = \min(v_p(a_j), v_p(a_k)) < m$ and $v_p(\operatorname{lcm}) = \max(v_p(a_j), v_p(a_k)) < m$: neither replacement value is divisible by p^m either, and the count of p^m-divisible terms among this pair stays 0. If both are divisible by p^m, both $v_p(a_j), v_p(a_k) \geq m$, so both min and max are $\geq m$: the count stays 2. If exactly one is divisible, say $v_p(a_j) < m \leq v_p(a_k)$, then $v_p(\gcd) = \min < m$ (not divisible) and $v_p(\operatorname{lcm}) = \max \geq m$ (divisible): the count stays 1. In every case the number of p^m-divisible terms among the pair being operated on is preserved, so the total count across the whole sequence is preserved at every step, for every (p, m).

Now let $h_{p,m}$ denote this invariant count for the original sequence $a_1, \ldots, a_n$. In the terminal sequence $a'_1, \ldots, a'_n$, the divisibility condition $a'_j \mid a'_k$ for all $j < k$ forces the set of indices i with $p^m \mid a'_i$ to be exactly the top $h_{p,m}$ indices, $\{n - h_{p,m} + 1, \ldots, n\}$: if p^m divided some a'_i but not some $a'_{i'}$ with $i < i'$, then since $a'_i \mid a'_{i'}$, p^m would divide $a'_{i'}$ too, a contradiction; so the set of p^m-divisible indices is an up-set, and its size is fixed at $h_{p,m}$, which pins it down as exactly the top $h_{p,m}$ indices. This holds for every prime p and every $m \geq 1$, and for a fixed index i, the exponent $v_p(a'_i)$ is exactly the number of m for which i lies among the top $h_{p,m}$ indices – a quantity determined entirely by the multiset $h_{p,1}, h_{p,2}, \ldots$, which in turn is determined entirely by the *original* sequence, not by any choice made during the process. So every prime-power valuation of every a'_i is forced, hence every a'_i is forced, hence the terminal sequence is unique.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains find the same first invariant (the product) and both eventually converge on the true shape of the terminal sequence (totally ordered by divisibility). They split on how to prove *uniqueness*. The natural attempt tries to track the integers themselves – performing an exchange argument on actual values a_j, a_k, case-by-case, hoping the bookkeeping closes up as an induction on step order. The model chain refuses to track the integers directly at all: it decomposes each integer into its prime-power divisibility pattern first, proves the invariant *there*, and only at the very end reassembles the integers from their (now fully pinned-down) valuations.

> *Tracking the actual numbers through an arbitrary interleaving of moves is going to require checking, case by case, that reordering any two nearby operations gives the same result – that is a statement about a specific combinatorial process, and it should generalise better if I can find some simpler, more rigid structure underneath the integers that the operation acts on predictably.*

> *The operation $(a, b) \mapsto (\gcd(a, b), \mathrm{lcm}(a, b))$ is, for a single fixed prime p, just replacing the pair of exponents $(v_p(a), v_p(b))$ by $(\min, \max)$ – and for a $\{0, 1\}$-valued question like "is $p^m \mid a_i$?", replacing $(v_p(a_j), v_p(a_k))$ by $(\min, \max)$ is exactly a compare-and-swap on a boolean pair. A sequence of 0s and 1s sorted by repeated compare-and-swaps has only one possible sorted order, determined purely by how many 1s it has – so if I track "is $p^m \mid a_i$?" as a boolean sequence for every (p, m) separately, uniqueness becomes a fact I don't have to prove about integers at all; it's automatic for sorted booleans.*

Layer 3 THE WHY

Why does the prime-power decomposition succeed where direct tracking of integers bogs down? Because the operation $(a, b) \mapsto (\gcd(a, b), \mathrm{lcm}(a, b))$ is not really one operation on integers – by unique factorization, it is infinitely many independent operations happening at once, one for each prime p, each acting on the pair of exponents $(v_p(a), v_p(b))$ by $(\min, \max)$. This is exactly the "Reinterpretation" Kedlaya's solution names explicitly: sort each prime's exponent-sequence separately by repeated compare-and-swap, the way a sorting network works. The reason this matters for *uniqueness* specifically is that a compare-and-swap sort of a fixed multiset has an answer that is uniquely determined the moment you know only *how many* of each value are present – there is exactly one way to arrange a fixed number of 1s and 0s (or, for a full exponent, a fixed multiset of nonnegative integers) into nondecreasing order. Uniqueness of the sorted arrangement of a single sequence of comparable values is close to definitional; it needs almost no proof once you are looking at the right object.

Integers do not have this property directly – two different integers are not simply "comparable values that get sorted," because the gcd/lcm operation does not act on the integers themselves as a linear order (there exist incomparable pairs, which is exactly the $a_j \nmid a_k$ condition that triggers a move in the first place). It is only after breaking each integer into its prime valuations that every coordinate of the resulting infinite-dimensional vector *is* linearly ordered (each valuation is a nonnegative integer along a single prime's axis), and $(\gcd, \mathrm{lcm})$ acts coordinatewise as exactly the two-element sort $(\min, \max)$ in each coordinate independently. This is why the invariant to look for is not a single symmetric numerical quantity (like $\sum a_i^2$, which only shows monotonicity, not uniqueness) but a whole family of counting invariants, indexed by (p, m), each of which is trivially preserved and, together, is rich enough to recover every digit of every prime factorization of every final term.

The general principle: when an operation on a compound object (here, an integer, itself a vector of prime exponents) reduces coordinatewise to a simple, well-understood operation (here, $(\min, \max)$, a comparison), decompose first and prove uniqueness in the coordinates, where it is close to automatic, rather than trying to prove it directly on the compound object, where the operation's structure is obscured.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the pair of operations gcd and lcm appearing together, replacing two numbers whose relationship is only "$a_j \nmid a_k$" (not a full order). On seeing gcd and lcm paired like this, the thing to write down immediately is: $\gcd(a, b) \cdot \mathrm{lcm}(a, b) = ab$, and also, for every prime p, $v_p(\gcd(a, b)) = \min(v_p(a), v_p(b))$ and $v_p(\mathrm{lcm}(a, b)) = \max(v_p(a), v_p(b))$ – the operation is a coordinatewise min/max in disguise.
2. That cue suggests treating the process, prime by prime, as a sorting network acting on exponent sequences – termination and uniqueness should each be provable separately in that coordinatewise picture, rather than by tracking the compound integers.
3. Hesitation sets in once termination is settled (via a bounded, eventually-stabilising last term, or a strictly increasing bounded quantity like $\sum a_i^2$) and the instinct is to keep using the same style of argument – a single global numerical quantity – to also prove uniqueness. What reinforces the switch away from that is noticing that a monovariant like $\sum a_i^2$ proves only

that the process cannot cycle; it says nothing about whether two different orders of moves reach the *same* endpoint, because many different terminal multisets could in principle share the same sum of squares.

4. The trap is believing that "the process terminates" and "the endpoint is unique" are the same kind of claim requiring the same kind of invariant. Termination is a statement about a well-ordered quantity decreasing (or increasing and bounded); uniqueness is a statement about the fibres of the process being singletons, and the correct tool for that is a set of invariants fine enough to pin down every coordinate of the answer – not one coarse invariant that merely bounds the search.

5. Recognise this shape faster by treating any hypothesis mixing gcd and lcm (or min and max, or union and intersection) as an instruction to factor the problem through unique factorization (or the relevant Boolean/lattice structure) first, then ask separately: does the process terminate (find a bounded monovariant), and is the result unique (find enough coordinatewise invariants to pin down every digit of the answer)?

6. Reusable rule: when an operation on compound objects (integers via prime factorization, sets via elements, vectors via coordinates) reduces, coordinate by coordinate, to a simple order-based operation like $(\min, \max)$ or compare-and-swap, prove termination with a single bounded monovariant on the whole object, but prove uniqueness by counting, for each coordinate/threshold, how many entries exceed it – that count is preserved by every step and, together over all thresholds, determines every entry of the final state exactly.

▷ **Cue:** an operation on integers is built from gcd and lcm (or a pair is repeatedly replaced by its min and max in some order), and termination + uniqueness of the final state are both required $\longrightarrow$ **Move:** decompose each integer into its prime valuations; the operation acts coordinatewise as $(\min, \max)$ in every valuation, so prove termination via a bounded monovariant on the whole sequence, and prove uniqueness via the invariant "number of terms with valuation at least m, for each prime p and threshold m" – preserved by every step and sufficient to reconstruct every final term

▷ **Cue:** a claimed invariant only shows a process cannot run forever (a monovariant), but the problem also asks whether the endpoint is unique $\longrightarrow$ **Move:** do not expect the same invariant to answer both questions; look separately for a family of finer, coordinatewise invariants (counts, or presence/absence at each threshold) rich enough to pin down every component of the final state

2008 A4

2008 A4. Define $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \begin{cases} x & \text{if } x \le e \\ xf(\ln x) & \text{if } x > e. \end{cases}$$

Does $\sum_{n=1}^{\infty} \frac{1}{f(n)}$ converge?

Fields: real analysis (series, integral test, iterated logarithms) **Position:** A4 — mid-paper, but it plays like an A2 once the right tool is chosen. The difficulty is refusing to compare term by term. **Answer:** The series diverges: on each band $[\exp^{k-1} 1, \exp^{k} 1]$ the integral $\int dx/f(x)$ equals exactly 1, so $\int_1^{\infty} dx/f(x) = \infty$ and the integral test gives divergence.

Layer 1 THE WHAT — surface analysis

The natural attempt. You unwind the definition. For $1 \le x \le e$, $f(x) = x$. For $e < x \le e^e$, $\ln x \in (1, e]$, so $f(x) = x \ln x$. For $e^e < x \le e^{e^e}$, $f(x) = x \ln x \ln\ln x$. And so on: on the k-th band $f(x) = x \cdot \ln x \cdot \ln\ln x \cdots \ln^{k-1} x$, where $\ln^j$ means the j-fold iterated logarithm.

Now you reach for comparison. You know $\sum 1/n$ diverges, $\sum 1/(n \ln n)$ diverges, $\sum 1/(n \ln n \ln\ln n)$ diverges — each by the integral test or by Cauchy condensation. So the natural plan is "$f(n)$ is one of these, so the sum diverges." But it is not one of these. On band k, f has k factors, and k grows without bound. For any fixed K, eventually $f(n) \geq n \ln n \cdots \ln^{K-1} n$, so $1/f(n)$ is *smaller* than each divergent comparison series' terms. Comparison in that direction proves nothing. The other direction — bounding $f(n)$ above by something whose reciprocal sum diverges — needs a single function that dominates every finite product of iterated logs, and you do not have one to hand.

You try numerics. The partial sum up to $n = 10^6$ is about 3.54. Up to 10^6 the terms are around 10^{-7}. Everything in your experience says a sum whose terms fall like that and whose partial sums have barely moved converges. This is the wall: the comparison tool is pointing the wrong way, and the numbers are lying.

The model chain. Step 1: shape of f. Write $\exp^0 1 = 1$, $\exp^k 1 = e^{\exp^{k-1} 1}$, so the band endpoints are $1, e, e^e, e^{e^e}, \ldots$. Write $\ln^1 x = \ln x$, $\ln^j x = \ln(\ln^{j-1} x)$. By induction on k: for $x \in (\exp^{k-1} 1,\ \exp^k 1]$ with $k \geq 2$, $\ln x \in (\exp^{k-2} 1,\ \exp^{k-1} 1]$, so

$$f(x) = x\, f(\ln x) = x \cdot \ln x \cdot \ln^2 x \cdots \ln^{k-1} x,$$

and on $[1, e]$, $f(x) = x$ (the $k = 1$ case, empty product of logs).

Step 2: f is positive, continuous and increasing on $[1, \infty)$. On the open band each factor $\ln^j x$ is positive (since $\ln^j x > \exp^{k-1-j} 1 \geq 1 > 0$ for $j \leq k-1$) and increasing, so f is positive and increasing there. At the band endpoint $x = \exp^k 1$ the new factor is $\ln^k x = \ln^k(\exp^k 1) = 1$, so the formula from band $k+1$ agrees with the formula from band k: f is continuous. (Check: $f(e) = e$ from both sides; $f(e^e) = e^e \cdot e$ from both sides.) Hence $1/f$ is positive, continuous and decreasing on $[1, \infty)$.

Step 3: integral test. Since $1/f$ is decreasing, $\frac{1}{f(n)} \geq \int_n^{n+1} \frac{dx}{f(x)}$ for each $n \geq 1$, so

$$\sum_{n=1}^{N} \frac{1}{f(n)} \;\geq\; \int_1^{N+1} \frac{dx}{f(x)}.$$

Step 4: each band contributes exactly 1. By the chain rule, $\frac{d}{dx} \ln^k x = \frac{1}{x \ln x \cdots \ln^{k-1} x}$, which on band k is exactly $\frac{1}{f(x)}$. Hence

$$\int_{\exp^{k-1} 1}^{\exp^k 1} \frac{dx}{f(x)} = \Big[\ln^k x\Big]_{\exp^{k-1} 1}^{\exp^k 1} = 1 - 0 = 1.$$

So $\int_1^{\exp^K 1} dx/f(x) = K$, which is unbounded in K. The partial sums are therefore unbounded, and the series diverges.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains unwind f into the band formula $f(x) = x \ln x \cdots \ln^{k-1} x$. They split at (†), the moment after that formula is written. The natural attempt compares $f(n)$ with a *fixed* product of logarithms, and finds every such comparison points the wrong way. The model notices that the band formula is exactly the derivative of $\ln^k x$ inverted, and integrates band by band.

> *Every fixed comparison series has finitely many logs and mine has more and more of them, so term-by-term comparison is doomed. But on each band the denominator $x \ln x \cdots \ln^{k-1} x$ is precisely what the chain rule produces when I differentiate $\ln^k x$. So the integral over a band is a difference of $\ln^k$ at the endpoints — and the endpoints are towers of e, so that difference is 1.*

> *The partial sum is* 3.5 *at a million. That tells me nothing: if each band contributes about* 1 *and the bands end at* e, 15, 3.8 *million,* $e^{3.8\ million}$, *then of course the sum has barely moved. Numerics cannot decide this; only the band integral can.*

Layer 3 THE WHY

Why is the integral test the natural tool, and not a lucky one? Look at how f is built. It is defined by a recursion in which x is replaced by $\ln x$. A recursion in $\ln x$ is a statement about the substitution $y = \ln x$, and the only operation that interacts cleanly with a substitution is integration: sums over integers do not change variables, integrals do. So as soon as the definition says "$f(x)$ in terms of $f(\ln x)$", the object to study is $\int dx/f(x)$, because under $y = \ln x$, $dx = x\,dy$ and

$$\int \frac{dx}{f(x)} = \int \frac{dx}{x\,f(\ln x)} = \int \frac{dy}{f(y)}.$$

That single line is the whole problem. It says the integral of $1/f$ over $(e^a, e^b]$ equals the integral over $(a, b]$. Iterating from $[1, e]$, where the integral is $\int_1^e dx/x = 1$, every band has integral 1. The band-by-band chain-rule computation in the model chain is this identity unrolled; the identity is the reason it works. Kedlaya's solution substitutes $y = \ln^k x$ on the k-th band and states the result. Here is where that substitution comes from: the definition of f is itself the substitution $x \mapsto \ln x$, so undoing it k times on the k-th band is the only thing to try.

Why does comparison fail structurally? Because f grows faster than every finite product $x \ln x \cdots \ln^K x$, and those products sit exactly on the convergence boundary: $\sum 1/(n \ln n \cdots \ln^K n)$ diverges for every K while $\sum 1/(n \ln n \cdots (\ln^K n)^{1+\varepsilon})$ converges. The series $\sum 1/f(n)$ lies beyond every member of that scale, so no member can serve as a comparator. When a series sits beyond a whole family of test series, you need a tool that is not "compare with a member of the family"; the integral computed from the defining recursion is that tool.

Why do the numerics lie? The integral up to $\exp^K 1$ is K, so the partial sum up to N grows like the number of times you must take ln to bring N below 1 — the iterated logarithm $\log^* N$. That is at most 5 for every N that could ever be written down in decimal. Divergence at this rate is invisible to any computer. The lesson is not "numerics are useless"; it is that a series whose terms are built from iterated logarithms lives on a scale where numerics carry no information.

Finally, why must you check that f is increasing and continuous? Because the integral test is a comparison of a sum with an integral, and that comparison is valid only for monotone integrands. Here monotonicity needs the continuity at the band joins; the joins are designed so the new factor $\ln^k x$ equals 1 at the join. That design is why the problem is posed with the threshold e rather than some other number.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the recursion $f(x) = xf(\ln x)$. Write down at once: *substitute* $y = \ln x$ *in* $\int dx/f(x)$, which gives $\int dy/f(y)$.

2. The integral test, with the integral evaluated band by band using the self-similarity of f under $x \mapsto \ln x$.

3. Hesitation occurs after the band formula is written, when you instinctively reach for a known divergent series to compare with and none fits. It is reinforced away by asking what the recursion does to an integral rather than to a term.

4. Two traps. First, term-by-term comparison with $\sum 1/(n \ln n \cdots)$: every such comparison goes the wrong way. Second, the numerical trap: partial sums near 3.5 after a million terms invite "converges". A third, smaller trap: invoking the integral test without checking that $1/f$ is decreasing, which needs continuity at the band joins.

5. Slogan: *a function defined by a recursion in* $\ln x$ *(or* $\sqrt{x}$*, or any substitution) is asking for an integral, because integrals are what change variables cleanly.*

6. Decision rule: **when a series sits beyond an entire scale of comparison series (more and more logs, more and more factors), stop comparing; compute an integral using the substitution that defines the terms, and read off the growth rate directly.**

▷ **Cue:** a function defined by $f(x) = x\,f(\ln x)$ or similar self-reference under a substitution $\longrightarrow$ **Move:** substitute $y = \ln x$ in $\int dx/f(x)$; the recursion becomes $\int dy/f(y)$, so the integral over each band equals the integral over the previous band

▷ **Cue:** a summand whose denominator is $x \ln x \ln\ln x \cdots$ with a varying number of factors $\longrightarrow$ **Move:** recognise the denominator as the reciprocal derivative of an iterated logarithm, and integrate band by band

▷ **Cue:** every standard comparison series gives an inequality in the useless direction $\longrightarrow$ **Move:** switch tool: integral test with the integral computed exactly, or condensation

▷ **Cue:** partial sums barely move but the terms involve iterated logarithms $\longrightarrow$ **Move:** distrust numerics; estimate the growth scale ($\log^* N$ here) analytically

2008 A5

2008 A5. Let $n \geq 3$ be an integer. Let $f(x)$ and $g(x)$ be polynomials with real coefficients such that the points $(f(1), g(1)), (f(2), g(2)), \ldots, (f(n), g(n))$ in $\mathbb{R}^2$ are the vertices of a regular n-gon in counterclockwise order. Prove that at least one of $f(x)$ and $g(x)$ has degree greater than or equal to $n-1$.

Fields: algebra (polynomials, complex numbers, finite differences) **Position:** A5 — hard end of the paper, but short once the complex packaging is chosen. The difficulty is entirely in choosing to merge f and g. **Answer:** Proof: $P = f + ig$ satisfies $P(k) = b + a\zeta^k$ with $\zeta = e^{2\pi i/n}$, $a \neq 0$; then $R(z) = Q(z+1) - \zeta Q(z)$ for $Q = (P-b)/a$ has the $n-1$ roots $1, \ldots, n-1$ and the same degree as Q, forcing $\deg P \geq n-1$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You set up coordinates. The polygon has a centre (x_0, y_0) and circumradius $r > 0$, and consecutive vertices are separated by angle $2\pi/n$ counterclockwise, so for some angle θ,

$$f(k) = x_0 + r\cos\left(\theta + \tfrac{2\pi k}{n}\right), \qquad g(k) = y_0 + r\sin\left(\theta + \tfrac{2\pi k}{n}\right), \qquad k = 1, \ldots, n.$$

Now you attack f. A polynomial of degree at most $n-2$ has vanishing $(n-1)$-th finite difference, so if $\deg f \leq n-2$ then $\sum_{j=0}^{n-1}(-1)^j\binom{n-1}{j}f(n-j) = 0$. You try to show this sum of cosines is nonzero, and it will not come out. Try $n = 3$: can f be linear? Yes. Take the triangle with vertices $(\frac{\sqrt{3}}{2}, -\frac{1}{2})$, $(0, 1)$, $(-\frac{\sqrt{3}}{2}, -\frac{1}{2})$ at $k = 1, 2, 3$ (counterclockwise: polar angles $-30°, 90°, 210°$). The x-coordinates $\frac{\sqrt{3}}{2}, 0, -\frac{\sqrt{3}}{2}$ are in arithmetic progression, so $f(x) = \sqrt{3} - \frac{\sqrt{3}}{2}x$ has degree $1 < n-1$. The y-coordinates $-\frac{1}{2}, 1, -\frac{1}{2}$ are not, so g is forced to be quadratic.

This is the wall, and it is informative: the statement says "at least one of" for a reason. No argument about f alone can work, because for a suitable rotation θ the cosine sum vanishes. You need a quantity that sees f and g together, and handling two trigonometric sums in tandem — showing they cannot both vanish — is a messy computation with cos and sin of sums.

The model chain. Step 1: complexify. Let $P(z) = f(z) + ig(z)$, a polynomial with complex coefficients. Its degree is at most $\max(\deg f, \deg g)$, so it suffices to prove $\deg P \geq n-1$.

Step 2: normalise. Let $b \in \mathbb{C}$ be the centre of the polygon and $\zeta = e^{2\pi i/n}$. Regular with vertices in counterclockwise order means each vertex is the previous one rotated by $2\pi/n$ about the centre: $P(k+1) - b = \zeta\,(P(k) - b)$ for $k = 1, \dots, n-1$. Hence $P(k) = b + a\zeta^k$ for $k = 1, \dots, n$, where $a = (P(1) - b)\zeta^{-1} \neq 0$ (the circumradius is positive). Put $Q(z) = (P(z) - b)/a$. Then $\deg Q = \deg P$ and $Q(k) = \zeta^k$ for $k = 1, \dots, n$.

Step 3: a shift operator that kills the data. Define $R(z) = Q(z+1) - \zeta\, Q(z)$. For $k = 1, \dots, n-1$,

$$R(k) = \zeta^{k+1} - \zeta \cdot \zeta^k = 0.$$

So R has the $n-1$ distinct roots $1, 2, \dots, n-1$.

Step 4: R has the same degree as Q. Q is not the zero polynomial ($Q(1) = \zeta \neq 0$). Let $d = \deg Q$ and $c \neq 0$ its leading coefficient. $Q(z+1)$ also has degree d and leading coefficient c (expanding $(z+1)^d$ changes only lower terms). Hence the z^d coefficient of R is $c - \zeta c = (1-\zeta)c$, which is nonzero because $\zeta \neq 1$ for $n \geq 2$. So R is a nonzero polynomial of degree exactly d.

Step 5: count roots. A nonzero polynomial of degree d has at most d roots; R has at least $n-1$. So $d \geq n-1$, i.e. $\deg P \geq n-1$, and one of f, g has degree at least $n-1$.

Sharpness. Lagrange interpolation through the n values ζ^k gives Q of degree at most $n-1$, so the bound is attained.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains write down the vertex coordinates. They split at (†), the choice of what to write: the natural attempt writes $f(k)$ and $g(k)$ as two separate trigonometric expressions; the model writes the single complex number $P(k) = b + a\zeta^k$.

> *"At least one of f and g" — so neither one alone is forced, and any argument about f by itself is doomed. I need one object that carries both. A point in the plane with a rotation acting on it is a complex number being multiplied by ζ. So let $P = f + ig$.*

> *Now the data say $Q(k+1) = \zeta\, Q(k)$. That is a relation between Q at z and at $z+1$, so the polynomial $Q(z+1) - \zeta Q(z)$ vanishes at $1, \dots, n-1$. It still has the degree of Q because $\zeta \neq 1$. Count roots.*

Layer 3 THE WHY

Why complex numbers? Drive to the definition of the object. A regular n-gon, traversed in order, is the orbit of one vertex under rotation by $2\pi/n$ about the centre. Rotation of the plane is, by definition of complex multiplication, multiplication by a unit complex number. So the definition of "regular n-gon in counterclockwise order" is literally "$P(k+1) - b = \zeta(P(k) - b)$". In real coordinates the same fact is a 2×2 rotation matrix mixing f and g — which is exactly why neither can be analysed alone. The complex packaging is not a trick; it is the coordinate system in which the problem's own definition is one equation instead of two coupled ones. The phrase "at least one of" is the problem telling you that the two coordinates must be kept together.

Why the operator $Q(z+1) - \zeta Q(z)$? Because the data are a geometric sequence, $Q(k) = \zeta^k$, and a geometric sequence is defined by the recurrence $u_{k+1} = \zeta u_k$. Any recurrence satisfied by the values of a polynomial at consecutive integers can be lifted to a polynomial identity evaluated at those integers: replace k by z. The resulting polynomial is zero at every point where the recurrence holds. That is the whole mechanism, and it is the same mechanism as finite differences: $Q(z+1) - Q(z)$ is the recurrence "constant sequence"; $Q(z+1) - \zeta Q(z)$ is the recurrence "geometric with ratio ζ". Kedlaya's first solution introduces $R(z) = Q(z+1) - tQ(z)$ without comment. Here is where it comes from: it is the defining recurrence of the data, written as a polynomial.

Why does the degree survive? The leading coefficient of $Q(z+1)$ equals that of $Q(z)$, so subtracting $\zeta Q(z)$ multiplies the leading coefficient by $1-\zeta$. The ordinary difference operator ($\zeta = 1$) lowers the degree, which is why polynomials have eventually-vanishing finite differences. A twisted difference with $\zeta \neq 1$ preserves the degree. That contrast is the heart of it: polynomials are "eigenfunctions" of the shift only with eigenvalue 1 (constants), so they cannot mimic a geometric sequence with ratio $\neq 1$ for long. The number of steps a polynomial of degree d can mimic it is at most $d+1$ values, which is exactly the bound.

Why is $n \geq 3$ in the statement? Only so that the polygon exists. The argument uses $\zeta \neq 1$, true for $n \geq 2$. When a hypothesis is not used in the proof, say which role it plays; here it is non-degeneracy of the object, not an ingredient of the argument.

The equivalent finite-difference version (Kedlaya's third solution) computes $\sum_{j=0}^{n-1}(-1)^j\binom{n-1}{j}Q(n-j) = \zeta(\zeta-1)^{n-1} \neq 0$. I checked this closed form numerically for $n = 3, \ldots, 12$; its modulus matches $|\zeta(1-\zeta)^{n-1}|$. It is the same proof in a less transparent form: the natural attempt was one step away from it, and the only missing move was to apply the difference to $f+ig$ instead of to f.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the combination "regular n-gon in counterclockwise order" with "at least one of f and g". Write down at once: $P = f + ig$ and $P(k) = b + a\zeta^k$.

2. Complex numbers to turn rotation into multiplication; then a polynomial version of the recurrence $u_{k+1} = \zeta u_k$, and a root count.

3. Hesitation occurs when you start working on f alone and the trigonometric sum refuses to be nonzero. It is reinforced away by testing $n = 3$: seeing that f can be linear tells you immediately that the two coordinates must be combined.

4. The trap is proving a statement about f alone, which is false (the triangle example). A second trap is forgetting to justify that R has degree exactly $\deg Q$; the whole count rests on $1-\zeta \neq 0$.

5. Slogan: *if a conclusion says "at least one of two coordinates", combine the coordinates into one object; for planar points with rotational structure that object is $f+ig$.*

6. Decision rule: **when a polynomial's values at consecutive integers satisfy a recurrence, write the recurrence as a polynomial in z (replace k by z); it vanishes at every index where the recurrence holds, and a degree count finishes.**

▷ **Cue:** "at least one of f and g" where $(f(k), g(k))$ are points in the plane $\longrightarrow$ **Move:** set $P = f + ig$; its degree is at most the larger degree, and it carries both coordinates

▷ **Cue:** points forming a regular polygon in order $\longrightarrow$ **Move:** write them as $b + a\zeta^k$ with $\zeta = e^{2\pi i/n}$; normalise by an affine map to ζ^k

▷ **Cue:** a polynomial Q with $Q(k+1) = \lambda Q(k)$ for $k = 1, \ldots, m-1$ $\longrightarrow$ **Move:** consider $Q(z+1) - \lambda Q(z)$: it has roots $1, \ldots, m-1$ and, if $\lambda \neq 1$, the same degree as Q

▷ **Cue:** a claim that seems to be about one coordinate refuses to yield $\longrightarrow$ **Move:** test the smallest case; if one coordinate can be degenerate, the proof must treat both together

2008 A6

2008 A6. Prove that there exists a constant $c > 0$ such that in every nontrivial finite group G there exists a sequence of length at most $c \log |G|$ with the property that each element of G equals the product of some subsequence. (The elements of G in the sequence are not required to be distinct. A *subsequence* of a sequence is obtained by selecting some of the terms, not necessarily consecutive, without reordering them; for example, $4, 4, 2$ is a subsequence of $2, 4, 6, 4, 2$, but $2, 2, 4$ is not.)

Fields: group theory, additive/multiplicative combinatorics (doubling, counting) **Position:** A6 — the hardest slot on the paper, and hard for a specific reason: every instinct from group theory points at

structure, and the problem is solved by counting alone. **Answer:** Proof, with $c = 2/\ln 2$ (natural log): greedily build $a_1, \ldots, a_k$ whose subsequence products H double in size at each step, choosing $a_{k+1} \notin H^{-1}H$; this stops after $k \le \log_2 |G|$ steps with $H^{-1}H = G$, and then $a_k^{-1}, \ldots, a_1^{-1}, a_1, \ldots, a_k$ works.

Layer 1 THE WHAT — surface analysis

The natural attempt. You start where you would for a cyclic group, because that case you can do. In $\mathbb{Z}/n$ the sequence $1, 2, 4, \ldots, 2^m$ with $2^{m+1} \ge n$ produces every residue by binary expansion: length about $\log_2 n$. So the problem is "binary expansion for an arbitrary group".

Next you use structure to reduce to that case. Take a composition series $G = G_0 \rhd G_1 \rhd \cdots \rhd G_r = \{e\}$. If every factor G_{i-1}/G_i is cyclic of prime order p_i (that is, G is solvable), pick $g_i \in G_{i-1}$ mapping to a generator of G_{i-1}/G_i and let B_i be the block $g_i, g_i^2, g_i^4, \ldots$ of length $\lceil \log_2 p_i \rceil$, whose subsequence products include g_i^m for every $0 \le m < p_i$. Any $x \in G$ is $g_1^{m_1} y_1$ with $y_1 \in G_1$, then $y_1 = g_2^{m_2} y_2$, and so on, so the concatenation $B_1 B_2 \cdots B_r$ works, with total length

$$\sum_i \lceil \log_2 p_i \rceil \le \sum_i (\log_2 p_i + 1) \le 2 \log_2 |G|,$$

using $r \le \log_2 |G|$. This is a correct proof for solvable groups, and it cost real effort.

Then the wall. For $G = A_5$, or any non-abelian simple group, the composition series is $G \rhd \{e\}$ and the factor is G itself. There is no cyclic quotient to expand in binary. You could try a subgroup chain instead of a normal one and use coset representatives, but coset representatives do not multiply like the cosets do: in a non-normal chain, a representative of G_{i-1}/G_i followed by elements of G_i does not stay organised. The structural route needs, for every simple group, some internal handle, and the problem says *every* finite group. Short of the classification of finite simple groups, this road ends here.

The model chain. Write $n = |G| \ge 2$. For a sequence S, let $P(S)$ denote the set of products of its subsequences, *including* the empty subsequence with product e; we remove this convenience at the end.

Step 1: what appending does. If S has $P(S) = H$ and $g \in G$, then a subsequence of S, g either omits the last term or ends with it, so

$$P(S, g) = H \cup Hg.$$

Step 2: when appending doubles. $|Hg| = |H|$ (right multiplication by g is a bijection). The union is disjoint unless $h_1 g = h_2$ for some $h_1, h_2 \in H$, i.e. unless $g = h_1^{-1} h_2 \in H^{-1}H$, where $H^{-1}H = \{h_1^{-1} h_2 : h_1, h_2 \in H\}$. So if $g \notin H^{-1}H$, then $|P(S, g)| = 2|H|$.

Step 3: the greedy loop. Start with S empty, $H = \{e\}$. While $H^{-1}H \ne G$, append some $g \notin H^{-1}H$. After k steps $|H| = 2^k$ exactly, and $|H| \le n$, so the loop runs at most $\log_2 n$ times. It ends with $S = a_1, \ldots, a_k$, $k \le \log_2 n$, and $H^{-1}H = G$. Since G is nontrivial and the first check has $H^{-1}H = \{e\} \ne G$, we have $k \ge 1$.

Step 4: realise $H^{-1}H$. Consider

$$T = a_k^{-1},\ a_{k-1}^{-1},\ \ldots,\ a_1^{-1},\ a_1,\ a_2,\ \ldots,\ a_k.$$

A subsequence of the first half picks indices $i_m > \cdots > i_1$ in that order, with product $a_{i_m}^{-1} \cdots a_{i_1}^{-1} = (a_{i_1} \cdots a_{i_m})^{-1}$, so the first half produces exactly H^{-1}. The second half produces H. Any subsequence of T splits into a part in each half, so T produces every element of $H^{-1}H = G$.

Step 5: nonempty subsequences. If the problem is read as requiring a nonempty subsequence, the only element whose representation above might use the empty subsequence twice is $e = e^{-1}e$; but $e = a_1^{-1} a_1$ is also a product of a nonempty subsequence of T (the two middle terms), available because $k \ge 1$. So no appended identity is needed.

Step 6: the constant. $|T| = 2k \leq 2\log_2 n = \frac{2}{\ln 2}\ln n$. So $c = 2/\ln 2$ works for natural logarithms (and $c = 2$ for $\log_2$).

Order of magnitude. A sequence of length L has 2^L subsequences, so $2^L \geq n$ and $L \geq \log_2 n$: the $\log|G|$ in the statement cannot be improved beyond the constant.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start from the cyclic case and the binary-expansion picture. They split at (†): the natural attempt asks "what structure does G have that lets me expand in binary?"; the model asks "what does appending one term do to the set of reachable products, and when does it double?"

> *Length* $\log|G|$ *and* 2^L *subsequences: the statement is telling me each new term has to roughly double what I can reach. I do not need structure for that. I need to know when appending one element doubles the reachable set.*

> *Appending g turns H into $H \cup Hg$. That fails to double only on a collision $h_1 g = h_2$, i.e. $g = h_1^{-1}h_2$. So either I can double, or every g is of the form $h_1^{-1}h_2$ — and "every element is $h_1^{-1}h_2$" is almost what I want. I can make h_1^{-1} by writing the inverses in reverse order in front.*

> *I checked solvable groups by composition series. Before I go further down that road, what happens for A_5? Nothing. Then structure is the wrong tool for a statement about every group.*

Layer 3 THE WHY

Why counting and not structure. Read the quantities in the statement: $|G|$ and $\log|G|$. No subgroup, no order of an element, no commutator. A statement whose only inputs are a cardinality and its logarithm is a counting statement, and the logarithm has one meaning in combinatorics: something doubles at each step. The number of subsequences of a length-L sequence is 2^L, and matching 2^L against n gives $L \geq \log_2 n$ for free. So the problem is asking you to show that the trivial lower bound is achievable up to a constant — that you can make the doubling *nearly* lossless. That is the "of course": once you see log, you ask what doubles.

Why $H \cup Hg$. This is the defining recursion of subsequences. Every subsequence of $a_1, \ldots, a_{k+1}$ either skips a_{k+1} or ends with it — the same split that gives $\binom{k+1}{j} = \binom{k}{j} + \binom{k}{j-1}$. Order matters in a non-abelian group, and the split puts the new term on the right, so the new products are Hg, not gH. Get this side wrong and the obstruction set comes out as HH^{-1}, which is the wrong set for Step 4.

Why $H^{-1}H$. Doubling fails exactly on a collision, and a collision is an equation $h_1 g = h_2$. Solving a group equation for g means multiplying by an inverse: $g = h_1^{-1}h_2$. So the "bad" set is not a clever object; it is the solution set of the collision equation. It is the multiplicative version of the difference set $H - H$ in additive combinatorics, and it appears for the same reason: x and $x + g$ both lie in H precisely when $g \in H - H$.

Why the greedy process is guaranteed to win. It is a dichotomy, and both branches are good. Branch one: some g avoids $H^{-1}H$, and $|H|$ doubles. This branch can occur at most $\log_2 n$ times because $|H| \leq n$. Branch two: $H^{-1}H = G$, and then the reversed-inverse trick converts $H^{-1}H$ into actual subsequence products. Kedlaya's first solution states both branches without saying why the second is acceptable. Here is why: the failure condition is itself a covering statement. The obstruction to progress ("every g is a quotient of two elements of H") is, read the other way round, a proof that H already covers G in quotients. Whenever a greedy argument's obstruction is a covering statement, you are one step from done.

Why reversed inverses. Subsequences preserve order, and $(a_{i_1} \cdots a_{i_m})^{-1} = a_{i_m}^{-1} \cdots a_{i_1}^{-1}$ reverses order. To produce H^{-1} as subsequence products you must write the inverses in

the reverse order; writing $a_1^{-1}, \ldots, a_k^{-1}$ produces something else in a non-abelian group. The parenthetical in the statement, with its $4, 4, 2$ versus $2, 2, 4$ example, is the problem flagging that order is the thing to be careful with.

The second route. Kedlaya's second solution averages over g: $\sum_g |H \cup Hg| = 2mn - m^2$ with $m = |H|$, so some g reaches at least $2m - m^2/n$, which squares the missing fraction $s = 1 - m/n$. This avoids the inverse trick at the price of an estimate, $(1 - 1/n)^{2^k} < 1/n$ for $k = \lfloor 2 \log_2 n \rfloor$. It is the same doubling idea, run probabilistically: a random g collides rarely when H is small. I prefer the first route under exam conditions because every step is an equality or a set inclusion; there is nothing to estimate.

Checked. I ran the greedy construction by computer on S_3, S_4, A_5 and $\mathbb{Z}/p$ for $p = 2, 7, 31, 97$, with random choices of $g \notin H^{-1}H$: in every run $|H| = 2^k$ exactly, $k \leq \log_2 n$, and the nonempty subsequence products of T were all of G. A_5 is included on purpose: it is the group that stopped the natural attempt.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "length at most $c \log |G|$" next to "product of some subsequence". Write down: 2^L subsequences, so $L \geq \log_2 |G|$; I must nearly double the reachable set with each term. Then write $P(S, g) = P(S) \cup P(S)g$.

2. A greedy doubling argument: extend one term at a time and measure the reachable set. The collision equation then names the obstruction set $H^{-1}H$.

3. Hesitation occurs when the greedy loop stops before $H = G$: it feels like failure. It is reinforced away by reading the stopping condition as a covering statement, $H^{-1}H = G$, and asking how to realise h_1^{-1} with subsequences. A second hesitation is abandoning the solvable-group proof; test A_5 early and the decision makes itself.

4. The trap is structure: Sylow subgroups, composition series, generators. They give a correct proof for solvable groups and nothing for simple ones, and the effort is sunk before the wall appears. Smaller traps are putting g on the wrong side (gH instead of Hg) and writing the inverses in forward order.

5. Slogan: *a bound of the form* $\log(\textit{size})$ *means "show something nearly doubles at each step"; find the one-step recursion and the exact condition for doubling.*

6. Decision rule: **in a greedy argument, write the condition under which the step fails as an equation and solve it; if the set of failures is itself a covering statement ("everything is a difference/quotient of two elements I already have"), finish by realising that difference directly.**

▷ **Cue:** a required length or number of steps of the form $c \log(\text{size})$ $\longrightarrow$ **Move:** count the configurations (2^L subsequences) to get the trivial bound, then build greedily so the reachable set doubles each step

▷ **Cue:** products of subsequences of $a_1, \ldots, a_k$ $\longrightarrow$ **Move:** use the recursion $P(S, g) = P(S) \cup P(S)g$: every subsequence either omits the last term or ends with it

▷ **Cue:** a union $H \cup Hg$ that must be disjoint $\longrightarrow$ **Move:** solve the collision $h_1 g = h_2$: disjoint exactly when $g \notin H^{-1}H$

▷ **Cue:** a greedy process stops with "every element is $h_1^{-1}h_2$" $\longrightarrow$ **Move:** prepend the inverses of the sequence in reverse order; the prefix produces H^{-1}, the suffix H

▷ **Cue:** a statement about *every* finite group, with only $|G|$ in it $\longrightarrow$ **Move:** test a non-abelian simple group such as A_5 before committing to structure; if structure gives nothing there, count

Putnam 2008 — Paper B

2008 B1

2008 B1. What is the maximum number of rational points that can lie on a circle in $\mathbb{R}^2$ whose center is not a rational point? (A *rational point* is a point both of whose coordinates are rational numbers.)

Fields: plane geometry, elementary number theory (rationality of intersections) **Position:** B1 – opener, a genuine two-line answer hiding a full existence-and-uniqueness argument **Answer:** Two. Achieved by e.g. $(0,0)$ and $(1,0)$ on any circle centered $(1/2, x)$ for irrational x; three rational points would force the center to be rational.

Layer 1 THE WHAT — surface analysis

The natural attempt. The natural first move is to write the circle's equation in the standard implicit form $x^2 + y^2 + Dx + Ey + F = 0$ and substitute three hypothetical rational points $(x_1, y_1), (x_2, y_2), (x_3, y_3)$, producing three linear equations in the unknowns D, E, F (linear because the $x_i^2 + y_i^2$ terms are known rational constants, not unknowns). This is honest and it does work: since every coefficient in the resulting 3×3 system is rational, Cramer's rule gives D, E, F as ratios of rational determinants, hence rational – provided the system's determinant is nonzero. The center $(-D/2, -E/2)$ then comes out rational, contradicting the hypothesis. The attempt stalls, however, on exactly the clause it glossed over: nothing in this route explains *why* the 3×3 determinant must be nonzero. A student who has only run Cramer's rule mechanically is left worrying about a degenerate case – what if the three points happen to make the system singular? – without a geometric reason to rule it out, and resolving that worry by direct determinant expansion is exactly the kind of heavy, unilluminating algebra the problem does not require.

The model chain. Any two distinct rational points P, Q determine a rational line in two senses at once: their midpoint $M = ((a+c)/2, (b+d)/2)$ is rational, and the slope of PQ is a ratio of rational differences, hence rational (or the line is vertical, still "rational" in the relevant sense of $ax + by = c$ with $a, b, c \in \mathbb{Q}$). The perpendicular bisector of PQ passes through the rational point M with slope $-1/\text{slope}(PQ)$, itself rational – so it too is a rational line, $ax + by = c$ with rational a, b, c. Given three rational points P, Q, R on the circle, form the perpendicular bisectors of PQ and QR: two rational lines, each of which the center must lie on (equidistant from P, Q and from Q, R respectively). Because P, Q, R lie on a genuine circle they are not collinear, so the segments PQ and QR point in different directions, and lines perpendicular to two different directions are themselves not parallel. Two nonparallel rational lines meet in exactly one point, and solving their two linear equations (rational coefficients, nonzero determinant because the lines truly intersect) gives that point's coordinates as ratios of rational determinants – rational. The center is rational, contradicting the hypothesis. Hence at most two rational points are possible, and two are achieved by, e.g., $(0,0)$ and $(1,0)$ on the circle centered at $(1/2, \sqrt{2})$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths end at the same mechanism – a unique solution of two linear

equations with rational coefficients is rational. They split on what supplies those two equations. The natural attempt manufactures them from the generic implicit circle equation, three points at once, and has no structural reason ready for why the resulting 3×3 system is nonsingular. The model manufactures them one pair of points at a time, as perpendicular bisectors, and the nonsingularity is not an algebraic fact to be checked afterward – it is a geometric fact (noncollinearity of three points on a genuine circle) visible before a single coefficient is written down.

> *I don't need three points to determine the center at once – two points already pin the center down to a line (their perpendicular bisector), so let me use the three points in two pairs, each pair giving one linear constraint, rather than folding all three into a single bigger system whose solvability I then have to justify separately.*

> *The reason these two perpendicular bisectors can't be parallel isn't an algebraic coincidence to verify by expanding a determinant – it's just that P, Q, R can't be collinear (a line meets a circle in at most two points), and perpendiculars to non-parallel lines are themselves non-parallel, so the two bisectors must cross exactly once.*

Layer 3 THE WHY

A circle's center is, by definition, the unique point equidistant from every point on the circle – so the center is genuinely determined by the equidistance condition, not merely computable from it. The perpendicular bisector of a chord PQ is exactly the locus of points equidistant from P and Q; reaching for it is not a trick borrowed from a geometry course, it is restating what "center" means, one chord at a time. This is why the model chain needs no case analysis to rule out a singular system: singularity would mean the two bisectors coincide or never meet, and a bisector is perpendicular to its chord, so two bisectors from chords PQ, QR fail to meet only if $PQ \parallel QR$ – impossible for three points on a circle, since that would make them collinear and a line cannot cross a circle in three places. The rationality itself descends from a single closed fact about $\mathbb{Q}$: it is a field, so any operation built from field operations (add, subtract, multiply, divide by a nonzero element) applied to rational data returns rational data. A perpendicular bisector's equation is built from midpoints (averages) and negative reciprocals of slopes (division) of rational points – all field operations – so it is rational by construction, and intersecting two rational lines is again solving a linear system with field-operation arithmetic. The official solution's remark that "any two nonparallel rational lines intersect at a rational point... holds with the rationals replaced by any field" names this explicitly: the whole argument is really a statement about fields, and $\mathbb{Q}$ only enters because the problem asks about $\mathbb{Q}$.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "three points determine a circle, and hence its center." On seeing it, the reflex to write down is: what linear condition does a *pair* of points, not all three at once, place on the center? – because equidistance from two points is already a full linear constraint (the perpendicular bisector), so three points give two independent linear constraints, which is exactly enough to pin down a point in the plane.
2. That cue suggests intersecting two perpendicular bisectors rather than solving a 3×3 system from the implicit circle equation: fewer unknowns, and the nonsingularity of the resulting 2×2 system is visible geometrically (non-collinearity) instead of needing to be checked algebraically.
3. Hesitation occurs at "why can't the two bisectors be parallel?" – reinforce it away by translating "parallel bisectors" back into a statement about the original chords: bisectors of PQ and QR are parallel exactly when $PQ \parallel QR$, and for three points sharing Q that only happens if P, Q, R are collinear, which a circle rules out immediately.

4. The trap is treating "rational coefficients ⇒ Cramer's rule ⇒ rational solution" as sufficient on its own, without separately confirming the determinant is nonzero – an algebraic tool applied without checking its hypothesis is holding, which the geometric route makes automatic.

5. Recognisable faster next time by treating "is this coordinate forced to be rational" problems as a search for the operations (midpoint, slope, perpendicular slope, line intersection) that stay inside $\mathbb{Q}$ by being pure field operations on rational inputs, rather than reaching first for the most general equation of the object in question.

6. Reusable rule: when a geometric point is defined as the intersection of two loci each determined by rational data, express those loci as literally as possible (fewest points/parameters per locus) so that (a) the coefficients are manifestly rational by field-closure and (b) the non-degeneracy of the intersection is visible as a geometric fact (e.g. non-collinearity, non-parallelism) rather than an algebraic one requiring separate verification.

▷ **Cue:** a point is defined as equidistant from several other points (a center, a circumcenter, etc.), and those other points are rational ⟶ **Move:** build the point as the intersection of pairwise perpendicular bisectors (each is a rational line by field-closure of midpoint and slope operations), not from the single equation of the whole family at once

▷ **Cue:** Cramer's rule or a linear system gives a candidate rational answer but its determinant's nonvanishing is unclear ⟶ **Move:** look for a geometric restatement of "the determinant is nonzero" (e.g. two lines are not parallel because two chords of the same circle are not parallel) instead of expanding the determinant algebraically

2008 B2

2008 B2. Let $F_0(x) = \ln x$. For $n \geq 0$ and $x > 0$, let $F_{n+1}(x) = \int_0^x F_n(t)\,dt$. Evaluate

$$\lim_{n\to\infty} \frac{n!F_n(1)}{\ln n}.$$

Fields: calculus (repeated integration, harmonic-series asymptotics) **Position:** B2 – second problem, a repeated-integration setup that rewards guessing a closed form before computing **Answer:** -1, via the closed form $F_n(x) = (\ln x - H_n)x^n/n!$ where $H_n = \sum_{k=1}^n 1/k$, giving $n!F_n(1) = -H_n \sim -\ln n$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The definition is a recursion, so the honest first move is to just run it: $F_1(x) = \int_0^x \ln t\,dt = x\ln x - x$ (integration by parts, with $t\ln t \to 0$ as $t \to 0^+$ handling the boundary term). Then $F_2(x) = \int_0^x (t\ln t - t)\,dt = \left(\frac{x^2}{2}\ln x - \frac{x^2}{4}\right) - \frac{x^2}{2} = \frac{x^2}{2}\ln x - \frac{3x^2}{4}$, another integration by parts on top of the previous result. Evaluating at $x = 1$: $F_1(1) = -1$, $F_2(1) = -3/4$, so $2!F_2(1) = -3/2$. The numbers $-1, -3/2$ are suggestive – they are -1 and $-(1 + 1/2)$ – but confirming that pattern honestly requires computing $F_3(1)$ by hand, then $F_4(1)$, each one built from an integration by parts on the previous *specific* polynomial-times-log expression, with the algebra growing by one term of bookkeeping each time. This works, but it does not scale: nothing about the third or fourth computation is conceptually different from the first two, so time is being spent re-deriving the same recursive relationship in increasingly cluttered concrete form, rather than proving it once, in general, for all n.

The model chain. Guess the closed form directly from the shape of F_1: it is $(\ln x - c_1)\cdot x^1/1!$ with $c_1 = 1$, and this "$(\ln x - \text{constant}) \times x^n/n!$" shape is exactly what a repeated antiderivative of $\ln x$ should look like, since differentiating $x^n \ln x$ produces $x^{n-1}\ln x$ plus lower-order terms and division by $n!$ is the natural normalisation for repeated integration. Conjecture $F_n(x) = (\ln x - H_n)x^n/n!$ where $H_n = \sum_{k=1}^n 1/k$ (with $H_0 = 0$), and prove it by induction. Base case

$n = 0$: $(\ln x - 0)x^0/0! = \ln x = F_0(x)$. Inductive step: compute $\int_0^x t^n \ln t\, dt = \frac{x^{n+1}}{n+1}\ln x - \frac{x^{n+1}}{(n+1)^2}$ once, in general (the boundary term at $t = 0$ vanishes since $t^{n+1}\ln t \to 0$), so

$$\int_0^x (\ln t - H_n)\frac{t^n}{n!}\,dt = \frac{x^{n+1}}{n!}\left[\frac{\ln x}{n+1} - \frac{1}{(n+1)^2} - \frac{H_n}{n+1}\right] = \frac{x^{n+1}}{(n+1)!}\left[\ln x - H_n - \frac{1}{n+1}\right],$$

which is exactly $(\ln x - H_{n+1})x^{n+1}/(n+1)!$ since $H_{n+1} = H_n + 1/(n+1)$. This single general computation proves the closed form for every n at once. Then $F_n(1) = (0 - H_n)/n! = -H_n/n!$, so $n!F_n(1) = -H_n$, and since $H_n = \ln n + \gamma + o(1)$ (standard: $\sum_{k=2}^n 1/k$ is sandwiched between $\int_2^n dx/x = \ln n - \ln 2$ and $\int_1^n dx/x = \ln n$, so $H_n/\ln n \to 1$), the limit is -1.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths correctly compute F_1 and F_2 by the identical integration-by-parts mechanism and both notice $-1, -3/2$ look like $-H_1, -H_2$. They split immediately after that observation: the natural attempt treats it as a pattern to be *confirmed* by computing $F_3, F_4, \ldots$ one concrete case at a time; the model treats it as a conjecture about the *general term* to be proved once by abstracting the specific computation already done for F_1 into an induction step that works for arbitrary n.

> *I already have the general-purpose piece of algebra I need – the integration by parts I just did on $t\ln t$ works exactly the same way on $t^n \ln t$ for any n, since only the exponent changes, not the mechanism – so instead of running it again on the next specific number, I should run it once with n left as a variable and turn today's computation into tomorrow's proof.*

> *The two numbers -1 and $-3/2$ are not just numbers to pattern-match against a lookup table of famous constants – they are partial sums of $1/k$, and recognising $1, 1+1/2, 1+1/2+1/3, \ldots$ as* the *harmonic numbers is what converts "the answer involves $\ln n$" from a mysterious coincidence in the problem statement into an expected consequence, since the harmonic numbers are the one everyday sequence whose growth rate is $\ln n$.*

Layer 3 THE WHY

The recursive definition $F_{n+1} = \int_0^x F_n$ is, structurally, "apply the same operator $n+1$ times to $\ln x$" – and whenever a problem asks about the n-fold iterate of a fixed operation, the productive question is never "what is the value after 1, 2, 3 iterations" but "what property is preserved, or accumulated, by one application of the operator, expressed in terms that make sense for arbitrary n." Integration by parts on $t^n \ln t$ is forced to look the same for every n because $\int t^n\,dt = t^{n+1}/(n+1)$ is a single formula valid for all $n \geq 0$, not a family of unrelated facts one per exponent – so there was never a genuine reason to expect the $n = 1$ and $n = 2$ computations to be different in kind, only in which number gets substituted into the same formula. This is exactly why guessing the closed form after one example is legitimate rather than reckless: the guess is that the recursion's action on the "$-H_n$" term mirrors its action on the $x^n/n!$ term, both driven by the same single antiderivative rule, and induction is the formal device for confirming that a pattern which holds for one application of a fixed operator holds under repeated application. The appearance of $\ln n$ in the problem statement is itself the tell: it is asking, in effect, "does $n!F_n(1)$ grow like $\ln n$", and since $H_n \sim \ln n$ is the single most standard asymptotic fact about an accumulating sum of reciprocals, the problem is really testing whether the closed form is found at all – once it is, the asymptotics are close to automatic, which is consistent with Kedlaya and Ng's solution stating the closed form as a "claim" verified by one line of induction and spending the rest of the write-up entirely on the $H_n \sim \ln n$ estimate.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the shape of $F_1(x) = x\ln x - x = (\ln x - 1)\cdot x$. On seeing it, write down explicitly: "this is (log term − constant) times $x^n/n!$" – naming the ansatz in words the moment it appears, rather than letting it sit as an unlabelled numerical coincidence.

2. That cue suggests proof by induction on a conjectured closed form, using one general integration by parts on $t^n \ln t$ as the inductive step, instead of repeated concrete integration by parts for successive specific n.

3. Hesitation occurs at trusting a pattern seen in only two data points (F_1, F_2) enough to conjecture it for all n. Reinforce it away by checking that the mechanism producing F_2 from F_1 (one integration by parts, one boundary term vanishing at 0) is manifestly the same mechanism that would produce F_3 from F_2 – if the *procedure* doesn't change with n, the pattern in its *outputs* is trustworthy to conjecture and then prove, not just observe.

4. The trap is treating each F_n as a fresh computation to be done from scratch, which is technically correct but produces answers only for the n actually computed – never establishing the general asymptotic behaviour the problem (dividing by $\ln n$, taking $n \to \infty$) actually requires.

5. Recognisable faster next time by treating "evaluate a limit involving the n-th iterate of a fixed operator" as an automatic cue to conjecture a closed form from the smallest case(s) and prove it by induction, rather than computing terms until the answer is "obvious" numerically.

6. Reusable rule: when a quantity is defined by repeatedly applying the same operation (integration, differentiation, a recursive substitution), find the closed form after the fewest examples needed to see the shape, prove it by induction using one general instance of the underlying computation, and only then take $n \to \infty$ – never chase the limit through concrete term-by-term computation.

▷ **Cue:** a sequence is defined by iterating a fixed operator (here, indefinite integration from 0) applied to itself repeatedly $\longrightarrow$ **Move:** conjecture a closed form for the general term from the first one or two explicit computations, then prove it by induction using a single general instance of the underlying computation (e.g. one integration by parts with exponent n left as a variable) rather than repeating the concrete computation for each new n

▷ **Cue:** a limit is stated relative to $\ln n$ (e.g. dividing by $\ln n$, or the problem visibly invites comparison to $\ln n$) $\longrightarrow$ **Move:** suspect the harmonic numbers $H_n = \sum 1/k$ are hiding in the quantity being studied, since $H_n \sim \ln n$ is the standard source of logarithmic growth from a discrete accumulation

2008 B3

2008 B3. What is the largest possible radius of a circle contained in a 4-dimensional hypercube of side length 1?

Fields: geometry (high-dimensional optimisation via per-coordinate constraints) **Position:** B3 – a genuine "hard" B-paper geometry problem, testing whether a 2-dimensional object is allowed to use all 4 available dimensions **Answer:** $\frac{\sqrt{2}}{2}$, achieved by the circle through $u = (1/2, 1/2, 0, 0)$ and $w = (0, 0, 1/2, 1/2)$ centred at the hypercube's centre.

Layer 1 THE WHAT — surface analysis

The natural attempt. The reflex, carried over from the familiar fact that the largest circle inscribed in a unit square has radius $1/2$, is to look for the largest circle lying flat inside a single coordinate 2-plane of the hypercube – say the plane where the last two coordinates are frozen at $1/2$. Restricted to that plane, the hypercube's cross-section is an ordinary unit square, and the largest

circle inside it has radius $1/2$. This is a completely valid construction: such a circle genuinely fits inside the hypercube, so $1/2$ is a legitimate lower bound on the answer. But it treats two of the four available dimensions as irrelevant scenery – the circle never uses any of the room the third and fourth coordinates have to offer, because it is confined to a 2-dimensional slice that happens to coincide with two of the axes. Nothing in the problem says the optimal circle has to respect the coordinate axes at all, and without a reason to believe $1/2$ is optimal, this natural attempt has produced a bound, not an answer.

The model chain. First reduce to the case where the circle is centred at the hypercube's own centre: if a circle C with centre P lies in the cube H, then, because a circle is centrally symmetric about its own centre, C also lies inside the reflection of H through P. So C lies in $H \cap \text{reflect}_P(H)$, a box centred at P whose side length in each coordinate direction is at most 1 (the side length of H itself); translating this box, and C with it, so that P moves to the centre of H, keeps C inside H (a box of side ≤ 1 centred at H's centre automatically fits inside H, which has side exactly 1 there). So it suffices to bound the radius of a circle centred at the hypercube's centre, taken to be the origin, with the cube $[-1/2, 1/2]^4$. Parametrise such a circle by two orthogonal vectors $u, w \in \mathbb{R}^4$ with $|u| = |w| = r$, so the circle's points are $u\cos\theta + w\sin\theta$. Containment in the cube means, coordinate by coordinate, $|u_i\cos\theta + w_i\sin\theta| \leq 1/2$ for all θ; the maximum over θ of $u_i\cos\theta + w_i\sin\theta$ is $\sqrt{u_i^2 + w_i^2}$, so the constraint is exactly $u_i^2 + w_i^2 \leq 1/4$ for each of the four coordinates $i = 1, \ldots, 4$. Summing these four constraints, $\sum_i (u_i^2 + w_i^2) = |u|^2 + |w|^2 = 2r^2 \leq 4 \cdot \frac{1}{4} = 1$, so $r \leq 1/\sqrt{2} = \sqrt{2}/2$. Equality is reached by taking $u = (1/2, 1/2, 0, 0)$, $w = (0, 0, 1/2, 1/2)$: each coordinate has $u_i^2 + w_i^2 = 1/4$ exactly, and $u \cdot w = 0$, $|u| = |w| = \sqrt{2}/2$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths correctly identify that the four coordinate constraints are what limit the circle. They split on how many of those four constraints the chosen circle actually uses to the full: at (†), the natural attempt's coordinate-plane circle has $u = (1/2, 0, 0, 0)$, $w = (0, 1/2, 0, 0)$, which makes $u_3^2 + w_3^2 = 0$ and $u_4^2 + w_4^2 = 0$ – two of the four per-coordinate budgets of $1/4$ sit entirely unspent. The model chain instead searches for a u, w that makes every one of the four constraints $u_i^2 + w_i^2 \leq 1/4$ tight simultaneously.

> *The bound I'm fighting against is* $u_i^2 + w_i^2 \leq 1/4$ per coordinate*, and I have four coordinates to spend that allowance in – if my circle only occupies two of them, the other two constraints are sitting there doing nothing for me, so the true optimum should draw equally on all four, not confine itself to a coordinate 2-plane just because that is the shape I know how to visualise.*

> *Before optimising anything, I should ask whether the optimal circle even has to be centred at the cube's centre – and it does, because a circle is symmetric under reflection through its own centre, so an off-centre circle is automatically contained in the reflected copy of the cube too, and that intersection can always be recentred without shrinking it.*

Layer 3 THE WHY

The hypercube is a product of four independent intervals, so containment inside it decomposes, with no loss of information, into four independent one-dimensional constraints – this is exactly what it means for a region to be a Cartesian product, and it is the reason $2r^2 = \sum_i (u_i^2 + w_i^2)$ can be bounded by summing four separate, per-coordinate caps rather than needing some genuinely 4-dimensional argument. Once containment is seen as a resource of "1/4 per coordinate, four coordinates," the radius-squared $2r^2$ is manifestly a sum being maximised subject to four independent upper bounds, and elementary optimisation says a sum of bounded nonnegative quantities is maximised by making every bound tight at once,

not by spending the whole budget in two of the four slots and leaving the rest at zero. The reason the coordinate-plane circle looks natural is that it is the answer to a different, easier question – the largest circle inside a 2-dimensional square – and that question only ever had two coordinates to work with in the first place; importing its answer unexamined into four dimensions silently discards the extra room the higher-dimensional problem actually offers. The centring reduction has its own axiom underneath it: a circle is, by definition, the set of points at fixed distance from its centre, hence symmetric under the point-reflection through that centre – so "does this circle fit in the cube" is invariant under simultaneously reflecting the circle and the cube through the circle's own centre, and that invariance is exactly what licenses sliding the whole picture so the two centres coincide. Kedlaya and Ng's solution states this centring step in a single sentence before the real optimisation begins; the reason it can be dispatched so quickly is that centrality-of-a-circle is not a geometric coincidence about circles specifically, it is what distinguishes a circle from an arbitrary closed curve.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "2-dimensional object inside a 4-dimensional box." On seeing it, write down immediately: containment will decompose into one constraint per coordinate (since the box is a product of intervals), so there are four separate budgets to spend, not one combined geometric picture to visualise directly.

2. That cue suggests parametrising the circle by two orthogonal radius vectors u, w and converting the containment condition into the per-coordinate inequality $u_i^2 + w_i^2 \leq (\text{half-side})^2$, then treating $2r^2 = \sum_i (u_i^2 + w_i^2)$ as a quantity to maximise subject to those four independent caps.

3. Hesitation occurs at trusting the familiar 2D answer ($1/2$, from the coordinate-plane circle) is optimal in 4D. Reinforce it away by checking, for that candidate, whether every one of the four per-coordinate constraints is tight – here two of them are slack ($= 0 \neq 1/4$), which is the signal that dimensions are being wasted and a better configuration exists.

4. The trap is defaulting to axis-aligned, coordinate-plane objects as "the" natural embedding of a lower-dimensional shape into a higher-dimensional box, when the box's symmetry has no preference for which 2 of the 4 coordinates carry the shape – the optimum instead spreads the shape's extent across every available coordinate.

5. Recognisable faster next time by asking, for any "largest k-dimensional object inside an n-dimensional box" problem with $k < n$, whether the proposed configuration uses all n coordinates' worth of room, and treating an unused (zero) constraint as immediate evidence the configuration is not yet optimal.

6. Reusable rule: when embedding a k-dimensional object into an n-dimensional box with $k < n$, a sum-of-squares quantity like the object's own squared size decomposes into n independent per-coordinate constraints; the optimal embedding makes every one of those n constraints tight simultaneously, so search for configurations that spread the object's extent equally across all n coordinates rather than confining it to a k-dimensional coordinate subspace.

▷ **Cue:** a lower-dimensional object (circle, ball) must be fit inside a higher-dimensional axis-aligned box $\longrightarrow$ **Move:** decompose containment coordinate-by-coordinate (the box is a product of intervals), express the object's squared size as a sum over all box-coordinates, and maximise by making every per-coordinate constraint tight rather than confining the object to a coordinate subspace matching its own dimension

▷ **Cue:** an object to be optimised (here, a circle) has its own internal symmetry (central symmetry about its centre) and the containing region is also symmetric (a box about its own centre) $\longrightarrow$ **Move:** recentre so the object's centre coincides with the region's centre; the object's own point-symmetry guarantees this recentring cannot shrink the maximal achievable size

2008 B4

2008 B4. Let p be a prime number. Let $h(x)$ be a polynomial with integer coefficients such that $h(0), h(1), \ldots, h(p^2-1)$ are distinct modulo p^2. Show that $h(0), h(1), \ldots, h(p^3-1)$ are distinct modulo p^3.

Fields: number theory; polynomials modulo prime powers; Taylor expansion over $\mathbb{Z}$ (Hensel-type lifting) **Position:** B4 — middle of the paper; short once the right expansion is written, but a student who does not extract the derivative condition from the hypothesis is stuck at the last line **Answer:** Proof. The hypothesis forces $h'(x) \not\equiv 0 \pmod{p}$ for every integer x, and that condition lifts every residue class mod p^2 to exactly p distinct classes mod p^3.

Layer 1 THE WHAT — surface analysis

The natural attempt. The natural move is to argue by contradiction on a collision. Suppose $0 \le a, b \le p^3-1$ and $h(a) \equiv h(b) \pmod{p^3}$. Reduce modulo p^2: $h(a) \equiv h(b) \pmod{p^2}$. Since h has integer coefficients, $h(x) \bmod p^2$ depends only on $x \bmod p^2$, and the hypothesis says h is injective on residues mod p^2; so $a \equiv b \pmod{p^2}$. Write $b = a + tp^2$ with $|t| \le p-1$. The goal is now $t = 0$.

Expand. For a single monomial, $(a+tp^2)^k = a^k + k\,a^{k-1}tp^2 + (\text{terms divisible by } p^4)$, so summing over the monomials of h,

$$h(a+tp^2) \equiv h(a) + tp^2\,h'(a) \pmod{p^3}.$$

The collision says the left side is $\equiv h(a)$, so $p^3 \mid tp^2h'(a)$, i.e. $p \mid t\,h'(a)$.

And here the student stops. If $p \nmid h'(a)$ then $p \mid t$, so $t = 0$ and we are done. But why should $h'(a)$ be a unit mod p? The hypothesis was about *values* of h; it has already been spent (in the step $a \equiv b \pmod{p^2}$) and says nothing visible about h'. A common next move is to argue "h is a bijection mod p, so its derivative is nonzero mod p" — which is false: $h(x) = x^p$ permutes $\mathbb{Z}/p$ by Fermat, but $h'(x) = px^{p-1} \equiv 0$. Another is to give up on the collision and try to count, which leads nowhere because counting cannot see the derivative either. This is the wall: the proof is one fact short, and the fact is not where the student is looking.

The model chain. **Step 0: an integer Taylor expansion.** Write $h(x) = \sum_k c_k x^k$ with $c_k \in \mathbb{Z}$. By the binomial theorem, for integers x, y,

$$h(x+y) = \sum_k c_k \sum_i \binom{k}{i} x^{k-i}y^i = h(x) + y\,h'(x) + y^2R(x,y),$$

where $R(x,y) = \sum_k c_k \sum_{i\ge 2} \binom{k}{i} x^{k-i}y^{i-2}$ has integer coefficients. (Equivalently, $h^{(i)}(x)/i! = \sum_k c_k \binom{k}{i} x^{k-i}$ is an integer polynomial.) Hence, for integers x, y and any $m \ge 1$,

$$h(x+yp^m) \equiv h(x) + yp^m\,h'(x) \pmod{p^{2m}}. \tag{T}$$

Step 1: read the derivative off the hypothesis. Take $m = 1$, $y = 1$ in (T): for each $x \in \{0, \ldots, p-1\}$,

$$h(x+p) \equiv h(x) + p\,h'(x) \pmod{p^2}.$$

Both x and $x+p$ lie in $\{0, \ldots, p^2-1\}$ (as $2p-1 \le p^2-1$) and are different, so by hypothesis $h(x+p) \not\equiv h(x) \pmod{p^2}$. Therefore $p\,h'(x) \not\equiv 0 \pmod{p^2}$, i.e. $h'(x) \not\equiv 0 \pmod{p}$. Since h' has integer coefficients, $h'(x) \bmod p$ depends only on $x \bmod p$; so $h'(x) \not\equiv 0 \pmod{p}$ for *every* integer x.

Step 2: lift. Every $n \in \{0, \ldots, p^3-1\}$ is uniquely $n = x + yp^2$ with $x \in \{0, \ldots, p^2-1\}$, $y \in \{0, \ldots, p-1\}$. Take $m = 2$ in (T) (so $p^{2m} = p^4$, a fortiori mod p^3):

$$h(x+yp^2) \equiv h(x) + p^2\,y\,h'(x) \pmod{p^3}.$$

Fix x. As y runs over $0, \ldots, p-1$, the numbers $y\,h'(x)$ run over all residues mod p (multiplication by the unit $h'(x)$ permutes $\mathbb{Z}/p$), so $h(x+yp^2)$ runs over the p distinct residues mod p^3 that are $\equiv h(x) \pmod{p^2}$.

Step 3: count. Different x give different $h(x)$ mod p^2 by hypothesis, so their families of lifts are disjoint. That is $p^2 \cdot p = p^3$ pairwise distinct residues mod p^3: $h(0), \ldots, h(p^3-1)$ are distinct mod p^3. □

Caveat. Not needed for the proof, which is complete as written; recorded as a check. A brute-force search over 2244 random integer polynomials of degree at most 5 that permute $\mathbb{Z}/p^2$ for $p \in \{2,3,5\}$ found every one permuting $\mathbb{Z}/p^3$; and x^p permutes $\mathbb{Z}/p$ but not $\mathbb{Z}/p^2$ for $p = 2,3,5$, confirming the trap described below.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains write the same expansion $h(a+tp^2) \equiv h(a) + tp^2 h'(a) \pmod{p^3}$ and both arrive at the same need: $h'(a) \not\equiv 0 \pmod{p}$. They split at the question "where does that come from?"

(N) Look for it in the *conclusion's* world — mod p bijectivity, counting, or the specific collision a, b. None of these contains derivative information.

(M) Run the *same expansion one level down*, with step p instead of p^2, and feed it the hypothesis: $h(x+p) \equiv h(x) + ph'(x) \pmod{p^2}$, and $h(x+p) \neq h(x) \bmod p^2$, so $h'(x)$ is a unit mod p.

> *Going from p^2 to p^3 I needed h' to be nonzero mod p. The hypothesis is the same kind of statement one level down, from p to p^2. So whatever the step from p^2 to p^3* needs, *the step from p to p^2 must* supply. *Let me write the same expansion with step p and see what distinctness mod p^2 tells me.*

> *Moving x by a multiple of p^m changes $h(x)$ by that multiple times $h'(x)$, up to p^{2m}. So h' mod p is exactly the thing that decides whether the p points above a residue get spread out or collapsed.*

- ▷ **Cue:** a polynomial with integer coefficients evaluated at $x +$ (multiple of p^m) $\longrightarrow$ **Move:** expand $h(x+y) = h(x) + yh'(x) + y^2R(x,y)$ with R integral; mod p^{2m} only the linear term survives
- ▷ **Cue:** a hypothesis mod p^j and a conclusion mod p^{j+1} about the same polynomial $\longrightarrow$ **Move:** find what the lift from p^j to p^{j+1} needs, then extract it from the hypothesis by running the same lift one level lower
- ▷ **Cue:** a map on $\{0, \ldots, p^3-1\}$ to prove injective mod p^3 $\longrightarrow$ **Move:** write $n = x + yp^2$ and show each fibre over a residue mod p^2 is spread over all p lifts; then count $p^2 \cdot p$

Layer 3 THE WHY

Why is "run the expansion one level down" the natural move rather than a lucky one? Start from what the problem is *shaped* like. It gives injectivity at level p^2 and asks for it at level p^3. That is a lifting statement, and a lifting statement has one moving part: how h behaves on the p integers $x, x+p^2, \ldots, x+(p-1)p^2$ that all reduce to the same residue mod p^2. Along such a fibre h is linear to first order — that is the content of (T), and (T) is nothing more than the binomial theorem with the observation that $(p^2)^2$ is already 0 mod p^3. A linear map $y \mapsto h(x) + p^2 y h'(x)$ on $\mathbb{Z}/p$ is a bijection onto the fibre exactly when its slope $h'(x)$ is a unit. So the *entire* content of going from p^2 to p^3 is "h' is never 0 mod p". Nothing else can matter.

Now look at the hypothesis through the same lens. Injectivity mod p^2 is itself a lift, from p to p^2, and along the fibres $x, x+p, \ldots$ the same linearisation holds (T with $m=1$: $p^2 = (p^1)^2$ is

exactly the modulus). So injectivity mod p^2 says two things: h permutes residues mod p, *and* h' is never 0 mod p. The student in (N) used only the first of these. That is why the hypothesis felt spent: half of it was never opened.

Kedlaya and Ng's solution begins by quoting Taylor's theorem with $h^{(i)}(x)/i!$ integral and then deduces $h'(x) \not\equiv 0$ from x and $x+p$. Here is where that comes from: it is the observation that the problem is a two-level instance of a single phenomenon, and the lower level is given to you. Their closing remark says exactly this in general form — for $d, e > 1$, h permutes $\mathbb{Z}/p^d$ iff it permutes $\mathbb{Z}/p^e$ — and names Hensel's lemma. The level p^1 is excluded precisely because at level p there is no fibre structure yet, so no derivative information: x^p shows it.

One more why: why is the step from p to p^2 enough to control *all* x, not only $x \in \{0, \dots, p-1\}$? Because h' has integer coefficients, so $h'(x) \bmod p$ is a function of $x \bmod p$. Checking the p residues checks everything. The proof uses this fact twice (for $h \bmod p^2$ and for $h' \bmod p$); it is the reason polynomial problems mod p^k are finite problems at all.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that hypothesis and conclusion are the same statement at consecutive prime-power levels. On seeing it, write the fibre: every $n < p^3$ is $x + yp^2$ with $x < p^2$, $y < p$, and ask how h moves along y.
2. The cue suggests linearisation along the fibre: $h(x+yp^2) \equiv h(x) + yp^2 h'(x) \pmod{p^3}$, i.e. a Hensel-type lifting argument driven by the unit-ness of h'.
3. Hesitation occurs at "I need $h'(x) \not\equiv 0 \pmod p$ but the hypothesis is about values". It is reinforced away by remembering that the hypothesis is itself a lift and contains derivative information; apply the same expansion to x and $x+p$.
4. The trap is believing that bijectivity mod p forces $h' \not\equiv 0 \pmod p$. It does not: x^p permutes $\mathbb{Z}/p$ and has $h' \equiv 0$; it fails to permute $\mathbb{Z}/p^2$. The problem starts at p^2 for exactly this reason. A smaller trap is writing Taylor's theorem with $h^{(i)}/i!$ and forgetting to justify that these are integer-valued.
5. Recognise it faster by treating any "true mod $p^j \Rightarrow$ true mod p^{j+1}" problem as a question about the p points above one residue, and asking what first-order quantity governs how those points spread.
6. Reusable decision rule: *when a proof needs a fact the hypothesis does not visibly state, run the argument you are making one level lower and let the hypothesis answer the same question there.* In inductive structures (prime powers, dimensions, depths of a tree) the hypothesis at level j is usually the lift from $j-1$, and it carries the same auxiliary information the lift to $j+1$ needs.

2008 B5

2008 B5. Find all continuously differentiable functions $f : \mathbb{R} \to \mathbb{R}$ such that for every rational number q, the number $f(q)$ is rational and has the same denominator as q. (The denominator of a rational number q is the unique positive integer b such that $q = a/b$ for some integer a with $\gcd(a,b) = 1$.) (Note: gcd means greatest common divisor.)

Fields: real analysis (mean value theorem, continuity of f'); elementary number theory (denominators, gcd) **Position:** B5 — the answer is guessed in a minute; the difficulty is converting a pointwise arithmetic condition into information about f' without getting tangled in gcds **Answer:** $f(x) = x + n$ or $f(x) = -x + n$ with $n \in \mathbb{Z}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. First the guess. $f(q)$ must have denominator b when q does; translations by integers and $x \mapsto -x$ preserve denominators, since $\gcd(a+nb, b) = \gcd(\pm a, b) = 1$. So $f(x) = \pm x + n$ ($n \in \mathbb{Z}$) work, and the task is to prove nothing else does.

The natural way to bring in differentiability is to fix a rational $q = a/b$ (lowest terms, $b > 0$) and approach it through rationals whose denominators are controlled. The obvious approximants are

$$q_n = \frac{a}{b} + \frac{1}{bn} = \frac{an+1}{bn} \qquad (n \geq 1).$$

The student reasons: "q_n has denominator bn, so $bn\, f(q_n)$ is an integer; also $bn\, f(q) = n \cdot bf(q)$ is an integer; so the difference quotient

$$\frac{f(q_n) - f(q)}{q_n - q} = bn\big(f(q_n) - f(q)\big)$$

is an integer, and it tends to $f'(q)$. A convergent sequence of integers is eventually constant, so $f'(q) = c \in \mathbb{Z}$ and $f(q_n) = f(q) + \frac{c}{bn}$ for large n." Then, to pin c down: "$f(q_n)$ must have denominator bn; if $c = 0$ it has denominator b; if $|c| \geq 2$, take $n = k|c|$, and $f(q) + \frac{c}{bn} = f(q) \pm \frac{1}{bk}$ has denominator dividing $bk < bn$. So $c = \pm 1$."

This is essentially the official solution, and it contains a false sentence: q_n need *not* have denominator bn. With $q = \frac{1}{3}$ and $n = 2$, $q_n = \frac{3}{6} = \frac{1}{2}$. In general $\gcd(an+1, n) = 1$ but $\gcd(an+1, b)$ can exceed 1. The integrality of the difference quotient survives (the true denominator divides bn), but the final step — "$f(q_n)$ must have denominator exactly bn" — is only valid for n with $\gcd(an+1, b) = 1$, e.g. n a multiple of b. The repair works (see Layer 3), but the student is now doing gcd bookkeeping at a point, which is exactly where marks leak.

The model chain. **Step 1: f' is integer-valued on a dense set.** Let $p \geq 3$ be prime and k an integer with $p \nmid k$ and $p \nmid k+1$. Then $\frac{k}{p}$ and $\frac{k+1}{p}$ both have denominator p, so $f(\frac{k}{p}) = \frac{A}{p}$ and $f(\frac{k+1}{p}) = \frac{B}{p}$ with $A, B \in \mathbb{Z}$. By the mean value theorem there is $\xi \in (\frac{k}{p}, \frac{k+1}{p})$ with

$$f'(\xi) = \frac{f(\frac{k+1}{p}) - f(\frac{k}{p})}{1/p} = B - A \in \mathbb{Z}.$$

Given any open interval (u, v), choose a prime $p \geq 3$ with $\frac{4}{p} < v - u$. Then (u, v) contains three consecutive points $\frac{k}{p}, \frac{k+1}{p}, \frac{k+2}{p}$. At most one of $k, k+1, k+2$ is divisible by p, so one of the pairs $(k, k+1)$, $(k+1, k+2)$ avoids multiples of p, and the corresponding ξ lies in (u, v). So the set $E = \{\xi : f'(\xi) \in \mathbb{Z}\}$ is dense in $\mathbb{R}$.

Step 2: f is linear with integer slope. Take any $x \in \mathbb{R}$ and $\xi_j \in E$ with $\xi_j \to x$. By continuity of f', $f'(\xi_j) \to f'(x)$; a convergent sequence of integers is eventually constant, so $f'(x) \in \mathbb{Z}$. Thus f' is a continuous function $\mathbb{R} \to \mathbb{Z}$. By the intermediate value theorem it is constant: if $f'(x_1) \neq f'(x_2)$, f' would take a non-integer value between them. So $f' \equiv c \in \mathbb{Z}$ and $f(x) = cx + d$ with $d = f(0)$.

Step 3: identify c and d. $0 = \frac{0}{1}$ has denominator 1, so $d = f(0) \in \mathbb{Z}$.

- If $c = 0$: $f(\frac{1}{2}) = d$ has denominator $1 \neq 2$. Impossible.
- If $|c| \geq 2$: $f(\frac{1}{|c|}) = \pm 1 + d$ has denominator $1 \neq |c|$. Impossible.

So $c = \pm 1$ and $f(x) = \pm x + d$, $d \in \mathbb{Z}$.

Step 4: these work. For $q = \frac{a}{b}$ in lowest terms, $f(q) = \frac{\pm a + db}{b}$ and $\gcd(\pm a + db, b) = \gcd(a, b) = 1$. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree that differentiability must enter through a difference quotient of rationals, and that the quotient will be an integer if the denominators cooperate. They split at the choice of the *pair* of rationals.

(N) One endpoint fixed at $q = \frac{a}{b}$, the other a nearby rational q_n with a *different, larger* denominator. The arithmetic now depends on $\gcd(an+1, b)$, and the argument is local: it learns $f'(q)$ at one point at a time.

(M) Two neighbours $\frac{k}{p}, \frac{k+1}{p}$ with the *same* prime denominator. The quotient is $p \cdot \frac{B-A}{p} = B - A$, an integer with no gcd in sight; the mean value theorem converts it into an integer value of f' somewhere, and continuity of f' does the rest.

> *If both points have the same denominator p and sit $1/p$ apart, the difference of the values is (integer)$/p$ and the gap is $1/p$, so the slope is an integer, full stop. The mean value theorem turns that slope into a value of f'. I don't need to know which point it is — continuity of f' will spread it.*

> *Once f' is integer-valued and continuous it is constant, so f is a line $cx + d$. After that I only need two test inputs: $x = 0$ for d, and $x = 1/|c|$ to kill $|c| \geq 2$.*

▷ **Cue:** a condition on f at rationals plus "f is C^1" $\longrightarrow$ **Move:** form difference quotients between well-chosen rationals and let the mean value theorem convert them into values of f'

▷ **Cue:** a continuous function that takes integer values on a dense set $\longrightarrow$ **Move:** it takes integer values everywhere, and a continuous integer-valued function on $\mathbb{R}$ is constant (IVT)

▷ **Cue:** you need a difference quotient to be an integer $\longrightarrow$ **Move:** choose the two points with a common denominator D and spacing $1/D$

▷ **Cue:** a functional equation reduced to $f(x) = cx + d$ $\longrightarrow$ **Move:** stop proving and test: plug in the one or two inputs that make each wrong (c, d) fail

Layer 3 THE WHY

Why is "same denominator, adjacent" the natural pair and not a trick? Go back to the definition of the derivative: it is a limit of $(f(y) - f(x))/(y - x)$. The problem controls numerators and denominators of $f(x)$ and $f(y)$. The quotient's denominator $y - x$ is the thing you get to choose, and its numerator $f(y) - f(x)$ is what the hypothesis constrains. If x and y have denominators $b_1 \neq b_2$, then $f(y) - f(x)$ lives in $\frac{1}{\operatorname{lcm}(b_1,b_2)}\mathbb{Z}$ and $y - x$ is some fraction whose reduced form depends on gcds — two unknown quantities fighting. If x, y both have denominator D and $y - x = \frac{1}{D}$, the numerator lives in $\frac{1}{D}\mathbb{Z}$ and the denominator is exactly $\frac{1}{D}$: they cancel by construction. The choice is dictated by asking *which pair makes the hypothesis and the quotient speak the same unit*.

Why is it then legitimate to lose control of *where* the integer derivative is? Because the problem gave you more than differentiability: it gave *continuous* differentiability. That word is not decoration; it is the licence to learn f' on a dense set and extend. Without it the local approach (N) is forced, because a merely differentiable f could have f' jump. When a problem says C^1, it is telling you that "f' at some point in each small interval" is as good as "f' everywhere" — which is precisely what the mean value theorem produces.

Why must the endgame be two test values? Once f' is a constant integer c, the problem has become a question about the family $cx + d$, and the arithmetic condition, applied to a single well-chosen rational, eliminates each wrong member. $x = \frac{1}{|c|}$ is chosen so that cx becomes an integer and the output's denominator collapses to 1; $x = \frac{1}{2}$ does the same when $c = 0$.

On the official solution. Kedlaya and Ng fix $q = a/b$ and use $q_n = \frac{an+1}{bn}$, asserting that $f(q_n)$ "has denominator bn". That is true only when $\gcd(an + 1, b) = 1$; for $q = \frac{1}{3}$, $n = 2$ it fails. The argument is repaired by taking $n = bm$: then $an + 1 = abm + 1 \equiv 1 \pmod{b}$ and $\gcd(abm + 1, m) = 1$, so $q_n = \frac{abm+1}{b^2m}$ is in lowest terms. The difference quotient is again an integer, eventually a constant $c = f'(q)$, and $f(q_n) = f(q) + \frac{c}{b^2m} = \frac{Abm+c}{b^2m}$ (where $f(q) = \frac{A}{b}$)

must be in lowest terms for all large m. If $c = 0$, bm divides both numerator and denominator; if a prime $\ell \mid c$, choose m divisible by ℓ and ℓ divides both. So $c = \pm 1$. Their closing remark ("f' is a continuous function from the rationals to the integers, so must be constant") is the model chain in embryo. The archive's point: the global mean-value route avoids the gcd question entirely, which is why it is the one to reach for.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "continuously differentiable" sitting next to a condition on rational inputs. On seeing it, write a difference quotient between two rationals and ask what the hypothesis says about it.
2. The cue suggests the mean value theorem on a grid $\frac{1}{p}\mathbb{Z}$: integer slopes between same-denominator neighbours, hence integer values of f' on a dense set, hence (by continuity) everywhere.
3. Hesitation occurs at "but the MVT point ξ is uncontrolled". It is reinforced away by the word *continuously*: a continuous f' is determined by its values on any dense set.
4. The trap is assuming $\frac{an+1}{bn}$ is in lowest terms. It is not in general ($\frac{1\cdot 2+1}{3\cdot 2} = \frac{1}{2}$); the official solution relies on it without comment. A second trap is forgetting to verify that $\pm x + n$ actually satisfy the condition.
5. Recognise it faster by asking, for any condition on f at special points, *which pair of special points makes the difference quotient automatically nice*. The answer is usually "same type, minimal spacing".
6. Reusable decision rule: *to extract derivative information from a discrete condition, compare two points of the same kind at the natural spacing, apply the mean value theorem, and let continuity of f' carry the result from a dense set to everywhere; only then pin the remaining finitely many parameters with test values.* The same rule shows at once that a C^1 function with $f(\frac{k}{n}) \in \frac{1}{n}\mathbb{Z}$ for all integers k and $n \geq 1$ must be $cx + d$ with $c \in \mathbb{Z}$, and it is the first thing to try on any "f maps special rationals to special rationals, and f is smooth" problem.

2008 B6

2008 B6. Let n and k be positive integers. Say that a permutation σ of $\{1, 2, \ldots, n\}$ is *k-limited* if $|\sigma(i) - i| \leq k$ for all i. Prove that the number of k-limited permutations of $\{1, 2, \ldots, n\}$ is odd if and only if $n \equiv 0$ or $1 \pmod{2k+1}$.

Fields: combinatorics (parity of a count); linear algebra over $\mathbb{F}_2$ (permanent $\equiv$ determinant, kernels of banded matrices) **Position:** B6 — genuinely hard: the count itself has no usable formula for $k \geq 2$, so the parity must be reached without counting **Answer:** Proof. Over $\mathbb{F}_2$ the count is the determinant of the $0/1$ band matrix of half-width k, whose kernel has dimension equal to the overlap of two arcs of length k in $\mathbb{Z}/(2k+1)$; the arcs are disjoint exactly when $n \equiv 0, 1 \pmod{2k+1}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. Write $F_{n,k}$ for the number of k-limited permutations. The natural first move is to compute and look for a recurrence. For $k = 1$ it goes beautifully: $\sigma(n) \in \{n-1, n\}$; if $\sigma(n) = n$ the rest is a 1-limited permutation of $\{1, \ldots, n-1\}$; if $\sigma(n) = n-1$ then n must be hit by $n-1$ (nobody else can reach it), so σ swaps $n-1, n$. Hence $F_{n,1} = F_{n-1,1} + F_{n-2,1}$: Fibonacci, $1, 2, 3, 5, 8, 13, \ldots$. Fibonacci numbers mod 2 run $1, 0, 1, 1, 0, 1, \ldots$ with period 3, and $F_{n,1}$ is odd iff $n \equiv 0, 1 \pmod 3$. Exactly the claim for $k = 1$.

Encouraged, the student tries $k = 2$ and conditions on $\sigma(n) \in \{n-2, n-1, n\}$. It no longer closes. If $\sigma(n) = n-2$, then n and $n-1$ must be hit from $\{n-3, n-2, n-1\}$, and what is left is not a 2-limited permutation of a shorter initial segment but a permutation with a ragged boundary. One is forced into a transfer matrix whose states record which of the last $2k$ values

have been used — $\binom{2k}{k}$ states. The numbers are $F_{n,2} = 1, 2, 6, 14, 31, 73, 172, 400, 932, \ldots$, odd exactly at $n = 1, 5, 6, 10, 11, \ldots$, matching the claim, but no recurrence the student can find by hand explains why. For general k this route is a wall: it computes, it does not prove.

The instinct that goes further is "I only need the parity, so pair things up." Pairing σ with σ^{-1} (both are k-limited) reduces to counting k-limited involutions mod 2; this is the first official solution (Tsimerman), and it works, but it still needs a careful structural induction on how the involution acts on $\{1, \ldots, k+1\}$.

The model chain. Let $T = 2k+1$ and let M be the $n \times n$ matrix with $M_{ij} = 1$ if $|i-j| \le k$ and 0 otherwise.

Step 1: parity is a determinant. By definition

$$\det M = \sum_{\sigma} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} M_{i\sigma(i)},$$

and the product is 1 if σ is k-limited and 0 otherwise. Since $\operatorname{sgn}\sigma = \pm 1 \equiv 1 \pmod 2$, $\det M \equiv F_{n,k} \pmod 2$. So $F_{n,k}$ is odd iff M is invertible over $\mathbb{F}_2$, iff $Ma = 0$ has only the solution $a = 0$ in $\mathbb{F}_2^n$.

Step 2: rewrite the kernel as a sequence problem. Extend $a \in \mathbb{F}_2^n$ by $a_i = 0$ for $i \notin \{1, \ldots, n\}$ and put $S_m = a_{m-k} + \cdots + a_{m+k}$. Then $(Ma)_m = S_m$, so $Ma = 0$ iff $S_1 = \cdots = S_n = 0$. This is equivalent to $S_1 = 0$ together with $S_{m+1} - S_m = 0$ for $1 \le m \le n-1$, and $S_{m+1} - S_m = a_{m+k+1} - a_{m-k}$. So $Ma = 0$ iff

$$a_{j+T} = a_j \quad \text{for } 1-k \le j \le n-1-k, \qquad \text{and} \qquad a_{1-k} + \cdots + a_{1+k} = 0. \tag{$*$}$$

Let $W = \{1-k, \ldots, n+k\}$, a window of $n + 2k \ge T$ consecutive integers. The pairs $(j, j+T)$ with both entries in W are exactly those with $1-k \le j \le n-1-k$. So $(*)$ says: the sequence $(a_i)_{i \in W}$ is T-periodic on W, it vanishes on the k left-end positions $\{1-k, \ldots, 0\}$ and the k right-end positions $\{n+1, \ldots, n+k\}$ (by the extension convention), and the sum over one length-T window is 0.

Step 3: pass to residues. Because $|W| \ge T$, every residue class mod T meets W, and a T-periodic sequence on W is the same thing as a vector $c \in \mathbb{F}_2^{\mathbb{Z}/T}$ via $a_i = c_{i \bmod T}$; the sum over any length-T window is $\sum_r c_r$. Let

$$R_1 = \{1-k, \ldots, 0\} \bmod T, \qquad R_2 = \{n+1, \ldots, n+k\} \bmod T,$$

two arcs of k consecutive residues. Then $\ker M \cong \{c : c_r = 0 \text{ for } r \in R_1 \cup R_2,\ \sum_r c_r = 0\}$. Let F be the set of residues outside $R_1 \cup R_2$; $|F| = T - |R_1 \cup R_2| \ge T - 2k = 1$. The solutions are the vectors supported on F with an even number of ones, a space of dimension $|F| - 1 = |R_1 \cap R_2|$. Hence

$$M \text{ invertible over } \mathbb{F}_2 \iff R_1 \cap R_2 = \varnothing.$$

Step 4: when are the arcs disjoint? Modulo T, $R_1 = \{k+2, k+3, \ldots, 2k, 0\}$ (since $-j \equiv T - j$). Write $n \equiv r \pmod T$ with $0 \le r \le 2k$, so $R_2 = \{r+1, \ldots, r+k\}$.

- $r = 0$: $R_2 = \{1, \ldots, k\}$; $r = 1$: $R_2 = \{2, \ldots, k+1\}$. Both avoid $\{k+2, \ldots, 2k, 0\}$: disjoint.
- $2 \le r \le k+1$: the largest element $r + k$ satisfies $k + 2 \le r + k \le 2k + 1$, so it is in R_1 (with $2k + 1 \equiv 0$).
- $k + 2 \le r \le 2k$: then $r + 1 \le 2k + 1 \le r + k$, so R_2 contains $2k + 1 \equiv 0 \in R_1$.

So $R_1 \cap R_2 = \varnothing$ iff $n \equiv 0$ or $1 \pmod T$, and by Steps 1 and 3 that is exactly when $F_{n,k}$ is odd. □

Caveat. Proof complete as written; recorded as a check. For $k = 1, 2, 3$ and $1 \le n \le 9$ a brute-force count confirmed the parity claim, and Gaussian elimination over $\mathbb{F}_2$ confirmed $\dim \ker M = |R_1 \cap R_2|$ in every case.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains begin by writing down the 0/1 condition $|\sigma(i) - i| \le k$. They split at what to do with "is odd".

(N) Treat "odd" as a property of a number you must first compute: find $F_{n,k}$ (or a recurrence for it) and reduce mod 2. Works for $k = 1$; for $k \ge 2$ there is no hand-sized recurrence.

(M) Treat "odd" as a statement over $\mathbb{F}_2$, where sign disappears: the count is a permanent, the permanent is the determinant mod 2, and "determinant $\neq 0$" is "trivial kernel" — a linear problem with a local structure (band) that can be solved outright.

> *I'm counting permutations σ with $\sigma(i)$ in an allowed set for each i. That's the permanent of the 0/1 matrix of allowed pairs. And I only want it mod 2, where $+1 = -1$, so the permanent is the determinant. A determinant I don't have to compute — I only have to decide whether the matrix is singular.*

> *The kernel condition says every window of length $2k+1$ sums to zero. Two overlapping windows differ by one entry at each end, so $a_{j+2k+1} = a_j$: the kernel vectors are periodic, and the only freedom left is how the period meets the zeros at the two ends.*

- ▷ **Cue:** "prove that the number of … is odd" $\longrightarrow$ **Move:** look for an $\mathbb{F}_2$ object whose value is that count mod 2 (a determinant, a fixed-point count of an involution, a coefficient mod 2) rather than computing the count
- ▷ **Cue:** counting permutations with $\sigma(i) \in A_i$ for each i $\longrightarrow$ **Move:** it is the permanent of the 0/1 matrix $[j \in A_i]$; mod 2 it equals the determinant
- ▷ **Cue:** a linear condition "every window of length L sums to 0" $\longrightarrow$ **Move:** subtract consecutive windows: the sequence is L-periodic, and the boundary conditions become conditions on residues mod L
- ▷ **Cue:** an answer depending on n mod m $\longrightarrow$ **Move:** expect an object with period m inside the proof; find it before trying anything else

Layer 3 THE WHY

Why is "permanent $\equiv$ determinant mod 2" the natural move and not a trick? Start from what a determinant *is*: the signed sum, over all permutations, of products of one entry from each row and column. Put 0s and 1s in the matrix and each term is ± 1 exactly when the permutation uses only allowed positions. So the determinant of the "allowed positions" matrix is a signed count of exactly the objects in the problem. Signs are the only obstruction to it *being* the count, and in characteristic 2 signs vanish. A problem that asks for the parity of a count of permutations with position constraints is therefore, verbatim, asking whether a specific matrix is singular over $\mathbb{F}_2$. That is not an insight about this problem; it is the definition of the determinant read with the problem's word "odd" in mind.

Why is the band structure the right thing to exploit next? The constraint $|\sigma(i) - i| \le k$ is translation-invariant: position i may go to $i - k, \dots, i + k$ whatever i is. So M is a Toeplitz band matrix, and Ma is the sequence of moving sums of a with window $T = 2k + 1$. A moving sum that is identically zero forces, by subtracting neighbours, T-periodicity. The answer's own shape — a condition on n mod $(2k + 1)$ — is the problem announcing that a period $2k + 1$ lives somewhere; here it is.

Why do the two arcs appear, and why is disjointness the criterion? A periodic sequence is a vector c on the T residues. The finiteness of $\{1,\dots,n\}$ enters only through the k forced zeros just before position 1 and the k just after position n; these are $2k$ residue-constraints from two arcs, and a single parity constraint (window sum) is added. $T = 2k+1$ residues minus $2k$ constraints (if independent) leaves one free residue, and one free coordinate with sum zero forces zero. Any overlap of the arcs wastes a constraint and frees a second residue, and two free residues give the nonzero kernel vector $e_r + e_{r'}$. The count $T = 2k+1 = 2k+1$ is exact: that is why the answer is so sharp. The right-hand arc sits at $n+1,\dots,n+k$, so its position mod T — that is, n mod T — decides everything.

Kedlaya and Ng's second solution (Zhao) row-reduces M block by block to show $\det M_n \equiv \det M_{n-T}$, and their third (Belulovich) finds explicit kernel vectors or proves the kernel trivial. Both begin, without comment, with "$\det M \equiv F_{n,k} \pmod 2$". Here is where that comes from: the word "odd" in the statement plus the definition of the determinant. The version above sharpens Belulovich's argument into a single computation, $\dim_{\mathbb{F}_2} \ker M = |R_1 \cap R_2|$, which handles invertibility and non-invertibility at once.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "the number of … is odd" attached to permutations with a per-position constraint. On seeing it, write the $0/1$ matrix $M_{ij} = [|i-j| \le k]$ and the line $F_{n,k} \equiv \det M \pmod 2$.

2. The cue suggests linear algebra over $\mathbb{F}_2$: decide whether M is invertible by solving $Ma = 0$. The band shape of M suggests moving-sum and periodicity arguments.

3. Hesitation occurs twice: first at "the determinant has signs, the count does not" (reinforced away by characteristic 2), and then at the boundary — how the periodic kernel meets the ends of $\{1,\dots,n\}$. Extending a by zeros to the window $\{1-k,\dots,n+k\}$ turns the boundary into two clean arcs of residues.

4. The trap is the $k = 1$ success: Fibonacci makes it look as though a recurrence will do for all k, and a student can spend the whole session on $k = 2$. A second trap in the kernel argument is forgetting the single window-sum equation ($S_1 = 0$), which is what kills the one remaining free residue.

5. Recognise it faster by reading "odd/even number of permutations (or perfect matchings) with allowed positions" as "is this $0/1$ matrix singular over $\mathbb{F}_2$?" — and by expecting, whenever the answer depends on n mod m, a period-m object in the proof.

6. Reusable decision rule: *when a problem asks only for the parity of a count, move the whole problem into $\mathbb{F}_2$ before counting anything; a signed sum (determinant, alternating sum, character) that equals the count mod 2 is often computable when the count is not.* Then for any translation-invariant linear system, subtract shifted copies of the equations to expose periodicity and reduce the boundary to a finite condition on residues.