DAWN PUTNAM SESSION

Putnam 2012 — Papers A and B

Putnam 120 Archive

Contents

Putnam 2012 — Paper A **5**
2012 A1 5
2012 A2 7
2012 A3 9
2012 A4 12
2012 A5 15
2012 A6 18

Putnam 2012 — Paper B **23**
2012 B1 23
2012 B2 25
2012 B3 27
2012 B4 30
2012 B5 32
2012 B6 35

Putnam 2012 — Paper A

2012 A1

2012 A1. Let $d_1, d_2, \ldots, d_{12}$ be real numbers in the open interval $(1, 12)$. Show that there exist distinct indices i, j, k such that d_i, d_j, d_k are the side lengths of an acute triangle.

Fields: combinatorics, elementary inequalities, induction (Fibonacci growth) **Position:** A1 — the opener
Answer: Such a triple always exists; sorting the twelve numbers and assuming no three consecutive ones form an acute triangle forces $d_{12}^2 \geq 144\, d_1^2$, which contradicts $d_1 > 1$, $d_{12} < 12$.

Layer 1 THE WHAT — surface analysis

The natural attempt. Twelve numbers give $\binom{12}{3} = 220$ possible triples, so testing triples directly is not a plan, it is a confession that you have not yet found the structure. The honest first move is to sort: relabel so $d_1 \leq d_2 \leq \cdots \leq d_{12}$. A triangle with side lengths $a \leq b \leq c$ is acute exactly when $c^2 < a^2 + b^2$ (the largest side is the one that can fail the condition, by the Law of Cosines applied to the angle opposite c). So for each sorted triple the only inequality that can fail is the one involving its largest member.

The instinct at this point is to say "surely, among the ten consecutive triples (d_i, d_{i+1}, d_{i+2}), one of them must be acute – they're the closest together, so they're the likeliest candidates" and stop there, treating it as intuitively obvious. That is not a proof. To make it one you have to ask what happens if none of the ten consecutive triples is acute, and that requires tracking exactly how fast $d_i^2 + d_{i+1}^2 \leq d_{i+2}^2$ forces the sequence to grow – which is the model chain, not a shortcut around it.

The model chain. Sort as above: $1 < d_1 \leq d_2 \leq \cdots \leq d_{12} < 12$. Suppose, for contradiction, that no three indices give an acute triangle. In particular, for every $i = 1, \ldots, 10$, the consecutive triple (d_i, d_{i+1}, d_{i+2}) is not acute. Since d_{i+2} is the largest of the three, non-acuteness means

$$d_{i+2}^2 \geq d_i^2 + d_{i+1}^2.$$

(The other two triangle-inequality-type conditions, $d_i^2 < d_{i+1}^2 + d_{i+2}^2$ and $d_{i+1}^2 < d_i^2 + d_{i+2}^2$, hold automatically once $d_i \leq d_{i+1} \leq d_{i+2}$, since the left side is then at most the largest square, which appears on the right; the only condition genuinely in play is the one on the largest side.)

Write $e_i = d_i^2$. The inequality is $e_{i+2} \geq e_i + e_{i+1}$ for $i = 1, \ldots, 10$. Let $F_1 = F_2 = 1, F_3 = 2, \ldots$ be the Fibonacci numbers. Claim: $e_i \geq F_i\, e_1$ for all $i = 1, \ldots, 12$.

Base cases: $e_1 = F_1 e_1$ trivially, and $e_2 \geq e_1 = F_2 e_1$ since $d_2 \geq d_1$. Inductive step: if $e_i \geq F_i e_1$ and $e_{i+1} \geq F_{i+1} e_1$, then

$$e_{i+2} \geq e_i + e_{i+1} \geq (F_i + F_{i+1}) e_1 = F_{i+2} e_1.$$

By induction the claim holds up to $i = 12$, and $F_{12} = 144$, so $e_{12} \geq 144\, e_1$, i.e. $d_{12}^2 \geq 144\, d_1^2$.

But $d_1 > 1$ gives $144\, d_1^2 > 144$, while $d_{12} < 12$ gives $d_{12}^2 < 144$. These contradict each other. So the assumption was false: some triple among the d_i is acute.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree up to sorting and reducing to the ten consecutive triples (d_i, d_{i+1}, d_{i+2}). They split at the word "surely." The natural attempt treats "consecutive triples are the likeliest to be acute" as self-evidently sufficient and looks for one directly. The model chain instead assumes the worst case – that every single one of the ten fails – and asks what that assumption, taken literally and chased forward, forces.

> *I don't need to find the acute triple by inspection. I need to show that assuming none exists is impossible – so let me take the failure condition at face value, as an inequality on squared lengths, and see how fast it forces the sequence to grow.*

> *The failure condition $d_{i+2}^2 \geq d_i^2 + d_{i+1}^2$ is a sum-of-two-previous-terms recursion, which is exactly the shape of the Fibonacci recursion – so instead of hand-checking a few terms, I should prove by induction that d_i^2 grows at least as fast as $F_i \cdot d_1^2$, and then just look at what the numbers 1, 12, and 12 (the count of terms) force at $i = 12$.*

Layer 3 THE WHY

Why does non-acuteness of a sorted triple reduce to a single inequality, $c^2 \geq a^2 + b^2$, rather than three? Because acuteness of a triangle with sides $a \leq b \leq c$ is controlled entirely by its largest angle, and the Law of Cosines gives $\cos C = \frac{a^2+b^2-c^2}{2ab}$ for the angle C opposite c: this is negative (obtuse or right) exactly when $c^2 \geq a^2 + b^2$, and the other two angles, being opposite shorter sides, are automatically acute once the triangle inequality $a + b > c$ holds and c is the longest side. So checking non-acuteness of the whole triangle collapses to checking one inequality on the two smallest members against the largest – there is no case work to do beyond identifying which side is largest, which sorting has already done.

Why look at consecutive triples in the sorted order rather than all 220? Because the claim to be refuted is a *universal* one – "no triple is acute" – and a universal claim about all triples in particular implies the same claim about the ten consecutive ones. Working with only the consecutive triples is not a simplification that risks missing cases; it is choosing the ten inequalities that are logically implied by, and hence no weaker a consequence of, the full 220-triple hypothesis. This is the standard move when a claim quantifies over "some choice of k objects from n": pick the sub-collection whose failure is easiest to chase (here, an index-adjacent chain), because if the universal claim is false even restricted to that sub-collection, it was false to begin with.

Why does the recursion $e_{i+2} \geq e_i + e_{i+1}$ get compared to Fibonacci numbers specifically, rather than bounded some other way? Because Fibonacci is, by definition, the slowest-growing sequence satisfying $F_{i+2} = F_i + F_{i+1}$ with $F_1 = F_2 = 1$ – any sequence satisfying $e_{i+2} \geq e_i + e_{i+1}$ with $e_1 \geq e_1$, $e_2 \geq e_1$ dominates it termwise, by the same induction that defines Fibonacci itself. And the reason $144 = F_{12}$ is the number that appears is not an accident to be marvelled at: the problem gives exactly 12 numbers and exactly the interval $(1, 12)$, and $12^2 = 144 = F_{12}$ is precisely tight. The problem was built so that the Fibonacci bound at $i = 12$ lands exactly on the boundary of the given interval – recognising that the count of terms (12) and the width of the interval (12) are the same number is the cue that a growth-rate argument reaching a factor of 12^2 over 11 steps is the intended mechanism, not a coincidence to be surprised by.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that the interval width (12) equals the number of quantities (12), while the exponent in "acute triangle" conditions is always 2 (squares). On seeing "12 numbers, interval up to 12," write down: $12^2 = 144$ and ask whether some induction reaches exactly that number.

2. That cue suggests a growth-rate/induction argument on squared quantities, with the recursion supplied by the failure condition for acuteness – i.e., proof by contradiction plus induction, not direct search.

3. Hesitation occurs right after sorting, at the temptation to test small triples by hand and hope one works. What reinforces the switch to contradiction is recognising that "no acute triple exists" is a clean universal statement, and universal statements that are hard to attack directly are usually easiest to attack by assuming them and deriving a numerical contradiction.

4. The trap is believing that because acute triangles feel "generic" (most triangles are, informally), one should appear without a real argument. The problem is adversarial: it is asking you to rule out a specific, deliberately near-extremal configuration (numbers exactly Fibonacci-spaced in square), and only a quantitative bound closes that off.

5. Recognise this family faster by treating "n real numbers in an interval, find k satisfying a comparison condition on some function of them" as a cue to sort first, reduce to consecutive tuples, and look for a recursion; then match the recursion's growth rate against the interval's numeric bounds.

6. Reusable rule: when a problem gives n ordered quantities in a fixed range and asks you to find some sub-tuple avoiding a "bad" comparison (here: non-acute), assume every consecutive sub-tuple is bad, turn that into a recursive inequality on a natural monotone function of the quantities (here, the square), and check whether the resulting worst-case growth rate is incompatible with the given range. If it is, the assumption fails and the desired sub-tuple exists.

▷ **Cue:** comparison condition on the largest of three ordered lengths (e.g. "acute triangle") $\longrightarrow$ **Move:** rewrite it as a single inequality between squares of the largest side and the sum of squares of the other two, via the Law of Cosines

▷ **Cue:** universal claim over all k-subsets of n sorted quantities is to be refuted $\longrightarrow$ **Move:** restrict to the $n-k+1$ consecutive k-tuples; refuting the claim there is sufficient and turns the count of cases from $\binom{n}{k}$ into $n-k+1$

▷ **Cue:** a failure condition has the shape $e_{i+2} \geq e_i + e_{i+1}$ $\longrightarrow$ **Move:** compare against the Fibonacci recursion by induction to get $e_i \geq F_i e_1$, then check whether the resulting bound at the last index is compatible with the problem's stated range

2012 A2

2012 A2. Let $*$ be a commutative and associative binary operation on a set S. Assume that for every x and y in S, there exists z in S such that $x * z = y$. (This z may depend on x and y.) Show that if a, b, c are in S and $a * c = b * c$, then $a = b$.

Fields: abstract algebra (commutative semigroups) **Position:** A2 **Answer:** Cancellation holds: $a * c = b * c \Rightarrow a = b$. Proved by constructing an element $f = c * e$ that acts as a local identity on both a and b, then absorbing f into a "divisor" g of b by a.

Layer 1 THE WHAT — surface analysis

The natural attempt. The hypothesis – for every x, y there is z with $x * z = y$ – looks exactly like the definition of a group with the identity and inverses stripped out, and the remark after the official solution confirms S is in fact an abelian group under $*$. So the natural first move is to try to build that structure from scratch: find an identity element e_0 (solve $x * e_0 = x$ for some fixed x, using the hypothesis with $y = x$), then show it works for every element, then find inverses, then cancel exactly as in a group. This is not unreasonable, and it can be pushed through, but it front-loads three separate existence claims (identity, its universality, inverses)

before cancellation is even in sight, and each one needs its own short argument. It is solving a harder, more general problem than the one asked, when only a, b, c are actually in play.

The model chain. Only a, b, c need an identity-like element, not the whole set S, so build one that works for exactly these three. Let $d = a * c = b * c$. By hypothesis (with $x = d, y = a$), there is $e \in S$ with $d * e = a$. Set $f = c * e$.

Now compute $a * f$ using associativity: $a * f = a * (c * e) = (a * c) * e = d * e = a$. And, since $b * c = d$ as well, $b * f = b * (c * e) = (b * c) * e = d * e = a$. So f behaves as an identity on a (that's forced, since $d * e = a$ was chosen to make it so) and, remarkably, also sends b to a under $* f$ – not to b, but that turns out to be exactly what is needed.

By hypothesis again (with $x = a$, $y = b$), there is $g \in S$ with $g * a = b$. Substitute $a = a * f$ into this:

$$b = g * a = g * (a * f) = (g * a) * f = b * f.$$

But $b * f = a$ was already computed above. So $b = a$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains recognise the hypothesis as a weakened group axiom. They split at the decision of *scope*: the natural attempt tries to establish an identity element for all of S before touching a, b, c; the model chain builds a single element f engineered to satisfy exactly the two equations needed – $a * f = a$ and $b * f = a$ – and nothing more general.

> *I don't need S to have an identity element. I need one specific element that fixes a under $*$ – and the hypothesis lets me manufacture exactly that, starting from the one piece of data I'm actually given, $a * c = b * c$.*

> *Once I have f with $a * f = a$, associativity means f passed through the shared factor c: it inherits its "fixes a" property from the equation $d * e = a$, and inherits its effect on b for free, from the same equation, because $b * c$ is also d. I don't have to compute $b * f$ separately – it's forced to equal $a * f$ by the very equation I started from.*

> *To finish I need to move from "$b * f = a$" to "$b = a$", and I have no cancellation law yet – that's what I'm proving. But I do have another instance of the same solvability hypothesis: some g satisfies $g * a = b$. If I substitute $a * f$ for a inside that equation, associativity lets me pull the f outside and reduce it to $b * f$, which I already know equals a.*

Layer 3 THE WHY

Why does building a "local identity" f for just a and b succeed where trying to build a global identity for all of S merely stalls? Because the goal is a statement about three specific elements, $a * c = b * c \Rightarrow a = b$, and the hypothesis is only ever usable as an existence statement about a triple of elements at a time ($x, y \mapsto z$ with $x * z = y$). Nothing forces the z found for one pair (x, y) to interact well with the z found for another pair unless the elements x that are used are chosen to overlap deliberately. The model chain's e is deliberately built from $d = a * c = b * c$, the one element that already ties a, b, c together – so that every subsequent equation inherits that tie through associativity, rather than introducing a fresh, unrelated existence witness at each step.

Why does $f = c * e$, rather than e itself, turn out to be the useful object? Because $a * e$ and $b * e$ are not controlled by anything – the hypothesis only promises $d * e = a$, a statement about d, not about a or b individually. Multiplying e by c reintroduces the factor that a and b share ($a * c = b * c = d$), so that $a * f = a * (c * e) = (a * c) * e = d * e$ and $b * f = (b * c) * e = d * e$ collapse to the *same* value by associativity alone – the value a is not chosen, it falls out of the

definition of e. This is the substantive use of associativity in the problem: it is not decorative, it is the mechanism that lets the shared factor $c * e$ act identically on a and b.

Why does the final substitution work? Because $g * a = b$ says nothing yet about f, but $a = a * f$ is now an established fact, and associativity again lets a $*$-expression be regrouped without changing its value: $g * (a * f) = (g * a) * f$. This is the same move as the previous paragraph – push a known element (f) through a product via associativity, into a slot where its effect is already known ($b * f = a$) – deployed a second time on a different pair. The whole proof is one idea, applied twice: manufacture an element with a known effect on the elements in play, then relocate it inside a larger product using associativity to transfer that known effect to a new equation.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that the hypothesis is stated about arbitrary $x, y \in S$, but the conclusion is about three fixed elements a, b, c. On seeing that mismatch, write down: "use the hypothesis only with inputs built from a, b, c themselves, so every witness element inherits a relationship to them."

2. That cue suggests constructing an auxiliary element from $d = a * c = b * c$ specifically, rather than attempting to establish global structure (identity, inverses) for all of S.

3. Hesitation sets in once e with $d * e = a$ has been found: it's tempting to try to compute $a * e$ and $b * e$ directly and hope they're useful, since e feels "close to an inverse of c." What reinforces the correct move is noticing $a * e$ and $b * e$ are genuinely unconstrained by the hypothesis, while $a * (c * e)$ and $b * (c * e)$ both reduce, via associativity, to $d * e$ – the one quantity actually pinned down.

4. The trap is assuming the problem wants a group-theoretic proof (identity, then inverses, then cancellation) because the hypothesis resembles group axioms. That route works eventually (the remark confirms S is a group) but is strictly more than the problem asks, and building it first buries the two-line cancellation argument that alone answers the question.

5. Recognise this pattern faster by treating "for all x, y there exists z with $x * z = y$" as licence to manufacture, on demand, an element with any single desired effect on a chosen element – and to look for such an element built from data already tying the elements of the conclusion together, rather than from an arbitrary fresh pair.

6. Reusable rule: in a semigroup with universal solvability ($\forall x, y\, \exists z,\ x * z = y$) but no stated identity, cancellation-type statements are proved by manufacturing a local identity for the specific elements in the conclusion – built from a quantity the hypothesis already forces to be shared between them – and then relocating that local identity inside further instances of solvability via associativity, never by first establishing a global identity or inverses.

▷ **Cue:** hypothesis of the form "for all x, y exists z with $x * z = y$" (universal solvability, no stated identity) $\longrightarrow$ **Move:** build a witness element from a quantity already shared between the elements of the desired conclusion, rather than from an arbitrary pair

▷ **Cue:** two elements a, b satisfy $a * c = b * c$ in an associative structure with universal solvability $\longrightarrow$ **Move:** find e with $(a * c) * e = a$; then $f = c * e$ satisfies $a * f = a$ and, by the shared value $a * c = b * c$, also $b * f = a$

▷ **Cue:** a known equation $u = u * f$ needs to be combined with a second solvability witness $g * u = w$ $\longrightarrow$ **Move:** substitute $u * f$ for u inside $g * u = w$ and regroup via associativity to move f outside, landing on an already-known value

2012 A3

2012 A3. Let $f : [-1, 1] \to \mathbb{R}$ be a continuous function such that (i) $f(x) = \frac{2-x^2}{2} f\left(\frac{x^2}{2-x^2}\right)$

for every x in $[-1,1]$, (ii) $f(0)=1$, and (iii) $\lim_{x\to 1^-}\frac{f(x)}{\sqrt{1-x}}$ exists and is finite. Prove that f is unique, and express $f(x)$ in closed form.

Fields: real analysis (functional equations), contraction/iteration arguments **Position:** A3 **Answer:** $f(x)=\sqrt{1-x^2}$, and it is the unique continuous function on $[-1,1]$ satisfying (i)-(iii).

Layer 1 THE WHAT — surface analysis

The natural attempt. The functional equation relates $f(x)$ to f at a new point $x^2/(2-x^2)$, so the reflex is to hunt for a substitution that makes the relation look familiar – try $x=\sin\theta$, hoping $x^2/(2-x^2)$ simplifies to something like sin or tan of a related angle. It doesn't collapse cleanly: $2-x^2=1+\cos^2\theta$ is not a standard identity, and the substitution stalls without producing a recognisable recursion in θ.

A second natural move: plug in specific values. $x=0$ gives $f(0)=1\cdot f(0)$, no information. Try $x=1$: the map sends $1\mapsto 1$, a fixed point, and the coefficient $(2-x^2)/2\to 1/2$, giving $f(1)=\frac{1}{2}f(1)$, forcing $f(1)=0$ – useful, but it only pins down one value, and condition (iii) is about behaviour *near* $x=1$, not at it, so this alone doesn't finish anything. What's missing is a way to use the functional equation *repeatedly*, and the direct route doesn't suggest one until you notice that the map $x\mapsto x^2/(2-x^2)$ can be iterated.

The model chain. Strip out the singular behaviour near ±1 by defining $g:(-1,1)\to\mathbb{R}$, $g(x)=f(x)/\sqrt{1-x^2}$ (well-defined and continuous on the open interval, since $1-x^2>0$ there). Substituting $f(x)=g(x)\sqrt{1-x^2}$ into (i) and simplifying (using the identity, checked directly, that $1-\left(\frac{x^2}{2-x^2}\right)^2=\frac{4(1-x^2)}{(2-x^2)^2}$, so that $\sqrt{1-(x^2/(2-x^2))^2}=\frac{2\sqrt{1-x^2}}{2-x^2}$) collapses the coefficient $(2-x^2)/2$ against this square root exactly, leaving

$$g(x)=g\left(\frac{x^2}{2-x^2}\right)\qquad\text{for all } x\in(-1,1).$$

Now fix any $x\in(-1,1)$ and iterate the map $\varphi(t)=t^2/(2-t^2)$: set $a_1=x$, $a_{n+1}=\varphi(a_n)$. For $t\in(-1,1)$, $|\varphi(t)|=\frac{t^2}{2-t^2}\le t^2$ since $2-t^2\ge 1$. So $|a_{n+1}|\le|a_n|^2$, hence $|a_n|\le|x|^{2^{n-1}}\to 0$ (this needs $|x|<1$, which holds). So $a_n\to 0$.

Because $g(a_n)=g(a_{n+1})=\cdots=g(x)$ for every n by the equation above, and g is continuous at 0 (it's continuous on all of $(-1,1)$), taking $n\to\infty$ gives $g(x)=\lim g(a_n)=g(0)$. And $g(0)=f(0)/\sqrt{1-0}=f(0)=1$ by (ii). So $g(x)=1$ for every $x\in(-1,1)$, i.e. $f(x)=\sqrt{1-x^2}$ on $(-1,1)$.

To extend to $x=\pm1$: f is given continuous on all of $[-1,1]$, and $\sqrt{1-x^2}\to 0$ as $x\to\pm1$, matching the value forced above ($f(1)=0$ from the direct substitution, consistent with $\sqrt{1-1^2}=0$); continuity of f at ±1 then forces $f(\pm1)=\lim_{x\to\pm1}f(x)=\lim_{x\to\pm1}\sqrt{1-x^2}=0$, matching the formula there too. So $f(x)=\sqrt{1-x^2}$ on the closed interval.

Caveat. Kedlaya's published solution notes that condition (iii) is in fact unnecessary for uniqueness – continuity of f on the closed interval already forces $g(0)=1$ to propagate to the endpoints. I have not independently reverified that (iii) is redundant beyond taking the remark's word for it; the argument above uses only (i), (ii), and continuity, and never actually invokes (iii), which is consistent with the remark but I have not constructed the alternative derivation that Kedlaya says (iii) enables (running the recursion in the opposite direction).

Finally, existence: direct substitution (done in the Layer $+\alpha$ discussion of the algebraic identity above) confirms $f(x)=\sqrt{1-x^2}$ satisfies (i) exactly, and (ii), (iii) are immediate ($f(0)=1$; $f(x)/\sqrt{1-x}=\sqrt{1+x}\to\sqrt{2}$ as $x\to 1^-$, finite).

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains try to exploit the recursive structure of (i). They split at the choice of *what to iterate*. The natural attempt tries to iterate or simplify the substitution $x \mapsto x^2/(2-x^2)$ itself, in the original variable f, and gets stuck on messy coefficients at every step. The model chain first strips the expected singular factor $\sqrt{1-x^2}$ out of f by defining $g = f/\sqrt{1-x^2}$, so that the coefficient $(2-x^2)/2$ in (i) cancels *exactly* against the change in $\sqrt{1-x^2}$ under the substitution, leaving a bare fixed-point equation $g(x) = g(\varphi(x))$ with no coefficient left to track.

> *The coefficient $(2-x^2)/2$ in front of $f(\varphi(x))$ is not arbitrary – it looks exactly like the ratio $\sqrt{1-x^2}/\sqrt{1-\varphi(x)^2}$ would need to be, if f were $\sqrt{1-x^2}$ times something simpler. Let me factor that guess out as g and see if the equation for g is cleaner.*

> *Once the equation is $g(x) = g(\varphi(x))$ with no coefficient, the problem becomes purely about the dynamics of the map $\varphi(t) = t^2/(2-t^2)$: if I can show φ has a globally attracting fixed point at 0 that every starting point reaches under iteration, continuity of g finishes the argument in one line, because g is constant along every orbit.*

Layer 3 THE WHY

Why does dividing by $\sqrt{1-x^2}$, rather than any other function, make the coefficient vanish? Because the coefficient $(2-x^2)/2$ is exactly the factor that converts $\sqrt{1-\varphi(x)^2}$ back into $\sqrt{1-x^2}$: this is not guessed blindly, it is read off the structure of the problem, since condition (iii) – the finite limit of $f(x)/\sqrt{1-x}$ at $x=1$ – is a direct hint that f behaves like $\sqrt{1-x^2} = \sqrt{1-x}\sqrt{1+x}$ near the endpoint, and the natural thing to do with a functional equation whose solution is suspected to have a known singular factor is to divide that factor out before hunting for the rest. Condition (iii) is not merely a technical closing condition; it is the clue that tells you what to divide by.

Why does the resulting equation $g(x) = g(\varphi(x))$ reduce to "g is constant" rather than something weaker? Because $g(x) = g(\varphi(x)) = g(\varphi(\varphi(x))) = \cdots$ says g is constant along every forward orbit of φ, and the map $\varphi(t) = t^2/(2-t^2)$ has 0 as an attracting fixed point for every starting point in $(-1,1)$ – the bound $|\varphi(t)| \le t^2$ is not a coincidence but follows directly from $2-t^2 \ge 1$ on that interval, which is itself forced by $t^2 \le 1$. So *every* orbit is eventually forced arbitrarily close to 0, and continuity of g (guaranteed since f is continuous and $\sqrt{1-x^2} > 0$ strictly inside the interval) converts "constant along an orbit converging to 0" into "equal to $g(0)$." The single value $g(0) = f(0) = 1$, supplied by condition (ii), therefore propagates to every point of the domain – uniqueness of f is really uniqueness of the one free parameter $g(0)$, forced by (ii).

Why is this the same mechanism as a fixed-point/contraction argument, and why does recognising that matter? Because $|\varphi(t)| \le t^2$ is a contraction estimate (stronger than a standard linear contraction – it's quadratic, so convergence to 0 is very fast, $|a_n| \le |x|^{2^{n-1}}$), and any functional equation of the form "g equals itself composed with a self-map having a globally attracting fixed point" forces g to be constant on the basin of attraction, by the same continuity argument, regardless of the specific formula for φ. Seeing $\varphi(t) = t^2/(2-t^2)$ and immediately checking whether it is a contraction toward a fixed point is the transferable move; the specific bound $|\varphi(t)| \le t^2$ is just the verification that it is one here.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is condition (iii): a limit of $f(x)/\sqrt{1-x}$ existing finitely at an endpoint is a strong hint that f has $\sqrt{1-x^2}$ (or a comparable vanishing factor) built into it. On seeing it, write down: "try dividing f by this suspected factor before manipulating the functional equation directly."

2. That cue suggests defining $g = f/\sqrt{1-x^2}$ and substituting into the functional equation, hoping the coefficient cancels – which converts a coefficient-bearing functional equation into a bare invariance equation $g(x) = g(\varphi(x))$.

3. Hesitation sets in right after the functional equation for g appears clean: it's tempting to try to solve $g(x) = g(\varphi(x))$ by finding an explicit closed form for iterates of φ. What reinforces the better move is recognising that an explicit formula for $\varphi^{(n)}(x)$ is not needed – only that the iterates converge to a point where g's value is already known.

> *I don't need to know $\varphi^{(n)}(x)$ in closed form. I only need $\varphi^{(n)}(x) \to 0$, because g is continuous and takes the same value along the whole orbit – convergence alone forces $g(x) = g(0)$.*

4. The trap is trying to force the substitution $x = \sin\theta$ or similar trigonometric guesses to work on the original equation for f, without first isolating the singular factor – this buries the clean fixed-point structure under coefficients that never simplify.

5. Recognise this pattern faster by treating "$f(x) = (\text{coefficient}) \cdot f(\varphi(x))$, plus a boundary condition hinting at f's singular behaviour" as an instruction to normalise by the suspected singular factor first, converting the equation into $g(x) = g(\varphi(x))$, then study φ's dynamics.

6. Reusable rule: for a functional equation $f(x) = c(x)\, f(\varphi(x))$ where a side condition hints at f's behaviour near a boundary point, guess the singular factor $s(x)$ suggested by that hint, set $g = f/s$, and check whether $c(x)$ is exactly the ratio $s(x)/s(\varphi(x))$ needed to cancel. If it is, $g(x) = g(\varphi(x))$; then show φ has a globally attracting fixed point on the domain and use continuity of g to force g constant, pinned by any known value of f.

▷ **Cue:** functional equation $f(x) = c(x)f(\varphi(x))$ plus a boundary limit condition hinting at f's vanishing rate near an endpoint $\longrightarrow$ **Move:** normalise $g = f/s$ for the suggested singular factor s; check $c(x) = s(x)/s(\varphi(x))$ collapses the equation to $g(x) = g(\varphi(x))$

▷ **Cue:** invariance equation $g(x) = g(\varphi(x))$ for a continuous g $\longrightarrow$ **Move:** show φ's iterates converge to a fixed point in the domain, then conclude g is constant, equal to its value at that fixed point

▷ **Cue:** self-map φ with $\varphi(0) = 0$ and $|\varphi(t)| \le t^2$ near 0 $\longrightarrow$ **Move:** recognise this as a (super-linear) contraction toward 0; iterates converge doubly-exponentially fast

2012 A4

2012 A4. Let q and r be integers with $q > 0$, and let A and B be intervals on the real line. Let T be the set of all $b + mq$ where b and m are integers with b in B, and let S be the set of all integers a in A such that ra is in T. Show that if the product of the lengths of A and B is less than q, then S is the intersection of A with some arithmetic progression.

Fields: number theory (congruences); geometry of numbers (lattices and area); discrete convexity **Position:** A4 — mid-paper. Nothing deep is needed, but the problem is unreadable until you draw the right picture, and the official proof never draws it. **Answer:** S is empty or a (finite) arithmetic progression of integers; in the plane, the pairs (a, b) that witness membership of S all lie on one line, because three non-collinear such pairs would span a triangle of area at least $q/2$, and none fits in a box of area less than q.

Layer 1 THE WHAT — surface analysis

The natural attempt. You unpack the definitions. T is the set of integers congruent mod q to some integer of B, so

$$a \in S \iff a \in A \cap \mathbb{Z} \text{ and } ra \equiv b \pmod{q} \text{ for some integer } b \in B.$$

Then you try examples. $q = 7$, $r = 3$, $B = [0,1]$, $A = [0,3]$ (product $3 < 7$): $3a \bmod 7$ for $a = 0,1,2,3$ is $0,3,6,2$, so $S = \{0\}$. Fine. $A = [0,6]$, $B = [0,1]$ (product 6): residues $0,3,6,2,5,1,4$, so $S = \{0,5\}$ — two points, an AP. Now you want a proof. The natural move is to take three elements $a_1 < a_2 < a_3$ of S with witnesses b_1, b_2, b_3 and aim for $a_2 - a_1 = a_3 - a_2$, or at least for the gaps to be multiples of a common step. You subtract congruences: $r(a_i - a_j) \equiv b_i - b_j \pmod q$. If $\gcd(r,q) = 1$ you can invert r and write $a_i - a_j \equiv r^{-1}(b_i - b_j)$, and now you are lost: r^{-1} is an arbitrary residue, the b-differences are small but the a-differences need not be, and nothing forces equal gaps. You try the pigeonhole principle on residues, then case analysis on signs of $b_i - b_j$. After twenty minutes you have a page of congruences and no mechanism. The wall is that you are treating a and b as separate one-dimensional objects when the hypothesis is about a two-dimensional one: the *product* of two lengths.

The model chain. *Step 0: the picture.* Put

$$L = \{(a,b) \in \mathbb{Z}^2 : b \equiv ra \pmod q\}, \qquad P = L \cap (A \times B).$$

Then $a \in S$ exactly when some point of P has first coordinate a. L is closed under addition and subtraction (a lattice), and for any $u = (u_1, u_2)$, $w = (w_1, w_2)$ in L, writing $u_2 = ru_1 + qk$, $w_2 = rw_1 + q\ell$,

$$\det(u,w) = u_1(rw_1 + q\ell) - w_1(ru_1 + qk) = q(u_1\ell - w_1 k),$$

a multiple of q. Hence **any triangle with vertices in L has area 0 or at least $q/2$.**

Step 1: a box of area less than q holds no fat triangle. Let the three vertices lie in the closed box $\bar{A} \times \bar{B}$, with widths $|A|$ and $|B|$, and order them by first coordinate, $x_1 \le x_2 \le x_3$. The vertical line $x = x_2$ cuts the triangle into two triangles sharing a vertical side of length $h \le |B|$, with horizontal extents $x_2 - x_1$ and $x_3 - x_2$, so the area is $\frac{1}{2}h(x_3 - x_1) \le \frac{1}{2}|A||B| < \frac{q}{2}$. Therefore **all points of P are collinear** (if $|A|$ or $|B|$ is zero, the box is a segment and this is immediate).

Step 2: one witness per element. If $|B| \ge q$ then $|A| < 1$, A contains at most one integer, and S has at most one element — trivially an AP. Otherwise $|B| < q$; two integers of B congruent mod q would differ by at least q, so each $a \in S$ has exactly one witness b. So P is the graph of a function on S, and P lies on a non-vertical line ℓ.

Step 3: collinear plus convex gives an AP. Assume $|S| \ge 2$; let $s_0 < s_1$ be the two smallest elements, with points $p_0 = (s_0, b_0)$, $p_1 = (s_1, b_1)$, and put $v = p_1 - p_0 \in L$. The points of L on ℓ are closed under differences and all lie on one line through p_0; the one closest to p_0 on the side of p_1 gives a generator $w' \in L$ with positive first coordinate, so that $L \cap \ell = \{p_0 + kw' : k \in \mathbb{Z}\}$ and $v = mw'$ for an integer $m \ge 1$. If $m \ge 2$, $p_0 + w'$ is a point of L strictly inside the segment p_0p_1, hence in P, and its first coordinate lies strictly between s_0 and s_1 — contradicting the choice of s_1. So v generates $L \cap \ell$, and with $d = s_1 - s_0$:

$$P \subseteq \{p_0 + kv\} \quad \Longrightarrow \quad S \subseteq \{s_0 + kd : k \ge 0\}.$$

Conversely, if $s_0 + Kd = \max S$, every $p_0 + kv$ with $0 \le k \le K$ lies on the segment from p_0 to the top point of P, so is in $A \times B$ and in L: $s_0 + kd \in S$. Hence $S = \{s_0, s_0 + d, \ldots, s_0 + Kd\}$, an arithmetic progression, and $S = A \cap \{s_0, \ldots, s_0 + Kd\}$.

Caveat. Two points of reading. (i) "Intersection of A with some arithmetic progression" must allow *finite* progressions. Kedlaya–Ng give $q = 5$, $r = 1$, $A = [1,3]$, $B = [0,2]$: product $4 < 5$, and $S = \{1,2\}$, which is not A meet any infinite AP. So the provable statement is "S is empty or a finite AP", and that is what is proved here. (ii) The lengths are taken finite. Verified numerically: 2×10^5 random instances with $q \le 12$, $|r| \le 15$, rational endpoints and $|A||B| < q$, no counterexample.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start from "$ra \equiv b \pmod q$ with $b \in B$". They split at (†): *what is the hypothesis $|A| \cdot |B| < q$ about?* The natural attempt reads it as two separate size bounds and works with congruences on a and on b separately. The model reads a product of two lengths as an *area*, puts the witnesses (a, b) in the plane, and asks what area a lattice with "determinant" q forces.

> *The hypothesis multiplies a length in a by a length in b. That is an area. So the right object is not a alone but the pair (a, b), living in the rectangle $A \times B$ of area less than q.*

> *The pairs with $b \equiv ra \bmod q$ form a lattice with one point per q units of area. A rectangle smaller than one cell cannot hold three of them in a proper triangle, so they must all be on a line — and equally spaced points on a line project to an arithmetic progression.*

> *I only need to rule out a gap in the middle: if the two smallest witnesses are p_0, p_1, any lattice point between them on the line would be in the box by convexity, so $p_1 - p_0$ is already the smallest step.*

Layer 3 THE WHY

Why should a product of lengths be read as an area? Because nothing else in the problem produces a product. Every other quantity — q, r, a, b — enters linearly, through $ra - b \in q\mathbb{Z}$. A condition shaped like $|A| \cdot |B| < q$ is the signature of a comparison between the area of a region and the area per point of a lattice; this is the entire content of the geometry of numbers (Minkowski, Pick, Blichfeldt), and it is how the problem is built.

Why is L the lattice, and why is its cell area q? L is defined by one linear congruence in two integer unknowns. Among the q^2 residue pairs (a, b) mod q, exactly q satisfy $b \equiv ra$, so L contains a fraction $1/q$ of all integer points: one point per q units of area. The determinant calculation in Step 0 is the same fact written as an identity. Once you have "one point per area q", the triangle bound is forced: a lattice triangle is half a lattice parallelogram, and a lattice parallelogram with non-parallel sides has area a nonzero multiple of the cell area.

Why does collinearity give an arithmetic progression? Because the lattice points on a line through two lattice points are themselves equally spaced (they form a one-dimensional lattice: differences of points of L on ℓ are in L and parallel to ℓ). A box is convex, so the set of such points inside it is an unbroken run. Projecting an unbroken run of equally spaced points onto the a-axis gives an unbroken run of an AP. That is the whole proof; the only thing to be careful about is which step is the generator, which is why Step 3 picks the two *smallest* elements.

Kedlaya–Ng's solution begins by proving a lemma — a finite set of integers closed under $a + c - b$ for $a \le b \le c$ is an AP — and then shows collinearity by the congruence $(a_i - a_j)(b_k - b_l) \equiv (b_i - b_j)(a_k - a_l) \pmod q$ with both sides of absolute value less than q and the same sign. Here is where that comes from: $(a_i - a_j)(b_k - b_l) - (b_i - b_j)(a_k - a_l)$ is the determinant of two difference vectors of L — twice a triangle's area — so it is divisible by q; and each product is less than $|A||B| < q$ in size. Their "same sign" device is a hand-made version of Step 1. Their remark that the problem is "somewhat confusing" is the finite-AP issue in the caveat above. Elkies' "cross products" reinterpretation, which they mention without detail, is the determinant here.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the hypothesis "the product of the lengths of A and B is less than q". On seeing it, write: *points (a, b) in the rectangle $A \times B$, area $< q$; lattice $b \equiv ra \pmod q$ has one point per area q.*

2. Geometry of numbers in its most elementary form: lattice triangles have area at least half the cell area, so a small convex region's lattice points are collinear.

3. Hesitation occurs after the congruences $r(a_i - a_j) \equiv b_i - b_j$ are written and seem to lead nowhere. What reinforces it away is refusing to divide by r: dividing destroys the smallness of the b-differences. Keep both coordinates and cross-multiply — the cross-multiplied quantity is a determinant.

4. Three traps. Inverting r modulo q, which scatters everything. Reading "arithmetic progression" as infinite, which makes the statement false (the $q = 5$ example). And in the final step, assuming the difference of the two smallest witnesses might be a multiple of a smaller lattice step — rule it out by convexity, not by hope.

5. Slogan: *a product of two lengths compared with a modulus is an area compared with a lattice cell.* Whenever a congruence links two small unknowns and the hypothesis bounds the product of their ranges, draw the pairs in the plane.

6. Decision rule: **if a hypothesis multiplies the sizes of two ranges and compares the product with a modulus or a count, promote the pair of unknowns to a point in the plane and compare areas; three non-collinear lattice points cost half a cell of area.** Companion rule: collinear lattice points in a convex set are an unbroken equally spaced run, so their projections are an AP.

▷ **Cue:** a congruence $b \equiv ra \pmod{q}$ with a and b each confined to an interval, and a hypothesis on the product of the interval lengths $\longrightarrow$ **Move:** plot the pairs (a, b): they form a lattice of cell area q inside a rectangle; compare areas

▷ **Cue:** need to show three lattice points are collinear $\longrightarrow$ **Move:** compute twice the triangle's area as a determinant; show it is a multiple of the cell area yet smaller than one cell

▷ **Cue:** a triangle inside an axis-parallel box of width W and height H $\longrightarrow$ **Move:** cut vertically through the middle vertex: area $\leq \frac{1}{2}WH$

▷ **Cue:** a finite set of lattice points on a line inside a convex region $\longrightarrow$ **Move:** the points are consecutive terms $p_0 + kv$ of a one-dimensional lattice; choose v from the two extreme-adjacent points and use convexity to show no gaps

2012 A5

2012 A5. Let $\mathbb{F}_p$ denote the field of integers modulo a prime p, and let n be a positive integer. Let v be a fixed vector in $\mathbb{F}_p^n$, let M be an $n \times n$ matrix with entries of $\mathbb{F}_p$, and define $G : \mathbb{F}_p^n \to \mathbb{F}_p^n$ by $G(x) = v + Mx$. Let $G^{(k)}$ denote the k-fold composition of G with itself, that is, $G^{(1)}(x) = G(x)$ and $G^{(k+1)}(x) = G(G^{(k)}(x))$. Determine all pairs p, n for which there exist v and M such that the p^n vectors $G^{(k)}(0)$, $k = 1, 2, \ldots, p^n$ are distinct.

Fields: linear algebra over finite fields; binomial coefficients mod p; permutations and orbits **Position:** A5 — hard. The answer is guessable from small cases; the proof needs one structural reduction (the linear part is unipotent) and one binomial computation. **Answer:** Exactly the pairs with $n = 1$ (every prime p), together with the single pair $(p, n) = (2, 2)$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You start small. For $n = 1$, $G(x) = x + 1$ visits $1, 2, \ldots, p-1, 0$: works for every p. For $n = 2$, $p = 2$ there are only 16 matrices, and a short search finds $M = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, $v = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ with orbit $(0,1), (1,0), (1,1), (0,0)$. For $(3,2)$ you try the same shape, $M = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$,

and compute $M^k = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}$, so

$$G^{(k)}(0) = \sum_{j<k} M^j v = \begin{pmatrix} kv_1 + \binom{k}{2} v_2 \\ kv_2 \end{pmatrix},$$

and at $k = 3$ both entries are 0 mod 3: period 3, not 9. Other shapes — diagonalisable M, M with no eigenvalues in $\mathbb{F}_3$ — you try one at a time, and each fails for a reason that seems different. The strong student now conjectures "$n = 1$ and $(2,2)$" and is stuck on the proof, because there are p^{n^2} matrices and the case-by-case failures do not visibly share a cause. The wall is: *what property must M have?*

The model chain. *Step 1: the condition means G is one p^n-cycle.* Write $x_k = G^{(k)}(0)$, $x_0 = 0$. If $x_1, \ldots, x_{p^n}$ are distinct they are all of $\mathbb{F}_p^n$, so $0 = x_j$ for some $1 \le j \le p^n$; since $x_{k+1} = G(x_k)$, the sequence then has period j, and $j < p^n$ would give $x_1 = x_{1+j}$ with both indices in range. So $j = p^n$: every vector lies on a single cycle of G of length p^n, and in particular $G^{(p^n)}(x) = x$ for *every* x.

Step 2: the linear part is unipotent. By induction $G^{(k)}(x) = M^k x + G^{(k)}(0)$. Putting $k = p^n$ and using Step 1 for all x: $M^{p^n} x = x$, so $M^{p^n} = I$. Now M and I commute, and $\binom{p^n}{j} \equiv 0 \pmod{p}$ for $0 < j < p^n$, so in characteristic p

$$(M - I)^{p^n} = M^{p^n} - I = 0.$$

Hence $N := M - I$ is nilpotent, and a nilpotent $n \times n$ matrix satisfies $N^n = 0$ (the kernels of $N, N^2, \ldots$ strictly increase until they stop, and they must reach the whole space within n steps).

Step 3: a closed form. $M^j = (I + N)^j = \sum_i \binom{j}{i} N^i$, and $\sum_{j=0}^{k-1} \binom{j}{i} = \binom{k}{i+1}$ (hockey stick). Therefore

$$G^{(k)}(0) = \sum_{j=0}^{k-1} M^j v = \sum_{i=0}^{n-1} \binom{k}{i+1} N^i v.$$

Step 4: the obstruction. Take $k = p^{n-1}$. Every coefficient $\binom{p^{n-1}}{i+1}$ with $1 \le i+1 \le n$ is divisible by p provided $n < p^{n-1}$. Then $G^{(p^{n-1})}(0) = 0$, so the orbit of 0 has length dividing $p^{n-1} < p^n$ — contradiction. For $n \ge 2$, $n < p^{n-1}$ fails only when $p^{n-1} \le n$: $p = 2$ needs $2^{n-1} \le n$, i.e. $n \le 2$; $p \ge 3$ needs $3^{n-1} \le n$, i.e. $n = 1$. So the only survivors are $n = 1$ and $(2,2)$.

Step 5: constructions. $n = 1$: $M = (1)$, $v = (1)$. $(2,2)$: $M = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, $v = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, so $G(x_1, x_2) = (x_1 + x_2,\ x_2 + 1)$ and the orbit is $(0,1), (1,0), (1,1), (0,0)$: four distinct vectors.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains compute small cases and find the answer. They split at (†): "I need a property every good M shares." The natural attempt keeps sampling matrices. The model asks what the single-cycle condition says about $G^{(p^n)}$ *as a map*, reads off $M^{p^n} = I$, and turns that into nilpotence of $M - I$ by the characteristic-p binomial theorem.

> *If the orbit of 0 is everything, then every point is on the same p^n-cycle, so $G^{(p^n)}$ is the identity map — not just at 0. The linear part of $G^{(p^n)}$ is M^{p^n}, so $M^{p^n} = I$.*

> *p^n is a power of p, and I am in characteristic p: $(M - I)^{p^n} = M^{p^n} - I = 0$. So $M = I + N$ with N nilpotent, and everything becomes binomial coefficients.*

> *The orbit formula is $\sum \binom{k}{i+1} N^i v$ with $i < n$. The first k that kills all these binomials mod p is a power of p bigger than n. If that power is below p^n, the orbit closes early.*

Layer 3 THE WHY

Why look at $G^{(p^n)}$ rather than at G? Because the condition is about the *length* of an orbit, and length is controlled by the smallest power of G that fixes things. The question "when is $G^{(k)}(0) = 0$?" is the right one to ask, and the natural candidate for k is the cycle length itself. Having that, the step from "fixes 0" to "fixes everything" is where the hypothesis is spent: one cycle through all points means no point is special.

Why does unipotence appear, and why is it not a trick? Homogenise. An affine map $x \mapsto Mx + v$ on $\mathbb{F}_p^n$ is the linear map

$$\widehat{G} = \begin{pmatrix} M & v \\ 0 & 1 \end{pmatrix}$$

on $\mathbb{F}_p^{n+1}$ restricted to the slice "last coordinate $= 1$". A linear map of order a power of p in characteristic p is always of the form $I + (\text{nilpotent})$, because $X^{p^m} - 1 = (X-1)^{p^m}$ over $\mathbb{F}_p$: the only p-power roots of unity are 1. So the moment the problem asks for an orbit of length p^n, the matrix has to be unipotent. That is the structural fact; Step 2 is its two-line proof.

Why is the answer "$n = 1$ or $(2,2)$"? Because the order of a unipotent $m \times m$ matrix $I + N'$ in characteristic p is the least p^s with $p^s \geq$ (nilpotency index of N'), and that index is at most m. For $\widehat{G}$, $m = n+1$, so the order of G is at most the least power of p that is $\geq n+1$. We need order p^n, so $p^{n-1} < n+1$. For $p = 2$, $n = 2$ this gives $2 < 3$; for anything bigger it fails. Concretely: the $(2,2)$ example is a two-bit binary counter. $G(x_1, x_2) = (x_1 + x_2, x_2 + 1)$ flips the low bit and adds the carry into the high bit — and the carry is the low bit itself, which is linear. A three-bit counter needs the carry $x_2 x_3$ into the top bit, a product, which no affine map can produce. That is why $(2,3)$ fails, and it is a way to *see* the answer before proving it.

Kedlaya–Ng begin with a Fitting decomposition $\mathbb{F}_p^n = V \oplus W$, $M - I$ nilpotent on V and invertible on W, and show $W \neq 0$ yields a coset $w + V$ that the orbit of a point w never leaves. Here is where that comes from: it is the "no eigenvalue other than 1" half of unipotence, proved by exhibiting a short orbit. Step 2 above reaches the same conclusion without the decomposition, from $M^{p^n} = I$ and Frobenius on binomial coefficients, which uses only facts a Further Maths student can check.

> **Caveat.** Verified by exhaustive search over all (M, v): $(2,1), (3,1), (5,1), (2,2)$ admit a full orbit; $(3,2), (5,2), (2,3)$ do not. The step "nilpotent implies $N^n = 0$" is standard linear algebra and is quoted rather than proved in full.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "the p^n vectors $G^{(k)}(0) \ldots$ are distinct" — an orbit that exhausts a set of size p^n. Write down: *G is a single p^n-cycle, so $G^{(p^n)} = \mathrm{id}$ and $M^{p^n} = I$.*
2. Characteristic-p algebra: $(A+B)^{p^m} = A^{p^m} + B^{p^m}$ for commuting A, B, reducing to $M = I + N$ with N nilpotent; then expand orbits in binomial coefficients and use Lucas/Kummer divisibility $p \mid \binom{p^m}{j}$ for $0 < j < p^m$.
3. Hesitation occurs after the small cases, facing the full space of matrices. What reinforces it away is looking for a *necessary* algebraic identity satisfied by every good M — and the cheapest identity available is $G^{(\text{order})} = \mathrm{id}$.
4. The trap is to believe the problem is about choosing M cleverly (diagonalisable, companion matrices of primitive polynomials, which give orbit length $p^n - 1$ for linear maps). Those are the right tool for *linear* maps and a dead end here: the affine translation is essential, and it forces eigenvalue 1 only.

5. Slogan: *a cycle of length p^m in characteristic p means unipotent; then everything is binomial coefficients mod p.* Recognise it faster by homogenising any affine map into a linear map one dimension up.

6. Decision rule: **when a map on a finite structure must have a single orbit, write "$G^{(\text{size})} = \mathrm{id}$" and extract what this forces algebraically; in characteristic p with p-power orders, expect $(X-1)^{p^m}$.** Companion rule: when $M = I + N$ with N nilpotent, sums of powers of M collapse to binomial coefficients by the hockey-stick identity.

▷ **Cue:** an orbit of a map is required to hit every element of a finite set $\longrightarrow$ **Move:** the map is a single cycle; the size-th iterate is the identity everywhere, not just at the starting point

▷ **Cue:** a matrix over $\mathbb{F}_p$ with $M^{p^m} = I$ $\longrightarrow$ **Move:** $(M-I)^{p^m} = 0$ by the binomial theorem in characteristic p: write $M = I + N$ with N nilpotent

▷ **Cue:** an affine map $x \mapsto Mx + v$ $\longrightarrow$ **Move:** homogenise to the $(n+1) \times (n+1)$ matrix $\begin{pmatrix} M & v \\ 0 & 1 \end{pmatrix}$ and apply linear-algebra facts

▷ **Cue:** $\sum_{j<k}(I+N)^j$ $\longrightarrow$ **Move:** $= \sum_i \binom{k}{i+1} N^i$; then ask for which k all these binomials vanish mod p

▷ **Cue:** an iteration that behaves like counting in base p $\longrightarrow$ **Move:** ask whether the carry is linear; carries are products, so only very short counters are affine

2012 A6

2012 A6. Let $f(x,y)$ be a continuous, real-valued function on $\mathbb{R}^2$. Suppose that, for every rectangular region R of area 1, the double integral of $f(x,y)$ over R equals 0. Must $f(x,y)$ be identically 0?

Fields: real analysis (differentiation under the integral sign, fundamental theorem of calculus); integral geometry **Position:** A6 — the hardest slot on the paper, but it yields to one systematic move; the difficulty is that the hypothesis looks like it gives only "zero" and nothing to differentiate. **Answer:** Yes. In fact the axis-parallel rectangles of area 1 already force $f \equiv 0$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You first look for a counterexample, which is the right instinct for a "Must …?" question. Periodic functions come to mind: $f(x,y) = \sin(2\pi x)$ integrates to 0 over every axis-parallel unit *square*, since each horizontal slice covers one full period. So you try to kill every rectangle $[x, x+c] \times [y, y+1/c]$ as well: you would need $\sin(2\pi x)$ to integrate to zero over every interval of length c, for every c — impossible. Then $\sin(2\pi x)\sin(2\pi y)$, $\sin(2\pi(x+y))$, sums with different periods: each one is defeated by some aspect ratio c. Now you suspect the answer is yes, and you try to prove it. The natural first proof idea is to take an antiderivative F with $F_{xy} = f$ and write the hypothesis as

$$F(x+c, y+\tfrac{1}{c}) - F(x+c, y) - F(x, y+\tfrac{1}{c}) + F(x,y) = 0 \quad \text{for all } x, y,\ c > 0,$$

then differentiate in x and y to get $f(x+c, y+\frac{1}{c}) - f(x+c, y) - f(x, y+\frac{1}{c}) + f(x,y) = 0$: the corners of every area-1 axis-parallel rectangle satisfy $f(A) + f(C) = f(B) + f(D)$. That is correct and is Kedlaya–Ng's Lemma 1. But from here to $f \equiv 0$ requires rotating rectangles, an irrational-angle density argument, circles of diameter at least $\sqrt{2}$ and a further lemma — a long road that a student without the plan will not find. The wall is: corner identities are four-point statements, and turning them into "f is constant" needs a lot of geometry.

The model chain. Step 1: differentiate in the only free parameter. Fix a corner and vary the aspect ratio. With the bottom-left corner at (x,y),

$$0 = I(c) := \int_x^{x+c} \int_y^{y+1/c} f(u,w)\, dw\, du \qquad (c > 0).$$

Since f is continuous, Leibniz's rule applies (inner integral $H(u,c) = \int_y^{y+1/c} f(u,w)\,dw$ is continuous with continuous $\partial_c H = -c^{-2} f(u, y+1/c)$):

$$0 = I'(c) = \int_y^{y+1/c} f(x+c,w)\,dw \;-\; \frac{1}{c^2}\int_x^{x+c} f(u, y+\tfrac{1}{c})\,du.$$

Multiply by c. The right side has length $1/c$ and the top side has length c, so this says

$$\text{(average of } f \text{ on the right side)} = \text{(average of } f \text{ on the top side).}$$

Fixing instead the top-left corner (rectangle $[x, x+c] \times [Y-1/c, Y]$) and differentiating in c gives right side = bottom side; so all four sides of every axis-parallel area-1 rectangle carry the same average of f.

Step 2: slide along a row of rectangles. Fix $c > 0$ and the rectangles $R_m = [mc, (m+1)c] \times [0, 1/c]$, $m \in \mathbb{Z}$. R_m and R_{m+1} share a vertical side. Bottom of R_m = right of R_m = left of R_{m+1} = bottom of R_{m+1} (averages). Translating the whole row, for every real x and every $c > 0$:

$$\int_x^{x+c} f(u,0)\,du = \int_{x+c}^{x+2c} f(u,0)\,du, \;\text{ and so }\; \int_x^{x+c} f(u,0)\,du = \int_{x+mc}^{x+(m+1)c} f(u,0)\,du \;\; (m \in \mathbb{Z}).$$

Step 3: shrink. Fix $t > 0$, let $n \geq 1$ and use Step 2 with $c = t/n$ and $m = n$:

$$\frac{n}{t}\int_x^{x+t/n} f(u,0)\,du = \frac{n}{t}\int_{x+t}^{x+t+t/n} f(u,0)\,du.$$

As $n \to \infty$, continuity makes the left side tend to $f(x,0)$ and the right side to $f(x+t,0)$. So $f(\cdot,0)$ is constant. Nothing singled out the line $y = 0$ (translate the rows vertically), nor the horizontal direction (swap the roles of the axes), so f is constant on every horizontal and every vertical line, hence constant on $\mathbb{R}^2$: any two points are joined by a horizontal segment followed by a vertical one.

Step 4: the constant is 0. A constant K integrates to K over a region of area 1, so $K = 0$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains believe the answer is yes. They split at (†): *what do I do with a family of identities $I(\text{rectangle}) = 0$?* The natural attempt differentiates in the *position* (x, y) and obtains a four-corner identity, which is true but a long way from the goal. The model differentiates in the *shape* c, and gets identities about whole sides — one-dimensional averages — which chain together along a row of rectangles.

> *My counterexample attempts all died to changing the aspect ratio. So the aspect ratio is where the information is. I'll differentiate the hypothesis with respect to c.*

> *Differentiating a double integral in its limits leaves single integrals over the edges. So the conclusion will be about averages along sides, and two rectangles glued along a side share that side.*

> *I now know all intervals of length c on a line, spaced by c, have the same integral. Make c tiny and the integrals become point values: the function cannot tell x from x + t.*

Layer 3 THE WHY

Why differentiate in c and not in x or y? Look at what the hypothesis actually fixes. A rectangle of area 1 with sides parallel to the axes is described by three numbers: its position (x, y) and its aspect ratio c (the other side is forced to be $1/c$). Differentiating in x or y is available

for *any* family of rectangles — unit squares included — and unit squares do *not* force $f = 0$ (the function $\sin(2\pi x)$ survives them). So an argument that only moves rectangles around cannot possibly finish the job without further input; the extra input has to come from c. That is not a matter of taste: the counterexample tells you which parameter carries the content. This is the concrete version of the rule "use the hypothesis where it is strongest".

Why do edge averages appear, and why are they *averages*? When a limit of integration moves, the fundamental theorem of calculus says the integral changes by the integrand at that limit — here the integrand is an edge integral. The right side moves at speed 1 and the top at speed $-1/c^2$; multiplying through by c divides each edge integral by its own length. The area constraint $c \cdot \frac{1}{c} = 1$ is exactly what makes these weights the reciprocals of the side lengths. So "average right = average top" is the hypothesis "area = 1" differentiated, nothing more.

Why tile? An identity between two sides of one rectangle does not move you along a line. Two rectangles that share a side do. A row of congruent rectangles is the simplest configuration in which consecutive bottom sides are linked through shared vertical sides, and Step 2 is the resulting telescoping.

Why the limit? "Equal integrals over consecutive intervals of every length" is a statement about f integrated; to get a pointwise statement, shrink the length and divide by it — the definition of the derivative of $x \mapsto \int_0^x f$, i.e. the fundamental theorem of calculus once more. The whole proof is FTC applied three times: to the c-derivative, implicitly in the averages, and in the final limit.

Kedlaya–Ng's first solution differentiates in position and needs four lemmas, including an irrational-rotation density argument; this is also where the problem's permission to rotate rectangles is used. Their second solution (Eric Larson's) is the chain above: it uses only axis-parallel rectangles. Here is where Larson's opening move comes from — it is the unit-square counterexample, read backwards.

Caveat. The official second solution writes the final step as $\int_0^{c/n} f(u,0)\,du = \int_c^{c+1/n} f(u,0)\,du$; the upper limit should be $c + c/n$ (replacing c by c/n and using n copies). The version in Step 3 above is the corrected one. The use of Leibniz's rule is justified by continuity of f and of $\partial_c H$ on compact sets.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the word "area 1" in "every rectangular region R of area 1": the rectangles come in a one-parameter family of shapes, $c \times \frac{1}{c}$. Write $I(c) = \int_x^{x+c} \int_y^{y+1/c} f = 0$ immediately.
2. Differentiation under the integral sign with respect to the shape parameter, followed by a tiling (telescoping) argument and a limit that converts interval averages into point values.
3. Hesitation occurs after the corner identity $f(A) + f(C) = f(B) + f(D)$ has been found and seems to be the natural next object. Reinforce with the counterexample test: unit squares satisfy the same corner-type identities for $\sin(2\pi x)$, so anything derivable without varying c cannot finish.
4. Two traps. The first is the unit-square counterexample generalised too hopefully — believing some periodic function works; test every candidate against every aspect ratio. The second is the long road through rotations: the problem allows tilted rectangles, which tempts you to use them, when the axis-parallel family already suffices.
5. Slogan: *a family of vanishing integrals with a free parameter: differentiate in the parameter your counterexamples could not survive.* Moving limits produce boundary terms; glue regions along shared boundaries; shrink to points.
6. Decision rule: **when a hypothesis says an integral vanishes for a whole family of regions, identify the parameters of the family, find which one a failed counterexample was killed**

by, and differentiate in that one; the fundamental theorem of calculus turns the region statement into a boundary statement. Companion rule: equal integrals over intervals of every small length at two places mean equal values there — divide by the length and let it go to 0.

- ▷ **Cue:** an integral over a region is known for a family of regions depending on a continuous parameter $\longrightarrow$ **Move:** differentiate with respect to the parameter; the region identity becomes an identity on its boundary
- ▷ **Cue:** a "must it be zero?" question where a periodic candidate survives some but not all regions $\longrightarrow$ **Move:** the parameter that kills the candidate is the one carrying the information; build the proof around varying it
- ▷ **Cue:** identities relating two sides of a single rectangle $\longrightarrow$ **Move:** tile a row (or grid) of congruent rectangles so shared sides telescope the identities along a line
- ▷ **Cue:** $\int_x^{x+h} g = \int_{x+t}^{x+t+h} g$ for arbitrarily small h $\longrightarrow$ **Move:** divide by h and let $h \to 0$: $g(x) = g(x+t)$

Putnam 2012 — Paper B

2012 B1

2012 B1. Let S be a class of functions from $[0,\infty)$ to $[0,\infty)$ that satisfies:

(i) The functions $f_1(x) = e^x - 1$ and $f_2(x) = \ln(x+1)$ are in S;

(ii) If $f(x)$ and $g(x)$ are in S, the functions $f(x)+g(x)$ and $f(g(x))$ are in S;

(iii) If $f(x)$ and $g(x)$ are in S and $f(x) \geq g(x)$ for all $x \geq 0$, then the function $f(x)-g(x)$ is in S.

Prove that if $f(x)$ and $g(x)$ are in S, then the function $f(x)g(x)$ is also in S.

Fields: algebra, functional equations (closure systems) **Position:** B1 – opener, but the construction it demands is the whole paper's B-side signature move **Answer:** Yes: compose through $\ln(x+1)$ and $e^x - 1$ to turn the sum $\ln(f+1)+\ln(g+1)$ back into a product, then subtract off $f+g$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The three closure rules on offer are addition, composition, and conditional subtraction – nothing multiplicative in sight. The honest first move is to try to build multiplication out of these three operations directly and algebraically, the way a student reaches for $ab = \frac{1}{4}\big((a+b)^2-(a-b)^2\big)$ or $ab = \frac{1}{2}\big((a+b)^2-a^2-b^2\big)$. Both identities need squaring, and squaring is not a listed operation – $x \mapsto x^2$ is not given to be in S, and trying to manufacture it from f_1, f_2 by repeated addition and composition is itself an unsolved sub-problem, arguably harder than the original. A second natural attempt tries to get traction from the specific generators: compose f_1 and f_2 with each other, notice $f_1(f_2(x)) = x$ and $f_2(f_1(x)) = x$, and feel that this identity ought to matter – without yet seeing what to do with it. Left here, the attempt has correctly identified the only special structural fact available (the generators are inverse to one another) but has not converted that fact into an operation that produces fg from f and g, and can burn a large amount of time trying squaring-based identities that this closure system simply does not support.

The model chain. Let $f, g \in S$. By (ii) applied to f_2 and f, and to f_2 and g: $\ln(f(x)+1) \in S$ and $\ln(g(x)+1) \in S$. By (ii) (addition), their sum $\ln(f(x)+1)+\ln(g(x)+1) \in S$. By (ii) (composition) with $f_1(x) = e^x - 1$ as the outer function:

$$f_1\big(\ln(f+1)+\ln(g+1)\big) = e^{\ln(f+1)+\ln(g+1)} - 1 = (f+1)(g+1) - 1 = fg + f + g,$$

so $fg+f+g \in S$. By (ii) (addition), $f+g \in S$. Since $f, g \geq 0$ on $[0,\infty)$, $fg \geq 0$, hence $fg+f+g \geq f+g$ pointwise, so by (iii), $(fg+f+g)-(f+g) = fg$ is in S.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both attempts notice, correctly, that f_1 and f_2 are compositional inverses of one another on $[0,\infty)$: $f_1(f_2(x)) = x$. They split on what that observation is *for*. The natural

attempt treats it as a curiosity while hunting for a squaring-based identity. The model treats it as the entire mechanism: it is the identity map disguised as a composition, and the only reason to write the identity map as a composition of these two specific functions is that $f_2 = \ln(x+1)$ turns sums into logs of products.

> *I am given, as generators, a function and its own compositional inverse – that is never accidental in a closure problem.* $\ln(x+1)$ *is a shifted logarithm; logarithms turn products into sums. I don't need to build multiplication from scratch out of* $+, \circ, -$*; I need to route* f *and* g *through* $\ln(x+1)$*, add there (where I do have closure), and route back out through* $e^x - 1$*, its inverse.*

> *Once I've composed my way back to* $(f+1)(g+1)-1$*, I'm not finished – that's* $fg+f+g$*, not* fg*. But* $f+g$ *is separately available in* S *by the addition rule, and* $fg \geq 0$ *means the extra term only ever makes* $fg+f+g$ *larger than* $f+g$*, never smaller – so rule (iii)'s hypothesis* $f(x) \geq g(x)$ *for all* x *is satisfied exactly, and subtraction is licensed, not just plausible.*

Layer 3 THE WHY

The three closure rules define S as, in effect, a semiring-like structure under $+$ that is also closed under composition and (conditionally) under $-$: a purely additive-and-compositional toolkit. Multiplication of two arbitrary nonnegative functions is not an additive operation, so no amount of rearranging $+, \circ, -$ alone can reach it *unless* some element of the toolkit already encodes the bridge between addition and multiplication. That bridge is one of the oldest facts in mathematics: $\ln(ab) = \ln a + \ln b$, equivalently $ab = \exp(\ln a + \ln b)$. The problem does not ask the solver to discover logarithms; it hands them over as f_2, already shifted by $+1$ to map $[0,\infty) \to [0,\infty)$ (so that composing stays inside the codomain the whole problem is stated on), together with its inverse $f_1 = e^x - 1$, similarly shifted. Seeing f_1 and f_2 as inverse functions is not merely a nice fact to log – it is the statement, in the vocabulary of this problem, "here is your log-exp bridge, already closed under composition and shifted to respect the domain constraint $[0,\infty)$." Once that is recognised, the composition step is forced, not guessed. The residual $+f+g$ left over from $(f+1)(g+1)-1$ is likewise forced by the shift: shifting by 1 before taking logs (to keep everything nonnegative) necessarily reintroduces linear terms when the shift is undone, and rule (iii) exists in the axiom list *precisely* to let the solver discard exactly that residue once it is shown nonnegative. Nothing about this problem is arbitrary: every one of the three closure rules earns its keep in the six-line proof, which is itself evidence the axioms were chosen backwards from this argument.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is structural, not computational: the two named generators $e^x - 1$ and $\ln(x+1)$ are inverse functions of one another on $[0,\infty)$. On seeing two named "seed" functions in a closure problem, write down immediately whether they are related (inverses, one a shift of the other, one the derivative of the other) before trying to use them as raw material for algebra.

2. That cue suggests the method: if the generators are exp-shifted and log-shifted inverses, the toolkit is secretly "addition of logs equals log of product," so route the target operation (multiplication) through the pair via composition, using the addition rule where you actually have closure.

3. Hesitation occurs when squaring-based identities for ab come to mind first, because they are the standard reflex for building products from sums; that reflex has to be overruled by checking, explicitly, whether $x \mapsto x^2$ is available in the given system. It is not, and confirming that in one line prevents a long detour.

4. The trap is treating $f_1 \circ f_2 = \mathrm{id}$ as a throwaway observation rather than the design of the

problem; a solver who logs the fact and moves on without asking "what is an identity-via-composition normally used for" will not reach the construction.

5. Recognisable faster next time by treating any closure system whose generators include a function and its explicit inverse as a signal that the target operation is meant to be reached by conjugating through that inverse pair – $h \mapsto \phi^{-1}(\phi(h_1) + \phi(h_2))$ for some ϕ in the generator list.

6. Reusable rule: to build a nonlinear target operation (multiplication, in this case) from a toolkit closed only under linear operations $(+, -, \circ)$, look for a generator pair that is mutually inverse – it is almost always there to serve as a log/exp-style conjugation bridge, converting the target operation into one the toolkit already has, with only a residual correction term (here $f + g$) left to remove via the toolkit's subtraction rule.

▷ **Cue:** a closure system whose generator list includes two functions that are compositional inverses of one another $\longrightarrow$ **Move:** treat the pair as a conjugation bridge: apply the "inward" generator to each input, combine with the operation the system already supports, then apply the "outward" (inverse) generator to recover a target operation the system does not directly support

▷ **Cue:** a construction produces the target quantity plus an unwanted nonnegative residual term (e.g. $fg + f + g$ when the goal is fg) $\longrightarrow$ **Move:** check whether the residual is itself already known to lie in the system and is pointwise dominated the right way; if so, the conditional subtraction rule removes it in one step rather than requiring a fresh construction

2012 B2

2012 B2. Let P be a given (non-degenerate) polyhedron. Prove that there is a constant $c(P) > 0$ with the following property: If a collection of n balls whose volumes sum to V contains the entire surface of P, then $n > c(P)/V^2$.

Fields: geometry, inequalities (Hölder's inequality) **Position:** B2 – early in the B-side but requires spotting a dimension mismatch, not just applying a formula **Answer:** Yes: with A the area of any fixed face of P, the constant $c(P) = \frac{16A^3}{9\pi}$ works, so $n > \frac{16A^3}{9\pi V^2}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The instinct on seeing "n balls, total volume V, must cover a surface" is an area/volume comparison: total surface area of the n balls is at least the area of what they cover, so bound $\sum 4\pi r_i^2$ below by (something related to) the polyhedron's surface area A_{tot}, and separately note $\sum \frac{4}{3}\pi r_i^3 = V$. The natural next step is Cauchy-Schwarz, because Cauchy-Schwarz is the reflex tool for relating a sum of squares to a sum of things: $\left(\sum r_i^2\right)^2 \le n \sum r_i^4$. This is where the attempt stalls – it has produced a relation between $\sum r_i^2$ and $\sum r_i^4$, but the given data is $\sum r_i^2$ (from area) and $\sum r_i^3$ (from volume). The exponents 2 and 4 do not match the exponents actually in hand, 2 and 3; the attempt has reached for the wrong tool because it reached for the *familiar* tool rather than the one whose exponents fit the problem. Trying to force r_i^4 into the argument (e.g. via $r_i^4 \le (\max r_i)\, r_i^3$) reintroduces an unbounded quantity ($\max r_i$) that the problem gives no control over, and the argument collapses.

The model chain. Fix any face F of P with area $A > 0$. Since the n balls (radii $r_1, \ldots, r_n$) cover the entire surface of P, they cover F. The intersection of a ball of radius r_i with the plane containing F is a disc of radius at most r_i, covering area at most πr_i^2 of that plane, so

$$\sum_{i=1}^{n} \pi r_i^2 \ge A.$$

Also $\sum_{i=1}^{n} \frac{4}{3}\pi r_i^3 = V$. Write $n = \sum_{i=1}^{n} 1$ and apply Hölder's inequality with exponents $p = 3$, $q = \frac{3}{2}$ (so $\frac{1}{p} + \frac{1}{q} = 1$), to $a_i = 1$, $b_i = \left(\frac{4}{3}\pi r_i^3\right)^{2/3}$:

$$\sum_{i=1}^{n} \left(\tfrac{4}{3}\pi r_i^3\right)^{2/3} \le n^{1/3} \left(\sum_{i=1}^{n} \tfrac{4}{3}\pi r_i^3\right)^{2/3} = n^{1/3} V^{2/3},$$

i.e. $nV^2 \ge \left(\sum_i \left(\frac{4}{3}\pi r_i^3\right)^{2/3}\right)^3 = \left(\frac{4}{3}\pi\right)^2 \left(\sum_i r_i^2\right)^3 \ge \frac{16}{9}\pi^2 A^3$, so $nV^2 \ge \frac{16}{9\pi} A^3$. Any $c(P) < \frac{16A^3}{9\pi}$ therefore satisfies $n > c(P)/V^2$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths reach the same two sums: $\sum \pi r_i^2 \ge A$ (from area) and $\sum \frac{4}{3}\pi r_i^3 = V$ (from volume). They split on which inequality is used to relate the two sums to n. The natural attempt reaches for Cauchy-Schwarz, which relates $\left(\sum r_i^2\right)^2$ to $\sum r_i^4$ – a pairing of *equal* exponents (2 and 2, doubled to 4). The model asks what exponent pairing the problem actually contains – 2 (area) against 3 (volume) – and reaches for Hölder's inequality with the specific conjugate pair $(3, \frac{3}{2})$ that produces exactly 2 as a weighted combination.

> *Cauchy-Schwarz is the special case of Hölder with $p = q = 2$; it relates a sum to itself, same exponent on both sides. I have two* different *exponents here,* 2 *from area and* 3 *from volume, with a counting variable n sitting between them. I should not force this into Cauchy-Schwarz's shape; I should ask which exponents p, q with $\frac{1}{p} + \frac{1}{q} = 1$ let me pair "1" (from $n = \sum 1$) against "r_i^3" (from volume) and get "r_i^2" (area) to fall out.*

> *I want $(a_i b_i)$ to equal r_i^2 with $a_i = 1$. If $b_i = (r_i^3)^{2/3} = r_i^2$, then I need q such that $b_i^q = (r_i^2)^q = r_i^3$, forcing $q = \frac{3}{2}$, and then $p = 3$ by the conjugate relation. The exponents are not a guess; they are solved for directly from the requirement that Hölder's output land exactly on $\sum r_i^2$.*

Layer 3 THE WHY

The problem has a built-in dimensional mismatch: covering a surface is a two-dimensional demand (area A, scaling as r^2), while the resource being rationed is three-dimensional (volume V, scaling as r^3), and the quantity being bounded, n, is zero-dimensional (a count). Any inequality connecting these three has to reconcile three different homogeneity degrees at once, which is precisely what a single exponent pair in Cauchy-Schwarz cannot do – Cauchy-Schwarz only ever compares a quantity to itself at doubled exponent, so it is blind to genuine three-way scaling mismatches. Hölder's inequality is the general form built for exactly this: given any target exponent split between "how many terms" and "what power of the terms," it supplies the conjugate pair that realises it. Here the dimensional bookkeeping is transparent if you run it heuristically first: if n balls of roughly equal radius r split the volume evenly, $r \sim (V/n)^{1/3}$, and to cover area A needs $nr^2 \gtrsim A$, i.e. $n(V/n)^{2/3} \gtrsim A$, i.e. $n^{1/3}V^{2/3} \gtrsim A$, i.e. $n \gtrsim A^3/V^2$ – the A^3/V^2 shape of the answer is forced by dimensional analysis alone, before any inequality is proved rigorously. Kedlaya's write-up states the Hölder step without this motivation; the reason the exponents $(3, \frac{3}{2})$ are the right ones to reach for is that they are the unique conjugate pair compatible with the r^2-vs-r^3 mismatch the geometry hands you, and the correct response to "I have sums of different powers of the same variable" is always to solve for the Hölder exponents that connect them, not to reach for the symmetric special case.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that the problem supplies two sums over the same radii r_i at two *different* powers – $\sum r_i^2$ (area, via a face) and $\sum r_i^3$ (volume). On seeing mismatched powers

of the same variable that need to be related through a count n, write down immediately: this is a Hölder problem, not a Cauchy-Schwarz problem.

2. That cue suggests the method: solve for the conjugate exponent pair (p, q) with $\frac{1}{p} + \frac{1}{q} = 1$ that reproduces the target power (r^2) from the given power (r^3) and the counting term (1), then apply Hölder directly with that pair.

3. Hesitation occurs at the reflex step "relate a sum of squares to n" – Cauchy-Schwarz answers that reflex instantly but produces r_i^4, a power that is not in the problem. Reinforce the hesitation away by checking, explicitly, whether the exponent produced matches an exponent actually given before trusting the inequality chosen.

4. The trap is defaulting to Cauchy-Schwarz because it is the most memorised inequality involving a sum of squares and a count n, when the presence of a *third*, different exponent (the volume term, degree 3) is the signal that the symmetric special case cannot be the right tool.

5. Recognisable faster next time by treating "a bound must relate quantities of two different homogeneity degrees (area vs. volume, length vs. area, etc.) through a count" as an automatic cue for Hölder with solved-for exponents, and reaching for a rough dimensional-analysis estimate of the final exponent on V and A as a sanity check before writing the proof.

6. Reusable rule: when a bound must connect two sums of the same variable raised to different powers via a counting variable n, do not reach for Cauchy-Schwarz by reflex; solve for the Hölder conjugate exponents that produce the target power from the given power, and sanity-check the resulting scaling law by a one-line dimensional estimate before proving it rigorously.

▷ **Cue:** a geometric bound must relate quantities of different homogeneity degree (e.g. area $\sim r^2$ and volume $\sim r^3$) through a count n $\longrightarrow$ **Move:** solve for the Hölder conjugate exponent pair (p, q), $\frac{1}{p} + \frac{1}{q} = 1$, that reproduces the target power from the given power and the counting term $\sum 1$; do not default to the equal-exponent Cauchy-Schwarz special case

▷ **Cue:** a "must exist a constant such that..." existence claim in a geometry problem $\longrightarrow$ **Move:** fix one convenient sub-object (here, a single face) rather than working with the full polyhedron; the constant only needs to work for *some* valid choice, and a single face reduces a 3D covering problem to a 2D one

2012 B3

2012 B3. A round-robin tournament of $2n$ teams lasted for $2n - 1$ days, as follows. On each day, every team played one game against another team, with one team winning and one team losing in each of the n games. Over the course of the tournament, each team played every other team exactly once. Can one necessarily choose one winning team from each day without choosing any team more than once?

Fields: combinatorics (Hall's marriage theorem / systems of distinct representatives) **Position:** B3 – mid paper; the difficulty is entirely in recognising the right theorem, not in computation **Answer:** Yes, always: the $2n - 1$ days and $2n$ teams admit a system of distinct representatives (one distinct winning team per day).

Layer 1 THE WHAT — surface analysis

The natural attempt. The question asks for "one winning team per day, all different" – an assignment problem, and the natural first move is to try to build the assignment directly. Process the days in some order; on each day, pick any team that won that day and has not yet been used. This works for a while, because early on there are plenty of unused winners to choose from. The trouble is what happens near the end: on a late day, every team that won *that*

day might already have been used up on earlier days, and the greedy process has no mechanism to go back and re-route an earlier choice. A student pursuing this concretely (perhaps on small cases, $n = 2$ or $n = 3$) can usually patch it by hand-checking and backtracking, and will correctly guess the answer is always "yes" – but a greedy-with-backtracking argument does not generalise into a proof; it shows the answer works out on examples, not that it must work out for every possible schedule of $2n - 1$ days. The natural attempt has produced the right answer without a mechanism that scales, and is at risk of writing up "it always seems to work" as if that were a proof.

The model chain. Model the situation as a bipartite graph B: one side is the $2n - 1$ days, the other is the $2n$ teams, with an edge from day d to team t iff t won its game on day d. The question is exactly whether B has a matching saturating the days side – a system of distinct representatives for the "winners of day d" sets. By Hall's marriage theorem, such a matching exists iff for every subset D of the days, the set of teams that won on at least one day in D has size $\geq |D|$.

Fix any m with $1 \leq m \leq 2n - 1$ and any collection of m days. Suppose, for contradiction, that fewer than m distinct teams won on those m days combined. Since there are $2n > m$ teams in total (as $m \leq 2n - 1$), some team T won none of the games it played on those m days – i.e. T lost on each of those m days. Because the tournament pairs every two teams exactly once across the whole $2n - 1$ days, T's opponent is different on each day; so T lost to m *distinct* teams, one per day in the collection. Each of those m teams is therefore among the winners on some day in the collection – giving at least m distinct winning teams, contradicting the assumption of fewer than m. Hall's condition holds for every collection of days, so B has a matching saturating the days side: a system of distinct representatives exists, i.e. one can choose one winning team per day with no repeats.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths correctly land on the answer "yes" and both correctly sense that the problem is about avoiding collisions in an assignment. They split on how the "yes" is to be established. The natural attempt tries to *construct* the assignment procedurally (greedily, then patch failures by hand). The model instead reframes "choose one winner per day, no repeats" as an existence question about a bipartite graph – a system of distinct representatives – and reaches for the one theorem built to answer exactly that existence question without ever constructing the assignment explicitly.

> *"Pick one item from each of several sets, with no repeats across sets" is not a problem to solve constructively from scratch – it is Hall's marriage theorem's exact hypothesis, phrased in tournament language instead of set language. I should stop trying to build the matching by hand and instead verify Hall's condition: does every collection of m days have at least m distinct teams winning among them?*

> *To verify Hall's condition I need a lower bound on the number of distinct winners across any m days, and the tournament structure gives me exactly one fact I haven't used yet: every pair of teams meets* exactly once *in the whole tournament, so any team's opponents across m different days are m genuinely distinct teams. If a team lost every one of those m days, that alone produces m distinct winners – I don't need to inspect who won each individual game, only that some team's m distinct losses each certify a winner.*

Layer 3 THE WHY

The phrase "choose one X from each of several groups so that no choice repeats" is, definitionally, the search for a system of distinct representatives, and Hall's marriage theorem is

not a technique invoked by analogy here – it is the precise, general-purpose answer to that exact question, stated as a graph-theoretic existence criterion so that no construction needs to be exhibited by hand. Recognising the tournament's "days vs. teams, edge = won" structure as a bipartite graph is therefore not a modelling choice made for convenience; it is the only structure under which the actual question being asked (existence of a matching saturating one side) has a name and a theorem. The one piece of work left after that recognition is verifying Hall's condition, and the tournament's defining property – every pair of teams meets exactly once across the whole event – is precisely what is needed to prove it: it guarantees that a team which loses on m different days does so to m genuinely different opponents (never the same opponent twice), so a shortfall in distinct winners on some collection of days would force a single team to have distinct losses without distinct winners, which is a direct contradiction of what "distinct opponent per day" means. The round-robin's "exactly once" clause, which looks like flavour text describing the tournament format, is in fact the load-bearing hypothesis that makes Hall's condition provable at all; a tournament where teams could replay each other would not obviously satisfy it.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the phrasing itself: "choose one [team] from each [day] without choosing any [team] more than once." On seeing that shape of question – one representative per group, no repeats across groups – write down immediately "system of distinct representatives / Hall's marriage theorem," before attempting any construction.

2. That cue suggests the method: build the bipartite graph (days vs. teams, edge = won), and reduce the problem entirely to verifying Hall's condition on that graph, rather than constructing an explicit assignment.

3. Hesitation occurs when the greedy construction seems to be working on small cases and it feels like "just keep going" should be a proof; reinforce it away by asking explicitly whether the greedy argument would survive an adversarial worst-case ordering of the $2n - 1$ days, which it visibly would not without backtracking that has no termination guarantee.

4. The trap is treating the problem as a constructive combinatorics exercise (build the assignment) rather than an existence question (does a valid assignment exist), which sends the solver toward case-by-case verification instead of toward the one theorem that settles existence questions of this exact shape in one line.

5. Recognisable faster next time by treating any "each of several days/committees/rounds must be paired to a distinct representative from a corresponding collection" as an automatic cue for Hall's theorem, and by immediately asking what structural fact of the problem will supply the needed lower bound on set sizes.

6. Reusable rule: "select one representative per group with no repeats" is always a Hall's marriage theorem question; the entire proof reduces to finding, inside the problem's own hypotheses, the fact that forces every subcollection of groups to have at least as many distinct available representatives as groups – here, "every pair meets exactly once" turning a team's m losses into m distinct certified winners.

▷ **Cue:** a problem asks to choose one representative from each of several groups (days, committees, rounds) with no repeats across groups $\longrightarrow$ **Move:** recognise this as a system-of-distinct-representatives question; model it as a bipartite graph (groups vs. candidates, edge = eligible) and reduce the problem to verifying Hall's condition rather than constructing an assignment by hand

▷ **Cue:** verifying Hall's condition requires a lower bound on how many distinct candidates a subcollection of groups reaches $\longrightarrow$ **Move:** look inside the problem's own uniqueness hypotheses (here, "every pair of teams meets exactly once") for the fact that forces distinctness; a single excluded item's distinct failures across the subcollection often directly certifies enough distinct successes elsewhere

2012 B4

2012 B4. Suppose that $a_0 = 1$ and that $a_{n+1} = a_n + e^{-a_n}$ for $n = 0, 1, 2, \ldots$. Does $a_n - \log n$ have a finite limit as $n \to \infty$? (Here $\log n = \log_e n = \ln n$.)

Fields: real analysis; asymptotics of recursive sequences; comparison with a differential equation **Position:** B4 — mid-paper; the answer is guessable in a minute, and the marks are all in making the guess rigorous without circularity **Answer:** Yes. In fact $e^{a_n} = n + O(\log n)$, so $a_n - \log n \to 0$; the sequence $a_n - \log n$ $(n \geq 1)$ is positive and strictly decreasing.

Layer 1 THE WHAT — surface analysis

The natural attempt. A strong student reads "$a_{n+1} - a_n = e^{-a_n}$" and hears a derivative. Replace a_n by $y(x)$ and the difference by y': $y' = e^{-y}$, so $e^y y' = 1$, $e^y = x + C$, $y = \log(x + C)$. So the guess is $a_n \approx \log n$, and $a_n - \log n = \log \frac{n+C}{n} \to 0$. The answer is almost certainly "yes".

Now the student tries to make it rigorous by working directly with $c_n = a_n - \log n$:

$$c_{n+1} - c_n = e^{-a_n} - \log\Big(1 + \frac{1}{n}\Big) = \frac{e^{-c_n}}{n} - \log\Big(1 + \frac{1}{n}\Big).$$

Since $\log(1 + \frac{1}{n}) = \frac{1}{n} - \frac{1}{2n^2} + \cdots$, this is roughly $\frac{e^{-c_n}-1}{n}$: if $c_n > 0$ it pushes c_n down, if $c_n < 0$ it pushes c_n up. That is a stable fixed point at $c = 0$, which is encouraging. But to turn it into a proof the student needs to know the sign of c_n to control e^{-c_n}, and to know how big c_n is to control the error in $e^{-c_n} \approx 1 - c_n$. Every estimate on c_{n+1} asks for an estimate on c_n of the same kind. Summing the differences,

$$c_N = c_1 + \sum_{n=1}^{N-1} \Big(\frac{e^{-c_n}}{n} - \log\big(1 + \tfrac{1}{n}\big) \Big),$$

the student sees a harmonic-type sum whose convergence depends on $c_n \to 0$ at some rate, which is what is being proved. The attempt is heading into a bootstrap: possible, but long and easy to get wrong, and the student does not yet know which inequality to induct on.

The model chain. **Step 1: linearise.** Put $b_n = e^{a_n}$. Then $e^{-a_n} = 1/b_n$ and

$$b_{n+1} = e^{a_n + e^{-a_n}} = b_n\, e^{1/b_n}.$$

Write $b_{n+1} = b_n + 1 + R_n$ with $R_n = b_n\big(e^{1/b_n} - 1 - \frac{1}{b_n}\big)$.

Step 2: bound the error. For $u > 0$, $e^u \geq 1 + u$, so $R_n \geq 0$. For the upper bound, the Taylor remainder gives $e^u - 1 - u = \frac{u^2}{2} e^{\xi}$ for some $\xi \in (0, u)$, so $e^u - 1 - u \leq \frac{u^2}{2} e^u$. Since $R_n \geq 0$ we have $b_{n+1} \geq b_n + 1$, hence $b_n \geq b_0 + n = n + e \geq e$ and $u = 1/b_n \leq 1/e$. Then $\frac{e^u}{2} \leq \frac{e^{1/e}}{2} < 1$, and

$$0 \leq R_n \leq b_n \cdot \frac{1}{b_n^2} = \frac{1}{b_n} \leq \frac{1}{n + e}.$$

Step 3: sum. Telescoping $b_n = b_0 + n + \sum_{i=0}^{n-1} R_i$ gives, for $n \geq 1$,

$$n + e \;\leq\; b_n \;\leq\; n + e + \sum_{i=0}^{n-1} \frac{1}{i + e} \;\leq\; n + e + \frac{1}{e} + \log n,$$

using $\sum_{i=0}^{n-1} \frac{1}{i+e} \leq \frac{1}{e} + \int_0^{n-1} \frac{dx}{x+e} = \frac{1}{e} + \log \frac{n-1+e}{e}$ and $n - 1 + e \leq en$.

Step 4: conclude. For $n \geq 1$,

$$0 < \log\Big(1 + \frac{e}{n}\Big) \leq a_n - \log n = \log \frac{b_n}{n} \leq \log\Big(1 + \frac{e + 1/e + \log n}{n}\Big).$$

Both bounds tend to 0 because $(\log n)/n \to 0$. So $a_n - \log n$ has the finite limit 0. The answer is *yes*. □

The official first solution, for comparison. Kedlaya and Ng prove $a_n > \log(n+1)$ by induction (using that $x \mapsto x + e^{-x}$ is increasing for $x > 0$ and $\log(n+1) + \frac{1}{n+1} \geq \log(n+2)$), then show $c_{n+1} - c_n = e^{-a_n} - \log\frac{n+1}{n} < \frac{1}{n+1} - \log\frac{n+1}{n} \leq 0$. A decreasing sequence of positive numbers converges. That proves existence of the limit without identifying it; their second solution is Steps 1–4.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree on the guess $a_n \approx \log n$, which comes from $y' = e^{-y}$. They split at the line

$$e^y\, y' = 1 \qquad (\dagger)$$

in the solution of the differential equation. The natural attempt uses $(\dagger)$ only to get the answer $y = \log(x + C)$ and then goes back to the original variable a_n, where the error terms are nonlinear in the unknown. The model notices that $(\dagger)$ says $\frac{d}{dx}e^y = 1$: in the variable e^y the equation is *linear with constant right-hand side*, and it transfers that change of variable to the recursion. In $b_n = e^{a_n}$ the recursion is "add 1, plus a small nonnegative error", which telescopes.

> *When I solved $y' = e^{-y}$ I multiplied by e^y and the equation became $(e^y)' = 1$. So e^y is the quantity that grows at a constant rate. I'll use $b_n = e^{a_n}$ in the recursion itself: then I expect $b_{n+1} \approx b_n + 1$, and all I have to control is one error term, which telescopes.*

> *The error $b_n(e^{1/b_n} - 1 - 1/b_n)$ is of size $1/b_n$, and I already know $b_n \geq n + e$ from the lower bound, so the errors add up to a harmonic sum — $O(\log n)$, which is small compared with n. That is all I need, because the question is about $\log(b_n/n)$.*

- ▷ **Cue:** a recursion $a_{n+1} = a_n + f(a_n)$ with f small $\longrightarrow$ **Move:** write the ODE $y' = f(y)$, solve it by separation, and use the separating function G with $G' = 1/f$ as the new variable: $G(a_{n+1}) \approx G(a_n) + 1$
- ▷ **Cue:** a limit of $a_n - \log n$ is asked for $\longrightarrow$ **Move:** exponentiate: the question is whether e^{a_n}/n converges, a ratio question about a nearly-linear sequence
- ▷ **Cue:** a telescoped sum of errors each bounded by C/b_i with $b_i \geq i$ $\longrightarrow$ **Move:** bound the total by a harmonic sum $O(\log n)$, then compare with the main term n

Layer 3 THE WHY

Why is e^{a_n} the right variable and not a trick?

Go back to what the recursion says. The step $a_{n+1} - a_n$ equals e^{-a_n}: the higher you are, the slower you climb. A sequence whose step depends on where it is has no simple closed form; a sequence whose step is *constant* is an arithmetic progression, the one recursion everyone can sum. So the question to ask of any autonomous recursion $a_{n+1} = a_n + f(a_n)$ is: is there a function G of the position such that G increases by (approximately) 1 per step? To first order,

$$G(a_{n+1}) - G(a_n) \approx G'(a_n)\, f(a_n),$$

so you want $G' = 1/f$. Here $f(a) = e^{-a}$, so $G'(a) = e^a$ and $G = e^a$. This is exactly separation of variables, $\int \frac{dy}{f(y)} = \int dx$, read as a statement about which quantity the process advances at unit speed. It is the discrete version of choosing the "natural clock" of the process. The official remark calls this "studying a sufficiently similar-looking differential equation"; what that remark leaves out is that the ODE hands you not only the answer but the *coordinate* in which the discrete proof is a telescoping sum.

Then the error. Because G is not linear, $G(a_{n+1}) - G(a_n)$ is not exactly 1; the second-order term is $\frac{1}{2}G''f^2 = \frac{1}{2}e^a e^{-2a} = \frac{1}{2}e^{-a} = \frac{1}{2b_n}$. That is why $R_n \approx \frac{1}{2b_n}$, and it tells you in advance that the errors sum to about $\frac{1}{2}\log n$: large in absolute terms (so $b_n - n$ is unbounded) but negligible against n. The problem asks about $a_n - \log n = \log(b_n/n)$, a *ratio*, so an additive error of $O(\log n)$ is harmless. Numerically, $a_n - \log n$ is about 1.03×10^{-6} at $n = 10^7$, consistent with $b_n \approx n + \frac{1}{2}\log n$.

Why the natural attempt stalls: in the variable $c_n = a_n - \log n$ the recursion is still nonlinear (e^{-c_n} appears), and all its error terms are functions of the unknown. The official first solution escapes that circularity in another way — by finding the exact comparison $a_n > \log(n+1)$ that makes $e^{-a_n} < \frac{1}{n+1}$ — but that inequality has to be guessed, and it is proved by comparing with the same exponential, $\log(n+1) + \frac{1}{n+1} \geq \log(n+2)$, which is $e^u \geq 1 + u$ again. Both proofs rest on $e^u \geq 1 + u$; the b_n route makes that visible.

Caveat. The claim $e^{a_n} = n + \frac{1}{2}\log n + O(1)$ is a heuristic from the second-order term and was checked only numerically ($n \leq 10^7$, where $n(a_n - \log n)/\log n$ drifts slowly down towards 0.5). The proof above needs only the cruder bound $b_n \leq n + e + 1/e + \log n$, which is fully proved. Positivity and strict decrease of $a_n - \log n$ for $1 \leq n < 2 \times 10^5$ were also checked numerically.

Layer $+\alpha$ BACKTESTING

1. The cue is the shape $a_{n+1} - a_n = f(a_n)$ together with the target $\log n$. On seeing it, write the ODE $y' = e^{-y}$ and solve it, keeping the intermediate line $(e^y)' = 1$.

2. The cue suggests changing to the variable in which the ODE is trivial, $b_n = e^{a_n}$, and then telescoping: $b_n = n + e + \sum R_i$ with $0 \leq R_i \leq 1/b_i$.

3. Hesitation occurs at the error sum: $\sum R_i$ is unbounded, which can look like failure. It is reinforced away by rereading the question: the target is $\log(b_n/n)$, a ratio, so an $O(\log n)$ error against n is enough.

4. The trap is working in $c_n = a_n - \log n$, where every bound requires the bound being proved, and a secondary trap is to answer "yes" from the ODE heuristic alone, with no rigorous error control. Another is to forget that $\log 0$ is undefined: the statement concerns $n \geq 1$.

5. Recognise it faster by asking of any recursion "next = current + small(current)": *what quantity does this process increase by exactly* 1 *each step, to first order?* The answer is $\int \frac{dy}{f(y)}$.

6. Reusable decision rule: *for a slowly moving recursion, solve the matching ODE by separation, then use the separated variable $G(a_n)$ with $G' = 1/f$ as the new unknown; the recursion becomes $G(a_{n+1}) = G(a_n) + 1 +$ (error of size $\frac{1}{2}G''f^2$), and the proof is a telescoped error sum.* The same rule gives $a_n \sim \sqrt{2n}$ for $a_{n+1} = a_n + 1/a_n$ (use a_n^2) and $a_n \sim 1/n$ for $a_{n+1} = a_n - a_n^2$ (use $1/a_n$).

2012 B5

2012 B5. Prove that, for any two bounded functions $g_1, g_2 : \mathbb{R} \to [1, \infty)$, there exist functions $h_1, h_2 : \mathbb{R} \to \mathbb{R}$ such that, for every $x \in \mathbb{R}$,

$$\sup_{s \in \mathbb{R}} (g_1(s)^x g_2(s)) = \max_{t \in \mathbb{R}} (x h_1(t) + h_2(t)).$$

Fields: real analysis; convex functions; supporting lines and convex duality **Position:** B5 — late in the paper, but short once the right object is seen; the difficulty is entirely in reading the right-hand side correctly **Answer:** Let $F(x) = \sup_s g_1(s)^x g_2(s)$. It is finite and convex; take $x\,h_1(t) + h_2(t)$ to be a supporting line of F at t. Then every line lies below F and the one at $t = x$ touches it.

Layer 1 THE WHAT — surface analysis

The natural attempt. The strong student looks for h_1, h_2 built from g_1, g_2, because the left side is indexed by s and the right side by t, and it seems natural to let t play the role of s. For fixed s the function $x \mapsto g_1(s)^x g_2(s) = g_2(s)e^{x \log g_1(s)}$ is an exponential in x, and the right side is made of lines in x. So the student tries to replace each exponential by one of its tangent lines: pick a point $x_0(s)$ and set

$$h_1(s) = g_2(s) \log g_1(s)\, g_1(s)^{x_0}, \qquad h_2(s) = g_1(s)^{x_0} g_2(s) - x_0 h_1(s).$$

An exponential lies above its tangent lines, so $xh_1(s) + h_2(s) \le g_1(s)^x g_2(s) \le F(x)$, where $F(x)$ denotes the left side. The inequality "$\ge$" is therefore fine. The wall is equality: the right side must *attain* $F(x)$ for each x. The line from s touches F only if the s-th exponential achieves the supremum at the point of tangency. But the supremum need not be achieved by any s at all — for instance $g_1 \equiv 1$, $g_2(s) = 2 - \frac{1}{1+s^2}$ gives $F \equiv 2$, never attained. No choice of $x_0(s)$ fixes that; indexing by s cannot produce a max. The student is stuck trying to repair a construction whose index set is wrong.

The model chain. Write $F(x) = \sup_{s \in \mathbb{R}} g_1(s)^x g_2(s)$.

Step 1: F is finite. Let $1 \le g_1 \le M_1$ and $1 \le g_2 \le M_2$. For $x \ge 0$, $g_1(s)^x \le M_1^x$; for $x < 0$, $g_1(s)^x \le 1$ because $g_1(s) \ge 1$. So $1 \le F(x) \le M_2 \max(1, M_1^x) < \infty$ for every x.

Step 2: F is convex. For each s, $\varphi_s(x) = g_2(s)e^{x \log g_1(s)}$ is convex in x (a nonnegative multiple of e^{cx}). A pointwise supremum of convex functions is convex: for $\lambda \in [0,1]$,

$$\varphi_s(\lambda x + (1-\lambda)y) \le \lambda \varphi_s(x) + (1-\lambda)\varphi_s(y) \le \lambda F(x) + (1-\lambda)F(y),$$

and taking the supremum over s on the left gives $F(\lambda x + (1-\lambda)y) \le \lambda F(x) + (1-\lambda)F(y)$.

Step 3: supporting lines exist. Fix t and let $\sigma(x) = \frac{F(x)-F(t)}{x-t}$ for $x \ne t$. Convexity is equivalent to: for $u < v < w$, the chord slopes satisfy $\frac{F(v)-F(u)}{v-u} \le \frac{F(w)-F(u)}{w-u} \le \frac{F(w)-F(v)}{w-v}$ (write v as a convex combination of u and w and rearrange). Applying this to the three points $\{x, t, y\}$ in their natural order shows σ is nondecreasing on $\mathbb{R} \setminus \{t\}$; in particular $\sigma(x) \le \sigma(y)$ whenever $x < t < y$. Hence $m(t) = \sup_{x<t} \sigma(x)$ is a finite real number with $\sigma(x) \le m(t) \le \sigma(y)$ for $x < t < y$. Multiplying out (and noting the sign of $x - t$),

$$F(x) \ge F(t) + m(t)(x - t) \qquad \text{for all } x \in \mathbb{R}.$$

Step 4: define h_1, h_2. Put $h_1(t) = m(t)$ and $h_2(t) = F(t) - t\,m(t)$. By Step 3, $xh_1(t) + h_2(t) \le F(x)$ for all x, t; at $t = x$ the two sides are equal. Therefore for each x the supremum over t of $xh_1(t) + h_2(t)$ equals $F(x)$ and is attained at $t = x$, so it is a maximum. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree that the right-hand side is "lines in x" and the left side is "exponentials in x". They split at the question of what t should index.

(N) "t plays the role of s": one line per exponential. This gives $\le$ everywhere but cannot force attainment, because the left side's supremum need not be attained.

(M) "t is a point on the x-axis": one line per point of the graph of F, the line that touches F at $x = t$.

The split is at the word max on the right. Once you insist the maximum be attained at every x, the natural index for the line that achieves it at x is x itself.

> *The right side is a maximum of lines in x. A maximum of lines is always convex. So before I construct anything, the left side had better be convex — and it is, because each $g_1(s)^x g_2(s)$ is an exponential in x and a supremum of convex functions is convex. Then the problem is the statement that a convex function is the maximum of its supporting lines, and I should index the lines by the point where they touch: $t = x$.*

> *Why are the hypotheses $g_i \geq 1$ and bounded there? Boundedness above makes F finite for $x > 0$; $g_1 \geq 1$ makes $g_1^x \leq 1$ for $x < 0$, so F is finite there too. A convex function that takes the value $+\infty$ somewhere has no supporting lines, so this is exactly what the hypotheses are for.*

- ▷ **Cue:** an identity of the form "(something) $= \max_t\big(x\,h_1(t) + h_2(t)\big)$" $\longrightarrow$ **Move:** recognise the right side as an upper envelope of lines, hence convex; prove the left side is convex and finite, and take h to be supporting lines indexed by the touching point
- ▷ **Cue:** a supremum over a parameter of functions each convex in x $\longrightarrow$ **Move:** the supremum is convex in x (one line of proof), provided it is finite
- ▷ **Cue:** a problem hypothesis that looks arbitrary (here $g \geq 1$, bounded) $\longrightarrow$ **Move:** ask which step fails without it; here it is finiteness of the supremum for $x < 0$ and $x > 0$

Layer 3 THE WHY

Why is convexity the object, and why is that not a leap?

Start from the definition of the right side. For each t, $x \mapsto xh_1(t) + h_2(t)$ is a straight line. The right side is the upper envelope of a family of straight lines — the function whose graph is the top edge of all of them. Any upper envelope of lines is convex, because each line is convex and "the chord lies above the graph" survives taking a supremum. So the problem is *necessarily* asking you to show the left side is convex; no construction can succeed otherwise. That is the circle-and-centre moment: the right side is a max of lines, therefore convexity is the object of interest.

Conversely, every finite convex function on $\mathbb{R}$ *is* the upper envelope of its supporting lines: at each point the chord slopes to the left are all at most the chord slopes to the right (the three-chord inequality is a rearrangement of the definition), so a line with slope between them passes through the point and stays below the graph. That is the whole content of Step 3, and it is the one piece of machinery beyond A-level. So the problem is the statement "convex $\iff$ max of lines" with a specific convex function supplied.

Kedlaya and Ng take logarithms first: $f(x) = \sup_s(x \log g_1(s) + \log g_2(s))$ is a supremum of *affine* functions, hence convex, and then they prove e^f is convex using weighted AM–GM. Here is where that comes from: in log form each member of the family is already a line, so convexity of f is immediate, and the only cost is transferring convexity through exp. The route above skips the logarithm, because $x \mapsto g_2 e^{x \log g_1}$ is already convex and "sup of convex is convex" costs one line. Both are correct; the direct one avoids a step. Their remark names the structure: this is convex duality. The left side is a Laplace-transform-like supremum over s; the right side is its representation by its tangent lines, the Legendre picture.

Now the hypotheses. Why $[1, \infty)$? If g_1 could take values near 0, then for $x < 0$ the terms $g_1(s)^x$ would be unbounded and $F(x) = +\infty$; a function with infinite values cannot equal a maximum of real numbers. Why bounded? The same failure for $x > 0$. The requirement $g_2 \geq 1$ is stronger than this argument needs: it uses only $g_2 \geq 0$ (for convexity of each member) and g_2 bounded (for finiteness). The problem setters chose one clean hypothesis for both functions. Noticing which hypotheses are load-bearing is how you find the proof: the only thing the hypotheses can be for is finiteness, so finiteness is a step, and the step after finiteness is convexity.

Caveat. The claim that $g_2 \geq 1$ is stronger than necessary refers to this proof only (it uses $0 \leq g_2 \leq M_2$); the problem as stated is proved in full above. The example $g_1 \equiv 1$, $g_2(s) = 2 - \frac{1}{1+s^2}$ is included to show the supremum over s need not be attained; it satisfies the hypotheses and gives $F \equiv 2$.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "$\max_t(xh_1(t) + h_2(t))$": an upper envelope of lines in x. On seeing it, write "RHS is convex in x, so LHS must be convex; supporting lines indexed by $t = x$".
2. The cue suggests convex duality: show F is finite and convex, then take $h_1(t)$ a subgradient of F at t and $h_2(t) = F(t) - th_1(t)$.
3. Hesitation occurs at existence of a supporting line for a convex function that is not assumed differentiable. It is reinforced away by the three-chord inequality: slopes from t increase, so a slope between the left and right families exists.
4. The trap is letting t index the same thing as s. Tangent lines to the individual exponentials give "$\geq$" but not attainment, and the supremum over s need not be attained at all.
5. Recognise it faster: whenever an expression is a max or sup of affine functions of a variable, say the word *convex* out loud, and check whether the other side of the identity is convex.
6. Reusable decision rule: *an identity "$A(x) = \max_t(\text{line in } x)$" is equivalent to "$A$ is finite and convex"; prove convexity (sup of convex is convex), prove finiteness (this is where the hypotheses go), and use supporting lines indexed by the point of contact.* The same rule handles Legendre transforms, support functions of convex sets, and any "min over t of a line" (concave case).

2012 B6

2012 B6. Let p be an odd prime number such that $p \equiv 2 \pmod 3$. Define a permutation π of the residue classes modulo p by $\pi(x) \equiv x^3 \pmod p$. Show that π is an even permutation if and only if $p \equiv 3 \pmod 4$.

Fields: number theory; group theory (cyclic groups, signs of permutations); Chinese remainder theorem
Position: B6 — the last problem, and hard in the sense that the natural computation is long; the structural proof is short once you see that only the 2-part of $p - 1$ can matter **Answer:** Via a primitive root, π is multiplication by 3 on $\mathbb{Z}/(p-1)$; by the Chinese remainder theorem its sign equals the sign of multiplication by 3 on $\mathbb{Z}/2^k$, where $2^k \parallel p - 1$, and that sign is $+1$ for $k = 1$ and -1 for $k \geq 2$.

Layer 1 THE WHAT — surface analysis

The natural attempt. A strong student first checks that π is a permutation: $\gcd(3, p-1) = 1$ because $p \equiv 2 \pmod 3$, so cubing is injective on the nonzero residues, and $0 \mapsto 0$. Then small cases. For $p = 5$: $0, 1, 2, 3, 4 \mapsto 0, 1, 3, 2, 4$, a single transposition, odd; and $5 \equiv 1 \pmod 4$. For $p = 11$: $2 \mapsto 8 \mapsto 6 \mapsto 7 \mapsto 2$ and $3 \mapsto 5 \mapsto 4 \mapsto 9 \mapsto 3$, two 4-cycles, with $0, 1, 10$ fixed; two odd cycles make an even permutation, and $11 \equiv 3 \pmod 4$. The pattern holds.

Now the general case. Cubing modulo p is irregular as a map on $\{0, \ldots, p-1\}$, so inversions cannot be counted directly. The student reaches for a primitive root g: writing $x = g^i$, cubing becomes $i \mapsto 3i \bmod (p-1)$. This is a good move. Now they count inversions of $\sigma(i) = 3i \bmod m$ on $\{0, \ldots, m-1\}$, $m = p - 1$. Write $\sigma(i) = 3i - m\lfloor 3i/m \rfloor$; the blocks $0 \leq i < m/3$, $m/3 < i < 2m/3$, $2m/3 < i < m$ each map increasingly, and inversions occur between blocks. Counting them by residue classes mod 3 is what the official first solution does, arriving at $(p-2)(p+1)/6$ inversions. It is correct, but it needs care with the floor functions and with m mod 3 (here $m \equiv 1 \pmod 3$), and the student does not see, until the final line, why p mod 4 is what decides the parity. Under exam conditions this is where errors enter.

The model chain. **Step 1: reduce to $\mathbb{Z}/m$.** The residue 0 is fixed and does not affect the sign. Let $m = p - 1$ and g a primitive root mod p. The map $L : \mathbb{Z}/m \to (\mathbb{Z}/p)^\times$, $i \mapsto g^i$, is a bijection with $\pi(L(i)) = L(3i)$. So π on the nonzero residues is conjugate to $\mu_3 : i \mapsto 3i$ on $\mathbb{Z}/m$, and conjugate permutations have the same sign (same cycle type). So $\operatorname{sgn}\pi = \operatorname{sgn}\mu_3$.

Step 2: split off the odd part. Since p is odd, m is even; write $m = 2^k m'$ with $k \geq 1$ and m' odd. By the Chinese remainder theorem, $i \mapsto (i \bmod 2^k,\ i \bmod m')$ is a bijection $\mathbb{Z}/m \to \mathbb{Z}/2^k \times \mathbb{Z}/m'$, and it carries μ_3 to $\alpha \times \beta$, where α, β are multiplication by 3 on $\mathbb{Z}/2^k$ and on $\mathbb{Z}/m'$ (3 is invertible mod both, as $\gcd(3, m) = 1$).

Lemma. If α permutes a set A and β a set B, then $\operatorname{sgn}(\alpha \times \beta) = \operatorname{sgn}(\alpha)^{|B|} \operatorname{sgn}(\beta)^{|A|}$. Indeed $\alpha \times \beta = (\alpha \times \mathrm{id})(\mathrm{id} \times \beta)$, and $\alpha \times \mathrm{id}$ acts as a copy of α on each of the $|B|$ disjoint rows $A \times \{b\}$, so its cycle decomposition is $|B|$ copies of that of α; likewise for $\mathrm{id} \times \beta$.

Here $|A| = 2^k$ is even and $|B| = m'$ is odd, so $\operatorname{sgn}\mu_3 = \operatorname{sgn}(\alpha)^{m'} \operatorname{sgn}(\beta)^{2^k} = \operatorname{sgn}\alpha$. Write s_k for the sign of $x \mapsto 3x$ on $\mathbb{Z}/2^k$.

Step 3: compute s_k.

- $k = 1$: on $\mathbb{Z}/2$, $3x = x$, the identity, $s_1 = +1$.
- $k = 2$: on $\mathbb{Z}/4$, $0, 1, 2, 3 \mapsto 0, 3, 2, 1$, one transposition, $s_2 = -1$.
- $k \geq 3$: split $\mathbb{Z}/2^k$ into even and odd residues; $x \mapsto 3x$ preserves each. On the evens, $2y \mapsto 2(3y)$, and $y \bmod 2^{k-1}$ determines $2y \bmod 2^k$, so this part is conjugate to $x \mapsto 3x$ on $\mathbb{Z}/2^{k-1}$, with sign s_{k-1}. On the 2^{k-1} odd residues, the cycle through x is $x, 3x, 9x, \ldots$, of length $L =$ the order of 3 modulo 2^k (since x is invertible, $3^j x \equiv x \iff 3^j \equiv 1$). All cycles have this same length, so there are $2^{k-1}/L$ of them. Claim: $3^{2^{k-2}} \equiv 1 \pmod{2^k}$. By induction $3^{2^j} = 1 + 2^{j+2}u_j$ with u_j odd for $j \geq 1$: $9 = 1 + 8$; and $(1 + 2^{j+2}u)^2 = 1 + 2^{j+3}(u + 2^{j+1}u^2)$. Taking $j = k - 2 \geq 1$ gives the claim, so L divides 2^{k-2}. Then $L \geq 2$ is a power of 2 (as $3 \not\equiv 1$) and the number of cycles $2^{k-1}/L \geq 2$ is a power of 2, hence even. An even number of even-length cycles has sign $+1$. So $s_k = s_{k-1}$.

Therefore $s_k = -1$ for all $k \geq 2$ and $s_1 = +1$.

Step 4: conclude. $k = 1$ means $4 \nmid p - 1$, i.e. $p \equiv 3 \pmod 4$. So π is even iff $p \equiv 3 \pmod 4$. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree through the primitive root: π is conjugate to $i \mapsto 3i$ on $\mathbb{Z}/(p-1)$. They split at the next line.

(N) "Count the inversions of $i \mapsto 3i \bmod m$ on $\{0, \ldots, m-1\}$." The map is treated as a function on an ordered list of integers.

(M) "$\mathbb{Z}/m$ is a product; the sign of a product permutation is controlled by the factor of even size." The map is treated as an endomorphism of a group, and the group is decomposed.

On (N) the answer emerges from arithmetic with no visible reason why p mod 4 matters. On (M) the answer p mod 4 is predicted before any computation: only the 2-part of $p - 1$ can affect the sign, so only $v_2(p-1)$ can matter.

> *The answer is supposed to depend on p mod 4, i.e. on how many factors of 2 are in $p - 1$. So I should separate the 2-part of $p - 1$ from the rest. The Chinese remainder theorem does that for $\mathbb{Z}/(p-1)$, and multiplication by 3 respects the splitting. A permutation that acts on a product with one side of odd size and the other of even size — the sign of the odd-size factor gets raised to an even power and disappears.*

> *On $\mathbb{Z}/2^k$ I peel off the odd residues: they form orbits of equal length under multiplication by 3, and if the number of orbits is even they contribute $+1$. So everything reduces to $\mathbb{Z}/2^{k-1}$, and I only have to look at $\mathbb{Z}/2$ and $\mathbb{Z}/4$ by hand.*

- ▷ **Cue:** a permutation of $\mathbb{Z}/p$ given by a power map $x \mapsto x^a$ $\longrightarrow$ **Move:** fix 0, take a primitive root, and replace it with multiplication by a on $\mathbb{Z}/(p-1)$
- ▷ **Cue:** the sign of a permutation that respects a product decomposition $A \times B$ $\longrightarrow$ **Move:** use $\operatorname{sgn}(\alpha \times \beta) = \operatorname{sgn}(\alpha)^{|B|} \operatorname{sgn}(\beta)^{|A|}$; a factor on an odd-size set is killed by an even-size partner
- ▷ **Cue:** the answer depends on p mod 4 (or on the 2-adic valuation of something) $\longrightarrow$ **Move:** split off the 2-part by the Chinese remainder theorem and work in $\mathbb{Z}/2^k$
- ▷ **Cue:** sign of a permutation made of many cycles of equal even length $\longrightarrow$ **Move:** the sign is $(-1)^{\text{number of cycles}}$; count cycles as (set size)/(cycle length)

Layer 3 THE WHY

Why must the 2-part be separated, and why does that come from the definition of sign? The sign of a permutation is $(-1)^{(\text{number of elements})-(\text{number of cycles})}$, equivalently $(-1)^{\text{number of even-length cycles}}$. Everything about sign is parity, and parity is about the prime 2. Now look at the permutation $i \mapsto 3i$ on $\mathbb{Z}/m$. Cycle lengths are orders of 3 acting on various subgroups; the odd prime factors of m contribute cycle structure that comes in copies, and whether those copies change the sign depends on how many copies there are — which is the size of the other factor. The Lemma in Step 2 is the precise form of that remark: a sign on one factor is raised to the power of the size of the other factor. When the other factor has even size (and 2^k does, since p is odd), the odd part's sign is erased. So it is *forced* that only $\mathbb{Z}/2^k$ matters. That is the "of course" here: sign is a parity, parity lives at the prime 2, and the Chinese remainder theorem is the tool that isolates a prime.

Then $\mathbb{Z}/2^k$. Why is $k = 1$ versus $k \geq 2$ the dividing line? On $\mathbb{Z}/2$ multiplication by any odd number is the identity. On $\mathbb{Z}/4$, $3 \equiv -1$ and $x \mapsto -x$ swaps 1 and 3: one transposition. Beyond that, each new layer of odd residues comes in an even number of equal cycles, because 3 has order at most 2^{k-2}, not 2^{k-1}, modulo 2^k — the group of units mod 2^k is not cyclic for $k \geq 3$, and so no single element (in particular 3) sweeps all of the odd residues in one cycle. That is the structural fact behind the recursion $s_k = s_{k-1}$.

The official first solution counts inversions and gets $(p-2)(p+1)/6$. Here is what that number is hiding: $(p-2)(p+1)/6$ is odd exactly when $p \equiv 1 \pmod 4$ (with $p \equiv 2 \pmod 3$), a fact you verify after the count, not predict before it. Their third solution (van Hoeij) counts cycles of even length by divisors d of $p-1$ and pairs d with $2d$ for d odd, then treats $4 \mid d$ — that pairing is the Chinese remainder theorem written divisor by divisor, and the route above is the same idea done once at the level of groups. Kedlaya's remark cites the Duke–Hopkins reciprocity law; the special case relevant to odd moduli is Zolotarev's lemma (the sign of $x \mapsto ax$ on $\mathbb{Z}/n$, n odd, is the Jacobi symbol $(\frac{a}{n})$), mentioned here only as context.

Caveat. Verified by computation: for every prime $p < 3000$ with $p \equiv 2 \pmod 3$, the sign of $x \mapsto x^3$ on $\mathbb{Z}/p$ is $+1$ exactly when $p \equiv 3 \pmod 4$, and it equals the sign of $x \mapsto 3x$ on $\mathbb{Z}/2^k$ with $2^k \| p-1$; the parity of the official count $(p-2)(p+1)/6$ agrees. The order of 3 modulo 2^k was confirmed to be 2^{k-2} for $3 \leq k \leq 14$. The Zolotarev reference is context only; the proof does not use it.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "x^3 modulo p with $p \equiv 2 \pmod 3$": a power map that is a bijection because $\gcd(3, p-1) = 1$. Write down "primitive root: $\pi \cong (i \mapsto 3i)$ on $\mathbb{Z}/(p-1)$".

2. The cue "the answer depends on p mod 4" suggests isolating the 2-part of $p-1$: Chinese remainder theorem plus the sign-of-a-product lemma, then a direct analysis of $\mathbb{Z}/2^k$.

3. Hesitation occurs at the product lemma: it is not obvious in the arena that the sign of $\alpha \times \beta$ has such a clean formula. It is reinforced away by factoring $\alpha \times \beta = (\alpha \times \mathrm{id})(\mathrm{id} \times \beta)$ and seeing $\alpha \times \mathrm{id}$ as $|B|$ disjoint copies of α.

4. The trap is inversion counting on $\{0, \ldots, p-2\}$: it is correct but long, error-prone with floors and residues mod 3, and gives no reason for the answer. A second trap is forgetting that 0 is a fixed point and must be dropped before using the primitive root.

5. Recognise it faster: whenever a sign or parity question is asked about a map on a cyclic group, ask *which prime can affect a parity?* Only 2. Then split off the 2-part.

6. Reusable decision rule: *to compute the sign of a structure-respecting permutation, decompose the underlying set as a product and use* $\operatorname{sgn}(\alpha \times \beta) = \operatorname{sgn}(\alpha)^{|B|} \operatorname{sgn}(\beta)^{|A|}$*; factors of odd size paired with factors of even size drop out.* The same rule computes signs of multiplication maps on $\mathbb{Z}/n$ (Zolotarev), of linear maps on $\mathbb{F}_q^n$, and of permutations of grids acting row by row.