# DAWN PUTNAM SESSION

Putnam 2010 — Papers A and B

Putnam 120 Archive

# Contents

**Putnam 2010 — Paper A** **5**
2010 A1 . . . . . 5
2010 A2 . . . . . 7
2010 A3 . . . . . 9
2010 A4 . . . . . 12
2010 A5 . . . . . 14
2010 A6 . . . . . 17

**Putnam 2010 — Paper B** **21**
2010 B1 . . . . . 21
2010 B2 . . . . . 23
2010 B3 . . . . . 25
2010 B4 . . . . . 28
2010 B5 . . . . . 31
2010 B6 . . . . . 34

# Putnam 2010 — Paper A

## 2010 A1

**2010 A1.** Given a positive integer $n$, what is the largest $k$ such that the numbers $1, 2, \ldots, n$ can be put into $k$ boxes so that the sum of the numbers in each box is the same? [When $n = 8$, the example $\{1,2,3,6\}, \{4,8\}, \{5,7\}$ shows that the largest $k$ is *at least* 3.]

**Fields:** combinatorics, elementary bounding arguments **Position:** A1 — the opener **Answer:** $k = \left\lfloor \frac{n+1}{2} \right\rfloor$, achieved by pairing $i$ with $n+1-i$ for each $i$ (plus the leftover singleton $\{n\}$ when $n$ is odd).

**Layer 1 THE WHAT — surface analysis**

*The natural attempt.* Small cases first. $n = 1$: only $k = 1$. $n = 2$: $\{1,2\}$, $k = 1$. $n = 3$: $\{1,2\}, \{3\}$ both sum to 3, $k = 2$. $n = 4$: $\{1,4\}, \{2,3\}$, $k = 2$. $n = 8$ is given, $k = 3$. The pattern $1, 1, 2, 2, \ldots$ suggests $k = \lceil n/2 \rceil = \lfloor (n+1)/2 \rfloor$, and pairing $i \leftrightarrow n+1-i$ gives a construction for every $n$ (with $\{n\}$ left over when $n$ is odd, since $n+1-i = i$ forces $i = (n+1)/2$, an integer exactly when $n$ is odd). So the lower bound is settled by direct construction.

The upper bound is the real content, and the natural line of attack is counting: with $k$ boxes and $n$ items distributed among them, if $k$ is large, most boxes must be small. Concretely: could two boxes each hold exactly one number? If box $\{x\}$ and box $\{y\}$ are both singletons with $x \neq y$, they cannot have the same sum, since $x \neq y$ — contradiction. So at most one box is a singleton. Every other box holds at least 2 elements. If there are $k$ boxes and at most one singleton, the total count of elements satisfies

$$n \;\geq\; 1 \cdot 1 + (k-1) \cdot 2 = 2k - 1$$

(the "1" only appears if a singleton box actually exists; without one the bound is even more generous, $n \geq 2k$, which is stronger, so the displayed inequality covers both cases). Rearranging, $k \leq \dfrac{n+1}{2}$, and since $k$ is an integer, $k \leq \left\lfloor \dfrac{n+1}{2} \right\rfloor$. This closes the problem, but it took two separate pieces of casework — ruling out two singletons, then converting "at most one small box" into a global count — before the bound fell out, and the argument had to be checked wasn't off by one in either direction.

*The model chain.* The number $n$ itself is one of the $n$ numbers being distributed, so it lies in some box. Every number in that box is a positive integer, so the box's sum is at least the value of any single element it contains — in particular, it is at least $n$. Since *every* box has the same sum $s$, this one observation about the box containing $n$ gives $s \geq n$ immediately.

Now $ks = 1 + 2 + \cdots + n = \dfrac{n(n+1)}{2}$, so

$$k = \frac{n(n+1)/2}{s} \leq \frac{n(n+1)/2}{n} = \frac{n+1}{2}.$$

As $k$ is a positive integer, $k \leq \left\lfloor \dfrac{n+1}{2} \right\rfloor$. Combined with the pairing construction above (which achieves exactly this value), the largest $k$ is $\left\lfloor \dfrac{n+1}{2} \right\rfloor$.

### Layer 2 THE HOW — the bifurcation

**Bifurcating point.** Both chains agree that the construction gives $k = \lfloor (n+1)/2 \rfloor$ and that the remaining work is the upper bound. They also agree that $ks = n(n+1)/2$. They part company at the question each one asks next. The natural attempt asks: *how many elements can each box hold*, and reasons about box sizes via pigeonhole. The model chain instead asks: *where does the number n itself end up*, and reasons about box value via one specific element.

> *n is somewhere. It's the single largest of the n numbers I'm distributing, and it has to belong to exactly one box — I don't need to know how many other numbers share that box with it, or reason about box sizes at all. Every element in a box is positive, so that box's total is at least the value of any one thing inside it. That gives me $s \geq n$ for free, from the existence of n alone, without any counting argument.*

### Layer 3 THE WHY

Why does fixing attention on the single element $n$ beat reasoning about box *sizes*? Because the quantity the problem actually constrains is the common sum $s$, and the cheapest possible lower bound on a sum of positive numbers is the value of any one of its terms — that is the definition of "sum of positive numbers", not a trick imported from elsewhere: if $a_1, \ldots, a_m > 0$ then $a_1 + \cdots + a_m \geq a_i$ for every $i$, with no further argument required. The pigeonhole route also arrives at a true bound, but it does so by first converting the problem into a statement about *cardinalities* (how many elements per box) and only later back into a statement about sums (via "$x \neq y$ forces two singletons to disagree"). That conversion is unnecessary work: the problem was already stated in terms of sums, and $n$ is available as a term in one of those sums the instant the numbers $1, \ldots, n$ are named. The reusable habit is to ask, before building any counting argument: is there a single distinguished element (the largest, the smallest, the one appearing in every construction) whose location alone already forces the bound I want?

The construction side needs no separate justification beyond verifying $i + (n+1-i) = n+1$ for every pair — this is arithmetic, not insight, and both chains do it identically. The asymmetry between the two chains lives entirely in how the upper bound is found, which is exactly where the Blueprint's "of course" standard should be applied: once you see that the numbers are positive and that $n$ is among them, $s \geq n$ is not a discovery, it is a restatement of positivity plus membership.

### Layer $+\alpha$ BACKTESTING

**1.** The first useful cue is "$1, 2, \ldots, n$ distributed into boxes with equal sum $s$" together with the fact that $n$ (the largest value) is one of the items. On seeing it, write down: "some box contains $n$; that box's sum is therefore at least $n$" before attempting any counting argument.

**2.** That cue suggests bounding $s$ from below by the value of the largest available element, then combining $s \geq n$ with the fixed total $ks = n(n+1)/2$ to bound $k$ from above — an extremal-element argument, not a counting argument.

**3.** Hesitation sets in right after the pairing construction is found: with a working construction in hand for the lower bound, the pull is to prove the upper bound by the same combinatorial register (box sizes, pigeonhole), because that register is already "hot" from finding the pairing. What reinforces the switch is asking explicitly which single number in the problem is both fixed and extremal — $n$ is always the largest value distributed, regardless of how the boxes are chosen, and that fact does not depend on how many boxes there are.

**4.** The trap is treating "the pigeonhole argument is correct" as license to stop looking for a shorter one. It is correct, and it generalises worse: it needs the extra step of noting two singleton boxes cannot share a sum, which is a second idea layered on top of the counting, whereas the extremal-element bound needs only positivity.

**5.** Recognise it faster by treating any "distribute $1, \ldots, n$ into parts with equal [sum/size/some additive statistic]" problem as a standing invitation to ask what the extremal element (here $n$, the largest) forces about that statistic, before reaching for pigeonhole on the part sizes.

**6.** Reusable rule: when $k$ parts must share a fixed total $S$ under a common per-part value $s$, and the parts are built from a known finite set of positive numbers, first check whether the largest number in that set already forces a lower bound on $s$ (namely $s \geq$ that number, since it sits in one part and every part is a sum of positives). Use $s \geq$ (largest element) together with $S = ks$ to bound $k$, before building any argument about how many elements each part can hold.

▷ **Cue:** a fixed finite set of positive numbers is split into parts with equal sum $s$, and $k$ is to be bounded $\longrightarrow$ **Move:** find the largest element in the set; it lies in one part, so $s \geq$ that element, giving $k = S/s \leq S/(\text{largest element})$

▷ **Cue:** a bound is available either via counting how many objects occupy each container, or via the value of one extremal object $\longrightarrow$ **Move:** prefer the extremal-object bound first — it needs only positivity/order, not a two-step counting-then-contradiction argument

## 2010 A2

**2010 A2.** Find all differentiable functions $f : \mathbb{R} \to \mathbb{R}$ such that

$$f'(x) = \frac{f(x+n) - f(x)}{n}$$

for all real numbers $x$ and all positive integers $n$.

**Fields:** real analysis, functional equations **Position:** A2 **Answer:** $f(x) = cx + d$ for constants $c, d \in \mathbb{R}$; these are exactly the solutions.

**Layer 1 THE WHAT — surface analysis**

*The natural attempt.* The hypothesis holds for every $n$, so start by testing named families to see which are even candidates. Try $f(x) = x^2$: $f'(x) = 2x$, while $\dfrac{f(x+n) - f(x)}{n} = \dfrac{(x+n)^2 - x^2}{n} = 2x + n$. These agree only if $n = 0$, so no quadratic works. Try $f(x) = e^{ax}$: $f'(x) = ae^{ax}$, while $\dfrac{f(x+n) - f(x)}{n} = e^{ax} \cdot \dfrac{e^{an} - 1}{n}$. Matching requires $a = \dfrac{e^{an} - 1}{n}$ for *every* positive integer $n$ simultaneously — the left side is a constant and the right side genuinely depends on $n$ unless $a = 0$. So the only surviving exponential is the constant function. Try $f(x) = cx + d$: $f'(x) = c$, and $\dfrac{f(x+n) - f(x)}{n} = \dfrac{cn}{n} = c$. This works for every $n$, no restriction needed.

So among the families tested, only linear functions survive, and the conjecture $f(x) = cx + d$ is natural. But this is guess-and-check over a handful of named shapes (polynomial, exponential); it gives strong evidence, not a proof. To finish, this route would need to rule out every other differentiable function — including ones with no closed form at all — and testing more families one at a time cannot do that. Some argument that uses the hypothesis *itself*, for a general $f$, is needed, not a larger list of examples.

*The model chain.* Fix $x \in \mathbb{R}$. Apply the hypothesis with $n = 2$:

$$2f'(x) = f(x+2) - f(x).$$

Split the right side using the two consecutive gaps of length 1, and rewrite each using the hypothesis with $n = 1$ (once at $x$, once at $x+1$):

$$f(x+2) - f(x) = \big(f(x+2) - f(x+1)\big) + \big(f(x+1) - f(x)\big) = f'(x+1) + f'(x).$$

Combining, $2f'(x) = f'(x+1) + f'(x)$, so

$$f'(x+1) = f'(x) \quad \text{for every } x \in \mathbb{R}.$$

Now define $g(x) = f(x+1) - f(x)$. Since $f$ is differentiable, so is $g$, and $g'(x) = f'(x+1) - f'(x) = 0$ identically, so $g$ is a constant, $g(x) \equiv c$ where $c = g(0)$. But by the hypothesis at $n = 1$, $f'(x) = f(x+1) - f(x) = g(x) = c$ for every $x$. So $f' \equiv c$, a constant, giving $f(x) = cx + d$ for $d = f(0)$. Conversely every such $f$ was already checked to satisfy the hypothesis, so these are exactly the solutions.

**Layer 2 THE HOW — the bifurcation**

**Bifurcating point.** Both chains begin by noticing the hypothesis quantifies over every positive integer $n$. The natural attempt uses that universality to *filter candidate families* — plug in a guessed $f$ and see whether the equation can hold for all $n$ at once. At the point (†) where a candidate has survived, the natural attempt has no way to promote "survives every family I tried" to "is the only solution." The model chain instead uses the universal quantifier on $n$ directly: it takes *two particular values of n (1 and 2) applied at two related points (x and x + 1)* and combines the resulting identities algebraically, for an arbitrary differentiable $f$ — no family is ever guessed.

> *The equation holds for every n at once, for the* same *function f — so instead of picking one n and asking what functions satisfy it, I should pick two different instances of the equation (say n = 2 at x, and n = 1 at both x and x + 1) and see what purely algebraic relation falls out when I compare them, without assuming any form for f at all.*

> *I've reduced to $f'(x+1) = f'(x)$ for all $x$ — but that alone only says $f'$ is 1-periodic, not that it's constant; periodic non-constant functions certainly exist. The extra fact that closes the gap is that the hypothesis itself (at $n = 1$) already expresses $f'(x)$ as the difference $f(x+1) - f(x)$ — so if I can show that difference is constant, I get $f'$ constant directly, without needing any separate argument against periodicity.*

**Layer 3 THE WHY**

Why does combining two instances of the hypothesis outperform testing families? Because "holds for all $n$" is not a single piece of information about $f$ — it is infinitely many equations sharing the one unknown function, and the content of the problem lives in how those equations constrain each other, not in what any single one of them looks like in isolation. Testing named families only ever uses the equation at whichever $n$ makes the arithmetic clean for that family; it never forces two instances of the hypothesis to interact. The model chain's move — writing $f(x+2) - f(x)$ two different ways, once via $n = 2$ directly and once as a telescoping sum of two $n = 1$ steps — is exactly forcing two instances of the same universal hypothesis to describe the same quantity, so that they can be set equal to each other. This is the general lesson for "for all $n$" functional equations: look for a value of $n$ that can be reconstructed as a sum or composition of smaller instances (here, $n = 2$ as two copies of $n = 1$), then equate the two descriptions.

The second load-bearing fact is the shift from "$f'$ is 1-periodic" to "$f'$ is constant." This works only because the *original* hypothesis, not a derived consequence of it, already identifies $f'(x)$ with the difference $g(x) = f(x+1) - f(x)$ at $n = 1$. Differentiating $g$ and using periodicity of $f'$ gives $g' \equiv 0$, hence $g$ constant — and because $g = f'$ (from the hypothesis itself, not from a new argument), $f'$ inherits that constancy for free. The definition being leaned on is nothing more exotic than "a function with derivative zero everywhere is constant" (mean value theorem), applied to $g$; the insight is recognising that $g$ was secretly $f'$ all along, so proving $g$ constant *is* proving $f'$ constant, collapsing what looked like two separate goals into one.

**Layer $+\alpha$ BACKTESTING**

**1.** The first useful cue is a functional/differential equation stated "for all positive integers $n$" with the same unknown function on every instance. On seeing it, write down: "find a value of $n$ expressible as a sum of smaller values of $n$, and force the two resulting descriptions of the same quantity to agree" before testing candidate functions.

**2.** That cue suggests the method of combining $n = 2$ (directly) with two copies of $n = 1$ (telescoped) at shifted arguments, rather than guessing closed forms for $f$.

**3.** Hesitation sets in right after $f'(x+1) = f'(x)$ is derived: periodicity of $f'$ feels like it should require a separate boundedness or integral argument to rule out non-constant periodic derivatives. What reinforces the correct move is re-reading the $n = 1$ instance of the hypothesis itself and noticing it already writes $f'(x)$ as $f(x+1) - f(x)$ — the object whose derivative was just shown to vanish *is* $f'$, not merely related to it.

**4.** The trap is treating a handful of confirmed families (linear works, quadratic and exponential fail) as if they constitute a proof of uniqueness. They constitute evidence for the conjecture and nothing more; the "for all $n$" hypothesis has to be used on an arbitrary differentiable $f$, not tested against named shapes, to actually finish.

**5.** Recognise it faster by treating any "holds for every positive integer $n$" functional equation as a standing instruction to find an additive relation among the values of $n$ (e.g. $2 = 1 + 1$) and exploit it algebraically, before reaching for guess-and-check on candidate functions.

**6.** Reusable rule: for a functional equation depending on a free parameter $n$ that ranges over all positive integers, look first for a small identity among values of $n$ (such as $2 = 1 + 1$) and substitute it to produce two descriptions of the same quantity; equating them yields structural information (here, periodicity of $f'$) for free, and often the hypothesis itself — reapplied at the smallest parameter value — then upgrades that structural information into the full answer.

▷ **Cue:** a differential/functional equation is imposed "for all positive integers $n$" with one unknown function $\longrightarrow$ **Move:** find a value of $n$ that decomposes as a sum of smaller values (e.g. $2 = 1 + 1$), substitute, and equate the two resulting descriptions of the same quantity

▷ **Cue:** a derived fact says some auxiliary function has zero derivative, i.e. is constant, and that auxiliary function was defined using the same expression the hypothesis already sets equal to the object of interest $\longrightarrow$ **Move:** recognise the auxiliary function is not merely related to the target but literally equals it via the original hypothesis, collapsing two claims into one

## 2010 A3

**2010 A3.** Suppose that the function $h : \mathbb{R}^2 \to \mathbb{R}$ has continuous partial derivatives and satisfies the equation

$$h(x,y) = a\frac{\partial h}{\partial x}(x,y) + b\frac{\partial h}{\partial y}(x,y)$$

for some constants $a, b$. Prove that if there is a constant $M$ such that $|h(x,y)| \leq M$ for all $(x,y) \in \mathbb{R}^2$, then $h$ is identically zero.

**Fields:** partial differential equations reduced to ordinary differential equations, directional derivatives **Position:** A3 **Answer:** $h \equiv 0$. Restricting $h$ to the line through any point in direction $(a,b)$ turns the PDE into the ODE $f' = f$; boundedness of $f(t) = Ce^t$ for all real $t$ forces $C = 0$.

**Layer 1 THE WHAT — surface analysis**

*The natural attempt.* The equation $h = ah_x + bh_y$ is a first-order linear PDE, and the standard machinery for that is the method of characteristics: find new coordinates in which the equation becomes an ODE in one variable. Set $u = ax + by$ (the direction along which the right-hand

side differentiates) and $v = bx - ay$ (a direction orthogonal to it, chosen so $(u, v)$ is a genuine change of coordinates when $(a, b) \neq (0, 0)$). By the chain rule,

$$\frac{\partial h}{\partial u} = \frac{ah_x + bh_y}{a^2 + b^2}, \qquad \frac{\partial h}{\partial v} = \frac{bh_x - ah_y}{a^2 + b^2}.$$

The hypothesis $h = ah_x + bh_y$ becomes $h = (a^2 + b^2)\,\partial h/\partial u$, i.e. $\partial h/\partial u = h/(a^2 + b^2)$, an ODE in $u$ with $v$ carried along as a parameter. Solving it treating $v$ as fixed gives

$$h(u, v) = \phi(v)\, e^{u/(a^2+b^2)}$$

for some function $\phi$ (the "constant" of integration in $u$, which may depend on the parameter $v$). Translating back, $h(x, y) = \phi(bx - ay)\, e^{(ax+by)/(a^2+b^2)}$ for some function $\phi$ of one real variable.

Now impose boundedness. Along any fixed line $bx - ay = v_0$ (constant), as $u = ax + by \to +\infty$ the exponential factor blows up unless $\phi(v_0) = 0$; since this must hold for every $v_0$, $\phi \equiv 0$ and hence $h \equiv 0$. This works, but it required setting up a full change of variables, converting the PDE into an ODE in the new coordinates via the chain rule on both partials (not just the one matching the equation), solving that ODE with a coordinate-dependent "constant", translating the general solution back into $(x, y)$, and only then invoking boundedness — considerably more machinery than the conclusion (a single point, $h \equiv 0$) seems to need.

*The model chain.* The problem only requires showing $h(x_0, y_0) = 0$ for an arbitrary fixed point $(x_0, y_0)$ — there is no need to describe $h$ everywhere first. Restrict $h$ to the single line through $(x_0, y_0)$ in the direction $(a, b)$: define $L(t) = (x_0, y_0) + (a, b)t$ for $t \in \mathbb{R}$, and $f(t) = h(L(t))$. By the chain rule,

$$f'(t) = a\, h_x(L(t)) + b\, h_y(L(t)) = h(L(t)) = f(t),$$

where the middle equality is exactly the hypothesis, evaluated at the point $L(t)$ — no rescaling or normalisation is needed, because the coefficients in the chain rule for $L$ are precisely $a$ and $b$, matching the hypothesis term for term. So $f$ satisfies the ODE $f' = f$ on all of $\mathbb{R}$, giving $f(t) = f(0)\, e^t$.

Since $|h(x, y)| \leq M$ everywhere, $|f(t)| \leq M$ for every $t \in \mathbb{R}$, including as $t \to +\infty$. But $|f(0)e^t| \to \infty$ unless $f(0) = 0$. So $f(0) = 0$, i.e. $h(x_0, y_0) = 0$. As $(x_0, y_0)$ was an arbitrary point of $\mathbb{R}^2$, $h$ is identically zero. (If $a = b = 0$, the hypothesis reads $h \equiv 0$ directly, so this case is immediate and needs no line at all.)

### Layer 2 THE HOW — the bifurcation

**Bifurcating point.** Both chains recognise the equation as a first-order linear PDE in the direction $(a, b)$. They split at the question of *scope*: the natural attempt asks "what is $h$, in general, everywhere?" and reaches for the full change-of-variables machinery that answers that broader question. The model chain asks only "what is $h$ at one arbitrary point?" — and restricts to the single line through that point, which is exactly enough data to answer the question actually being asked, no more.

> *I don't need the general solution of this PDE — the problem only asks me to show h vanishes at every point, one point at a time. If I restrict to the line through a single point in the direction $(a, b)$, the hypothesis $h = ah_x + bh_y$ becomes, by the chain rule for that exact line, an ODE in one variable with no leftover parameter to track — because the line's direction vector is literally $(a, b)$, the same pair of constants that appear in the hypothesis, so nothing needs to be normalised or carried along as a second coordinate.*

### Layer 3 THE WHY

Why does restricting to one line beat solving the PDE in general? Because the object $ah_x + bh_y$ *is*, by definition, the derivative of $h$ along the direction $(a, b)$ — the chain rule for

$h(x_0 + at,\, y_0 + bt)$ produces exactly $ah_x + bh_y$ evaluated on the line, term for term, with no extra factor. The hypothesis is not merely "related to" a directional derivative; restricted to any line in direction $(a, b)$, it *becomes* the statement "the directional derivative of $h$ along this line equals $h$ itself," which is the single-variable equation $f' = f$ verbatim. Once that is seen, solving the PDE in full — finding $h$ as a function of two variables — is solving a strictly harder problem than the one asked, since the question only quantifies over points, not over the whole function's shape. The natural attempt's extra work (the orthogonal coordinate $v = bx - ay$, the coordinate-dependent constant $\phi(v)$) is entirely spent describing how $h$ varies *across* lines, information the conclusion $h \equiv 0$ never uses: every line gets killed by the same one-line argument independently, so there was never a need to relate different lines to each other.

The second fact doing the work is that boundedness is a statement about *all* of $\mathbb{R}$, including $t \to +\infty$ along every characteristic line — and $Ce^t$ is unbounded on $[0, \infty)$ for any $C \neq 0$. This is not a subtle limiting argument; it is the defining growth property of the exponential, and the only reason it applies so cleanly here is that the ODE reduction produced a pure exponential with no polynomial or oscillatory factor to complicate the limit. The general lesson: when a PDE hypothesis states a function equals a specific directional derivative of itself, restricting to a line in exactly that direction converts the hypothesis into an ODE with the same growth-rate constant as the equation's own coefficient, and boundedness on the whole line kills any nonzero growth constant immediately, one point at a time, with no need to solve for $h$ as a two-variable object at all.

### Layer $+\alpha$ BACKTESTING

**1.** The first useful cue is a PDE stated as "$h$ equals a fixed linear combination $ah_x + bh_y$ of its own first partials." On seeing it, write down: "$ah_x + bh_y$ is exactly the directional derivative of $h$ along $(a, b)$ — restrict to a line in that direction before doing anything else."

**2.** That cue suggests reducing to a single-variable ODE along the characteristic line through an arbitrary point, rather than solving the PDE in general via a full change of variables.

**3.** Hesitation sets in at the very start, when the equation is first written down: PDEs "look like" they call for characteristics and a general solution, and it can feel like cutting a corner to only look at one point at a time. What reinforces the shortcut is rereading exactly what the conclusion asks for — "$h$ is identically zero," i.e. a statement about every point separately — and noticing that a general formula for $h$ was never actually required to prove a pointwise statement.

**4.** The trap is assuming that because a general solution formula is available (and it is, via characteristics), it must be the intended route. It works, but it computes and stores information — how $h$ varies from one characteristic line to the next, via $\phi(v)$ — that the final conclusion discards entirely, since every line is shown to vanish by the identical one-line argument.

**5.** Recognise it faster by treating "a function equals a specific linear combination of its own partial derivatives, plus a boundedness hypothesis" as a standing cue for restrict-to-a-line-and-use-boundedness, reserving full characteristics for problems that actually ask for the general shape of the solution.

**6.** Reusable rule: if a PDE states $h = c_1 h_{x_1} + \cdots + c_n h_{x_n}$ for constants $c_i$, and the goal is a pointwise conclusion (vanishing, a bound, a sign) rather than a description of $h$ everywhere, restrict $h$ to the line through the point of interest in direction $(c_1, \ldots, c_n)$: the chain rule turns the hypothesis into the exact one-variable ODE $f' = f$ with no leftover parameter, because the line's direction matches the hypothesis's coefficients term for term. Combine with the growth or boundedness hypothesis on that one line to finish, one point at a time.

▷ **Cue:** a PDE states that $h$ equals a fixed linear combination $ah_x + bh_y$ of its own first partials $\longrightarrow$ **Move:** restrict $h$ to the line through the point of interest in direction $(a, b)$; the chain rule converts the hypothesis into the single-variable ODE $f' = f$ with no extra normalisation

▷ **Cue:** a solution of $f' = f$ (an unbounded exponential unless the constant vanishes) is required to stay bounded on all of $\mathbb{R}$ $\longrightarrow$ **Move:** conclude the constant of integration is zero, since

$Ce^t$ is unbounded as $t \to \infty$ for any $C \neq 0$

▷ **Cue:** the conclusion needed is pointwise (a value vanishes, a bound holds, at each point separately) rather than a description of the function everywhere ⟶ **Move:** prefer a one-point / one-line reduction over solving the full PDE or functional equation in general

## 2010 A4

**2010 A4.** Prove that for each positive integer $n$, the number $10^{10^{10^n}} + 10^{10^n} + 10^n - 1$ is not prime.

**Fields:** number theory (modular arithmetic, orders of elements, 2-adic valuation) **Position:** A4 — mid-paper, but one of the gentler A4s: the whole difficulty is choosing the modulus, after which the verification is four lines. **Answer:** For $n = 2^m k$ with $k$ odd, the number is divisible by $10^{2^m} + 1$ and strictly larger than it, hence composite.

### Layer 1 THE WHAT — surface analysis

Throughout, write $N = 10^{10^{10^n}} + 10^{10^n} + 10^n - 1$.

*The natural attempt.* A number this size can only be attacked by a modulus, so you run through small primes. Modulo 3: $10 \equiv 1$, so $N \equiv 1 + 1 + 1 - 1 = 2$. Dead. Modulo 9: same, $N \equiv 2$. Modulo 7: the order of 10 is 6, and $10^j \equiv 4 \pmod 6$ for every $j \geq 1$, so both towers reduce to $10^4 \equiv 4 \pmod 7$ and $N \equiv 4 + 4 + 10^n - 1 \equiv 10^n \pmod 7$, never 0. Dead.

Modulo 11 something happens. Since $10 \equiv -1 \pmod{11}$, a power of 10 is $\pm 1$ according to the parity of the exponent. The exponents $10^{10^n}$ and $10^n$ of the two top terms are even, so both are $\equiv 1$, and

$$N \equiv 1 + 1 + (-1)^n - 1 = 1 + (-1)^n \pmod{11}.$$

For odd $n$ this is 0, and $N > 11$, so $N$ is composite. Half the problem is done in five minutes.

For even $n$ you are back to searching. Modulo 13, 37, 101, … each prime needs its own order computation, and a strong student tries 13 (order 6), finds nothing uniform, and starts to wonder whether even $n$ needs a different factorisation altogether. Trying 101 does work for $n = 2$ — but you tried it because it was on a list, and without a reason you have no idea what to do at $n = 4$.

*The model chain.* *Step 1: why 11 worked.* $11 = 10^1 + 1$, and modulo $10^1 + 1$ the base 10 is $-1$. So each power of 10 reduces to $(-1)^{\text{exponent}}$. The top two terms have huge even exponents and become $+1$; the term $10^n$ becomes $(-1)^n$; the $-1$ stays. The sum is $1 + 1 + (-1)^n - 1$.

*Step 2: generalise the modulus.* Replace $10^1 + 1$ by $M = 10^d + 1$. Then $10^d \equiv -1 \pmod M$, so $10^{2d} \equiv 1$, and for any exponent $e$ that is a multiple of $d$,

$$10^e \equiv (-1)^{e/d} \pmod M.$$

I want $10^n \equiv -1$ (so $e/d = n/d$ odd) and the two tower terms $\equiv +1$ (so $10^n/d$ and $10^{10^n}/d$ even). Then $N \equiv 1 + 1 - 1 - 1 = 0$.

*Step 3: choose d.* Write $n = 2^m k$ with $k$ odd and take $d = 2^m$. Then $n/d = k$ is odd, so $10^n \equiv -1 \pmod M$. For the towers I need $2^{m+1} \mid 10^n$ and $2^{m+1} \mid 10^{10^n}$. Since $2^n \mid 10^n$ and $2^{10^n} \mid 10^{10^n}$, it suffices that $m + 1 \leq n$. That holds because $n \geq 2^m \geq m + 1$ (the second inequality by induction, or since $2^m = (1 + 1)^m \geq 1 + m$). Hence

$$N \equiv 1 + 1 + (-1) + (-1) = 0 \pmod{10^{2^m} + 1}.$$

*Step 4: the divisor is proper.* $M = 10^{2^m} + 1 \leq 10^n + 1 < 10^{10^n} < N$, and $M > 1$. So $M$ is a divisor of $N$ strictly between 1 and $N$, and $N$ is not prime.

**Caveat.** Verified by machine for $1 \le n \le 40$: $N \bmod (10^{2^m}+1) = 0$ in every case, with the tower exponents reduced modulo the order $2^{m+1}$ of 10. The divisor $10^{2^m}+1$ need not be prime (e.g. $10^4+1 = 73 \cdot 137$); the proof does not need it to be.

### Layer 2 THE HOW — the bifurcation

**Bifurcating point.** Both chains reach (†): "$N \equiv 1+(-1)^n \pmod{11}$, done for odd $n$." At (†) the natural attempt treats 11 as a lucky prime and goes looking for another lucky prime for even $n$. The model asks *why* 11 was lucky, sees that the reason was $11 = 10+1$, and replaces $10+1$ by $10^d+1$ with $d$ tuned to $n$.

> *Mod* 11 *worked because* $10 \equiv -1$ *there, so every power of* 10 *became a sign. The only thing that went wrong for even* $n$ *is that the sign of* $10^n$ *flipped. I can make* $10^d$ *be* $-1$ *instead of* 10 *by working mod* $10^d+1$*; then I want* $n/d$ *odd. So* $d$ *is* $n$ *with its odd part stripped:* $d = 2^m$.

> *The towers are harmless as long as their exponents are divisible by* $2d$*. They are powers of* 10*, so they are divisible by huge powers of* 2*. I only need to check that* $2^{m+1}$ *is small compared with* $10^n$.

### Layer 3 THE WHY

Why is $10^d+1$ the natural modulus rather than a guess? Go back to what $N$ is made of: four terms, three of which are powers of 10 and one of which is $-1$. Any modulus reduces each power of 10 to $10^{e \bmod \mathrm{ord}}$, where ord is the order of 10. The problem is then entirely about exponents. For $N \equiv 0$ you need the four residues to cancel, and the most rigid way four quantities of size $\pm 1$ can cancel is $1+1-1-1$. So you want a modulus in which 10 has a power equal to $-1$ — and the modulus *defined* by that property is $10^d+1$. It is the circle-and-centre situation from the Blueprint: $10^d+1$ is literally the set of integers in which "$10^d = -1$" is true, so if you want $10^d = -1$, of course that is where you work.

Why $d = 2^m$, the 2-part of $n$? Because the exponents present are $n$, $10^n$ and $10^{10^n}$, and the sign $(-1)^{e/d}$ depends only on the parity of $e/d$. The towers have exponents enormously divisible by 2 and by nothing else useful about their oddness, so they are even over any small power of 2. The only exponent whose parity you can steer is $n$ itself, and $n/d$ is odd precisely when $d$ contains exactly the full power of 2 in $n$. The odd part $k$ is irrelevant: this is the structural fact behind $a+1 \mid a^k+1$ for odd $k$, which you already know as the factorisation $x^k+1 = (x+1)(x^{k-1}-x^{k-2}+\cdots+1)$.

Kedlaya's solution opens with "Write $n = 2^m k \ldots$ modulo $10^{2^m}+1$." Here is where that comes from: it is the $n$-odd case (modulus $11 = 10^{2^0}+1$) with the exponent 1 replaced by the part of $n$ that stops $n$ from being odd. Once you have done the odd case honestly — noticing that the reason was the sign — the even case is a relabelling.

The one remaining place a reader can lose marks is the inequality $2^{m+1} \mid 10^n$. It needs $m+1 \le n$, which is true because $2^m$ divides $n$, so $n \ge 2^m \ge m+1$. Say this; do not wave at it.

### Layer $+\alpha$ BACKTESTING

**1.** The first useful cue is "is not prime" attached to a number built from powers of 10 and $\pm 1$. Write down: *find a modulus* $M$ *with* $1 < M < N$ *and* $N \equiv 0 \pmod{M}$, and reduce every term to a residue.

**2.** Small-modulus testing, and specifically moduli of the form $b^d \pm 1$ where $b$ is the base appearing in the powers, since there each power of $b$ collapses to a sign or to 1.

**3.** Hesitation occurs after 11 disposes of odd $n$ and you start testing $13, 37, 101$ at random. Reinforce with the habit: when a modulus works in one case, before trying another, write down *why* it worked in one line, and generalise that reason.

**4.** The trap is to believe the tower exponents need to be computed. They never do: modulo $10^d + 1$ only $e$ mod $2d$ matters, and a tower of 10s is divisible by any modest power of 2. A second trap is forgetting to show the divisor is proper ($M < N$), which is where "not prime" is actually proved.

**5.** Slogan: *a sum of powers of b plus or minus* 1 *that should factor: try* $b^d + 1$*, where* $b^d$ *is a sign.* The exponent $d$ is then read off from which exponents must be odd multiples of $d$.

**6.** Decision rule: **when a modulus disposes of some cases and not others, extract the property that made it work and parametrise the modulus by that property, instead of searching for a new modulus.** Companion rule: for $b^e \bmod (b^d + 1)$, only the parity of $e/d$ matters, so strip the odd part.

▷ **Cue:** a huge number of the form $\sum \pm b^{e_i} \pm 1$ must be shown composite $\longrightarrow$ **Move:** reduce modulo $b^d + 1$ (or $b^d - 1$), where each $b^e$ becomes $\pm 1$ according to the parity of $e/d$

▷ **Cue:** a modulus works for $n$ odd but not for $n$ even $\longrightarrow$ **Move:** replace $b + 1$ by $b^{2^m} + 1$ with $2^m \parallel n$, so that $n/2^m$ is odd again

▷ **Cue:** an exponent is itself a tower $10^{10^n}$ $\longrightarrow$ **Move:** do not compute it; only its residue modulo the order of the base matters, and a power of 10 is divisible by every small power of 2

▷ **Cue:** a divisibility argument proves "not prime" $\longrightarrow$ **Move:** finish by checking $1 < M < N$

## 2010 A5

**2010 A5.** Let $G$ be a group, with operation $*$. Suppose that

(i) $G$ is a subset of $\mathbb{R}^3$ (but $*$ need not be related to addition of vectors);

(ii) For each $\mathbf{a}, \mathbf{b} \in G$, either $\mathbf{a} \times \mathbf{b} = \mathbf{a} * \mathbf{b}$ or $\mathbf{a} \times \mathbf{b} = 0$ (or both), where $\times$ is the usual cross product in $\mathbb{R}^3$.

Prove that $\mathbf{a} \times \mathbf{b} = 0$ for all $\mathbf{a}, \mathbf{b} \in G$.

**Fields:** abstract algebra (group axioms) meeting vector geometry (cross product) **Position:** A5 — hard for most contestants because it looks like it needs structure theory; in fact it needs only the identity, one associativity check, and the geometry of $\times$. **Answer:** Every group of this kind lies on a single line through the origin; a pair with nonzero cross product forces the identity to be $\mathbf{0}$ and then contradicts associativity.

### Layer 1 THE WHAT — surface analysis

Read (ii) as: *whenever* $\mathbf{a} \times \mathbf{b} \neq \mathbf{0}$*, the group product* $\mathbf{a} * \mathbf{b}$ *is the cross product.* When $\mathbf{a} \times \mathbf{b} = \mathbf{0}$ (the vectors are parallel, or one is $\mathbf{0}$), nothing is known about $\mathbf{a} * \mathbf{b}$.

*The natural attempt.* Suppose, for contradiction, $\mathbf{a} \times \mathbf{b} \neq \mathbf{0}$. Then $\mathbf{c} = \mathbf{a} * \mathbf{b} = \mathbf{a} \times \mathbf{b} \in G$, and $\mathbf{a}, \mathbf{b}, \mathbf{c}$ are linearly independent. The instinct is to generate: $\mathbf{a} \times \mathbf{c}$, $\mathbf{b} \times \mathbf{c}$, $\mathbf{c} \times (\mathbf{a} \times \mathbf{c})$, … all lie in $G$ and are products. You now have a growing set of vectors with a partially known multiplication table, and the plan is to find two ways of computing the same product that disagree.

The difficulty: every product you know is a cross product, and the cross product, restricted to non-parallel pairs, is not associative — but you cannot use that directly, because the triple products you would compare typically involve some pair that *is* parallel (for instance $\mathbf{a} * \mathbf{a}$), whose product is unknown. Students try $(\mathbf{a} * \mathbf{b}) * \mathbf{b}$ versus $\mathbf{a} * (\mathbf{b} * \mathbf{b})$ and stall on $\mathbf{b} * \mathbf{b}$; they try inverses and stall on not knowing where $\mathbf{a}^{-1}$ is. The table grows, and every attempted contradiction runs into an unknown entry on the diagonal.

*The model chain.* *Step 1: the identity is* $\mathbf{0}$. Let $\mathbf{e}$ be the identity of $G$. For any $\mathbf{x} \in G$, if $\mathbf{e} \times \mathbf{x} \neq \mathbf{0}$ then by (ii) $\mathbf{x} = \mathbf{e} * \mathbf{x} = \mathbf{e} \times \mathbf{x}$, which is orthogonal to $\mathbf{x}$; so $\mathbf{x} \cdot \mathbf{x} = 0$, $\mathbf{x} = \mathbf{0}$, and then $\mathbf{e} \times \mathbf{x} = \mathbf{0}$

after all — contradiction. Hence $\mathbf{e} \times \mathbf{x} = \mathbf{0}$ for every $\mathbf{x} \in G$. Now suppose $\mathbf{a}, \mathbf{b} \in G$ with $\mathbf{a} \times \mathbf{b} \neq \mathbf{0}$; then $\mathbf{a}, \mathbf{b}$ are linearly independent and $\mathbf{e}$ is parallel to both (or zero), which forces $\mathbf{e} = \mathbf{0}$.

*Step 2: set up the associativity test.* Put $\mathbf{c} = \mathbf{a} \times \mathbf{b} = \mathbf{a} * \mathbf{b} \in G$. Since $\mathbf{c} \perp \mathbf{a}$ and both are nonzero, $\mathbf{a} \times \mathbf{c} \neq \mathbf{0}$, so

$$\mathbf{a} * (\mathbf{a} * \mathbf{b}) = \mathbf{a} * \mathbf{c} = \mathbf{a} \times (\mathbf{a} \times \mathbf{b}) = (\mathbf{a} \cdot \mathbf{b})\,\mathbf{a} - |\mathbf{a}|^2\,\mathbf{b} =: \mathbf{u}.$$

Its component along $\mathbf{b}$ satisfies

$$\mathbf{u} \cdot \mathbf{b} = (\mathbf{a} \cdot \mathbf{b})^2 - |\mathbf{a}|^2|\mathbf{b}|^2 < 0,$$

strictly, by Cauchy–Schwarz with $\mathbf{a}, \mathbf{b}$ independent. So $\mathbf{u} \neq \mathbf{b}$ (since $\mathbf{b} \cdot \mathbf{b} > 0$) and $\mathbf{u} \not\perp \mathbf{b}$.

*Step 3: locate* $\mathbf{s} = \mathbf{a} * \mathbf{a}$. Associativity gives $\mathbf{u} = (\mathbf{a} * \mathbf{a}) * \mathbf{b} = \mathbf{s} * \mathbf{b}$. If $\mathbf{s} \times \mathbf{b} \neq \mathbf{0}$, then $\mathbf{u} = \mathbf{s} \times \mathbf{b} \perp \mathbf{b}$, contradicting Step 2. So $\mathbf{s} \times \mathbf{b} = \mathbf{0}$.

The pair $(\mathbf{a}, \mathbf{c})$ satisfies every hypothesis the pair $(\mathbf{a}, \mathbf{b})$ did: $\mathbf{c} \in G$ and $\mathbf{a} \times \mathbf{c} \neq \mathbf{0}$. Steps 2–3 used only that, so they give $\mathbf{s} \times \mathbf{c} = \mathbf{0}$ too. Now $\mathbf{s}$ is either $\mathbf{0}$ or parallel to both $\mathbf{b}$ and $\mathbf{c}$, which are independent ($\mathbf{c} \perp \mathbf{b}$, both nonzero). Hence $\mathbf{s} = \mathbf{a} * \mathbf{a} = \mathbf{0} = \mathbf{e}$.

*Step 4: contradiction.* Then $\mathbf{u} = (\mathbf{a} * \mathbf{a}) * \mathbf{b} = \mathbf{e} * \mathbf{b} = \mathbf{b}$, but Step 2 showed $\mathbf{u} \neq \mathbf{b}$. So no pair with $\mathbf{a} \times \mathbf{b} \neq \mathbf{0}$ exists.

**Caveat.** This chain is the archive's own; it differs from Kedlaya–Ng's (three lemmas, working with the mutually orthogonal triple $\mathbf{b}, \mathbf{c}, \mathbf{b} \times \mathbf{c}$ and the fact that commuting elements have zero cross product). Both have been checked line by line. The identity $\mathbf{u} \cdot \mathbf{b} = (\mathbf{a} \cdot \mathbf{b})^2 - |\mathbf{a}|^2|\mathbf{b}|^2$ was also confirmed numerically on random vectors. The official Step 1 reaches $\mathbf{e} = \mathbf{0}$ via commutation; the direct argument above ($\mathbf{e} * \mathbf{x} = \mathbf{x}$ cannot be orthogonal to $\mathbf{x}$) is shorter and equally rigorous.

### Layer 2 THE HOW — the bifurcation

**Bifurcating point.** Both chains start at (†): "assume $\mathbf{a} \times \mathbf{b} \neq \mathbf{0}$, so $\mathbf{a} * \mathbf{b} = \mathbf{a} \times \mathbf{b} \in G$." At (†) the natural attempt starts manufacturing new elements of $G$ and hunting for a clash. The model instead asks what the group axioms say that the cross product cannot satisfy, starting with the one axiom that names a specific element: the identity.

> *The unknown entries in my table are exactly the products of parallel vectors. The identity is the one element whose products I know by the axioms, not by (ii). What does* $\mathbf{e} * \mathbf{x} = \mathbf{x}$ *say if it had to be a cross product? A cross product is orthogonal to* $\mathbf{x}$*, and* $\mathbf{x}$ *is not orthogonal to itself. So* $\mathbf{e}$ *is parallel to everything, so* $\mathbf{e} = \mathbf{0}$*.*

> *I keep getting stuck on* $\mathbf{a} * \mathbf{a}$*. Fine: make it the unknown. Associativity* $\mathbf{a} * (\mathbf{a} * \mathbf{b}) = (\mathbf{a} * \mathbf{a}) * \mathbf{b}$ *has a fully known left side, so it is an equation for* $\mathbf{a} * \mathbf{a}$*. If* $\mathbf{a} * \mathbf{a}$ *were not parallel to* $\mathbf{b}$*, the right side would be orthogonal to* $\mathbf{b}$*, and the left side is not.*

### Layer 3 THE WHY

Why the identity first? Look at what (ii) gives: it pins the product only on non-parallel pairs, and leaves the parallel pairs — including every square $\mathbf{x} * \mathbf{x}$ — completely free. So the problem is a partially filled multiplication table, and a contradiction must come from a group axiom forcing a value the table does not allow. Of the axioms (closure, associativity, identity, inverses), only the identity axiom specifies a product outright: $\mathbf{e} * \mathbf{x} = \mathbf{x}$. That is the circle-and-centre moment. A group is *defined* by the existence of an element whose product with $\mathbf{x}$ is $\mathbf{x}$; a cross product with $\mathbf{x}$ is *defined* to be orthogonal to $\mathbf{x}$; those two definitions collide the moment you put

them side by side. There is no trick in Step 1; it is the first line you write if you list the axioms and ask which one says something concrete.

Why associativity with $\mathbf{a}$ repeated? Because the free entries are the squares, you want an associativity equation in which exactly one side contains a square and the other side is entirely determined. $\mathbf{a} * (\mathbf{a} * \mathbf{b})$ is determined (two non-parallel products in succession), $(\mathbf{a} * \mathbf{a}) * \mathbf{b}$ contains the square. The equation then constrains $\mathbf{a} * \mathbf{a}$, and the only information a cross product carries is orthogonality, so the constraint is "$\mathbf{a} * \mathbf{a} \parallel \mathbf{b}$". Doing it again with $\mathbf{c}$ in place of $\mathbf{b}$ traps $\mathbf{a} * \mathbf{a}$ on two different lines, so it is $\mathbf{0}$. The whole proof is: *isolate the unknown entry, then corner it geometrically.*

Kedlaya's solution begins with the lemma "if $\mathbf{x}, \mathbf{y}$ are nonzero and orthogonal then $\mathbf{x} * \mathbf{x}$ is parallel to $\mathbf{y}$". Here is where that comes from: it is Step 3 above, specialised to an orthogonal pair so that $\mathbf{x} \times (\mathbf{x} \times \mathbf{y}) = -|\mathbf{x}|^2\mathbf{y}$ is visibly parallel to $\mathbf{y}$ rather than merely non-orthogonal to it. Their Lemma 3 (commuting elements have zero cross product) comes from $\mathbf{y} \times \mathbf{x} = -\mathbf{x} \times \mathbf{y}$: the cross product is anticommutative, so any group identity that forces commutation — the identity element, or elements of order 2 — breaks it. Same collision, different axiom.

Note what the answer means: every such $G$ lies on a line through $\mathbf{0}$, and on a line (ii) is vacuous, so any group structure on any subset of a line qualifies. The problem is not claiming $G$ is small, only that the cross product can never be doing the multiplying.

### Layer $+\alpha$ BACKTESTING

**1.** The first useful cue is "(but $*$ need not be related to addition of vectors)" together with "either $\dots$ or $\mathbf{a} \times \mathbf{b} = 0$". Write down: *non-parallel pairs multiply by $\times$; parallel pairs are unknown*. Then list the group axioms and ask which one determines a product without using (ii).

**2.** Contradiction via axioms: test each axiom (identity, associativity) against the known properties of $\times$ — orthogonality to both factors, anticommutativity, the expansion $\mathbf{x} \times (\mathbf{x} \times \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y})\mathbf{x} - |\mathbf{x}|^2\mathbf{y}$.

**3.** Hesitation occurs when products like $\mathbf{a} * \mathbf{a}$ appear and are unknown. Reinforce: an unknown quantity inside an equation is not a dead end, it is the variable you are solving for. Choose the associativity instance in which it appears once and everything else is known.

**4.** The trap is generating elements. Closure makes $G$ contain ever more vectors, and a student can fill a page with triple products without ever confronting an axiom. A second trap: assuming $\mathbf{a} * \mathbf{a}$ is something natural like $\mathbf{0}$ or $\mathbf{a}$ — it must be derived.

**5.** Slogan: *a hybrid structure where the operation is known only on part of the domain: find the axiom that pins a value outright (usually the identity), then use associativity to corner the unknown entries.*

**6.** Decision rule: **when a problem forces a group law to agree with a known operation only some of the time, compare each group axiom against the structural properties of the known operation (orthogonality, anticommutativity, non-associativity) and look for the collision; do not expand the set of elements.**

▷ **Cue:** a group operation is said to coincide with a known operation except in degenerate cases $\longrightarrow$ **Move:** write the identity axiom $\mathbf{e} * \mathbf{x} = \mathbf{x}$ in terms of the known operation and see what it forces on $\mathbf{e}$

▷ **Cue:** an unknown product (such as a square $\mathbf{a} * \mathbf{a}$) blocks every computation $\longrightarrow$ **Move:** pick an associativity instance with that product on one side and only known products on the other; solve for it

▷ **Cue:** a vector is known to be a cross product $\mathbf{s} \times \mathbf{b}$ $\longrightarrow$ **Move:** it is orthogonal to $\mathbf{b}$; test the other side of the equation for orthogonality to $\mathbf{b}$

▷ **Cue:** a vector is parallel to two linearly independent vectors $\longrightarrow$ **Move:** it is $\mathbf{0}$

▷ **Cue:** $\mathbf{x} \times (\mathbf{x} \times \mathbf{y})$ appears $\longrightarrow$ **Move:** expand as $(\mathbf{x} \cdot \mathbf{y})\mathbf{x} - |\mathbf{x}|^2\mathbf{y}$; its dot product with $\mathbf{y}$ is $-|\mathbf{x} \times \mathbf{y}|^2$

## 2010 A6

**2010 A6.** Let $f : [0,\infty) \to \mathbb{R}$ be a strictly decreasing continuous function such that $\lim_{x\to\infty} f(x) = 0$. Prove that $\int_0^\infty \frac{f(x)-f(x+1)}{f(x)}\,dx$ diverges.

**Fields:** real analysis (improper integrals, series with ratio terms, telescoping) **Position:** A6 — the hardest slot on the paper, and hard in the specific way that the obvious comparison (with a logarithm) almost works and then fails on a set you cannot control. The short proof is three lines once the right discrete fact is in view. **Answer:** The integral is $+\infty$: over suitably chosen integer blocks $[a,b]$ it is at least $\frac{1}{2}$ each, and there are infinitely many disjoint blocks.

### Layer 1 THE WHAT — surface analysis

Set $g(x) = \dfrac{f(x)-f(x+1)}{f(x)} = 1 - \dfrac{f(x+1)}{f(x)}$. Since $f$ strictly decreases to 0, $f(x) > 0$ for all $x$ and $0 < f(x+1) < f(x)$, so $g$ is continuous with $0 < g(x) < 1$. The integral $\int_0^T g$ is increasing in $T$; "diverges" means it is unbounded.

*The natural attempt.* The integrand is $1 - r(x)$ with $r(x) = f(x+1)/f(x) \in (0,1)$. The quantity $-\log r(x) = \log f(x) - \log f(x+1)$ telescopes under integration:

$$\int_0^T \log\frac{f(x)}{f(x+1)}\,dx = \int_0^1 \log f(x)\,dx - \int_T^{T+1} \log f(x)\,dx \longrightarrow +\infty,$$

because $\log f \to -\infty$. So if $1 - r \geq c\,(-\log r)$ for some constant $c > 0$, you are done. You check: $h(r) = (1-r)/(-\log r)$ increases from 0 (as $r \to 0$) to 1 (as $r \to 1$). So the comparison holds with $c = h(\frac{1}{2}) = \frac{1}{2\log 2} > \frac{1}{2}$ whenever $r \geq \frac{1}{2}$ — and fails without bound as $r \to 0$.

Now the wall. Split $[0,\infty)$ into $E = \{x : r(x) < \frac{1}{2}\}$ and its complement. On $E$ the integrand exceeds $\frac{1}{2}$, so if $E$ has infinite total length you are done. If $E$ has finite length, you want to use the logarithm on the complement — but the telescoping identity needs the logarithm on *all* of $[0,T]$, and on $E$ the term $-\log r$ can be as large as it likes, so removing it can remove all the divergence. The official first solution rescues this with Lebesgue measure and a translation-invariant set $I \subseteq [0,1]$ of positive measure; it is correct and it is not something to invent in an exam hall.

*The model chain.* *Step 1: cut into unit strips and swap.* For integers $0 \leq a < b$,

$$\int_a^b g(x)\,dx = \sum_{k=a}^{b-1}\int_0^1 g(x+k)\,dx = \int_0^1 \sum_{k=a}^{b-1} \frac{f(x+k)-f(x+k+1)}{f(x+k)}\,dx.$$

(A finite sum, so the swap is only linearity of the integral.)

*Step 2: freeze the denominator.* Fix $x \in [0,1]$. Every numerator is positive, and every denominator satisfies $f(x+k) \leq f(x+a)$ because $k \geq a$ and $f$ decreases. So

$$\sum_{k=a}^{b-1} \frac{f(x+k)-f(x+k+1)}{f(x+k)} \;\geq\; \sum_{k=a}^{b-1} \frac{f(x+k)-f(x+k+1)}{f(x+a)} \;=\; \frac{f(x+a)-f(x+b)}{f(x+a)},$$

the last sum telescoping.

*Step 3: make the block uniform in $x$.* For $x \in [0,1]$, $f(x+b) \leq f(b)$ and $f(x+a) \geq f(a+1)$, so

$$\int_a^b g(x)\,dx \;\geq\; \int_0^1 \left(1 - \frac{f(x+b)}{f(x+a)}\right) dx \;\geq\; 1 - \frac{f(b)}{f(a+1)}.$$

Since $f(a+1) > 0$ and $f \to 0$, there is an integer $b > a$ with $f(b) \leq \frac{1}{2} f(a+1)$; for that $b$ the block contributes at least $\frac{1}{2}$.

*Step 4: infinitely many blocks.* Put $a_0 = 0$ and let $a_{j+1} > a_j$ be an integer with $f(a_{j+1}) \le \frac{1}{2} f(a_j + 1)$. The intervals $[a_j, a_{j+1}]$ are disjoint up to endpoints, so

$$\int_0^{a_J} g(x)\,dx \;\ge\; \frac{J}{2} \qquad \text{for every } J,$$

and since $g > 0$ the integral is unbounded: it diverges.

**Caveat.** Steps 1–4 are the second official solution (Paul Allen), checked line by line. Sanity checks: for $f = e^{-x}$ the integrand is the constant $1 - e^{-1}$; for $f = 1/(1+x)$ it is $1/(x+2)$, whose integral grows like $\log T$ — divergent, but only just, which shows the problem has no room for a lossy comparison.

### Layer 2 THE HOW — the bifurcation

**Bifurcating point.** Both chains start from $g = 1 - f(x+1)/f(x)$. At (†), "what can I compare this with?", the natural attempt compares $1 - r$ with $-\log r$, a *pointwise* inequality that fails where $r$ is small. The model instead cuts the integral into unit strips, sees that for each $x \in [0,1]$ the strips form a discrete sum $\sum(u_k - u_{k+1})/u_k$ with $u_k = f(x+k)$, and bounds *that sum* by telescoping — an inequality over a block of terms, which does not care how small any individual ratio is.

> *The shift by* 1 *inside* $f$ *means the integrand only talks to itself at integer spacing. So for each* $x \in [0,1]$ *the integral is really the series* $\sum_k (u_k - u_{k+1})/u_k$ *with* $u_k = f(x+k)$ *decreasing to* 0*. I know that series: replace every denominator by the largest one, and the block telescopes to* $1 - u_n/u_m$*.*

> *The log comparison is losing exactly when* $f$ *drops fast. But when* $f$ *drops fast, the numerators are big. So I should not compare term by term; I should compare a whole block, where a big drop is a gain, not a loss.*

### Layer 3 THE WHY

Why unit strips? The only structure in the integrand is the comparison of $f$ at $x$ and $x+1$. Nothing links $x$ to $x + \frac{1}{2}$. So the integral over $[0,\infty)$ is a family, indexed by $x \in [0,1]$, of independent discrete problems living on the lattices $x + \mathbb{Z}_{\ge 0}$. When a problem's only coupling is a shift by 1, the continuous variable modulo 1 is a parameter, not a variable — that is what the phrase $f(x+1)$ is telling you, as surely as "equidistant from the centre" tells you to draw the centre.

Why does the discrete sum diverge? This is the standard fact: *if* $u_k > 0$ *decreases to* 0*, then* $\sum(u_k - u_{k+1})/u_k = \infty$. Its proof is Step 2, and Step 2 is not a trick. The terms are "drop divided by current height". Replacing the current height by the starting height $u_m$ makes every term smaller (denominators only grow), and the numerators then add up to the total drop $u_m - u_n$, which is almost all of $u_m$ once $n$ is large. So every tail of the series contains a block worth at least $\frac{1}{2}$, and the Cauchy criterion fails. It is the same inequality as $\sum_{k=m}^{n-1} \frac{1}{k} \ge \frac{n-m}{n}$ read the other way round, and it is the discrete shadow of $\int \frac{-du}{u} = \infty$. The logarithm the natural attempt reached for is the right *answer*; the mistake was to use it pointwise. Freezing the denominator gets the same divergence with a bound that is uniform however unevenly $f$ falls.

Why is Step 3 needed at all? Because each $x$ gives its own series, and "diverges for each $x$" does not by itself say the $x$-integral diverges unless the divergence is controlled uniformly or you invoke monotone convergence. Step 3 is the uniform control: monotonicity of $f$ lets you replace $f(x+b)$ and $f(x+a)$ by the worst cases $f(b)$ and $f(a+1)$ across the whole strip, so a single $b$ serves every $x$ at once. (If you prefer, the pointwise version also closes: the partial

sums $S_K(x) = \sum_{k<K} g(x+k)$ are continuous and increase to $+\infty$ at every point of the compact $[0,1]$, so for each $M$ some single $S_K$ exceeds $M$ on all of $[0,1]$. That is Dini's argument, and it is exactly what Kedlaya's measure-theoretic first solution is doing the hard way.)

The third official solution (Rosenberg) is worth knowing as a different axiom: if the integral converged, then $\frac{f(a+1)}{f(a)} < \frac{1}{f(a)} \int_a^{a+1} f = \frac{1}{f(a)} \int_a^\infty (f(x) - f(x+1))\,dx \le \int_a^\infty g$, which tends to 0; so $g \to 1$, and a positive integrand tending to 1 has divergent integral. There the telescoping is in the integral of $f$ itself.

**Layer $+\alpha$ BACKTESTING**

**1.** The first useful cue is the shift $f(x+1)$ against $f(x)$ with a quotient: write down $\int_0^\infty g = \int_0^1 \sum_{k\ge 0} g(x+k)\,dx$ and name the inner object — a series $\sum(u_k - u_{k+1})/u_k$ with $u_k$ decreasing to 0.

**2.** Blockwise telescoping with a frozen denominator: bound a block of the series below by (total drop)/(starting height), then choose the block long enough that the drop is at least half the height.

**3.** Hesitation occurs after the logarithm comparison works "except where $f$ falls fast". Reinforce: a comparison that fails exactly where the terms are large is a comparison you are making at the wrong scale — move from individual terms to blocks.

**4.** The trap is the pointwise inequality $1 - r \ge c(-\log r)$, which is true only for $r$ bounded below and drives you into measure theory. A second trap is proving divergence for each $x$ and forgetting that an integral over $x$ needs a uniform (or monotone-convergence) argument.

**5.** Slogan: *drop over current height diverges; freeze the height and telescope the drops.* It recognises $\sum(u_k - u_{k+1})/u_k$, $\sum(1 - u_{k+1}/u_k)$, $\sum a_k/S_k$ with $S_k$ partial sums of a divergent series (Abel–Dini), and continuous versions with a shift.

**6.** Decision rule: **when an integrand compares $f(x)$ with $f(x+1)$ only, reduce modulo 1: the integral is an average over $x \in [0,1]$ of a series along $x + \mathbb{Z}$. Solve the series, then make the estimate uniform in $x$ using monotonicity.** Companion rule: when a termwise comparison fails on large terms, compare blocks instead.

▷ **Cue:** an integrand involving only $f(x)$ and $f(x+1)$ $\longrightarrow$ **Move:** write $\int_0^\infty = \int_0^1 \sum_{k\ge 0}(\cdot)(x+k)\,dx$ and study the series for fixed $x$

▷ **Cue:** a sum of terms (drop)/(current value) for a decreasing positive sequence $\longrightarrow$ **Move:** replace every denominator by the first one in the block; the numerators telescope to $1 - u_n/u_m$

▷ **Cue:** a comparison $1 - r \ge c(-\log r)$ that fails as $r \to 0$ $\longrightarrow$ **Move:** abandon termwise comparison; bound whole blocks where large drops help rather than hurt

▷ **Cue:** divergence is proved for each value of a parameter $x$ in a compact interval $\longrightarrow$ **Move:** make it uniform via monotonicity (worst-case endpoints) or via continuity and compactness (Dini) before integrating over $x$

▷ **Cue:** show $\sum$ or $\int$ of a positive quantity diverges $\longrightarrow$ **Move:** exhibit infinitely many disjoint blocks each contributing at least a fixed constant

# Putnam 2010 — Paper B

## 2010 B1

**2010 B1.** Is there an infinite sequence of real numbers $a_1, a_2, a_3, \ldots$ such that

$$a_1^m + a_2^m + a_3^m + \cdots = m$$

for every positive integer $m$?

**Fields:** algebra (inequalities), analysis (infinite series) **Position:** B1 – opener **Answer:** No such sequence exists.

### Layer 1 THE WHAT — surface analysis

*The natural attempt.* The instinct on seeing "$\sum a_i^m = m$ for every $m$" is to write down what the first few constraints actually say and hunt for a contradiction by direct estimation, since a construction that must work for *every* $m$ simultaneously feels too rigid to exist. So: $\sum a_i = 1$, $\sum a_i^2 = 2$, $\sum a_i^3 = 3$, $\sum a_i^4 = 4$. A natural next move is to bound the terms. Since $\sum a_i^2 = 2$ and every term $a_i^2$ is nonnegative, each individual $a_i^2 \leq 2$. Split the index set into those $k$ with $a_k^2 \in [0,1]$ and those with $a_k^2 \in (1,2]$. On the first set, $a_k^4 \leq a_k^2$ termwise (since $t^2 \leq t$ for $t \in [0,1]$), which is a clean, honest inequality. The trouble is the second set: it is not empty in general, and nothing forces it to be empty from $\sum a_i^2 = 2$ alone – a single index with $a_k^2$ close to 2 is entirely consistent with that one equation. So the case split "*if every* $a_k^2 \leq 1$" only resolves a special case, and finishing the general argument requires a second piece of casework: showing that if some $a_k^2 > 1$ then that term's high power eventually overwhelms the fixed target $m$ as $m$ grows. This works – it is exactly Kedlaya's second solution – but it costs two separate arguments glued together, and a solver who commits to the "$a_k^2 \leq 1$ for all $k$" branch first, without first checking whether that branch is the only one, has spent time on a proof that is only half the answer.

*The model chain.* Apply the Cauchy–Schwarz inequality to the two sequences $(a_i)$ and $(a_i^2)$:

$$\left(\sum_i a_i \cdot a_i^2\right)^2 \leq \left(\sum_i a_i^2\right)\left(\sum_i a_i^4\right),$$

that is, $\left(\sum_i a_i^3\right)^2 \leq \left(\sum_i a_i^2\right)\left(\sum_i a_i^4\right)$. If the sequence has the required property, the left side is $3^2 = 9$ and the right side is $2 \cdot 4 = 8$. But $9 \leq 8$ is false. So no such sequence exists.

### Layer 2 THE HOW — the bifurcation

**Bifurcating point.** Both paths reach exactly the same four numbers: $\sum a_i = 1$, $\sum a_i^2 = 2$, $\sum a_i^3 = 3$, $\sum a_i^4 = 4$. They split on what to do with them. The natural attempt treats $\sum a_i^2 = 2$ and $\sum a_i^4 = 4$ as separate facts to be bounded one at a time, term by term. The model treats them as the two halves of a single Cauchy–Schwarz product, with $\sum a_i^3 = 3$ sitting exactly where the cross term belongs.

> *I have four power sums and I am trying to find a relation* between *them rather than a bound on each one separately. Cauchy–Schwarz is the standard tool for turning two power sums into a bound on a third, so before I estimate anything termwise I should ask: do three of these four numbers fit the pattern* $(\sum uv)^2 \le (\sum u^2)(\sum v^2)$ *for some choice of* $u_i, v_i$ *built from* $a_i$*?*

> *Powers* 2, 3, 4 *are an arithmetic progression with common difference* 1*, and* $3 = \frac{2+4}{2}$*. That is exactly the shape Cauchy–Schwarz wants: set* $u_i = a_i$ *(power* 1*) and* $v_i = a_i^2$ *(power* 2*), so* $u_i v_i = a_i^3$ *(power* 3*, the average) and* $u_i^2, v_i^2$ *are powers* 2 *and* 4*. The inequality is available the moment I notice that* 3 *is the arithmetic mean of* 2 *and* 4 *among the exponents I am given values for.*

**Layer 3 THE WHY**

The reason Cauchy–Schwarz is not a lucky guess here is that the problem hands over four consecutive-ish power sums and asks for existence, and existence questions built on sums of powers are, structurally, questions about how fast a sequence of nonnegative numbers $(a_i^2, a_i^4, \dots)$ can grow relative to a sequence that is not sign-definite $(a_i, a_i^3, \dots)$. Cauchy–Schwarz is precisely the inequality that converts two even power sums into an upper bound on the square of the power sum halfway between them, because $(\sum u_i v_i)^2 \le (\sum u_i^2)(\sum v_i^2)$ is nothing more than the statement that a covariance-type quantity cannot exceed the product of two variance-type quantities – and here the "halfway" exponent falls exactly on an integer, 3, because 2 and 4 are two apart. This is the deeper fact the case-split proof obscures: the problem is not really asking you to control the size of individual terms $a_i$ at all. It is asking whether the four numbers $(1, 2, 3, 4)$ are jointly realisable as consecutive power sums of one sequence, and joint realisability of that kind is governed by an inequality on the sums, not by bounding the terms that produced them. Kedlaya's first solution states the Cauchy–Schwarz application as a two-line computation with no discussion of why it is the tool to reach for; the reason is that the moment three exponents sit in arithmetic progression, the middle one is available as a cross term, and testing whether the given values obey that one inequality is strictly cheaper than any argument that inspects the sequence term by term.

**Layer $+\alpha$ BACKTESTING**

**1.** The first useful cue is the shape of the exponents supplied by the problem's own hypothesis, $m = 1, 2, 3, 4$, taken together: three of them, $2, 3, 4$, are consecutive integers. On seeing several fixed power sums of the same unknown sequence, write down immediately whether any three exponents are in arithmetic progression.

**2.** That cue suggests the method: Cauchy–Schwarz applied to $(a_i^p)$ and $(a_i^q)$ for the two outer exponents, which bounds the power sum at the middle exponent by the geometric mean of the two outer power sums – a relation to test against the given numbers rather than a construction to attempt.

**3.** Hesitation occurs right after writing down the four power-sum values, at the decision of what to do with them: the reflex is to start bounding $a_k^2$ termwise (case-split on size), because that feels like "making progress" on the object itself. Reinforce it away by asking, before any casework: is there an inequality relating these sums directly, without reference to individual terms?

**4.** The trap is treating "does such a sequence exist" as a construction problem by default. Existence questions phrased as a list of fixed aggregate constraints (sums, power sums, integrals) are frequently disproved by a single inequality among the constraints, and reaching for construction first wastes the observation that the numbers $1, 2, 3, 4$ might simply be jointly inconsistent.

**5.** Recognisable faster next time by treating "several power sums (or moments) of one unknown sequence are prescribed" as an automatic cue to check Cauchy–Schwarz (or, more generally,

Hölder) among any three exponents in arithmetic progression before attempting any direct estimate on the terms.

**6.** Reusable rule: when a problem fixes several power sums $\sum a_i^{p_1}, \ldots, \sum a_i^{p_k}$ of one sequence and asks about existence, first test whether any three exponents $p < q < r$ in arithmetic progression violate $(\sum a_i^q)^2 \le (\sum a_i^p)(\sum a_i^r)$ (Cauchy–Schwarz with $u_i = a_i^p, v_i = a_i^r$, wait – more precisely with $u_i = a_i^{(p+r)/2-(r-q)}$-style splitting so that $u_i v_i = a_i^q$); this single inequality test is cheaper than any termwise bound and often settles existence outright.

▷ **Cue:** several power sums (or moments) of a single unknown sequence are fixed by hypothesis and existence is being questioned $\longrightarrow$ **Move:** check whether any three of the given exponents form an arithmetic progression $p, q, r$; if so, test the given values against Cauchy–Schwarz, $(\sum a_i^q)^2 \le (\sum a_i^p)(\sum a_i^r)$, before attempting a direct construction or termwise bound

▷ **Cue:** a case-split argument ("suppose every term satisfies $X$") resolves only part of the index set and a second, structurally different argument is needed for the rest $\longrightarrow$ **Move:** pause and check whether an inequality relating the given aggregate quantities directly (Cauchy–Schwarz, power-mean, rearrangement) replaces both cases with one computation

## 2010 B2

**2010 B2.** Given that $A$, $B$, and $C$ are noncollinear points in the plane with integer coordinates such that the distances $AB$, $AC$, and $BC$ are integers, what is the smallest possible value of $AB$?

**Fields:** number theory / geometry (lattice points, Pythagorean-type constraints) **Position:** B2 **Answer:** The smallest possible value is 3, achieved by $A = (0,0)$, $B = (3,0)$, $C = (0,4)$.

### Layer 1 THE WHAT — surface analysis

*The natural attempt.* The worked example format of this competition rewards guessing small and checking, so the natural attempt is to search directly: try $AB = 1$, then $AB = 2$, then $AB = 3$, looking for lattice triangles with all three integer side lengths. A 3-4-5 right triangle with legs on the axes is found almost immediately at $AB = 3$, and it is tempting to stop there and declare 3 the answer once a valid configuration is in hand. But the problem asks for the *smallest* possible value, which is a claim about *all* configurations with $AB = 1$ or $AB = 2$, not just the ones tried. A natural but incomplete attempt tries a handful of lattice points at distance 1 or 2 apart – $(0,0), (1,0)$; $(0,0), (2,0)$ – searches a few candidate $C$ nearby, fails to find integer $AC, BC$, and concludes by pattern (not proof) that 1 and 2 are impossible. This reaches the right answer but supplies no argument that would survive a rigorous mark scheme, because it has not ruled out every lattice point $C$, only the ones checked by hand. The gap is invisible precisely because the correct answer was found early and the impossibility direction was never actually forced to produce a proof.

*The model chain.* First, note $AB \neq 0$ (else the points cannot be noncollinear) and search whether $AB = 1$ or $AB = 2$ can be excluded outright.

$AB = 1$. An integer vector of length 1 has $(dx, dy)$ with $dx^2 + dy^2 = 1$, so $(dx, dy) \in \{(\pm 1, 0), (0, \pm 1)\}$; translate and rotate coordinates within the lattice so $A = (0,0)$, $B = (1,0)$. For any $C = (x, y)$ with integer $x, y$, the triangle inequality gives $|AC - BC| \le AB = 1$, with equality only if $A, B, C$ are collinear. Since $C$ is not collinear with $A, B$, $|AC - BC| < 1$; as $AC, BC$ are both integers their difference is an integer, forcing $AC = BC$. Equal distances to $A = (0,0)$ and $B = (1,0)$ force $x = \frac{1}{2}$, impossible for integer $x$. So $AB = 1$ is impossible.

$AB = 2$. An integer vector of length 2 satisfies $dx^2 + dy^2 = 4$; the only representation of 4 as a sum of two squares is $4 + 0$, so again $(dx, dy) \in \{(\pm 2, 0), (0, \pm 2)\}$ and we may take $A = (0,0), B = (2,0)$. For $C = (x, y)$, $AC^2 = x^2 + y^2$ and $BC^2 = (2 - x)^2 + y^2$, so $AC^2 - BC^2 = 4x - 4$ is even; hence $AC, BC$ have the same parity, so $AC - BC$ is even. Combined

with $|AC - BC| \leq 1$ (strict inequality $< 2$ from noncollinearity, and integer), this forces $AC - BC = 0$, i.e. $AC = BC$, hence $x = 1$. Then $AC^2 = 1 + y^2 = BC^2$ must be a perfect square $k^2$, so $k^2 - y^2 = 1$, i.e. $(k - y)(k + y) = 1$ over the integers, forcing $y = 0, k = 1$ – but $y = 0$ puts $C$ on line $AB$, collinear, a contradiction. So $AB = 2$ is impossible.

$AB = 3$ *is achievable*: take $A = (0,0)$, $B = (3,0)$, $C = (0,4)$. Then $AB = 3$, $AC = 4$, $BC = \sqrt{9+16} = 5$, all integers, and the three points are noncollinear. Hence the smallest possible value of $AB$ is 3.

**Layer 2 THE HOW — the bifurcation**

**Bifurcating point.** Both paths reach a 3-4-5 triangle with $AB = 3$ almost immediately. They split on what happens to the claim "$AB = 1$ and $AB = 2$ are impossible." The natural attempt treats a short, unsuccessful hand search over nearby lattice points as sufficient evidence. The model instead fixes $A, B$ completely (forced by the small integer distance itself), introduces a free lattice point $C = (x, y)$, and derives two hard algebraic constraints – the triangle inequality plus a parity argument – that hold for *every* choice of $C$, not just the ones checked.

> *I cannot search all lattice points $C$ by hand, so the impossibility claim for $AB = 1, 2$ needs an argument that works for an arbitrary $(x, y)$, not a finite check. The one relation guaranteed to hold for every $C$, integer or not, is the triangle inequality $|AC - BC| \leq AB$ – and it becomes an equality* exactly *when $A, B, C$ are collinear, which is precisely the case the problem has excluded. So the inequality is strict here, and since $AC, BC$ are both forced to be integers, "strictly less than an integer" is a genuinely restrictive statement.*

> *For $AB = 1$, strict inequality plus integrality immediately collapses $|AC - BC|$ to 0: there is no integer strictly between 0 and 1. For $AB = 2$ that collapse does not happen for free – $|AC - BC|$ could a priori be 0 or 1 – so I need a second, independent constraint to rule out the value 1. The natural place to find it is parity: $AC^2 - BC^2$ is a specific integer expression in $x$ that I can compute directly, and if it always comes out even, then $AC, BC$ share parity and $AC - BC = 1$ is impossible on its own.*

**Layer 3 THE WHY**

The triangle inequality's equality case is not a technicality to recite – it is doing the entire work of the problem. $|AC - BC| \leq AB$ is equality if and only if $C$ lies on the ray from $A$ through $B$ or beyond (collinearity with $A, B$ on the correct side), because the triangle inequality for three points is itself just the statement that a path through $C$ is never shorter than the direct path, with equality exactly when there is no genuine detour – i.e. when the three points are already collinear. The problem's hypothesis that $A, B, C$ are *noncollinear* is therefore not incidental scene-setting; it is the clause that converts a non-strict, always-true inequality into a strict one, and a strict inequality between integers is what forces $AC = BC$ (when $AB = 1$) or forces the search down to a single residual case (when $AB = 2$). Kedlaya's solution invokes the equality condition of the triangle inequality without pausing on why it is available: it is available because "noncollinear" is precisely the negation of "equality holds." For $AB = 2$, the triangle inequality alone leaves two candidates, $AC - BC \in \{0, 1\}$, and the parity computation $AC^2 - BC^2 = 4x - 4 \equiv 0 \pmod 2$ is what a lattice-point problem should be expected to supply: squares of integers of matching parity are themselves congruent mod 4 in the relevant sense, and the difference $AC^2 - BC^2$ being manifestly even is a fact about the coordinates alone, unrelated to the geometry, which is exactly why it is available as a second, independent constraint on top of the triangle inequality. The two arguments together – one purely metric (triangle inequality), one purely arithmetic (parity of a difference of squares) – are the two kinds of constraint a lattice-point-with-integer-distances problem always has on offer, and knowing to reach for both, rather than stopping once one produces *a* constraint, is the transferable content here.

**Layer $+\alpha$ BACKTESTING**

**1.** The first useful cue is the smallness of the target: the problem asks for the minimum of an integer-valued quantity ($AB$), and integer distances plus lattice points is exactly the setting where the triangle inequality's equality case is unusually powerful. On seeing "smallest integer distance, noncollinear lattice points," write down immediately that noncollinearity will be used to make a triangle inequality strict.

**2.** That cue suggests the method: fix the two points at the candidate distance (using that an integer-length lattice vector has very few representations for small lengths), introduce the third point as a free lattice point, and derive constraints from (a) the strict triangle inequality and (b) parity/algebraic identities in the coordinates, rather than searching finitely many nearby lattice points by hand.

**3.** Hesitation occurs at the moment a valid example (3-4-5) is found for $AB = 3$ – the pull is to stop, since the answer "looks right." Reinforce it away by asking explicitly: has the impossibility of every smaller value actually been proved, or only failed to be found by search?

**4.** The trap is conflating "I could not find a configuration with $AB = 1$ or 2 after a short search" with "no such configuration exists." A minimisation problem always has two halves – an example achieving the value, and a proof that nothing smaller works – and the second half is never discharged by a failed search.

**5.** Recognisable faster next time by treating any "find the minimum integer distance among lattice points with integer pairwise distances" problem as an automatic cue to reach for the triangle-inequality equality condition first (it is often the cheapest tool available), and to look for a parity or modular obstruction as the natural second constraint once the triangle inequality alone is not sharp enough.

**6.** Reusable rule: for lattice-point problems with prescribed integer distances, noncollinearity converts the triangle inequality from non-strict to strict, and a strict inequality between integers collapses the range of possibilities to finitely many (often just one) residual cases; resolve those residual cases with a parity or modular identity computed directly in the coordinates, never by finite hand search.

▷ **Cue:** noncollinear points with integer pairwise distances, minimise one distance $\longrightarrow$ **Move:** use the triangle inequality's equality-exactly-at-collinearity fact to turn noncollinearity into a strict inequality between the two remaining integer distances, collapsing the possibilities to a small residual set

▷ **Cue:** a strict inequality between two integers leaves more than one residual case (e.g. difference 0 or 1) $\longrightarrow$ **Move:** compute a parity or modular identity directly from the coordinates (e.g. $AC^2 - BC^2$ as a linear expression in $x$) to eliminate the remaining cases algebraically

## 2010 B3

**2010 B3.** There are 2010 boxes labeled $B_1, B_2, \ldots, B_{2010}$, and $2010n$ balls have been distributed among them, for some positive integer $n$. You may redistribute the balls by a sequence of moves, each of which consists of choosing an $i$ and moving *exactly* $i$ balls from box $B_i$ into any one other box. For which values of $n$ is it possible to reach the distribution with exactly $n$ balls in each box, regardless of the initial distribution of balls?

**Fields:** combinatorics (invariants, extremal / pigeonhole arguments) **Position:** B3 **Answer:** Exactly the integers $n \geq 1005$.

**Layer 1 THE WHAT — surface analysis**

*The natural attempt.* The problem supplies a move and asks for which $n$ that move can always reach a target, so the natural first instinct is constructive: design an algorithm that reaches "$n$

balls in every box" from an arbitrary start, and see what it needs. A reasonable first algorithm is greedy: whenever some box $B_i$ has at least $i$ balls (call such an index *eligible*), move all $i$ of those balls somewhere useful – say, funnel everything toward $B_1$, then redistribute from there one ball at a time. This is workable and, with care, becomes the correct construction. The natural attempt's real gap is on the *other* side of the "for which $n$" question. Having built a construction that works whenever an eligible box can always be found, the temptation is to declare "this works for all $n \geq 1$" without ever asking whether an eligible box is guaranteed to exist at every step, or whether some starting distribution can be built that never has one. A construction-only proof answers "if $n$ is large enough, here is how," but the problem demands an exact threshold, and an exact threshold is a two-sided claim: a working algorithm for $n$ above it, and a frozen, unmovable configuration for every $n$ at or below it. Skipping the second half produces a plausible-sounding but incomplete solution that never actually computes 1005.

*The model chain.* **Necessity ($n \leq 1004$ fails).** A move from $B_i$ requires $B_i$ to hold at least $i$ balls, so if $B_i$ holds at most $i-1$ balls for every $i = 1, \ldots, 2010$, no move is available at all, and the configuration is frozen. The maximum total the boxes can hold under this cap is $\sum_{i=1}^{2010}(i-1) = 0+1+\cdots+2009 = \frac{2009 \cdot 2010}{2} = 2010 \times 1004.5$. For $n \leq 1004$, the required total $2010n \leq 2010 \times 1004 < 2010 \times 1004.5$, so a distribution with $2010n$ balls respecting $B_i \leq i-1$ for all $i$ exists (there is slack, e.g. put everything into $B_{2010}$ up to its cap, then fill downward). Such a distribution is frozen and is not the target distribution (in particular $B_1$, capped at 0, cannot hold $n \geq 1$ balls), so it can never be transformed into "$n$ in every box." Hence $n \leq 1004$ fails.

**Sufficiency ($n \geq 1005$ works).** If $n \geq 1005$ then $2010n \geq 2010 \times 1005 > 2010 \times 1004.5 = \sum_{i=1}^{2010}(i-1)$, so no distribution can have every $B_i \leq i-1$: by pigeonhole, some index is always eligible (holds at least $i$ balls). The algorithm:

(a) Repeatedly find the largest eligible index $i > 1$ and move all $i$ balls from $B_i$ to $B_1$, until only $i = 1$ can possibly be eligible.

(b) While some $B_j$ ($j > 1$) is nonempty: move 1 ball from $B_1$ to $B_j$ (possible since $B_1$ is then eligible, having absorbed everything); if this makes $j$ eligible, immediately move those $j$ balls back to $B_1$. Repeat until every $B_j$ ($j > 1$) is empty and all balls sit in $B_1$.

(c) For $i = 2, \ldots, 2010$, move one ball at a time from $B_1$ to $B_i$, $n$ times each.

Each step is a legal move (moving from an eligible box, or moving 1 ball from a nonempty $B_1$), and the process terminates with exactly $n$ balls in every box. So $n \geq 1005$ always succeeds.

Combining both directions, the answer is exactly $n \geq 1005$.

**Layer 2 THE HOW — the bifurcation**

**Bifurcating point.** Both paths eventually build the same funnel-to-$B_1$-then-redistribute algorithm. They split much earlier, on what question gets asked first. The natural attempt asks "how do I reach the target?" and starts building machinery. The model first asks "what is the largest total that can be stuck with *no move available at all*?" – a question about the move rule's failure condition, not about reaching the target – and only then builds the algorithm, now knowing exactly which $n$ it needs to handle.

> *A move needs $B_i$ to have at least $i$ balls to move from it, so the real obstruction is not "can I reach the target," it is "can the process get permanently stuck with zero legal moves." I should find the worst-case frozen configuration – the one that squeezes the most balls into boxes while keeping every box under its own index – before I ever design an algorithm, because that number is exactly the boundary the problem is asking for.*

> *The cap "$B_i$ holds at most $i-1$ balls" is not arbitrary; it is the negation of "eligible," which is the negation of "a move is possible from $B_i$." So $\sum_{i=1}^{2010}(i-1)$ is the maximum number of balls that can sit in a completely frozen configuration, and any total below that value can be realised as a frozen configuration (by choosing which boxes get how much slack), while any total strictly above it* cannot *be frozen – some box must exceed its cap, by pigeonhole. That single sum, $\frac{2009\cdot 2010}{2}$, is the whole problem: the comparison against* $2010n$ *decides it.*

**Layer 3 THE WHY**

The reason to look for a frozen configuration before building an algorithm is that the move rule in this problem is inherently one-directional in its requirement: moving out of $B_i$ needs $B_i \geq i$, but nothing constrains how many balls a box may *receive*. A rule like this always has a natural "stuck" set – the configurations where the requirement fails everywhere simultaneously – and for any problem of the form "can this process always reach a target from any start," the stuck set is the first object to characterise, because if the stuck set can hold as many balls as the problem allows, the process is not universal, full stop, regardless of how well-designed the algorithm from unstuck states is. This is a specific instance of a general principle: in a reachability problem defined by a threshold move rule (do X only if some resource is $\geq$ a bound), the extremal configuration saturating every bound simultaneously is both the natural adversary (it is what a worst-case initial distribution would look like) and the exact number the sufficient-$n$ threshold must beat. Kedlaya's solution states the sum $\sum_{i=1}^{2010}(i-1) = 2010 \times 1004.5$ without dwelling on where it comes from; it falls out because $i-1$ is definitionally "the largest amount $B_i$ can hold while remaining ineligible," and summing that over every box computes the single largest island of complete paralysis the system can construct. Once that number is in hand, the direction of the inequality (whether $2010n$ exceeds it or not) is not incidental case-checking – it *is* the necessary-and-sufficient boundary, because exceeding it guarantees an eligible box by pigeonhole (which is precisely what the algorithm in part (a) needs at every stage to keep making progress), while sitting at or below it permits a configuration where the algorithm never even gets to start.

**Layer $+\alpha$ BACKTESTING**

**1.** The first useful cue is the asymmetry in the move rule: moving *out* of $B_i$ requires $B_i \geq i$, but there is no constraint on receiving balls. On seeing a move rule gated by a per-box threshold, write down immediately: what is the maximum total achievable while every box stays strictly under its own threshold?

**2.** That cue suggests the method: compute the extremal "frozen" sum ($\sum(i-1)$ here) before attempting any algorithm, since comparing the given total against that sum is what actually determines the answer – constructive machinery only needs to be built for the side of the comparison where it can possibly succeed.

**3.** Hesitation occurs at the moment a workable-looking algorithm is found for "large" $n$: the pull is to declare victory and move on, since the construction "feels like the content" of the problem. Reinforce it away by asking explicitly: have I shown a frozen configuration exists for every $n$ below my claimed threshold, using the exact same cap that defines eligibility?

**4.** The trap is answering a "for which $n$ does this always work" question with only a sufficiency argument (an algorithm), leaving the threshold's exact value unproven and, worse, often wrong by construction bias (an algorithm that happens to work for $n \geq 1$ in easy cases gives false confidence that no smaller threshold exists).

**5.** Recognisable faster next time by treating any move rule of the form "resource $r$ at location $\ell$ must be $\geq$ threshold $t(\ell)$ to act" as an automatic cue to compute $\sum_\ell (t(\ell)-1)$, the maximal simultaneously-sub-threshold total, as the first step – before any algorithm design.

**6.** Reusable rule: for a "reach the target from any start" question governed by a per-location threshold move rule, the answer is a strict inequality between the total resource and

$\sum_\ell (t(\ell) - 1)$; total resource exceeding this sum guarantees (by pigeonhole) an eligible location at every stage, which is what any funnel-and-redistribute algorithm needs, while a total at or below it admits a frozen counterexample built by capping each location just under its threshold.

▷ **Cue:** a redistribution/reachability problem where a move requires a location to hold at least a threshold amount before it can give resource away, with no cap on receiving $\longrightarrow$ **Move:** compute the maximum total achievable with every location strictly under its own threshold ($\sum(t(\ell) - 1)$); this sum is the exact boundary between "always reachable" and "a frozen counterexample exists"

▷ **Cue:** "for which $n$ does this process always succeed" is asked $\longrightarrow$ **Move:** treat it as a two-sided claim from the start: build the frozen extremal configuration (necessity) with the same attention as the algorithm (sufficiency), rather than treating construction as the whole answer

## 2010 B4

**2010 B4.** Find all pairs of polynomials $p(x)$ and $q(x)$ with real coefficients for which

$$p(x)q(x+1) - p(x+1)q(x) = 1.$$

**Fields:** algebra / polynomial identities; degree and leading-coefficient comparison; determinants and row operations **Position:** B4 — mid-paper; the answer is guessable in a minute, and the whole problem is ruling out higher degree without drowning in coefficients **Answer:** Exactly the pairs $p(x) = ax + b$, $q(x) = cx + d$ with $a, b, c, d \in \mathbb{R}$ and $bc - ad = 1$.

### Layer 1 THE WHAT — surface analysis

*The natural attempt.* Guess first. Constants fail: $p = b$, $q = d$ give $bd - bd = 0$. Linear polynomials work: for $p = ax + b$, $q = cx + d$,

$$(ax+b)(cx+c+d) - (ax+a+b)(cx+d) = bc - ad,$$

the $x^2$ and $x$ terms cancelling. So every pair with $bc - ad = 1$ is a solution, including those where one of $p, q$ is constant ($a = 0$ or $c = 0$). The real question is whether anything of higher degree works.

The strong student now compares coefficients. Write $m = \deg p$, $n = \deg q$, leading coefficients $p_m, q_n$. Expanding $q(x+1) = q(x) + q'(x) + \frac{1}{2}q''(x) + \cdots$ (Taylor's theorem is exact for polynomials),

$$p(x)q(x+1) - p(x+1)q(x) = \left(pq' - p'q\right) + \tfrac{1}{2}\left(pq'' - p''q\right) + \cdots,$$

and the top term of $pq' - p'q$ is $(n-m)p_m q_n\, x^{m+n-1}$. If $m + n \geq 2$ this coefficient must vanish, so $m = n$. Good progress: unequal degrees are dead.

Now $m = n \geq 2$, and the student tries $m = n = 2$ by brute force. With $p = p_0 + p_1 x + p_2 x^2$ and similarly $q$, the left side is

$$(p_1q_2 - p_2q_1)\,x^2 + \big(2(p_0q_2 - p_2q_0) + (p_1q_2 - p_2q_1)\big)\,x + \big(p_0q_1 - p_1q_0 + p_0q_2 - p_2q_0\big).$$

Setting the $x^2$ and $x$ coefficients to 0 forces $p_1q_2 = p_2q_1$ and $p_0q_2 = p_2q_0$, so $p$ and $q$ are proportional and the constant term is $0 \neq 1$. That settles degree 2 — after a page of algebra. Degree 3 is worse, and the general degree needs the $2 \times 2$ minors $p_iq_j - p_jq_i$ organised by a descending induction with binomial coefficients (this is exactly the official second solution). It works. It is also a long way round, and the student has not yet seen why proportionality keeps appearing.

*The model chain.* **Step 1: degrees must be equal.** Let $R(x) = p(x)q(x+1) - p(x+1)q(x)$. Neither $p$ nor $q$ is the zero polynomial, since then $R = 0$. With $m, n, p_m, q_n$ as above, the coefficient of $x^{m+n-1}$ in $R$ is $(n-m)p_m q_n$ (the other Taylor terms have degree at most $m+n-2$). If $m+n \geq 2$ then $R = 1$ has no $x^{m+n-1}$ term, so $m = n$.

**Step 2: read $R$ as a determinant.**

$$R(x) = \det \begin{pmatrix} p(x) & p(x+1) \\ q(x) & q(x+1) \end{pmatrix}.$$

Subtracting $t$ times the first row from the second does not change a determinant, so for every real $t$ the pair $(p,\ q - tp)$ is again a solution.

**Step 3: kill the high-degree case.** Suppose $m = n \geq 2$. Choose $t = q_m/p_m$; then $\tilde{q} = q - tp$ has degree less than $m$. It is not the zero polynomial, because $(p, \tilde{q})$ is a solution and the zero polynomial gives no solution. So $(p, \tilde{q})$ is a solution with $\deg p = m \neq \deg \tilde{q}$ and $m + \deg \tilde{q} \geq 2$, contradicting Step 1.

**Step 4: finish.** So $m = n \leq 1$, or $m \neq n$ with $m + n \leq 1$, i.e. $\{m, n\} = \{0, 1\}$. The case $m = n = 0$ gives $R = 0$. Every remaining case has $p = ax + b$, $q = cx + d$, and the computation in the natural attempt shows $R = bc - ad$. Hence the solutions are exactly those with $bc - ad = 1$. □

**Layer 2 THE HOW — the bifurcation**

**Bifurcating point.** Both chains reach "$\deg p = \deg q$" by the same leading-coefficient comparison. They split at the next line.

(N) "Now $m = n$; compare the next coefficient, and the next." The equal-degree case is attacked coefficient by coefficient.

(M) "Now $m = n$; can I change the pair so the degrees are no longer equal?" The equal-degree case is *reduced* to the case already killed.

On (N) every degree is a new system of bilinear equations. On (M) there is no second computation at all: one row operation sends the hard case back into Step 1.

> *The expression is $p(x)q(x+1) - p(x+1)q(x)$ — that's a $2 \times 2$ determinant with rows $p$ and $q$. Determinants don't change if I subtract a multiple of one row from another. So I'm free to replace $q$ by $q - tp$, and with the right $t$ that drops the degree of $q$ below the degree of $p$. Then the degrees are unequal, and I already know unequal degrees are impossible.*

> *Every time I grind the coefficients I end up with $p_i q_j = p_j q_i$, i.e. $p$ proportional to $q$. That's telling me the operation that matters is "$q$ minus a multiple of $p$", so I should use it directly instead of rediscovering it one coefficient at a time.*

- ▷ **Cue:** an expression of the shape $f(x)g(y) - f(y)g(x)$, or $pq' - p'q$ ⟶ **Move:** write it as a $2 \times 2$ determinant; row operations $g \mapsto g - tf$ preserve it
- ▷ **Cue:** a case analysis where one case (unequal degrees) dies by leading terms and the other (equal degrees) does not ⟶ **Move:** look for an operation preserving the identity that moves the stubborn case into the dead one
- ▷ **Cue:** a polynomial identity $F(p, q) =$ constant ⟶ **Move:** compare the top coefficient first; it usually forces a relation between the degrees
- ▷ **Cue:** a shift $q(x+1)$ inside a polynomial identity ⟶ **Move:** expand $q(x+1) = q + q' + \frac{1}{2}q'' + \cdots$ to see the leading behaviour as a derivative (Wronskian-type) expression

**Layer 3 THE WHY**

Why is the determinant the natural object here, and not a trick someone saw in a dream?

Start from the shape of what is given. The left side is antisymmetric under swapping $p$ and $q$: swap them and it changes sign. It is also linear in $p$ and linear in $q$. Bilinear and antisymmetric in two slots is exactly the defining behaviour of a $2 \times 2$ determinant as a function of its two rows. So the phrasing of the problem is already the phrasing of a determinant, with the "columns" being evaluation at $x$ and at $x+1$. Once you name it as a determinant, the whole toolkit of row operations comes with it, and the most basic tool is the one used: adding a multiple of one row to another changes nothing. That is the axiom-level fact doing the work: $\det(r_1,\, r_2 - t r_1) = \det(r_1, r_2) - t \det(r_1, r_1) = \det(r_1, r_2)$, because an alternating form vanishes on a repeated argument.

Next, why degree is the right measure. The target is a constant, the smallest possible degree. The left side is a difference of two products of degree $m+n$; the $x^{m+n}$ terms always cancel (both are $p_m q_n$), so the first place where anything can go wrong is $x^{m+n-1}$. Taylor expanding the shift turns that coefficient into the leading coefficient of $pq' - p'q$, which is $(n-m)p_m q_n$ — a number that vanishes exactly when the degrees agree. So leading terms kill unequal degrees and are *blind* to equal degrees. Any student who reaches that point has been told by the mathematics which case is hard and why: equal degrees let the top terms cancel. The row operation is precisely the tool that exploits that cancellation, because $q - tp$ with $t = q_m / p_m$ is the polynomial in which the top terms have been cancelled on purpose.

There is a second way to see the same structure, which is Kedlaya's first solution, and it deserves its motivation too. Kedlaya's solution begins by subtracting the equation at $x$ from the equation at $x-1$. Here is where that comes from: $R(x) = 1$ is a statement that the "Wronskian-like" quantity is *constant*, and a constant quantity has zero difference, which gives

$$p(x)\big(q(x+1) + q(x-1)\big) = q(x)\big(p(x+1) + p(x-1)\big).$$

Any common factor of $p$ and $q$ divides $R = 1$, so $p$ and $q$ are coprime, and $p$ must divide $p(x+1) + p(x-1)$. That polynomial has the same degree as $p$ and leading coefficient $2p_m$, so it equals $2p(x)$; the second difference of $p$ vanishes, so the first difference $p(x+1) - p(x)$ is a polynomial of period 1, hence constant, hence $p$ is linear (likewise $q$). This is the discrete version of a fact from differential equations: if $y_1, y_2$ satisfy $y'' = V(x)y$, their Wronskian is constant — and here the "potential" turns out to be forced to vanish. Both proofs are running on one structural fact: $R$ is an alternating form, so it is constant along families and invariant under row operations.

Cheapness matters here. The coefficient grind is "zombie maths": it is the most primal thing you can do with a polynomial identity, and it does work, but it rediscovers proportionality one index at a time. The determinant view gets proportionality for free and turns a descending induction into one line.

**Caveat.** Verified symbolically: for $p = ax + b$, $q = cx + d$ the left side is identically $bc - ad$; for general quadratics the $x^2$ and $x$ coefficients vanish only when $p \propto q$, in which case the constant term is 0. The top-coefficient formula $(n-m)p_m q_n$ was checked on $p = x$, $q = 1$ (left side $-1$). The official text writes this coefficient as $(m-n)p_m q_n$; the sign is immaterial since only its vanishing is used.

### Layer $+\alpha$ BACKTESTING

**1.** The first useful cue is the antisymmetric shape $p(x)q(x+1) - p(x+1)q(x)$. On seeing it, write the $2 \times 2$ determinant with rows $(p(x), p(x+1))$ and $(q(x), q(x+1))$.

**2.** The cue suggests two methods in sequence: leading-coefficient comparison (which the degree drop from the cancelled $x^{m+n}$ term invites), and row reduction of the determinant to transfer the equal-degree case into the unequal-degree one.

**3.** Hesitation occurs after $m = n$ is established, when the top-coefficient argument goes silent. It is reinforced away by asking what the top-coefficient argument *needs* (unequal degrees) and whether an allowed operation can produce that.

**4.** The trap is the equal-degree case: it invites a coefficient-by-coefficient grind that works for degree 2 and becomes a notational swamp by degree 4. A smaller trap is forgetting that $q - tp$ might be zero; it cannot be, because the zero polynomial never gives $R = 1$. A third is omitting the linear solutions with $p$ or $q$ constant; they are the cases $a = 0$ or $c = 0$ of $bc - ad = 1$.

**5.** Recognise it faster by asking of any expression in two unknown functions: is it bilinear and does it change sign when I swap them? If so it is a determinant, and it is invariant under "subtract a multiple of one from the other".

**6.** Reusable decision rule: *when one case of a problem is killed by an argument and another survives, look for a symmetry or invariance of the hypothesis that maps the surviving case into the killed one before attacking the surviving case directly.* Here it is a row operation; elsewhere it is a change of variable, a Euclidean-algorithm step, or a Vieta jump.

## 2010 B5

**2010 B5.** Is there a strictly increasing function $f : \mathbb{R} \to \mathbb{R}$ such that $f'(x) = f(f(x))$ for all $x$?

**Fields:** real analysis / functional-differential equations; monotonicity and convexity; growth rates and finite-time blow-up **Position:** B5 — hard end of the paper; nothing to compute, everything to organise **Answer:** No. $f' = f \circ f$ is positive and increasing, so $f \to \infty$, then $f' \to \infty$, then $f$ grows exponentially, then $f' \geq Ke^{2f}$, which forces $f$ to blow up at a finite $x$ — impossible on all of $\mathbb{R}$.

### Layer 1 THE WHAT — surface analysis

*The natural attempt.* A strong student's first instinct with a differential equation is to solve it. Try families. Exponentials: $f = e^{kx}$ gives $ke^{kx} = e^{ke^{kx}}$, hopeless. Linear: $f = ax + b$ gives $a = a(ax + b) + b$ for all $x$, so $a^2 = 0$, and $a = 0$ is not strictly increasing. Powers on $(0, \infty)$: $f = cx^\beta$ gives $c\beta x^{\beta-1} = c^{1+\beta}x^{\beta^2}$, so $\beta^2 = \beta - 1$, which has no real root. Nothing fits, and a Putnam "is there" question whose natural families all fail is usually a "no" — but a failed search is not a proof.

So the student switches to "no" and looks for a contradiction, usually by pinning down signs: $f$ increasing gives $f' \geq 0$; so $f(f(x)) \geq 0$ for all $x$; so $f \geq 0$ on the range of $f$… and then circles. Or the student tries to find a fixed point, or to show $f$ is bounded, or to differentiate again: $f'' = f'(f(x))\, f'(x) = f(f(f(x))) \cdot f(f(x))$, which grows more compositions without producing an inequality. The pieces are all true; there is no plan connecting them, because the student has not decided *what kind* of contradiction to aim for.

*The model chain.* Suppose such an $f$ exists. It is differentiable, since $f'$ appears.

**Step 1: $f'$ is positive and increasing.** Since $f$ is strictly increasing, so is $f \circ f$; so $f' = f \circ f$ is strictly increasing. Since $f$ is increasing, $f' \geq 0$ everywhere. A strictly increasing function that is $\geq 0$ everywhere is $> 0$ everywhere: $f'(x) > f'(x - 1) \geq 0$.

**Step 2: $f \to \infty$ and $f' \to \infty$.** For $x > 0$ the mean value theorem gives $f(x) - f(0) = xf'(\xi)$ with $\xi \in (0, x)$, and $f'(\xi) \geq f'(0) > 0$, so $f(x) \geq f(0) + f'(0)x \to \infty$. Then $f'(x) = f(f(x)) \to \infty$ as well, since $f(x) \to \infty$ and $f(y) \to \infty$ as $y \to \infty$.

**Step 3: $f$ grows exponentially.** Choose $X_0$ with $f' \geq 4$ on $[X_0, \infty)$. Then $f(x) \geq f(X_0) + 4(x - X_0)$, and this is $\geq 2x$ once $x \geq 2X_0 - \frac{1}{2}f(X_0)$. So there is $X > 0$ with $f(x) \geq 2x > 0$ for $x \geq X$. By Step 2 choose $X' \geq X$ with $f(x) \geq X$ for $x \geq X'$. For such $x$,

$$f'(x) = f\big(f(x)\big) \geq 2f(x),$$

so $(\ln f)' \geq 2$ on $[X', \infty)$ and $f(x) \geq Ke^{2x}$ there, with $K = f(X')e^{-2X'} > 0$.

**Step 4: blow-up.** Choose $X'' \geq X'$ with $f(x) \geq X'$ for $x \geq X''$. For such $x$,

$$f'(x) = f\big(f(x)\big) \geq Ke^{2f(x)}, \qquad \text{i.e.} \qquad \frac{d}{dx}\Big(-\tfrac{1}{2}e^{-2f(x)}\Big) = f'(x)e^{-2f(x)} \geq K.$$

Integrating from $X''$ to $x$,

$$K(x - X'') \leq \tfrac{1}{2}\big(e^{-2f(X'')} - e^{-2f(x)}\big) < \tfrac{1}{2}e^{-2f(X'')}.$$

The left side is unbounded as $x \to \infty$ and the right side is a fixed number. Contradiction. So no such $f$ exists. □

**Layer 2 THE HOW — the bifurcation**

**Bifurcating point.** Both chains write down the two monotonicity facts ($f' \geq 0$, and $f'$ increasing). They split at what to do with them.

(N) "Now find a contradiction" — with no target, so the student hunts for signs, fixed points, bounds, higher derivatives.

(M) "Now look at $x \to +\infty$ and compare growth rates." The equation says the derivative of $f$ is $f$ evaluated at a large point, so $f'$ is far bigger than $f$ — and a function whose derivative is much bigger than itself blows up in finite time.

The whole problem is the choice of which end of the real line to look at and which quantity (growth) to compare.

> *$f(f(x))$ with $f$ increasing and large means "$f$ evaluated at an even larger point". So far out to the right, $f'$ is bigger than $f$ — much bigger. A function whose derivative outruns the function itself is the $y' = y^2$ or $y' = e^y$ situation, and those blow up at a finite time. That's the contradiction to aim for: $f$ is supposed to exist on all of $\mathbb{R}$.*

> *I don't need the sharpest growth bound, only one good enough that $\int^\infty dy/F(y)$ converges. Each time I feed a lower bound for $f$ back into $f' = f(f(x))$ I get a stronger one: linear gives exponential, exponential gives $e^{2f}$. Two rounds and I'm done.*

- ▷ **Cue:** a functional-differential equation with $f'$ equal to something increasing, on all of $\mathbb{R}$ ⟶ **Move:** write down monotonicity of $f'$, deduce convexity and $f \to \infty$, then look at $x \to +\infty$
- ▷ **Cue:** $f'(x) \geq F(f(x))$ with $F$ positive and $\int^\infty dy/F(y) < \infty$ ⟶ **Move:** separate variables: $\frac{d}{dx}\int^{f(x)} \frac{dy}{F(y)} \geq 1$, so $f$ blows up in finite time
- ▷ **Cue:** a self-referential inequality ($f'$ bounded below by $f$ of something) ⟶ **Move:** bootstrap: insert the current lower bound on $f$ into the right side to get a better one, and repeat until an integrability threshold is crossed
- ▷ **Cue:** "is there a function such that…" where every natural family fails ⟶ **Move:** switch to proving "no", and decide in advance which kind of contradiction (sign, boundedness, growth) is being targeted

**Layer 3 THE WHY**

Why is "compare growth at $+\infty$" the natural move, and not a lucky guess?

Go back to what the equation says. The derivative is the rate at which $f$ increases. The right side is $f$ evaluated at the point $f(x)$. For a strictly increasing $f$ that is eventually larger than $x$, the point $f(x)$ lies to the right of $x$, so $f(f(x)) \geq f(x)$: the rate of increase of $f$ is at least as large as $f$ itself, and the further out you go the more it exceeds it. That is a statement about *growth*, and growth statements are decided at infinity. So the equation's phrasing already points

to $x \to +\infty$; the only question is whether $f(x)$ does eventually exceed $x$, and the monotonicity of $f'$ answers that — Step 3 shows $f(x) \geq 2x$ eventually.

Now why growth produces a contradiction at all. The axiom-level fact is the separation of variables for autonomous equations. If $y' = F(y)$ with $F > 0$, then $\frac{d}{dx}G(y) = 1$ where $G(y) = \int^y dt/F(t)$. So the time it takes $y$ to travel from $y_0$ to $\infty$ is exactly $\int_{y_0}^{\infty} dt/F(t)$. If that integral is finite, $y$ reaches $\infty$ at a finite time. With $F(y) = y$ the integral diverges (exponential growth, fine for all time); with $F(y) = y^2$ or $F(y) = e^{2y}$ it converges (blow-up). A function defined on all of $\mathbb{R}$ cannot blow up, so any differential inequality $f' \geq F(f)$ with $\int^{\infty} dt/F(t) < \infty$ on a half-line $[X, \infty)$ is an immediate contradiction. That is the whole of Step 4.

The equation $f' = f \circ f$ supplies the inequality through bootstrapping. Each lower bound $f \geq L$ gives $f' = f(f) \geq L(f)$ wherever $f$ is in range. Start with $L(y) = 2y$, the weakest non-trivial one; $\int dt/2t$ diverges, so no contradiction yet, but it upgrades $f$ to exponential. With $L(y) = Ke^{2y}$, $\int dt/(Ke^{2t})$ converges and the argument closes. The number of bootstrap rounds is not a mystery either; you stop when the integral crosses from divergent to convergent.

Kedlaya's first solution takes a different route: it shows $f(x_0) \leq x_0 + 1$ for $x_0 \geq -1$, then that $f' = f \circ f \leq 1$, contradicting $f' \to \infty$. That is an *upper* bound argument and is harder to motivate from scratch; the second solution (Zara) proves $f \geq \alpha x^2$ and then integrates $f'/f^2$ — the same separation-of-variables step with $F(y) = \alpha y^2$. Elkies's third solution makes the mechanism explicit: the inverse $g$ of $f$ satisfies $g'(y) = 1/f(y)$, so $g(y) = \int^y dt/f(t)$ must diverge (since $g$ takes arbitrarily large values, the domain of $f$ being all of $\mathbb{R}$), and any superlinear lower bound on $f$ makes it converge. All three are the same statement: the time to reach infinity is $\int dy/(\text{speed})$, and $f \circ f$ is too fast.

Note what the problem punishes. Trying to solve the equation is zombie maths here: there is no closed form to find, and the families fail for reasons unrelated to the answer. Differentiating again produces $f'' = f(f(f(x)))\, f(f(x))$, which is true and useless. The problem rewards the student who asks, before any manipulation, *what kind of object would make this impossible*.

**Caveat.** Each step of the model chain has been checked by hand: the choices of $X_0, X, X', X''$ exist by Steps 2–3, $f > 0$ on $[X', \infty)$ so $\ln f$ is defined, and $K > 0$. The heuristic claim that $\beta^2 = \beta - 1$ has no real root is used only to motivate "no", not in the proof.

### Layer $+\alpha$ BACKTESTING

**1.** The first useful cue is $f \circ f$ on the right with $f$ strictly increasing. On seeing it, write: "$f'$ is strictly increasing and $\geq 0$, so $f' > 0$ and $f$ is convex."

**2.** That cue suggests growth analysis at $+\infty$: convexity gives $f \to \infty$, the equation then gives $f' \to \infty$, and the self-referential right side gives a differential inequality $f' \geq F(f)$ to be integrated by separation of variables.

**3.** Hesitation occurs after the monotonicity facts, when there is no obvious target. It is reinforced away by deciding the *type* of contradiction first: "this is a growth equation, so I'm looking for finite-time blow-up."

**4.** The trap is trying to solve the equation (or to test families), and its cousin, differentiating repeatedly. A rigour trap: $\ln f$ and $f(f(x)) \geq 2f(x)$ are only valid once $f(x)$ is in the region where the previous bound holds; each step needs its own threshold $X, X', X''$.

**5.** Recognise it faster by reading any equation of the form $f' = \Phi(f)$ or $f' \geq \Phi(f)$ as "speed as a function of position" and asking immediately: is $\int^{\infty} dy/\Phi(y)$ finite?

**6.** Reusable decision rule: *for a function required to exist on all of* $\mathbb{R}$ *(or for all time), a lower bound* $f' \geq F(f)$ *with* $\int^{\infty} dy/F(y) < \infty$ *is a contradiction; if the available bound is too weak, feed it back through the equation until it is strong enough.* The same rule settles existence questions for ODEs like $y' = y^2 + x^2$, for recurrences $a_{n+1} = a_n + a_n^2$, and for global existence in population models.

## 2010 B6

**2010 B6.** Let $A$ be an $n \times n$ matrix of real numbers for some $n \geq 1$. For each positive integer $k$, let $A^{[k]}$ be the matrix obtained by raising each entry to the $k$th power. Show that if $A^k = A^{[k]}$ for $k = 1, 2, \ldots, n+1$, then $A^k = A^{[k]}$ for all $k \geq 1$.

**Fields:** linear algebra / Cayley–Hamilton; linear recurrences; entrywise (Hadamard) powers **Position:** B6 — last problem of the paper, but short once the right object is named: two sequences obeying one recurrence **Answer:** Let $P(x) = x^n + c_{n-1}x^{n-1} + \cdots + c_0$ be the characteristic polynomial. Cayley–Hamilton and the $n+1$ hypotheses show every entry of $A$ is a root of $xP(x)$; then both $k \mapsto A^k$ and $k \mapsto A^{[k]}$ obey the recurrence with characteristic polynomial $P$ for $k \geq 2$, and agree on $n$ consecutive terms, so they agree forever.

### Layer 1 THE WHAT — surface analysis

*The natural attempt.* A strong student's instinct is to understand *which* matrices satisfy the hypothesis. Start with $n = 2$, $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$:

$$A^2 = \begin{pmatrix} a^2 + bc & b(a+d) \\ c(a+d) & d^2 + bc \end{pmatrix}, \qquad A^{[2]} = \begin{pmatrix} a^2 & b^2 \\ c^2 & d^2 \end{pmatrix}.$$

So $bc = 0$, $b(a+d) = b^2$, $c(a+d) = c^2$. Say $c = 0$; then $b = 0$ or $b = a + d$. Now impose $k = 3$: with $c = 0$, $b = a + d \neq 0$, the top-right entry of $A^3$ is $b(a^2 + ad + d^2)$ while that of $A^{[3]}$ is $b^3 = b(a+d)^2$, so $ad = 0$. With patience the $2 \times 2$ case can be pushed to a full list of cases and each checked for $k \geq 4$. But $3 \times 3$ has nine unknowns and polynomial conditions of every degree up to 4; classification is not a method that scales to general $n$.

A second instinct: diagonalise. If $A = SDS^{-1}$ then $A^k = SD^kS^{-1}$ is controlled; but $A^{[k]}$ is defined entry by entry and has no relation to $S$ or $D$. Entrywise powers do not commute with change of basis, so diagonalisation says nothing about the right-hand side. The student is now stuck between an operation that likes eigenvalues and an operation that likes entries.

*The model chain.* Notation. For a polynomial $g$, write $g[A]$ for the matrix whose $(i, j)$ entry is $g(A_{ij})$; so $A^{[k]} = x^k[A]$, and $g \mapsto g[A]$ is linear in $g$. Let $P(x) = x^n + c_{n-1}x^{n-1} + \cdots + c_1x + c_0$ be the characteristic polynomial of $A$.

**Step 1: every entry of $A$ is a root of $xP(x)$.** By Cayley–Hamilton, $P(A) = 0$, so

$$0 = A\,P(A) = A^{n+1} + c_{n-1}A^n + \cdots + c_1A^2 + c_0A.$$

Only the powers $A^1, \ldots, A^{n+1}$ appear, and for those the hypothesis gives $A^k = A^{[k]}$. Substituting,

$$0 = A^{[n+1]} + c_{n-1}A^{[n]} + \cdots + c_0A^{[1]} = \big(xP(x)\big)[A].$$

So $A_{ij}P(A_{ij}) = 0$ for every $i, j$.

**Step 2: the same recurrence for both sequences.** Fix $m \geq n+1$. On one side, $A^{m+1-n}P(A) = 0$ gives

$$A^{m+1} = -\big(c_{n-1}A^m + c_{n-2}A^{m-1} + \cdots + c_0A^{m+1-n}\big). \tag{1}$$

On the other side, $x^{m+1-n}P(x) = x^{m-n} \cdot xP(x)$ with $m - n \geq 1$, so it vanishes at every entry of $A$ by Step 1; that is, $\big(x^{m+1-n}P(x)\big)[A] = 0$, which reads

$$A^{[m+1]} = -\big(c_{n-1}A^{[m]} + c_{n-2}A^{[m-1]} + \cdots + c_0A^{[m+1-n]}\big). \tag{2}$$

**Step 3: induction.** We prove $A^k = A^{[k]}$ for all $k \geq 1$ by strong induction; $k \leq n+1$ is the hypothesis. If it holds for all $k \leq m$ with $m \geq n+1$, then the right sides of (1) and (2) involve only exponents from $m + 1 - n \geq 2$ to $m$, where they agree term by term; so $A^{m+1} = A^{[m+1]}$. □

**Layer 2 THE HOW — the bifurcation**

**Bifurcating point.** Both chains start from the hypothesis as a list of matrix equations. They split on what kind of object the list is.

(N) "These are polynomial conditions on the entries; solve them." The hypothesis is treated as equations to be solved for $A$.

(M) "These are the first $n+1$ terms of two sequences, $A^k$ and $A^{[k]}$; show both sequences obey the same linear recurrence." The hypothesis is treated as matching initial conditions.

On (N) the difficulty grows with $n$ without limit. On (M) the proof has the same length for every $n$, and the $n+1$ in the statement turns out to be the order of the recurrence.

> *"For $k = 1, \ldots, n+1$ implies for all $k$" is the shape of a recurrence: finitely many initial terms determine everything. So I want a linear recurrence satisfied by both $A^k$ and $A^{[k]}$. For $A^k$ I have one for free — Cayley–Hamilton, order $n$. The question is whether the entrywise powers satisfy it too.*

> *Entrywise, $A^{[k]}_{ij} = a^k$ with $a = A_{ij}$, and $a^k$ satisfies the recurrence with characteristic polynomial $Q$ exactly when $a$ is a root of $Q$. So I need each entry to be a root of the Cayley–Hamilton polynomial. I can get that from the hypothesis: write the Cayley–Hamilton identity using only powers I'm allowed to swap for entrywise powers, and it becomes "$Q$ evaluated at every entry is zero".*

> *I can't use $P(A) = 0$ directly, because it contains $A^0 = I$, and the entrywise zeroth power is the all-ones matrix, not $I$. Multiply by $A$ first: $AP(A) = 0$ uses only $A^1, \ldots, A^{n+1}$. That's why the problem gives me $n+1$ equations and not $n$.*

- ▷ **Cue:** "if it holds for $k = 1, \ldots, N$ then it holds for all $k$" $\longrightarrow$ **Move:** look for a linear recurrence of order $N$ satisfied by both sides; matching initial terms then propagate
- ▷ **Cue:** powers of an $n \times n$ matrix $\longrightarrow$ **Move:** Cayley–Hamilton: $A^k$ satisfies the order-$n$ recurrence with characteristic polynomial $P$
- ▷ **Cue:** an entrywise (Hadamard) operation mixed with a matrix operation $\longrightarrow$ **Move:** work entry by entry: $a^k$ obeys the recurrence with polynomial $Q$ iff $a$ is a root of $Q$
- ▷ **Cue:** a recurrence identity that involves the index 0 where the two sides disagree ($I$ versus the all-ones matrix) $\longrightarrow$ **Move:** shift the identity by multiplying by $A$ so every index lies in the range where the hypothesis applies

**Layer 3 THE WHY**

Why is "two sequences, one recurrence" the natural reading, and not an insight from nowhere?

Start with the statement's logical shape: finitely many instances imply all instances. There are only a few mechanisms in mathematics that produce that, and the dominant one is a linear recurrence: if $u_{k+N}$ is a fixed linear combination of $u_k, \ldots, u_{k+N-1}$, then $N$ consecutive values determine the whole sequence. So the phrasing "$k = 1, 2, \ldots, n+1 \ldots$ for all $k \geq 1$" is a direct cue for recurrences, and the difference $A^k - A^{[k]}$ is the sequence we want to show is zero from its initial terms.

Next, why Cayley–Hamilton is the recurrence to use. The axiom-level fact is that $n \times n$ matrices form an $n^2$-dimensional space, so the powers $I, A, A^2, \ldots$ must be linearly dependent, and Cayley–Hamilton names an explicit dependence of length $n+1$: $P(A) = 0$. Multiplying by $A^j$ turns it into a recurrence for the sequence $A^k$. There is nothing to invent.

Then the entrywise side. The sequence $k \mapsto a^k$, for a single real number $a$, satisfies the recurrence with characteristic polynomial $Q$ precisely when $Q(a) = 0$ (because $\sum q_i a^{k+i} =$

$a^k Q(a)$). This is the defining fact about characteristic polynomials of recurrences, and it is the bridge: the matrix $A^{[k]}$ obeys the same recurrence as $A^k$ iff *every entry* is a root of the recurrence polynomial. Kedlaya's solution states "each entry of $A$ is a root of $xP(x)$" in one line; here is where that comes from — it is exactly the condition under which the entrywise powers obey the Cayley–Hamilton recurrence, and the hypothesis is what delivers it.

Finally, the $n+1$. Cayley–Hamilton gives an order-$n$ recurrence, so why are $n$ values not enough? Because the identity $P(A) = 0$ contains $A^0 = I$, while $A^{[0]}$ (entrywise zeroth power) is the all-ones matrix $J$. The two sequences do *not* agree at $k = 0$, so the first application of the recurrence — the one producing $k = n$ from $k = 0, \ldots, n-1$ — is not available. Shifting to $AP(A) = 0$ costs one more initial term. And the $n+1$ is genuinely needed: for $n = 2$ take

$$A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \quad A^2 = \begin{pmatrix} 1 & 4 \\ 0 & 1 \end{pmatrix} = A^{[2]}, \quad A^3 = \begin{pmatrix} 1 & 6 \\ 0 & 1 \end{pmatrix} \neq \begin{pmatrix} 1 & 8 \\ 0 & 1 \end{pmatrix} = A^{[3]}.$$

Here $P(x) = (x-1)^2$ and the entry 2 is not a root of $xP(x)$, which is precisely the failure. (For $n = 1$ the statement is empty: $a^k = a^k$.) The official remark attributes sharpness in general to David Feldman.

The problem punishes classification. Solving for $A$ is zombie maths — the most primal response to "here are equations" — and at B6 it is unbounded work. The model never learns what $A$ looks like; it only learns that $A^k$ and $A^{[k]}$ are the same kind of sequence.

**Caveat.** Verified by exact computation: for $A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$, $A^k = A^{[k]}$ for $k = 1, 2$ and fails at $k = 3$ (and $k = 4$); for $A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}$, with $P(x) = x(x-1)^2$, the identity holds for $k = 1, \ldots, 4$ and then, as the theorem predicts, for all $k \leq 11$ tested. The sharpness claim for general $n \geq 2$ is quoted from the official remark and was checked here only for $n = 2$.

### Layer $+\alpha$ BACKTESTING

**1.** The first useful cue is "for $k = 1, \ldots, n+1 \ldots$ for all $k$". On seeing it, write: "find a linear recurrence of order $n+1$ satisfied by both $A^k$ and $A^{[k]}$".

**2.** That cue suggests Cayley–Hamilton for the matrix powers, and the root criterion ($a^k$ obeys the recurrence for $Q$ iff $Q(a) = 0$) for the entrywise powers.

**3.** Hesitation occurs at the index 0: Cayley–Hamilton includes $I$, and the entrywise side has no $I$. It is reinforced away by multiplying by $A$, and by noticing that this is exactly what the extra hypothesis $k = n+1$ pays for.

**4.** The trap is classifying the matrices that satisfy $A^2 = A^{[2]}$, $A^3 = A^{[3]}, \ldots$; a secondary trap is diagonalising, which ignores that entrywise operations are basis-dependent.

**5.** Recognise it faster by treating any family of objects indexed by $k$, with an "initial range implies all" conclusion, as a sequence problem first and a structure problem second.

**6.** Reusable decision rule: *to prove two sequences are equal for all indices from equality on an initial stretch, show both satisfy one linear recurrence whose order is at most the length of the stretch; for power sequences, the recurrence is supplied by an annihilating polynomial and is inherited by $a^k$ exactly at its roots.* The same rule proves identities for power sums (Newton's identities), for linear recurrent sequences of matrices, and for polynomial identities checked at finitely many points.