DAWN PUTNAM SESSION

Putnam 1980 — Papers A and B

Putnam 120 Archive

Contents

Putnam 1980 — Paper A **5**
1980 A1 5
1980 A2 7
1980 A3 9
1980 A4 11
1980 A5 14
1980 A6 16

Putnam 1980 — Paper B **21**
1980 B1 21
1980 B2 23
1980 B3 26
1980 B4 28
1980 B5 31
1980 B6 34

Putnam 1980 — Paper A

1980 A1

1980 A1. Let b and c be fixed real numbers and let the ten points (j, y_j), $j = 1, 2, \ldots, 10$, lie on the parabola $y = x^2 + bx + c$. For $j = 1, 2, \ldots, 9$, let L_j be the point of intersection of the tangents to the given parabola at (j, y_j) and $(j+1, y_{j+1})$. Determine the polynomial function $y = g(x)$ of least degree whose graph passes through all nine points L_j.

Fields: coordinate geometry, quadratics **Position:** A1 (opener) **Answer:** $g(x) = x^2 + bx + c - \frac{1}{4}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You have nine points. Nine points, in general position, pin down a unique polynomial of degree at most eight, by Lagrange interpolation. The problem even asks for "the polynomial of least degree" through them, which sounds like an invitation to build that interpolant and then notice cancellation.

So you do the honest thing first: find L_j explicitly. The tangent to $y = x^2 + bx + c$ at $x = j$ has slope $2j + b$, and the line is

$$y = (2j + b)x - j^2 + c.$$

Intersecting this with the tangent at $x = j + 1$,

$$(2j + b)x - j^2 + c = (2j + 2 + b)x - (j + 1)^2 + c,$$

you solve for x and get $x = j + \frac{1}{2}$. Substituting back, $y = j^2 + (1 + b)j + \frac{b}{2} + c$. Nine explicit points $L_1, \ldots, L_9$ now sit in your hand, each a pair of messy expressions in j, b, c. The instinct at this point is to set up the interpolation determinant, or to guess the degree by finite differences on the y-values as j increases by 1 each time and hope the differences terminate early. Both routes work eventually. Both routes also require you to carry b and c as symbols through a system that was built for numbers, and the finite-differences route in particular tempts you to compute L_1 through L_9 as nine separate coordinate pairs before you're allowed to say anything about the shape of g.

The model chain. Do not compute nine points. Compute one, symbolically, as a function of the free index j, and read off its shape directly.

From the two tangent lines above, $x = j + \frac{1}{2}$ and $y = j^2 + (1 + b)j + \frac{b}{2} + c$. The pair (x, y) is already parametrised by j; you don't need $j = 1, \ldots, 9$ as separate cases, you need the map $j \mapsto (x, y)$ once. Invert it: $j = x - \frac{1}{2}$, and substitute into the y-formula.

$$y = \left(x - \tfrac{1}{2}\right)^2 + (1 + b)\left(x - \tfrac{1}{2}\right) + \tfrac{b}{2} + c.$$

Expand:

$$y = x^2 - x + \tfrac{1}{4} + (1 + b)x - \tfrac{1+b}{2} + \tfrac{b}{2} + c = x^2 + bx + \left(\tfrac{1}{4} - \tfrac{1}{2} - \tfrac{b}{2} + \tfrac{b}{2} + c\right) = x^2 + bx + c - \tfrac{1}{4}.$$

This is a single quadratic, independent of j. All nine points $L_1, \ldots, L_9$ lie exactly on the parabola $g(x) = x^2 + bx + c - \frac{1}{4}$ — the original parabola, translated down by $\frac{1}{4}$. That is the polynomial of least degree, and its degree is 2, not 8.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree up to the point where $x = j + \frac{1}{2}$, $y = j^2 + (1+b)j + \frac{b}{2} + c$ has been derived — that much is forced algebra, not a choice. The chains split at what happens next: the natural attempt treats these as nine numerical data points and reaches for interpolation across $j = 1, \ldots, 9$. The model treats j itself as the interpolating variable already in hand, inverts $x = j + \frac{1}{2}$ for j, and substitutes straight back into the y-formula.

> *I have x and y both written as functions of the same parameter j — I don't need nine instances of that relationship, I need to eliminate j once and see what curve is left.*

Layer 3 THE WHY

The problem's phrasing does the work for you, if you let it: it does not ask you to find $L_1, \ldots, L_9$, it asks for tangents at (j, y_j) and $(j+1, y_{j+1})$ "for $j = 1, \ldots, 9$" — the general term is stated before the range is. A relation given in terms of a running index j is a parametric curve with j as parameter, and the standard move for a parametric curve is to eliminate the parameter and get a single Cartesian equation, not to tabulate nine values of it. The finite-differences instinct is not wrong, exactly — it will terminate at degree 2 if you carry it through symbolically in j — but it disguises a fact that the elimination makes visible immediately: L_j isn't nine unrelated points that happen to lie on a common curve, it's literally one point, $(j + \frac{1}{2}, g(j + \frac{1}{2}))$, sampled nine times. The "least degree" phrasing in the statement is the tell: if the intended answer required genuinely fitting nine independent points, the least-degree polynomial would generically be forced to degree 8, and the problem would be a much longer computation with no clean closed form. A Putnam problem that promises a clean formula in b, c is promising you that the nine points are not independent, and the way three points cease to be independent is that they satisfy a lower-degree relation you can find directly, not one you can only discover after generating all nine.

The more permanent fact underneath this: for the standard parabola $y = x^2$, the tangents at $x = p$ and $x = q$ meet at $\left(\frac{p+q}{2}, pq\right)$ — a fact worth knowing on sight, since it is the reason two tangents to any parabola meet on a curve related to the original by a simple affine shift, never by a curve of higher degree. Once you know that fact for $y = x^2$, the general $y = x^2 + bx + c$ case is a horizontal/vertical translate of it, and no new computation is needed to believe the answer will again be degree 2.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the phrasing itself: the tangent points and the intersection are defined "for $j = 1, \ldots, 9$" as a single general construction, not as nine separate constructions. On seeing it, write down x and y as functions of j before computing a single numerical L_j.
2. That cue suggests parameter elimination: express $x(j)$ and $y(j)$, solve $x(j)$ for j, substitute into $y(j)$, and read off the resulting equation in x and y alone.
3. Hesitation sets in right after deriving $x = j + \frac{1}{2}$: it's tempting to plug in $j = 1, \ldots, 9$ immediately because you now have nine concrete-looking values of x. Reinforce against this by asking: "do I actually need nine numbers, or do I need the shape of one relationship?" The moment you have x and y both as explicit functions of the same j, the interpolation step is already unnecessary.
4. The trap is the phrase "nine points" combined with "polynomial of least degree" — it is worded to make Lagrange interpolation across nine points sound like the intended tool, when the actual content of the problem is a parametric-curve elimination that produces degree 2, far below the degree 8 a generic nine-point fit would need.

5. Recognise this faster next time by treating any family of objects indexed by a running variable j (points, lines, coefficients) as a parametrisation first, and asking whether j can be eliminated, before treating the family as a list of independent data.

6. Decision rule: when a problem hands you a construction "for $j = 1, \ldots, n$" and asks for a curve or formula through the resulting objects, write the construction as explicit functions of j and eliminate j between them before ever substituting a numerical value of j. Tabulation is a last resort for when elimination genuinely fails, not a first response to seeing several instances of one relationship.

1980 A2

1980 A2. Let r and s be positive integers. Derive a formula for the number of ordered quadruples (a, b, c, d) of positive integers such that

$$3^r \cdot 7^s = \operatorname{lcm}[a, b, c] = \operatorname{lcm}[a, b, d] = \operatorname{lcm}[a, c, d] = \operatorname{lcm}[b, c, d].$$

The answer should be a function of r and s. (Note that $\operatorname{lcm}[x, y, z]$ denotes the least common multiple of x, y, z.)

Fields: number theory, combinatorics **Position:** A2 **Answer:** $(1 + 4r + 6r^2)(1 + 4s + 6s^2)$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The number $3^r 7^s$ has only the primes 3 and 7 in it, so a, b, c, d must each be of the form $3^m 7^n$ with $0 \le m \le r$, $0 \le n \le s$ — that much is immediate, since any other prime factor in a would survive into some lcm and violate the equation. The temptation now is to treat the quadruple (a, b, c, d) as a single object and reason about it directly: write out the four lcm conditions, note there are four constraints on four unknowns, and start casing on which of a, b, c, d is "the biggest" in the sense of carrying $3^r 7^s$ itself, then subcase on the second-biggest, and so on. This is a correct route — you can grind through it — but it multiplies two separate structures (the 3-exponents and the 7-exponents) together inside every case, so each case you open forks again into sub-cases on m's and n's simultaneously, and by the time you've enumerated "exactly two of the four hold the max 3-exponent, exactly three of the four hold the max 7-exponent, and they're not the same two-and-three" you are tracking a genuinely ugly inclusion-exclusion across two intertwined variables at once.

The model chain. Separate the two primes before doing any counting at all. Write $a = 3^{m_a} 7^{n_a}$ and similarly for b, c, d. Because lcm of prime powers takes the maximum exponent of each prime independently, the condition

$$\operatorname{lcm}[a, b, c] = \operatorname{lcm}[a, b, d] = \operatorname{lcm}[a, c, d] = \operatorname{lcm}[b, c, d] = 3^r 7^s$$

splits cleanly into two conditions that never interact: the maximum of m_a, m_b, m_c (and likewise every other triple of the m's) must equal r, and separately the maximum of any triple of the n's must equal s. Solve each one-variable problem on its own.

For the m's: each $m_i \in \{0, \ldots, r\}$, and every triple drawn from $\{m_a, m_b, m_c, m_d\}$ must attain the maximum r. If only one of the four equalled r, removing that one index would leave a triple with maximum below r — forbidden. So at least two of the four must equal r. Count directly by complementary counting: of the $(r+1)^4$ unconstrained quadruples, subtract those with at most one coordinate equal to r. None equal to r: r^4 ways. Exactly one equal to r: choose which coordinate, 4 ways, the rest in $\{0, \ldots, r-1\}$, r^3 ways each, giving $4r^3$. So the count is

$$(r+1)^4 - r^4 - 4r^3 = 6r^2 + 4r + 1 = 1 + 4r + 6r^2,$$

using the binomial expansion of $(r+1)^4$ to cancel the r^4 and $4r^3$ terms exactly.

By identical reasoning the n's contribute $1 + 4s + 6s^2$. Since the choice of each m_i is entirely free of the choice of each n_i — a quadruple (a, b, c, d) is determined by choosing all four m_i's and all four n_i's independently — the total count is the product:

$$(1 + 4r + 6r^2)(1 + 4s + 6s^2).$$

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree on the reduction to prime-power form: $a, b, c, d \in \{3^m 7^n : 0 \le m \le r,\ 0 \le n \le s\}$. They split immediately after that, at the decision of what to case on. The natural attempt cases on the joint structure of (a, b, c, d) as single numbers. The model cases on the 3-exponents and 7-exponents as two independent counting problems and multiplies the results.

> *The lcm of prime powers acts on each prime separately — a condition stated as "*$\mathrm{lcm}[\ldots] = 3^r 7^s$*" is really two conditions in disguise, one about the exponents of* 3 *and one about the exponents of* 7*, and they don't talk to each other, so I should solve the one-prime version of this problem once and multiply.*

Layer 3 THE WHY

The factorisation is licensed by an actual theorem, not a convenient guess: for coprime $x = 3^{m_1}7^{n_1}$ and any prime powers, lcm acts coordinatewise on the exponent vector of each prime in the factorisation — $\mathrm{lcm}(3^{m_1}7^{n_1}, 3^{m_2}7^{n_2}) = 3^{\max(m_1,m_2)}7^{\max(n_1,n_2)}$. This is a direct consequence of unique factorisation: the multiplicity of a prime p in a lcm is the max of its multiplicities in the inputs, and multiplicities of different primes are recorded in different coordinates of the exponent vector, so they cannot influence each other. Once you see the problem's target $3^r 7^s$ as a lattice point (r, s) in an exponent lattice rather than as a single number, the four quadruple-lcm conditions become four conditions on $(m_a, m_b, m_c, m_d) \in \mathbb{Z}^4$ together with four identical conditions on $(n_a, n_b, n_c, n_d) \in \mathbb{Z}^4$, and — since the map $(a, b, c, d) \mapsto \big((m_a, n_a), \ldots, (m_d, n_d)\big)$ is a bijection with no cross-constraint — the counting problem factors as a product over the two independent coordinate problems. This is the general fact underneath "count something built from several coprime prime-power conditions": whenever the defining conditions are stated using lcm, gcd, or divisibility among numbers whose only prime factors are drawn from a fixed coprime set, the count factors over that set's primes, because those operations were defined, exponent by exponent, to have no cross-prime interaction in the first place.

The "at least two of the four attain the maximum" translation deserves its own justification, since it is the one place a wrong inequality direction would sink the whole proof. The condition is that every one of the four triples (each omitting exactly one of a, b, c, d) has maximum exponent r. If exactly one index, say d, held $m_d = r$ and all others were below r, the triple (a, b, c) — which omits d — would have every entry below r, contradicting the requirement that $\mathrm{lcm}[a, b, c] = 3^r \cdots$. So a single maximiser is not enough; you need the maximiser to survive being deleted, which forces at least two coordinates to equal r. This is worth stating as a reusable fact: "every $(n-1)$-subset attains the max" is equivalent to "at least two elements attain the max," for any $n \ge 2$.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that the target $3^r 7^s$ is built from two coprime primes and the defining relations are all lcm's. On seeing it, write a, b, c, d in exponent-vector form immediately rather than as single integers.
2. That cue suggests splitting the problem into two independent one-prime counting problems and multiplying the results — the standard move whenever lcm/gcd conditions are stated over a number built from a fixed coprime prime set.

3. Hesitation occurs when translating "lcm of every 3-subset equals r" into a clean condition on four exponents; it's tempting to case on which specific coordinates equal r number by number. Reinforce against this by first asking the equivalent, cleaner question: how many maximisers, at minimum, are forced? Answer that once, abstractly, before touching any casework.

4. The trap is doing the 3-exponent and 7-exponent bookkeeping simultaneously inside a single case split on (a, b, c, d), which multiplies the complexity of two independent one-dimensional problems into one intractable two-dimensional one.

5. Recognise this faster next time by treating any number defined as a product of coprime prime powers $(p_1^{e_1} p_2^{e_2} \cdots)$ as a vector of independent single-prime problems the instant lcm, gcd or divisibility conditions are involved.

6. Decision rule: if a counting problem is phrased using lcm or gcd equal to $\prod p_i^{e_i}$ for distinct primes p_i, solve the single-prime version of the problem for one prime, then multiply the answers across primes — never carry more than one prime's exponent structure through the same casework.

1980 A3

1980 A3. Evaluate

$$\int_0^{\pi/2} \frac{dx}{1 + (\tan x)^{\sqrt{2}}}.$$

Caveat. The OCR of the source book renders the exponent on $\tan x$ as an unreadable glyph ("$(\tan x)^{\sim}$"). The commonly attested text of this problem gives the exponent as $\sqrt{2}$, and I have used that reading here. It does not affect the mathematics below: the argument that follows evaluates $\int_0^{\pi/2} dx/(1 + (\tan x)^n)$ to $\pi/4$ for any fixed real n for which the integral converges (in particular for any $n > 0$, since $(\tan x)^n \to 0$ as $x \to 0^+$ keeps the integrand bounded, and the substitution below handles the endpoint at $\pi/2$ symmetrically), so the value $\pi/4$ is correct regardless of which specific exponent the OCR obscured.

Fields: integral calculus, symmetry **Position:** A3 **Answer:** $\pi/4$, independent of the exponent.

Layer 1 THE WHAT — surface analysis

The natural attempt. The exponent $\sqrt{2}$ on $\tan x$ is the first thing that stops you. There is no elementary antiderivative of $1/(1 + t^{\sqrt{2}})$ after the substitution $t = \tan x$, $dt = \sec^2 x\, dx$ — you'd be looking at $\int \frac{dt}{(1+t^{\sqrt{2}})(1+t^2)}$, and an irrational exponent inside a rational-looking integrand is not something partial fractions can touch. Numerically the answer is clearly some clean constant (Putnam integrals over $[0, \pi/2]$ involving $\tan x$ almost always are), so the honest first move for many solvers is to try small sanity cases: replace $\sqrt{2}$ with 1 mentally and see if $\int_0^{\pi/2} dx/(1 + \tan x)$ has a nice value, hoping to reverse-engineer the general pattern from an easy case, or to try differentiating under the integral sign with respect to the exponent as a parameter. Both are reasonable instincts, but the second invites a genuinely difficult differentiation of $(\tan x)^n$ with respect to n inside an integral, and the first only tells you the exponent-1 case is special without telling you why every exponent behaves the same way.

The model chain. Do not touch the exponent at all. Use the interval's own symmetry: substitute $x = \pi/2 - u$ in the integral $I = \int_0^{\pi/2} dx/(1 + (\tan x)^{\sqrt{2}})$. Since $\tan(\pi/2 - u) = \cot u$ and $dx = -du$, with the limits reversing to cancel the sign,

$$I = \int_0^{\pi/2} \frac{du}{1 + (\cot u)^{\sqrt{2}}}.$$

Write $\cot u = 1/\tan u$, so

$$\frac{1}{1+(\cot u)^{\sqrt{2}}} = \frac{1}{1+(\tan u)^{-\sqrt{2}}} = \frac{(\tan u)^{\sqrt{2}}}{(\tan u)^{\sqrt{2}}+1}.$$

So $I = \int_0^{\pi/2} \frac{(\tan u)^{\sqrt{2}}}{1+(\tan u)^{\sqrt{2}}}\,du$ — the same integral as the original, but with the integrand's numerator and denominator roles swapped. Add this expression for I to the original expression for I:

$$2I = \int_0^{\pi/2} \left(\frac{1}{1+(\tan x)^{\sqrt{2}}} + \frac{(\tan x)^{\sqrt{2}}}{1+(\tan x)^{\sqrt{2}}} \right) dx = \int_0^{\pi/2} 1\,dx = \frac{\pi}{2}.$$

Hence $I = \pi/4$. Nowhere did the exponent $\sqrt{2}$ get used for anything beyond being the same exponent on both occurrences — it could have been any fixed real power and the computation is identical.

Layer 2 THE HOW — the bifurcation

Bifurcating point. The chains do not really share a common stretch of algebra — they split at the very first decision, which is what to do with the integral before any manipulation: the natural attempt treats the exponent $\sqrt{2}$ as an obstacle to be worked around or approximated, while the model treats the exponent as irrelevant and instead exploits the interval $[0, \pi/2]$ being symmetric under $x \mapsto \pi/2 - x$, which swaps $\tan x$ with its reciprocal $\cot x$.

> *The exponent looks unfriendly, but I never actually need to integrate $t^{\sqrt{2}}$ against anything — I only need the substitution $x \mapsto \pi/2 - x$ to turn $\tan x$ into $1/\tan x$, and $1/(1+t^{-n})$ is always $t^n/(1+t^n)$ for any n, so adding the two expressions for the same integral I kills the exponent entirely and leaves 1.*

Layer 3 THE WHY

The structural fact underneath is that $[0, \pi/2]$ is exactly the domain on which $\tan x$ and $\cot x$ are the same function up to the reflection $x \mapsto \pi/2 - x$ — this is definitional (co-tangent is literally the tangent of the complementary angle), and it is why the substitution is not a trick but the natural response to seeing $\tan x$ on this particular interval. Once you recognise that reflecting the interval sends $\tan x \to 1/\tan x$, the second fact that makes the whole thing collapse is purely algebraic and has nothing to do with trigonometry: for any $t > 0$ and any real n,

$$\frac{1}{1+t^{-n}} + \frac{1}{1+t^n} = \frac{t^n}{t^n+1} + \frac{1}{1+t^n} = \frac{t^n+1}{t^n+1} = 1.$$

This identity is the actual engine of the problem — it says that $f(t) = 1/(1+t^n)$ and $f(1/t)$ always sum to exactly 1, for every exponent n — and the Putnam problem is simply this identity, dressed as a definite integral so that the substitution $x \mapsto \pi/2 - x$ is what produces the $t \mapsto 1/t$ swap for free. Seeing that the exponent never enters the identity is what tells you, before doing any calculus, that the answer cannot depend on $\sqrt{2}$ specifically: replacing $\sqrt{2}$ by 1, by 3, or by π would give exactly the same value $\pi/4$, because the argument only ever uses that the same exponent appears on both sides of the swap. An irrational or otherwise unfriendly-looking exponent in a Putnam integral over a symmetric interval is frequently a deliberate signal — it rules out direct antidifferentiation as a viable path and rewards the solver who asks what property of the integral is exponent-independent rather than what the antiderivative of $t^{\sqrt{2}}$ looks like.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the pairing of the interval $[0, \pi/2]$ with $\tan x$ inside the integrand. On seeing it, write down the substitution $x \mapsto \pi/2 - x$ immediately and note that it sends $\tan x \mapsto \cot x = 1/\tan x$, before attempting any direct antidifferentiation.
2. That cue suggests the "add I to itself under the swap" method: compute the integral under the substitution, observe the integrand becomes $t^n/(1+t^n)$ for $t = \tan x$, and add it to the original integrand to get a constant.
3. Hesitation occurs on seeing the exponent $\sqrt{2}$, which reads as a signal that heavy machinery (series expansion, parameter differentiation) is required. Reinforce against this by checking, before any such machinery, whether the exponent is used anywhere except as "the same power on both sides" — if the argument would go through with the exponent replaced by an arbitrary symbol n, the exponent's specific value is a red herring.
4. The trap is the exponent $\sqrt{2}$ itself: it is irrational and unfamiliar precisely so that a solver who tries to integrate $t^{\sqrt{2}}$ directly, or to reason about it via a series in integer powers, wastes the exam on a computation the problem never asked for.
5. Recognise this faster next time by treating any integral of $f(\tan x)$ over $[0, \pi/2]$ as a candidate for the $x \mapsto \pi/2 - x$ swap the instant you see $\tan x$ and a symmetric-looking interval together, regardless of how complicated f looks.
6. Decision rule: for $\int_0^{\pi/2} g(\tan x)\,dx$, always test the substitution $x \mapsto \pi/2 - x$ (equivalently $t \mapsto 1/t$ in $\tan x = t$) before attempting to antidifferentiate g directly; if $g(t) + g(1/t)$ simplifies to a constant, the integral is that constant times $\pi/4$, independent of any parameter buried inside g.

1980 A4

1980 A4. (a) Prove that there exist integers a, b, c, not all zero and each of absolute value less than one million, such that

$$|a + b\sqrt{2} + c\sqrt{3}| < 10^{-11}.$$

(b) Let a, b, c be integers, not all zero and each of absolute value less than one million. Prove that

$$|a + b\sqrt{2} + c\sqrt{3}| > 10^{-21}.$$

Fields: number theory (Diophantine approximation), algebra (conjugates and norms), combinatorics (pigeonhole) **Position:** A4 — mid-paper; two short proofs, each of which is one idea, and neither idea is reachable by computation **Answer:** (a) By pigeonhole on the 10^{18} numbers $r + s\sqrt{2} + t\sqrt{3}$, $0 \le r, s, t < 10^6$, two lie within 4.2×10^{-12} of each other; their difference works. (b) The product of the four conjugates $a \pm b\sqrt{2} \pm c\sqrt{3}$ is a non-zero integer, and the other three conjugates are each less than 10^7 in size, so $|a + b\sqrt{2} + c\sqrt{3}| > 10^{-21}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. Part (a) says "there exist", so you try to find them. The first plan: take good rational approximations $\sqrt{2} \approx p/q$, $\sqrt{3} \approx p'/q$ with a common denominator q, and set $b = q, c = \pm q, a = -(p \pm p')$ or similar. You compute $\sqrt{2} = 1.41421356\ldots$, $\sqrt{3} = 1.73205080\ldots$ and start looking for $q < 10^6$ making both $q\sqrt{2}$ and $q\sqrt{3}$ close to integers. Each is close to an integer to within about $1/q$ at best for a well-chosen q, but *simultaneously* for the same q you get only about $q^{-1/2}$, and the errors add. With $q \approx 10^6$ that is an error of order 10^{-3}, not 10^{-11}. You try mixing: choose b and c independently so that $b\sqrt{2} + c\sqrt{3}$ is close to an integer. Now you have two free parameters and one target, which feels better, but you have no algorithm

for it and no calculator in the room. This is the wall: an existence statement whose witness is a specific triple of six-digit integers that nobody could compute by hand.

Part (b) looks like the reverse problem, and the natural attempt is to bound it the same way: $a+b\sqrt{2}+c\sqrt{3}$ is non-zero because $1, \sqrt{2}, \sqrt{3}$ are linearly independent over $\mathbb{Q}$, and then…? Non-zero is not the same as bounded away from zero. You try $|a+b\sqrt{2}+c\sqrt{3}| \geq \big||a|-|b\sqrt{2}+c\sqrt{3}|\big|$ and get nothing, because the right side can be arbitrarily small — that is exactly what (a) says.

The model chain. *(a)* Let $N = 10^6$ and let S be the set of the $N^3 = 10^{18}$ numbers

$$x_{r,s,t} = r + s\sqrt{2} + t\sqrt{3}, \qquad r,s,t \in \{0,1,\ldots,N-1\}.$$

Every such number lies in $[0,d]$ with $d = (1+\sqrt{2}+\sqrt{3})(N-1) < 4.15\times 10^6$. Cut $[0,d]$ into $10^{18}-1$ closed intervals of equal length $\varepsilon = d/(10^{18}-1) < 4.2\times 10^{-12}$. There are more triples than intervals, so two different triples $(r,s,t) \neq (r',s',t')$ give numbers in the same interval. Put $a = r-r'$, $b = s-s'$, $c = t-t'$. These are integers, not all zero because the triples differ, each of absolute value at most $N-1 < 10^6$, and

$$|a+b\sqrt{2}+c\sqrt{3}| = |x_{r,s,t} - x_{r',s',t'}| \leq \varepsilon < 10^{-11}.$$

(b) Define the four conjugates

$$F_1 = a+b\sqrt{2}+c\sqrt{3}, \quad F_2 = a-b\sqrt{2}+c\sqrt{3}, \quad F_3 = a+b\sqrt{2}-c\sqrt{3}, \quad F_4 = a-b\sqrt{2}-c\sqrt{3}.$$

The product is an integer. $F_1F_3 = (a+b\sqrt{2})^2 - 3c^2 = (a^2+2b^2-3c^2) + 2ab\sqrt{2}$ and $F_2F_4 = (a^2+2b^2-3c^2) - 2ab\sqrt{2}$, so

$$P = F_1F_2F_3F_4 = (a^2+2b^2-3c^2)^2 - 8a^2b^2 \in \mathbb{Z}.$$

No factor vanishes. Each F_i has the form $a+\beta\sqrt{2}+\gamma\sqrt{3}$ with $\beta = \pm b$, $\gamma = \pm c$, not all of a,β,γ zero. Suppose $a+\beta\sqrt{2}+\gamma\sqrt{3} = 0$. If $\gamma = 0$ then $a+\beta\sqrt{2} = 0$ forces $\beta = 0$ (else $\sqrt{2} = -a/\beta$ is rational) and then $a = 0$, a contradiction. If $\gamma \neq 0$, then $\gamma\sqrt{3} = -(a+\beta\sqrt{2})$; squaring, $3\gamma^2 = a^2+2\beta^2+2a\beta\sqrt{2}$, so $a\beta = 0$ by irrationality of $\sqrt{2}$. If $a = 0$: $3\gamma^2 = 2\beta^2$, impossible for $\gamma \neq 0$ since $\sqrt{3/2}$ is irrational (the power of 3 dividing the left side is odd, on the right even). If $\beta = 0$: $3\gamma^2 = a^2$, impossible since $\sqrt{3}$ is irrational. So P is a non-zero integer and $|P| \geq 1$.

The other three are not large. $|F_i| \leq |a| + \sqrt{2}|b| + \sqrt{3}|c| < (1+\sqrt{2}+\sqrt{3})10^6 < 4.15\times 10^6 < 10^7$. Hence

$$|F_1| = \frac{|P|}{|F_2F_3F_4|} \geq \frac{1}{|F_2F_3F_4|} > \frac{1}{(4.15\times 10^6)^3} > 10^{-20} > 10^{-21}.$$

Layer 2 THE HOW — the bifurcation

Bifurcating point. In (a) both chains read "there exist". The natural attempt treats that as "construct"; the model treats it as "count". In (b) both chains know $F_1 \neq 0$; the natural attempt tries to bound F_1 from below directly, the model multiplies F_1 by its three conjugates to reach an integer.

> *I am asked for existence, not for the numbers. I have 10^{18} combinations $r+s\sqrt{2}+t\sqrt{3}$ and they all fit in an interval of length about 4×10^6. Somewhere two of them are within $4\times 10^{-18+6} = 4\times 10^{-12}$. Subtract.*

> *A non-zero integer is at least 1. That is the only lower bound I own. So I need to turn $a+b\sqrt{2}+c\sqrt{3}$ into an integer by multiplying by things I can bound from above — its conjugates.*

Layer 3 THE WHY

Why counting and not constructing in (a)? Look at the numbers. The triples number 10^{18}; the values they produce spread over a range of length $\approx 4 \times 10^{6}$. The average gap between neighbours is therefore $\approx 4 \times 10^{-12}$. That calculation is not a trick; it is the first thing to do with any "make a linear combination small" problem, because it tells you what is true. The requested bound 10^{-11} sits a little above 4×10^{-12} — the examiners have told you the pigeonhole bound by choosing the constant. When the target in an existence problem is barely weaker than "(number of objects) versus (size of the space)", the problem is pigeonhole. Dirichlet's approximation theorem is exactly this argument, and it is the standard source of good simultaneous approximations; the continued-fraction route in the natural attempt handles one irrational at a time and cannot see the second.

Why conjugates in (b)? Start from what makes a number unable to be small. Real numbers can be anything; *integers* have a gap: a non-zero integer has absolute value at least 1. So every lower bound on an algebraic expression comes from an integer somewhere. The question is which integer. F_1 lives in $\mathbb{Q}(\sqrt{2}, \sqrt{3})$; the operations $\sqrt{2} \mapsto -\sqrt{2}$ and $\sqrt{3} \mapsto -\sqrt{3}$ are the symmetries of that field, and a quantity invariant under all of them is rational. The product of all four images of F_1 is invariant, has integer coefficients, hence is an integer. You met the one-root version at A-level when rationalising a denominator: $(a + b\sqrt{2})(a - b\sqrt{2}) = a^2 - 2b^2$. With two square roots you need to flip each independently, giving four factors. Nothing here is special to $\sqrt{2}, \sqrt{3}$; this is the norm, and it is why the answer's exponent is what it is. The norm is ≥ 1; three factors are each $\lesssim N$; so $|F_1| \gtrsim N^{-3} = 10^{-18}$. The examiners asked for 10^{-21}, leaving room for the constant $4.15^3 \approx 71$.

Put (a) and (b) side by side and the problem's structure appears: the smallest non-zero value of $|a + b\sqrt{2} + c\sqrt{3}|$ with coefficients below N is between about N^{-3} and N^{-2}. Pigeonhole gives the upper bound by counting; the norm gives the lower bound by integrality. These are the two fundamental tools of Diophantine approximation, and this problem is a controlled first meeting with both.

The official solution cites "the demonstration of this was Problem A-1 of the 15th Competition" for $F_i \neq 0$. Here is why that step cannot be skipped: if any conjugate were zero, $P = 0$ and the integrality gap gives nothing. Linear independence of $1, \sqrt{2}, \sqrt{3}$ over $\mathbb{Q}$ is the whole reason the product is non-zero.

Layer $+\alpha$ BACKTESTING

1. The cue in (a) is "there exist integers … such that |linear combination of irrationals| $< \varepsilon$" with a huge coefficient bound. Write down: number of coefficient choices (10^{18}), range of values ($\approx 4 \times 10^{6}$), average spacing. In (b) the cue is "not all zero … prove $|\cdot| > \delta$"; write down the conjugates.

2. (a) Pigeonhole on values of the linear form, then subtract two nearby ones. (b) Multiply by all Galois conjugates to reach a non-zero integer; bound the other factors above.

3. Hesitation in (a) comes from believing an existence proof must produce the object. Reinforce: "show there exist" with no request for the value is permission to count. Hesitation in (b) comes from not seeing where an integer could come from; reinforce with the rationalising-the-denominator memory.

4. The trap in (a) is attempting explicit simultaneous approximation of $\sqrt{2}$ and $\sqrt{3}$. The trap in (b) is forgetting to prove that no conjugate vanishes: the integer might be 0, and then the argument collapses.

5. Any time an existence claim's constant looks like (range)/(number of objects), count before you construct. Any time you need a lower bound on something algebraic and non-zero, look for an integer it divides or is a factor of.

6. Reusable rule: small values are produced by pigeonhole and forbidden by integrality. Upper bounds on min |expression| come from counting many candidates in a small range; lower bounds come from multiplying by conjugates until the result is a non-zero integer.

- ▷ **Cue:** "there exist integers, bounded by N, making a linear form in irrationals smaller than ε" $\longrightarrow$ **Move:** count N^k values in a range of length $\approx CN$; pigeonhole two within CN^{1-k}; subtract
- ▷ **Cue:** an existence constant equal to (range)/(number of objects) up to a small factor $\longrightarrow$ **Move:** the problem is pigeonhole; stop trying to construct
- ▷ **Cue:** lower bound needed for a non-zero expression $a + b\sqrt{2} + c\sqrt{3}$ $\longrightarrow$ **Move:** multiply by all sign-flipped conjugates; the product is a non-zero integer, so ≥ 1
- ▷ **Cue:** "not all zero" in the hypothesis of a lower-bound problem $\longrightarrow$ **Move:** it is there so that the norm is non-zero; prove each conjugate is non-zero using irrationality
- ▷ **Cue:** two parts, one asking for $< \varepsilon$ and one for $> \delta$ $\longrightarrow$ **Move:** expect pigeonhole for the first and integrality for the second

1980 A5

1980 A5. Let $P(t)$ be a nonconstant polynomial with real coefficients. Prove that the system of simultaneous equations

$$0 = \int_0^x P(t) \sin t \, dt = \int_0^x P(t) \cos t \, dt$$

has only finitely many real solutions x.

Fields: analysis (integration by parts, polynomials), complex numbers **Position:** A5 — late-paper; the computation is A-level, the difficulty is seeing that the two equations must be combined into one amplitude equation **Answer:** Writing $Q = P - P'' + P^{(4)} - \cdots$, any common solution satisfies $Q(x)^2 + Q'(x)^2 = Q(0)^2 + Q'(0)^2$, a non-constant polynomial equation of degree $2 \deg P$; so there are at most $2 \deg P$ real solutions.

Layer 1 THE WHAT — surface analysis

The natural attempt. Two integrals of a polynomial times a trigonometric function: you integrate by parts. Take $P(t) = t$ to see what happens:

$$F(x) = \int_0^x t \sin t \, dt = \sin x - x \cos x, \qquad G(x) = \int_0^x t \cos t \, dt = x \sin x + \cos x - 1.$$

Now you study $F = 0$, i.e. $\tan x = x$. That has infinitely many solutions, one in each interval $(k\pi - \frac{\pi}{2}, k\pi + \frac{\pi}{2})$. And $G = 0$ has infinitely many solutions too. So each equation on its own is useless, and you are left asking whether the two infinite families of roots can meet infinitely often. You try to locate the roots of $\tan x = x$ asymptotically ($x \approx k\pi + \frac{\pi}{2} - \frac{1}{k\pi}$) and substitute into G. For $P(t) = t$ this can be pushed through; for a general P of degree n the integration by parts produces $n + 1$ terms in each of F and G, and root-by-root asymptotics are hopeless.

A second natural attempt uses Rolle: $F'(x) = P(x) \sin x$, which has infinitely many zeros, so F can have infinitely many zeros. Rolle alone cannot bound anything on the whole line. This is the wall: every tool you apply to one equation at a time sees infinitely many roots.

The model chain. Let $n = \deg P \geq 1$ and define

$$Q = P - P'' + P^{(4)} - P^{(6)} + \cdots,$$

a finite sum since high derivatives of P vanish. Then $Q + Q'' = P$ (the sum telescopes), and Q has the same degree n and leading coefficient as P. Differentiating,

$$\frac{d}{dx}\left(-Q \cos x + Q' \sin x\right) = (Q + Q'') \sin x = P \sin x, \qquad \frac{d}{dx}\left(Q \sin x + Q' \cos x\right) = (Q + Q'') \cos x = P \cos x.$$

So, with $F(x) = \int_0^x P(t)\sin t\,dt$ and $G(x) = \int_0^x P(t)\cos t\,dt$,

$$F(x) = -Q(x)\cos x + Q'(x)\sin x + Q(0), \qquad G(x) = Q(x)\sin x + Q'(x)\cos x - Q'(0).$$

Suppose $F(x) = G(x) = 0$. Then

$$Q(x)\cos x - Q'(x)\sin x = Q(0), \qquad Q(x)\sin x + Q'(x)\cos x = Q'(0).$$

Square and add; the cross terms $\mp 2QQ'\sin x\cos x$ cancel and $\sin^2 + \cos^2 = 1$, so

$$Q(x)^2 + Q'(x)^2 = Q(0)^2 + Q'(0)^2. \qquad (*)$$

The left side is a real polynomial in x of degree $2n$ with leading coefficient $q_n^2 > 0$, where $q_n \neq 0$ is the leading coefficient of P (the term Q'^2 has degree only $2n-2$). The right side is a constant. So $(*)$ is a non-constant polynomial equation of degree $2n$, which has at most $2n$ real roots. Every common solution of the system is among them, so the system has at most $2\deg P$ real solutions.

(Check with $P(t) = t$: $Q = t$, and $(*)$ reads $x^2 + 1 = 1$, so $x = 0$ is the only solution. Check the hypothesis: if $P \equiv 1$ then $Q \equiv 1$, $(*)$ reads $1 = 1$, and indeed $F = 1 - \cos x$, $G = \sin x$ vanish together at every $x = 2k\pi$. Non-constancy is exactly what makes $(*)$ informative.)

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains integrate by parts. The natural attempt then treats $F = 0$ and $G = 0$ separately; the model writes both, notices that they are the two components of a rotation of the vector (Q, Q'), and squares and adds.

> *Each equation alone has infinitely many roots, so any argument about one equation is doomed. The problem is the pair. Sine and cosine in a pair means one complex number, $\int_0^x P(t)e^{it}\,dt$. Set it to zero and take the modulus: the e^{ix} disappears and a polynomial is left.*

> *After integration by parts I have $Q\cos x - Q'\sin x$ and $Q\sin x + Q'\cos x$. That is the vector (Q, Q') rotated by x. Rotation preserves length. Square and add.*

Layer 3 THE WHY

Why combine the equations? Because the phrase "system of simultaneous equations" is the whole content of the problem, and the example $P \equiv 1$ shows it: $1 - \cos x = 0$ and $\sin x = 0$ each have infinitely many roots, and so does the pair. The problem says the pair becomes finite once P is non-constant. So the proof must use both equations at once, and must use non-constancy. Any approach that analyses one equation and then "intersects" is fighting the structure.

Why the complex exponential? The two integrals are the imaginary and real parts of $H(x) = \int_0^x P(t)e^{it}\,dt$, and $H(x) = 0$ is equivalent to the system. Integrating $P(t)e^{it}$ by parts repeatedly produces $e^{ix}R(x)$ for a polynomial R of degree n (solve $R' + iR = P$; the solution is a polynomial because $D + i$ is invertible on polynomials: $R = -i\sum_{k\geq 0}(iD)^k P$). So $H(x) = e^{ix}R(x) - R(0)$, and $H(x) = 0$ says $e^{ix}R(x) = R(0)$. The factor e^{ix} is the only non-polynomial thing in sight, and it has modulus 1. Taking moduli removes it: $|R(x)|^2 = |R(0)|^2$. For real x, $|R(x)|^2 = (\operatorname{Re} R)^2 + (\operatorname{Im} R)^2$ is a real polynomial of degree $2n$. That is $(*)$; with $R = -iQ + Q'$ one checks $|R|^2 = Q^2 + Q'^2$. The real computation in the model chain is this complex computation written out.

Why is the modulus the right thing to take, at axiom level? A complex number is determined by its modulus and argument. The argument of $e^{ix}R(x)$ winds round forever as x grows; that winding is what produces infinitely many roots of each separate equation. The modulus does

not wind. Equating moduli throws away the oscillating information and keeps the algebraic information, and algebraic equations have finitely many roots. That is the entire mechanism.

The MAA solution derives $Q(x) = Q'(0)\sin x + Q(0)\cos x$ (multiply the first equation by $\cos x$, the second by $\sin x$, add), deduces from boundedness of the right side that solutions lie in some $|x| \le M$, and then uses Rolle's theorem on F inside $[-M, M]$. That proof is correct: $F' = P\sin x$ has finitely many zeros in a bounded interval, and between consecutive ones F is strictly monotone, so F has finitely many zeros there. But it uses only one consequence of the system (boundedness) and then falls back to one equation. Squaring and adding uses the full content, gives an explicit bound $2\deg P$, and needs no Rolle. Where the official solution writes "$Q = P - P'' + P^{(4)} - \cdots$" without comment, here is where it comes from: you want a polynomial Q with $(Q + Q'') = P$, because the operator $D^2 + 1$ is what differentiating $-Q\cos + Q'\sin$ produces; inverting $1 + D^2$ as a geometric series $1 - D^2 + D^4 - \cdots$ is legitimate on polynomials because D is nilpotent there.

Layer $+\alpha$ BACKTESTING

1. The cue is a matched pair $\int P\sin$ and $\int P\cos$, set simultaneously to zero. Write down $\int_0^x P(t)e^{it}\,dt = 0$.

2. Integrate by parts until the integral is $e^{ix}\cdot(\text{polynomial})$ minus a constant, then take the modulus (equivalently, square and add the real pair).

3. Hesitation occurs after computing F and G for $P(t) = t$ and seeing that each has infinitely many roots. Reinforce: "infinitely many roots each, finitely many together" means the proof must eliminate the oscillating variable between the two equations.

4. The trap is asymptotic root-location for one equation, and the related trap of applying Rolle on the whole line, where F' has infinitely many zeros and Rolle gives no bound.

5. When two equations are $A\cos x - B\sin x = c$ and $A\sin x + B\cos x = d$, square and add. More generally, when sine and cosine appear as a pair, pass to e^{ix} and separate modulus (algebraic) from argument (oscillating).

6. Reusable rule: to show a system with oscillating terms has finitely many solutions, find a combination of the equations from which the oscillation cancels exactly (modulus, sum of squares, a resultant); a non-constant polynomial equation remains, and it has finitely many roots.

▷ **Cue:** $\int P(t)\sin t$ and $\int P(t)\cos t$ appearing together $\longrightarrow$ **Move:** combine as $\int P(t)e^{it}\,dt$

▷ **Cue:** $A\cos x - B\sin x = c$ and $A\sin x + B\cos x = d$ simultaneously $\longrightarrow$ **Move:** square and add: $A^2 + B^2 = c^2 + d^2$

▷ **Cue:** each equation of a system has infinitely many roots but the claim is finitely many for the system $\longrightarrow$ **Move:** eliminate the oscillating variable between the equations; never analyse one equation alone

▷ **Cue:** repeated integration by parts of polynomial times sin or cos $\longrightarrow$ **Move:** find Q with $Q + Q'' = P$, namely $Q = P - P'' + P^{(4)} - \cdots$, and write the antiderivative in one line

▷ **Cue:** a hypothesis "nonconstant" in a finiteness claim $\longrightarrow$ **Move:** test the constant case; it will show which quantity must be a non-constant polynomial

1980 A6

1980 A6. Let C be the class of all real valued continuously differentiable functions f on the interval $0 \le x \le 1$ with $f(0) = 0$ and $f(1) = 1$. Determine the largest real number u such that

$$u \le \int_0^1 |f'(x) - f(x)|\,dx$$

for all f in C.

Fields: analysis (integral inequalities, integrating factors) **Position:** A6 — last of the morning paper; a two-line lower bound once the integrating factor is seen, plus an extremal construction that must be done carefully because the infimum is not attained **Answer:** $u = 1/e$. Lower bound from $f' - f = e^x(e^{-x}f)'$ and $e^x \geq 1$; sharpness from functions with $e^{-x}f$ rising from 0 to $1/e$ on a short interval $[0, a]$ and constant after.

Layer 1 THE WHAT — surface analysis

The natural attempt. You try examples. $f(x) = x$: $\int_0^1 |1 - x|\,dx = \frac{1}{2}$. $f(x) = x^n$: $\int_0^1 x^{n-1}(n - x)\,dx = 1 - \frac{1}{n+1}$, worse. Since $f' = f$ is solved by e^x, try to make $f' - f$ small with exponentials: $f = (e^x - 1)/(e - 1)$ gives $f' - f = 1/(e - 1)$, integral ≈ 0.582; $f = \sinh x / \sinh 1$ gives $f' - f = e^{-x}/\sinh 1$, integral $2/(e + 1) \approx 0.538$. The numbers go down, but they do not tell you where they stop.

So you try for a lower bound. The triangle inequality for integrals:

$$\int_0^1 |f' - f|\,dx \;\geq\; \left|\int_0^1 (f' - f)\,dx\right| = \left|1 - \int_0^1 f\,dx\right|.$$

But $\int_0^1 f$ can be anything — make f rise steeply to about 1 near $x = 0$ and $\int f \approx 1$. The bound is 0. The wall: the absolute value was pulled out of the integral against the wrong function, and all information was lost. Calculus of variations does not help either, because $|f' - f|$ is not differentiable where $f' = f$, and the infimum turns out not to be attained, so there is no minimiser to find.

The model chain. *Lower bound.* For $f \in C$ put $g(x) = e^{-x}f(x)$, so g is continuously differentiable, $g(0) = 0$, $g(1) = 1/e$, and

$$f'(x) - f(x) = e^x\big(g'(x) + g(x)\big) - e^x g(x) = e^x g'(x).$$

Since $e^x \geq 1$ on $[0, 1]$,

$$\int_0^1 |f' - f|\,dx = \int_0^1 e^x |g'(x)|\,dx \;\geq\; \int_0^1 |g'(x)|\,dx \;\geq\; \left|\int_0^1 g'(x)\,dx\right| = g(1) - g(0) = \frac{1}{e}.$$

So $u = 1/e$ is a lower bound for every $f \in C$.

No larger number works. Fix $0 < a < 1$ and let $h(s) = 3s^2 - 2s^3$ for $0 \leq s \leq 1$, $h(s) = 1$ for $s \geq 1$. Then $h(0) = 0$, $h(1) = 1$, $h'(s) = 6s(1 - s) \geq 0$ on $[0, 1]$ and $h'(1) = 0$, so h is continuously differentiable and non-decreasing on $[0, \infty)$. Define

$$g_a(x) = \tfrac{1}{e}\,h(x/a), \qquad f_a(x) = e^x g_a(x).$$

Then f_a is continuously differentiable, $f_a(0) = 0$, $f_a(1) = e \cdot \frac{1}{e} = 1$, so $f_a \in C$. By the identity above, $f_a' - f_a = e^x g_a'$, and $g_a' \geq 0$ vanishes for $x \geq a$. Hence

$$\int_0^1 |f_a' - f_a|\,dx = \int_0^a e^x g_a'(x)\,dx \leq e^a \int_0^a g_a'(x)\,dx = e^a \cdot \frac{1}{e} = e^{a-1}.$$

As $a \to 0^+$, $e^{a-1} \to 1/e$. So for any $v > 1/e$ some f_a has integral less than v, and v is not a valid lower bound. Therefore $u = 1/e$.

(The bound is never attained: equality in $\int e^x|g'| \geq \int |g'|$ needs $g' = 0$ on $(0, 1]$, forcing g constant, contradicting $g(0) = 0 \neq g(1)$. This is why the question asks for the largest lower bound and not for a minimum.)

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains want a lower bound and both reach for "absolute value of integral $\leq$ integral of absolute value". The natural attempt applies it to $f' - f$ directly and gets $|1 - \int f|$. The model first rewrites $f' - f = e^x(e^{-x}f)'$ and applies it to $(e^{-x}f)'$, whose integral is fixed by the boundary data.

> *The only things I know about f are $f(0)$ and $f(1)$. An integral gives me boundary values only when the integrand is an exact derivative. $f' - f$ is not a derivative, but $e^{-x}(f' - f) = (e^{-x}f)'$ is — the integrating factor from first-order ODEs.*

> *The weight e^x is smallest at $x = 0$. To make the integral small, put all of the change in $e^{-x}f$ right next to 0, and make $f' - f = 0$ everywhere else, i.e. $f = e^{x-1}$ on the rest of the interval.*

Layer 3 THE WHY

Why the integrating factor? Drive it down to what the problem gives you. The class C is defined by two numbers, $f(0) = 0$ and $f(1) = 1$; everything else about f is free. A lower bound valid for all of C can depend only on those two numbers, and the only mechanism that converts an integral into boundary values is the fundamental theorem of calculus: $\int_0^1 \phi' = \phi(1) - \phi(0)$. So the question becomes: what function ϕ built from f has a derivative closely related to $f' - f$? The expression $f' - f$ is the left side of the first-order linear ODE $y' - y = q$, and the A-level method for that ODE is to multiply by the integrating factor e^{-x} to make the left side $(e^{-x}y)'$. The integrating factor exists precisely to turn $y' - y$ into an exact derivative. So the move is not a trick; it is the standard answer to "make this a derivative".

Why is the natural triangle inequality wrong but the model's right? Both use $|\int \psi| \le \int |\psi|$. The inequality is sharp when ψ does not change sign, and it is useful only when $\int \psi$ is known. For $\psi = f' - f$, $\int \psi = 1 - \int f$ is not known. For $\psi = (e^{-x}f)'$, $\int \psi = 1/e$ is known. The price of switching is the weight: $|f' - f| = e^x|\psi|$, and to compare with $\int |\psi|$ you must discard $e^x \ge 1$. That is where the loss happens, and the loss is small exactly when ψ lives near $x = 0$ — which tells you the extremal functions.

Why the answer is $1/e$ and not 1: the change in $e^{-x}f$ from 0 to $1/e$ is fixed. Pay for it where the exchange rate e^x is cheapest, at $x = 0$, and the cost is $1/e$. After that, ride the free solution $f = e^{x-1}$ of $f' = f$, which costs nothing and arrives at $f(1) = 1$. The natural attempt's examples were paying for the change spread over the whole interval at exchange rates up to e.

The MAA solution takes $f_a(x) = (e^{a-1}/a)x$ on $[0, a]$ and e^{x-1} on $[a, 1]$, computes $e^{a-1}(1 - a/2)$, and then says one can "smooth out the corner". That f_a is not in C (its derivative jumps at a), so the smoothing is a genuine step. The cubic $h(s) = 3s^2 - 2s^3$ above does the smoothing in advance: it has zero slope at $s = 1$, so g_a, and hence $f_a = e^x g_a$, is continuously differentiable, and the estimate e^{a-1} follows from $g_a' \ge 0$ without computing any integral.

Layer $+\alpha$ BACKTESTING

1. The cue is the combination $f' - f$ together with fixed boundary values. Write down $f' - f = e^x(e^{-x}f)'$.

2. Integrating factor, then triangle inequality on the exact derivative, then an extremal family concentrating the change where the weight is smallest.

3. Hesitation occurs after the naive bound $|1 - \int f|$ collapses to 0. Reinforce: a lower bound over a class defined by boundary values must come from $\int \phi' = \phi(1) - \phi(0)$; find the ϕ.

4. Two traps: pulling the absolute value out against $f' - f$ itself, and building a piecewise extremal function with a corner and forgetting that C requires a continuous derivative.

5. Whenever an integral involves $f' + \lambda f$ (or $f' + pf$), and the data are values of f at the endpoints, multiply by the integrating factor before estimating anything.

6. Reusable rule: to bound $\int |Lf|$ below over functions with prescribed boundary values, write $Lf = w \cdot \phi'$ with ϕ determined by f; then $\int |Lf| \ge (\min w)\,|\phi(1) - \phi(0)|$, and the bound is approached by functions whose ϕ' concentrates where w is smallest.

▷ **Cue:** $f' - f$ (or $f' + pf$) inside an integral, with $f(0)$ and $f(1)$ given $\longrightarrow$ **Move:** multiply by the integrating factor: $f' - f = e^x(e^{-x}f)'$

- ▷ **Cue:** a lower bound for $\int |\psi|$ is wanted $\longrightarrow$ **Move:** find a rewriting $\psi = w\phi'$ with $\int \phi'$ known; then $\int |\psi| \geq \min w \cdot |\phi(1) - \phi(0)|$
- ▷ **Cue:** "largest u such that $u \leq \ldots$ for all f" $\longrightarrow$ **Move:** prove the bound, then build a family approaching it; expect the infimum not to be attained
- ▷ **Cue:** an extremal function with a corner, but the class requires C^1 $\longrightarrow$ **Move:** replace the linear ramp by $3s^2 - 2s^3$, which has zero slope at both ends
- ▷ **Cue:** a weighted cost $\int w|\phi'|$ with fixed total change in ϕ $\longrightarrow$ **Move:** concentrate all of ϕ' where w is minimal

Putnam 1980 — Paper B

1980 B1

1980 B1. For which real numbers c is

$$\frac{e^x + e^{-x}}{2} \leq e^{cx^2}$$

for all real x?

Fields: analysis (power series, inequalities) **Position:** B1 – opening problem of the second session
Answer: $c \geq \frac{1}{2}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The left side is $\cosh x$, and $\cosh x \leq e^{cx^2}$ has to hold for *every* real x, so you reach for calculus. Set $f(x) = e^{cx^2} - \cosh x$; $f(0) = 0$ and f is even, so you differentiate: $f'(x) = 2cxe^{cx^2} - \sinh x$, and $f'(0) = 0$ too. You are now trying to show $f \geq 0$ on all of $\mathbb{R}$ by studying the sign of f', which itself needs f'', and $f''(0) = 2c - 1$ tells you $c \geq \frac{1}{2}$ is necessary for f to have a local minimum, not a local maximum, at 0. That much is clean. But necessity is not sufficiency: you still owe a global argument that $f(x) \geq 0$ for large $|x|$ and everywhere in between, and $f'(x) = 2cxe^{cx^2} - \sinh x$ does not factor or have an obvious sign pattern. You end up wanting to argue "for large x, e^{cx^2} beats $e^{|x|}$ because x^2 beats $|x|$", which is true but vague, and stitching the large-x case to the near-0 case into one rigorous argument means casework you have not actually carried out. You have the right constant and no proof.

The model chain. Both sides are entire functions of x that depend on x only through x^2, so write them as power series in x^2 and compare term by term.

$$\cosh x = \sum_{n=0}^{\infty} \frac{x^{2n}}{(2n)!}, \qquad e^{x^2/2} = \sum_{n=0}^{\infty} \frac{(x^2/2)^n}{n!} = \sum_{n=0}^{\infty} \frac{x^{2n}}{2^n n!}.$$

Split $(2n)! = \big(1 \cdot 3 \cdot 5 \cdots (2n-1)\big) \cdot \big(2 \cdot 4 \cdot 6 \cdots 2n\big) = \big(1 \cdot 3 \cdots (2n-1)\big) \cdot 2^n n!$. The first factor is a product of positive odd integers, each ≥ 1, so $(2n)! \geq 2^n n!$ for every $n \geq 0$ (equality at $n = 0, 1$). Hence $x^{2n}/(2n)! \leq x^{2n}/(2^n n!)$ term by term for every real x, and summing,

$$\cosh x \leq e^{x^2/2} \qquad \text{for all } x \in \mathbb{R}.$$

If $c \geq \frac{1}{2}$ then $cx^2 \geq x^2/2$ for every real x, so $e^{cx^2} \geq e^{x^2/2} \geq \cosh x$: sufficiency, with no case split on the size of x.

For necessity, suppose $\cosh x \leq e^{cx^2}$ holds for all real x. Both sides equal 1 at $x = 0$ and both are even, so

$$0 \leq \lim_{x \to 0} \frac{e^{cx^2} - \cosh x}{x^2} = \lim_{x \to 0} \frac{\big(1 + cx^2 + O(x^4)\big) - \big(1 + \frac{x^2}{2} + O(x^4)\big)}{x^2} = c - \frac{1}{2},$$

so $c \geq \frac{1}{2}$. Combined with sufficiency, the answer is exactly $c \geq \frac{1}{2}$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree that $c \geq \frac{1}{2}$ is forced by the behaviour at $x = 0$ – that half is identical. They split on how to prove the inequality holds *everywhere*. The natural attempt keeps $\cosh x$ and e^{cx^2} as functions of x and tries to control the sign of their difference (or its derivative) pointwise, which forces a case split between "near 0" and "large $|x|$". The model rewrites both sides as power series in the single variable x^2 and compares corresponding coefficients, which needs no case split at all.

> *Both sides are sums of nonnegative terms once I write them in x^2 – every term of $\cosh x$'s series and every term of e^{cx^2}'s series is ≥ 0 for real x, so I can compare the sums by comparing the terms, the same way I'd compare $\sum a_n$ and $\sum b_n$ when $a_n \geq 0, b_n \geq 0$ termwise. That is a single global argument, not a pointwise one, so it needs no split between small and large x.*

> *The constant is pinned down at $x = 0$, not at some worst-case x I would have to hunt for – an inequality between two smooth even functions that are equal and tangent at 0 can only fail arbitrarily close to 0 if their quadratic coefficients disagree, so I should read off the necessary condition from a limit at 0, not from testing candidate values of x.*

Layer 3 THE WHY

Why does the termwise comparison work at all – why is it not a trick that happens to succeed here? Because for real x, every term $x^{2n}/(2n)!$ and $x^{2n}/(2^n n!)$ is nonnegative: there is no sign to lose track of. Comparing two series of nonnegative numbers by comparing them term by term is not an inequality technique specific to this problem, it is what "termwise" means – if $a_n \leq b_n$ for every n and both series converge, then $\sum a_n \leq \sum b_n$, full stop. The only work is establishing $a_n \leq b_n$, i.e. $(2n)! \geq 2^n n!$, and that in turn is not an estimate to be proved by induction or bounded crudely: it falls out of what $(2n)!$ *is*. The factorial $(2n)!$ is the product of the first $2n$ positive integers; splitting that product by parity into odd and even factors is always available, and the even factors are exactly $2 \cdot 4 \cdots 2n = 2^n n!$. The odd factors are the "extra" multiplicative content that a general factorial has and $2^n n!$ does not, and each is at least 1, so the inequality is not close – it is a statement that a product of integers ≥ 1 is ≥ 1. Once the problem is phrased in x^2, the comparison is forced to be this cheap; the work of the model chain is entirely the decision to phrase it in x^2 at all, which is licensed by both functions depending on x only through x^2 (both are even).

Why is the limit at 0 the right place to extract necessity, rather than an inequality that must be checked at some specific worst x? Because the inequality $\cosh x \leq e^{cx^2}$, viewed as a statement about the single real parameter c, is monotonic in c (increasing c only helps), so there is a threshold value of c below which it fails somewhere and above which it holds everywhere; the threshold is exactly where the two sides first stop being tangent to second order at $x = 0$, since that is where they are forced to agree to lowest order (both are 1 at $x = 0$, both are even, so the difference has no constant or linear term). Reading the coefficient of x^2 is therefore reading the threshold directly, not searching for it.

Layer $+\alpha$ BACKTESTING

1. The cue is an inequality between two manifestly even, entire functions of x (here $\cosh x$ and e^{cx^2}) required to hold for *all* real x. On seeing it, the useful thing to write down is: rewrite both as power series in x^2.
2. That cue suggests termwise coefficient comparison for sufficiency, and a limit (equivalently, matching the lowest-order Taylor coefficient) at $x = 0$ for necessity, rather than differentiating the difference and studying its sign.

3. Hesitation occurs right after getting $c \geq \frac{1}{2}$ from the second derivative at 0 and mistaking that local necessary condition for a finished proof; it is reinforced away by asking explicitly "have I shown this for *every* x, or only near 0?"
4. The trap is treating a global inequality as a calculus problem about the sign of a derivative, which is true in spirit but requires unstated casework between small and large $|x|$ to finish rigorously.
5. Any time both sides of an inequality are power series (or can be written as one) with all-nonnegative terms for the relevant range of x, recognise the setup instantly and compare coefficients; do not open a derivative.
6. Reusable rule: to prove $f(x) \leq g(x)$ for all x in some domain where both are given by (or can be rewritten as) power series with nonnegative terms there, compare the series term by term for sufficiency, and extract any necessary condition on a parameter from the lowest-order term at which the two series can disagree, not from a search over x.

▷ **Cue:** an inequality between two even, entire functions must hold for all real x $\longrightarrow$ **Move:** rewrite both sides as power series in x^2 rather than studying the sign of their difference
▷ **Cue:** two series with all-nonnegative terms (for the x in question) are compared $\longrightarrow$ **Move:** compare them term by term; the sum inequality follows immediately once each term inequality is shown
▷ **Cue:** need to compare $(2n)!$ with $2^n n!$ $\longrightarrow$ **Move:** split the factorial by parity into its odd and even factors: $(2n)! = (1 \cdot 3 \cdots (2n-1)) \cdot 2^n n!$, and the odd factors are each ≥ 1
▷ **Cue:** a global inequality $f(x, c) \geq 0$ (all x) must pin down a threshold value of a parameter c $\longrightarrow$ **Move:** read the threshold off the lowest-order Taylor coefficient at the point where the two sides are forced to agree (often $x = 0$), rather than searching over x for a worst case

1980 B2

1980 B2. Let S be the solid in three-dimensional space consisting of all points (x, y, z) satisfying the following system of six simultaneous conditions:

$$x \geq 0, \qquad y \geq 0, \qquad z \geq 0, \qquad x + y + z \leq 11, \qquad 2x + 4y + 3z \leq 36, \qquad 2x + 3z \leq 24.$$

(a) Determine the number v of vertices of S. (b) Determine the number e of edges of S. (c) Sketch in the bc-plane the set of points (b, c) such that $(2, 5, 4)$ is one of the points (x, y, z) at which the linear function $bx + cy + z$ assumes its maximum value on S.

Fields: geometry (convex polyhedra), linear programming **Position:** B2 – linear-algebra-flavoured problem of the second session **Answer:** $v = 7$, $e = 11$; (c) the points with $b + c = 2$ and $\frac{2}{3} \leq b \leq 1$.

Layer 1 THE WHAT — surface analysis

The natural attempt. S is cut out by six half-spaces, so a vertex is a point satisfying three of the six defining equalities with equality (the other three as strict or non-strict inequalities). There are $\binom{6}{3} = 20$ triples of planes to check. You start listing them: $x = 0, y = 0, z = 0$ gives $(0, 0, 0)$, which is feasible. $x = 0, y = 0, x + y + z = 11$ gives $(0, 0, 11)$, but that fails $2x + 3z \leq 24$ since $33 > 24$ – infeasible, discard. You keep going, triple by triple, substituting, solving a 3×3 linear system each time, and checking it against the remaining three inequalities. This works and will eventually produce the right seven points, but it is twenty small linear systems and twenty feasibility checks, done with no structural help from the shape of the numbers – a purely mechanical sweep. Worse, for part (b) there is no comparable mechanical recipe at all: "count the edges" invites you to look at your list of seven vertices and ask, pair by pair ($\binom{7}{2} = 21$ pairs), whether the segment between them lies in ∂S, which is even less structured than the vertex search. Part (c) then asks about $(2, 5, 4)$, a point that is not even on your vertex list, and

the natural attempt has no framework ready for "a linear function's maximiser is not a single vertex."

The model chain. Treat S as a bounded convex polyhedron and use its combinatorics, not brute enumeration.

Vertices. Find the 20 triples' intersections that survive, using the structure of the constraints to prune fast: $x+y+z \le 11$ is the binding "ceiling" unless $2x+3z \le 24$ or $2x+4y+3z \le 36$ cuts it first. Solving systematically gives seven feasible vertices:

$$V_0 = (0,0,0),\ V_1 = (11,0,0),\ V_2 = (0,9,0),\ V_3 = (0,0,8),\ V_4 = (0,3,8),\ V_5 = (9,0,2),\ V_6 = (4,7,0).$$

(For instance V_4: on $x = 0$, $2x+3z = 24 \Rightarrow z = 8$, $2x+4y+3z = 36 \Rightarrow y = 3$; check $x+y+z = 11 \le 11$, satisfied with equality but not violated.) So $v = 7$.

Edges, via Euler. Rather than test $\binom{7}{2} = 21$ pairs, identify which of the six bounding planes actually carve out a face (a 2-dimensional piece of ∂S) by listing, for each plane, which of the seven vertices lie on it:

$$x=0 : \{V_0, V_2, V_3, V_4\}, \quad y=0 : \{V_0, V_1, V_3, V_5\}, \quad z=0 : \{V_0, V_1, V_2, V_6\},$$

$$x+y+z=11 : \{V_1, V_4, V_5, V_6\}, \quad 2x+4y+3z=36 : \{V_2, V_4, V_6\}, \quad 2x+3z=24 : \{V_3, V_4, V_5\}.$$

All six planes support at least three vertices, so S has $F = 6$ faces: four quadrilaterals and two triangles. Euler's formula for a convex polyhedron, $v - e + f = 2$, gives $e = v + f - 2 = 7 + 6 - 2 = 11$. (Consistency check: summing edges per face, $4+4+4+4+3+3 = 22 = 2e$, so $e = 11$ again.)

Part (c). The point $(2,5,4)$ satisfies $x+y+z = 11$ and $2x+4y+3z = 36$ with equality (both tight) but $2x+3z = 16 < 24$ (slack): it lies in the relative interior of the edge shared by those two faces, namely V_4V_6 (check: $(2,5,4) = \frac{1}{2}(V_4 + V_6)$). A linear functional $L(x,y,z) = bx + cy + z$ attains its maximum on a convex polytope on a face, and it attains it at an *interior* point of an edge only if it is constant on that whole edge – otherwise the maximum on the edge is at one endpoint only. So the condition is: L is constant along V_4V_6, and that constant value is the global maximum of L on S. Constancy along the edge direction $V_6 - V_4 = (4,4,-8)$, i.e. direction $(1,1,-2)$, means $b \cdot 1 + c \cdot 1 + 1 \cdot (-2) = 0$, i.e. $b + c = 2$. Given $b + c = 2$, the common value on the edge is $L(V_4) = 3c + 8 = 3(2-b) + 8 = 14 - 3b$; this must be $\ge L$ at every other vertex. Checking the remaining vertices (using $c = 2 - b$) forces $\frac{2}{3} \le b \le 1$. So the answer is the segment $b + c = 2$, $\frac{2}{3} \le b \le 1$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains must find the same seven vertices; they split on what to do with that list. The natural attempt treats parts (a) and (b) as two independent enumeration tasks – solve 20 systems for (a), then inspect 21 pairs for (b) – and has no tool ready when (c) asks about a non-vertex point. The model reads the problem as "a convex polyhedron" the moment three planes bound it, and reaches for the invariant relating v, e, and the number of faces f, which turns (b) into arithmetic once the faces (not the edges) are tallied; it also recognises that $(2,5,4)$'s two tight constraints identify it as lying on an edge, which is what makes "the maximiser" a whole segment condition rather than a vertex condition.

> *This is a bounded region cut out by finitely many linear inequalities in $\mathbb{R}^3$ – that is a convex polyhedron, and convex polyhedra satisfy $v - e + f = 2$. If I can count faces (six candidate planes, check which actually meet the region in a two-dimensional piece) that is far fewer objects to examine than pairs of vertices, and Euler's formula converts that count directly into the edge count I actually need.*

> $(2,5,4)$ *has exactly two of the six defining inequalities tight – that is the signature of a point on an edge, not a vertex (which needs three tight, generically) or a face interior (which needs one). A linear function maximised at an interior point of an edge must be constant on the whole edge, because a non-constant linear function restricted to a line segment is strictly monotone and so has its two endpoints as its only maximisers.*

Layer 3 THE WHY

Why is Euler's formula the natural tool the instant the region is recognised as a bounded polyhedron, rather than a fact to recall and hope applies? Because $v - e + f = 2$ is not a coincidence about this particular shape – it holds for every convex polyhedron topologically equivalent to a sphere, which every bounded solid cut out by finitely many half-spaces in $\mathbb{R}^3$ is (its boundary is a sphere with a graph of vertices and edges drawn on it, dividing it into faces). Once you know the identity holds, counting faces is cheap because faces correspond one-to-one with the bounding planes that actually contribute a two-dimensional piece of the boundary, and there are only six candidates to test – versus $\binom{v}{2} = 21$ vertex pairs to test for edges directly, most of which are not edges. The problem's own data (six inequalities) already hands you the upper bound on f; edges do not have a comparably short list handed to you, which is precisely why routing through f is the efficient path.

Why does a maximiser in the interior of an edge force constancy on the whole edge, and why is this not a special fact about this problem but a fact about linear functions on line segments? Because a linear function restricted to any line is an affine function of one real parameter, and an affine function of one variable is either constant or strictly monotone; if it is strictly monotone on the segment V_4V_6, its unique maximiser is one endpoint, and $(2,5,4)$, sitting strictly between them, cannot be it. So interior-point-is-a-maximiser is only possible in the degenerate (constant) case, and constancy of $L(x,y,z) = bx + cy + z$ along a direction vector d is exactly the condition $(b,c,1) \cdot d = 0$ – the gradient of L is perpendicular to the edge. This is the same fact that underlies why linear programs have vertex (or edge, or face) optima: the objective's level sets are hyperplanes, and a bounded polytope's intersection with a family of parallel hyperplanes is maximised where the hyperplane last touches the polytope, which is a vertex generically and a whole face exactly when the objective's gradient is parallel to that face.

Layer $+\alpha$ BACKTESTING

1. The first cue is six linear inequalities in three variables cutting out a bounded region: on seeing it, write down "convex polyhedron, so $v - e + f = 2$ applies" before starting to search for vertices by hand.

2. That cue suggests finding vertices from triples of tight constraints (necessary regardless), then finding faces from which single constraints actually meet the region in two dimensions, then getting edges from Euler rather than direct enumeration.

3. Hesitation occurs at part (c), when $(2,5,4)$ turns out not to be on the vertex list; it is reinforced away by checking directly which of the six inequalities are tight there and recognising two tight constraints as the signature of an edge point.

4. The trap is enumerating all $\binom{7}{2}$ vertex pairs to find edges, which is slower and more error-prone than Euler's formula, and separately assuming a maximiser of a linear function must be a vertex, missing that it can be an edge or a face when the gradient aligns with it.

5. Any time a solid is presented as an intersection of finitely many half-spaces in $\mathbb{R}^3$ and is bounded, recognise "convex polyhedron" immediately and reach for $v - e + f = 2$ rather than raw enumeration for whichever of v, e, f is not the cheapest to count directly.

6. Reusable rule: for a bounded polyhedron given by linear inequalities, count vertices from triples of tight constraints, count faces from single constraints that actually support the boundary, and get the third quantity from Euler's formula; and remember a linear objective is maximised at an edge or face, not just a vertex, exactly when its gradient is orthogonal to that edge or face's direction(s).

- ▷ **Cue:** a bounded region in $\mathbb{R}^3$ is cut out by finitely many linear inequalities $\longrightarrow$ **Move:** recognise it as a convex polyhedron and reach for Euler's formula $v - e + f = 2$ to relate vertex, edge, and face counts
- ▷ **Cue:** need the edge count of a polyhedron given its vertices $\longrightarrow$ **Move:** count faces instead (constraints supporting ≥ 3 vertices) and solve $e = v + f - 2$, rather than testing all $\binom{v}{2}$ vertex pairs
- ▷ **Cue:** a specific point has exactly two of several defining inequalities tight $\longrightarrow$ **Move:** recognise it as lying on the edge where those two faces meet, not at a vertex or in a face interior
- ▷ **Cue:** a linear function is maximised at an interior point of a segment $\longrightarrow$ **Move:** it must be constant on the whole segment, since a non-constant linear function restricted to a line is strictly monotone; the constancy condition is that the objective's gradient is perpendicular to the segment's direction

1980 B3

1980 B3. For which real numbers a does the sequence defined by the initial condition $u_0 = a$ and the recursion $u_{n+1} = 2u_n - n^2$ have $u_n > 0$ for all $n \geq 0$? (Express the answer in the simplest form.)

Fields: algebra (linear recurrences, difference equations) **Position:** B3 – closing problem of the second session's first half **Answer:** $a \geq 3$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You compute terms: $u_0 = a$, $u_1 = 2a$, $u_2 = 4a - 1$, $u_3 = 8a - 1 - 4 = 8a - 5$, $u_4 = 16a - 10 - 9 = 16a - 19$, and the coefficients of a are doubling as expected while the constants are some accumulating mess. You suspect a closed form $u_n = 2^n a -$ (something growing polynomially, roughly like n^2), and you try to guess it from the pattern of constants $0, 1, 5, 19, \ldots$: differences $1, 4, 14, \ldots$, not an obviously nice sequence at a glance, and pattern-matching from four terms is unreliable. Alternatively you try to bound things directly: since $u_{n+1} = 2u_n - n^2$, if u_n is ever large enough relative to n^2 it stays positive and even grows without bound, but if u_n is small the subtraction of n^2 can push it negative, and it is unclear from the recursion alone how large is large enough at each step – you would need a bound that keeps re-proving itself as n grows, i.e. an inductive hypothesis you have not yet found. Both routes want a closed form; neither is finding one efficiently by staring at the numbers.

The model chain. Recognise $u_{n+1} - 2u_n = -n^2$ as a linear difference equation with constant coefficient 2 and polynomial forcing term $-n^2$, exactly the discrete analogue of a linear ODE $y' - 2y = -x^2$: solve it as (particular solution) + (general homogeneous solution).

Homogeneous part. $v_{n+1} = 2v_n$ has general solution $v_n = K \cdot 2^n$.

Particular part. Since the forcing term is a degree-2 polynomial in n and $2 \neq 1$ is not a root of the homogeneous equation's characteristic behaviour, try a polynomial ansatz of the same degree, $u_n^{(p)} = An^2 + Bn + C$:

$$u_{n+1}^{(p)} - 2u_n^{(p)} = A(n+1)^2 + B(n+1) + C - 2(An^2 + Bn + C) = -An^2 + (2A - B)n + (A + B - C).$$

Matching this to $-n^2$ (coefficients of n^2, n^1, n^0 on the right are $-1, 0, 0$) gives $A = 1$, $2A - B = 0 \Rightarrow B = 2$, $A + B - C = 0 \Rightarrow C = 3$. So $u_n^{(p)} = n^2 + 2n + 3$, and one checks directly that this indeed satisfies the recursion.

General solution and the initial condition. $u_n = n^2 + 2n + 3 + K \cdot 2^n$; setting $n = 0$ gives $u_0 = 3 + K = a$, so $K = a - 3$ and

$$u_n = n^2 + 2n + 3 + (a - 3)2^n.$$

The condition $u_n > 0$ for all n. If $a \geq 3$, then $(a-3)2^n \geq 0$ for every n, so $u_n \geq n^2+2n+3 > 0$ for all $n \geq 0$. If $a < 3$, then $(a-3)2^n \to -\infty$, and since 2^n grows faster than any polynomial, $n^2+2n+3+(a-3)2^n \to -\infty$ as $n \to \infty$; more precisely $\lim_{n\to\infty} \dfrac{n^2+2n+3}{2^n} = 0$, so for n large enough $(a-3)2^n$ dominates and $u_n < 0$. Hence $u_n > 0$ for all $n \geq 0$ if and only if $a \geq 3$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree the recursion is linear with polynomial forcing; they split on whether to name that structure before computing. The natural attempt computes $u_1, u_2, u_3, \ldots$ first and tries to reverse-engineer a closed form from the numbers, which is pattern-matching against a target you have not characterised. The model names the equation type first – linear, constant-coefficient, polynomial forcing – and reaches immediately for the standard decomposition (particular + homogeneous), which *manufactures* the closed form rather than guessing it, and then reads off $a \geq 3$ from the sign of the one free constant $K = a-3$ multiplying the dominant term 2^n.

> *This is a first-order linear recursion $u_{n+1} = 2u_n + (\text{polynomial in } n)$, and I already know how to solve exactly this shape: a particular solution matching the forcing term's degree, plus $K \cdot 2^n$ for the homogeneous part, with K fixed by u_0. I should not compute terms and guess a pattern when I know the general method that produces the closed form outright.*

> *Once I have $u_n = n^2+2n+3+(a-3)2^n$, the entire question "for which a is $u_n > 0$ for all n" collapses to a question about a single real number $K = a-3$: the polynomial part is always positive and grows only polynomially, while 2^n grows without bound, so the sign of K – not its size – is what eventually decides the sign of u_n for large n, and if $K \geq 0$ the polynomial part alone already guarantees positivity for every n, including the small ones.*

Layer 3 THE WHY

Why is "particular plus homogeneous" the natural decomposition here, and not a technique borrowed from differential equations and hoped to transfer? Because the recursion $u_{n+1} - 2u_n = -n^2$ is linear in u: if u_n and w_n both satisfy the inhomogeneous recursion, their difference $u_n - w_n$ satisfies the homogeneous one $v_{n+1} = 2v_n$, and every solution of the homogeneous recursion is $K \cdot 2^n$ for some constant K (immediate by induction: $v_n/2^n$ is constant). So the set of *all* solutions of the original recursion is exactly (one particular solution) + (the one-parameter family of homogeneous solutions) – this is not an ansatz to be justified case by case, it is what linearity of the recursion in u means. The only genuine work is finding one particular solution, and trying a polynomial of the forcing term's own degree is justified because applying the linear operator $u_n \mapsto u_{n+1} - 2u_n$ to a degree-d polynomial in n returns a polynomial of degree d with a computable leading coefficient (here, degree drops by nothing since the coefficient $2 \neq 1$ does not cause the leading terms to cancel) – so matching degree-2 forcing with a degree-2 ansatz is guaranteed, by degree-counting, to have a solution for its three coefficients, which is exactly what the three linear equations for A, B, C confirm.

Why does the single sign of $K = a-3$ decide everything, rather than the actual size of a? Because 2^n eventually dominates every polynomial in n – $n^2+2n+3 = o(2^n)$ – so for large n the term $(a-3)2^n$ overwhelms the polynomial part regardless of how close a is to 3; there is no intermediate regime where a slightly negative K survives forever, because "slightly negative" is still eventually infinitely more negative than any fixed polynomial is positive. Conversely, $K \geq 0$ is not merely enough for large n, it is enough for *every* $n \geq 0$ simultaneously, because it makes the whole 2^n term non-negative rather than merely eventually beneficial – there is no case left to check for small n. This is the general reason "eventually dominant term" problems reduce to a threshold on a single coefficient: geometric growth beats polynomial growth in the

limit, with no exceptions surviving at any finite n once the sign is fixed correctly, and no rescue possible once it is fixed the other way.

Layer $+\alpha$ BACKTESTING

1. The cue is a recursion of the form $u_{n+1} = (\text{constant}) \cdot u_n + (\text{polynomial in } n)$. On seeing it, write down: this is a linear recursion with polynomial forcing, solvable as particular + homogeneous.

2. That cue suggests the standard method directly: find $K \cdot (\text{growth rate})^n$ as the homogeneous family, and a same-degree polynomial ansatz as the particular solution, rather than computing terms and pattern-matching.

3. Hesitation occurs when guessing the degree of the polynomial ansatz – match it to the degree of the forcing term (n^2 forcing needs a degree-2 ansatz), and it is reinforced away by checking that the linear operator $u_n \mapsto u_{n+1} - 2u_n$ does not drop the degree here (it would if the recursion's constant were 1, causing the linear terms to cancel and forcing a higher-degree ansatz).

4. The trap is computing several terms by hand, spotting a partial pattern in the constants, and trying to leverage it into a closed form by further guessing, instead of applying the general method for this recursion shape once and for all.

5. Any time a recursion is $u_{n+1} = r \cdot u_n + p(n)$ for a polynomial p and constant $r \neq 1$, recognise the particular-plus-homogeneous method immediately; and any time a closed form is a bounded (or polynomial) term plus K times an exponential, recognise that only the sign of K ultimately controls eventual positivity or negativity.

6. Reusable rule: solve linear recursions with polynomial forcing by (i) finding the homogeneous family from the recursion's constant, (ii) finding one particular solution via a same-degree polynomial ansatz, (iii) fixing the free constant from the initial condition, and (iv) for an "eventually true/false" question about the resulting closed form, decide it from the sign of the coefficient of the fastest-growing term, since geometric growth eventually swamps any polynomial term regardless of the polynomial's own size.

▷ **Cue:** a recursion has the form $u_{n+1} = r \cdot u_n + p(n)$ for constant r and polynomial p $\longrightarrow$ **Move:** solve as particular solution (polynomial ansatz of p's degree) plus homogeneous solution $K \cdot r^n$, then fix K from the initial value

▷ **Cue:** a linear recursion's forcing term is a polynomial of degree d and the recursion's constant is not 1 $\longrightarrow$ **Move:** try a polynomial ansatz of the same degree d; the degree survives because the constant term of the linear operator does not cancel the leading coefficient

▷ **Cue:** a closed form is (bounded or polynomial growth) + $K \cdot$ (exponential growth), and the question asks for a property to hold for all n $\longrightarrow$ **Move:** the answer reduces to a threshold on the sign (or value) of K, since the exponential term eventually dominates any polynomial term regardless of size

▷ **Cue:** computing the first few terms of a sequence to guess a closed form $\longrightarrow$ **Move:** prefer naming the recursion's type (linear, constant-coefficient, polynomial-forced) and solving it by the standard method over pattern-matching from a handful of terms

1980 B4

1980 B4. Let $A_1, A_2, \ldots, A_{1066}$ be subsets of a finite set X such that $|A_i| > \frac{1}{2}|X|$ for $1 \leq i \leq 1066$. Prove there exist ten elements $x_1, \ldots, x_{10}$ of X such that every A_i contains at least one of $x_1, \ldots, x_{10}$. (Here $|S|$ means the number of elements in the set S.)

Fields: combinatorics; double counting; greedy algorithms **Position:** B4 — middle of the paper; a covering problem whose number 1066 is chosen to sit just above $2^{10} = 1024$ **Answer:** Choose the x_j greedily. Double counting shows that some element lies in *strictly* more than half of the sets still

uncovered, so the number r of uncovered sets drops to at most $\lfloor (r-1)/2 \rfloor$ each step: $1066 \to 532 \to 265 \to 132 \to 65 \to 32 \to 15 \to 7 \to 3 \to 1 \to 0$. That is ten steps.

Layer 1 THE WHAT — surface analysis

The natural attempt. Every set is more than half of X, and ten elements have to hit all 1066 sets. A strong student who knows the probabilistic method sees "more than half" and "ten" and reaches for it straight away. Pick $x_1, \ldots, x_{10}$ independently and uniformly from X. A fixed A_i is missed by one choice with probability $1 - |A_i|/|X| < \frac{1}{2}$, so it is missed by all ten with probability less than 2^{-10}. So the expected number of missed sets is

$$\sum_{i=1}^{1066} \Pr(A_i \text{ missed}) < \frac{1066}{1024}.$$

And that is the wall. $1066/1024 > 1$, so the expectation bound does not force a missed count of 0. Sampling without replacement helps a little for small $|X|$, but as $|X| \to \infty$ the miss probability tends to 2^{-10} from below and the bound is still above 1. Tightening "$< \frac{1}{2}$" to "$\leq \frac{1}{2} - \frac{1}{2|X|}$" does not rescue it either, because that gain also vanishes as $|X|$ grows.

The second natural attempt is the greedy one done loosely: "some element is in at least half the sets, so each step halves the count." Carry it out: $1066 \to 533 \to 266 \to 133 \to 66 \to 33 \to 16 \to 8 \to 4 \to 2 \to 1$. Ten steps, and one set is still uncovered. Same wall, same place, and for the same reason: $1066 > 2^{10}$.

The model chain. Let $m = |X|$. Note first that $m \geq 1$ and every A_i is nonempty, since $|A_i| > m/2 \geq 0$ and $|A_i|$ is an integer.

Claim. If $\mathcal{F}$ is a nonempty family of r subsets of X (repetitions allowed), each of size $> m/2$, then some $x \in X$ lies in more than $r/2$ of them.

Proof. Count the pairs (x, A) with $A \in \mathcal{F}$ and $x \in A$. Counting by sets gives $\sum_{A \in \mathcal{F}} |A| > rm/2$. Counting by elements gives $\sum_{x \in X} n_x$, where n_x is the number of members of $\mathcal{F}$ that contain x. If every $n_x \leq r/2$, then $\sum_x n_x \leq mr/2$, a contradiction. □

Greedy step. Let $\mathcal{F}_0 = \{A_1, \ldots, A_{1066}\}$ and $r_0 = 1066$. Given $\mathcal{F}_{k-1}$ with $r_{k-1} \geq 1$ members, choose x_k by the Claim and let $\mathcal{F}_k$ be the members of $\mathcal{F}_{k-1}$ that do not contain x_k. Then

$$r_k < r_{k-1} - \frac{r_{k-1}}{2} = \frac{r_{k-1}}{2}, \qquad \text{so} \qquad r_k \leq \left\lfloor \frac{r_{k-1} - 1}{2} \right\rfloor.$$

(The members of $\mathcal{F}_k$ still have size $> m/2$, so the Claim applies again. We never need to shrink X.) If $r_{k-1} = 0$, let x_k be any element and $r_k = 0$.

Count. Iterating $r \mapsto \lfloor (r-1)/2 \rfloor$ from 1066:

$$1066 \to 532 \to 265 \to 132 \to 65 \to 32 \to 15 \to 7 \to 3 \to 1 \to 0.$$

So $r_{10} = 0$: every A_i contains one of $x_1, \ldots, x_{10}$. □

On distinctness. The problem does not ask for the x_j to be distinct. If they have to be, and $|X| \geq 10$, replace any repeats with unused elements. Adding elements cannot uncover a set. If $|X| < 10$, ten distinct elements do not exist, which is the official solution's remark that the statement "is not true if $|X| < 10$" when it is read that way.

Layer 2 THE HOW — the bifurcation

Bifurcating point. All three chains start from the same fact: on average, a random element hits more than half the sets. They split on the *strict* inequality. The probabilistic chain turns "$> \frac{1}{2}$" into "$< 2^{-10}$" multiplied by 1066, which throws the strictness away. The loose greedy chain writes "at least half" and throws it away too. The model chain keeps it: $r_k < r_{k-1}/2$, and because r_k is an integer this becomes $r_k \leq \lfloor (r_{k-1} - 1)/2 \rfloor$. That extra -1 at every step is what gets 1066 down to 0 in ten steps.

> *The random choice gives an expectation above* 1*, so it proves nothing here. Where can I get the extra slack from? Randomness averages over all ten choices at once. If I choose one element at a time, I can take the* best *element for the sets that are still uncovered, not an average one.*

> *"More than half" is a strict inequality between integers.* $r_k < r/2$ *means* $r_k \le (r-1)/2$*. I must carry that* -1 *through every step and not round it away, because* 1066 *is only a little above* 1024*.*

- ▷ **Cue:** "each set contains more than half the ground set", and a small number of points must hit every set $\longrightarrow$ **Move:** double count incidences to find a point in more than half of the sets; take it greedily; recurse on the sets it misses
- ▷ **Cue:** a union bound or expectation bound comes out just above 1 $\longrightarrow$ **Move:** switch from a random choice to a greedy (derandomised) choice, which uses the best point at each step instead of the average one
- ▷ **Cue:** a halving argument with a strict inequality on integers $\longrightarrow$ **Move:** write the recursion as $r' \le \lfloor (r-1)/2 \rfloor$ and compute the actual chain; do not approximate by $r/2^k$

Layer 3 THE WHY

Why is greedy the right tool, and not just a tool that happens to work? Look at what the hypothesis says. "$|A_i| > \frac{1}{2}|X|$" is a statement about each set separately, but it has a dual form, and the dual form is the one the conclusion needs. The conclusion is about *points* (ten of them) hitting *sets*. The hypothesis is about sets being large. Double counting is the standard way to move a size condition from one side of an incidence structure to the other: $\sum_A |A| = \sum_x n_x$. That identity turns "every set is big" into "some point is popular". It is not a trick. It is the only equation that relates the two sides of the problem.

Once you have "some point lies in more than half the remaining sets", the rest is forced. Choose that point. The sets it misses form a smaller instance of the same problem, with the same size condition, because a set that avoids x_1 is still a subset of X with more than $m/2$ elements. The hypothesis is inherited, so the argument can be repeated. Whenever a property survives passing to the leftover sets, you should expect an iteration.

Now the number. Why 1066? Partly because it is the year of the Battle of Hastings. Mostly because it lies between $2^{10} = 1024$ and $2^{11} - 2 = 2046$. Inverting $r' \le \lfloor (r-1)/2 \rfloor$ gives $r \le 2r' + 2$, so the largest family that ten greedy steps can clear is

$$0 \to 2 \to 6 \to 14 \to 30 \to 62 \to 126 \to 254 \to 510 \to 1022 \to 2046.$$

Any number up to 2046 would do. Any number above 1024 defeats both the naive expectation argument and the "at least half" greedy argument. The setters chose a number that punishes throwing away the strictness in the hypothesis. That is the point of the problem: the "$>$" in the statement has to be used, and integer rounding is how it gets used.

Why does greedy beat random when both start from the same average? The expectation argument charges each set its average miss probability. Greedy instead takes the point that is best for the sets still in play, and then updates. Choosing the best point instead of an average point, one step at a time, is exactly what derandomising means. So when an expectation bound misses by a constant factor, the first repair to try is the sequential version of the same count.

The official solution runs this same greedy argument, but it opens with the remark about $|X| < 10$ and gives the chain $532, 265, 132, \ldots$ without saying where the -1 comes from. It comes from $n_1 > 533$ forcing $n_1 \ge 534$, so that at most 532 sets are left: strictness plus integrality.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "$|A_i| > \frac{1}{2}|X|$" next to "ten elements". On seeing it, write the incidence double count $\sum_i |A_i| = \sum_x n_x$, and also note $2^{10} = 1024 < 1066$. That comparison warns you in advance that a plain halving argument will be one step short.

2. Double counting, then a greedy choice, then induction on the number of uncovered sets.
3. Hesitation comes when $1066/1024 > 1$ appears: the student decides the approach is wrong and drops it. What reinforces it away is noticing that the approach is only a constant factor off. A constant-factor miss is fixed by greed plus strictness, not by a new idea.
4. The trap is 1066 itself. It is set just above 2^{10} so that both the random argument and the "at least half" greedy argument leave one set uncovered. A secondary trap is worrying that the universe shrinks. It does not need to: count over all of X, since a chosen element lies in none of the remaining sets.
5. When a problem's constant sits just past a power of two, or just past the value the obvious bound needs, assume that a strict inequality or integer rounding is supposed to supply the difference. Find the ">" in the hypothesis and track it.
6. Reusable decision rule: *if an averaging or expectation argument for a covering problem misses by a constant, run it sequentially: choose the best point for what is still uncovered, recurse, and keep every strict inequality as an integer inequality ($r' < r/2 \Rightarrow r' \le \lfloor (r-1)/2 \rfloor$).*

1980 B5

1980 B5. For each $t \ge 0$, let S_t be the set of all nonnegative, increasing, convex, continuous, real-valued functions $f(x)$ defined on the closed interval $[0,1]$ for which

$$f(1) - 2f(2/3) + f(1/3) \ge t\,[f(2/3) - 2f(1/3) + f(0)].$$

Develop necessary and sufficient conditions on t for S_t to be closed under multiplication. (This closure means that, if the functions $f(x)$ and $g(x)$ are in S_t, so is their product $f(x)g(x)$. A function $f(x)$ is convex if and only if $f(su + (1-s)v) \le sf(u) + (1-s)f(v)$ whenever $0 \le s \le 1$.)

Fields: real analysis; inequalities; finite differences **Position:** B5 — late in the paper; the analysis is light, but the sufficiency argument needs a product rule for a discrete operator **Answer:** S_t is closed under multiplication exactly when $0 \le t \le 1$. Necessity: $x \in S_t$ for every t, but $x^2 \in S_t$ only for $t \le 1$. Sufficiency: a discrete Leibniz identity writes $E(fg)$ as $g(\frac{2}{3})E(f) + f(\frac{1}{3})E(g)$ plus a cross term, and convexity makes the cross term nonnegative when $t \le 1$.

Caveat. The OCR gives the convexity clause as "$0 < s < 1$" with the inequality sign garbled. The standard definition is used here. Including or excluding $s = 0, 1$ changes nothing.

Layer 1 THE WHAT — surface analysis

The natural attempt. Necessity comes quickly for anyone who tests the simplest members. Write $E_t(h) = [h(1) - 2h(\frac{2}{3}) + h(\frac{1}{3})] - t[h(\frac{2}{3}) - 2h(\frac{1}{3}) + h(0)]$. For $f(x) = x$ both brackets vanish, so $x \in S_t$ for all t. Then $x \cdot x = x^2$ gives $E_t(x^2) = \frac{2}{9} - \frac{2t}{9}$, so closure forces $t \le 1$.

Sufficiency is where the student gets stuck. The natural move is to write f and g through their values a_0, a_1, a_2, a_3 and $b_0, \ldots, b_3$ at $0, \frac{1}{3}, \frac{2}{3}, 1$. Then expand

$$E_t(fg) = a_3b_3 - 2a_2b_2 + a_1b_1 - t(a_2b_2 - 2a_1b_1 + a_0b_0)$$

and try to show it is ≥ 0 using $E_t(f) \ge 0$, $E_t(g) \ge 0$, monotonicity and convexity. That is eight unknowns, three families of constraints, and a quadratic target. The student tries $a_3b_3 \ge \ldots$, substitutes $a_3 \ge 2a_2 - a_1 + t(\ldots)$, and immediately has to multiply inequalities whose sides may not be nonnegative. After a page, the expression is not visibly signed, and it is not clear

which constraint to use next. The approach is not wrong, but without a plan for grouping terms it goes nowhere.

The model chain. Write $E = E_t$, which is linear in h. For $h \in \{f, g\}$ define the successive differences at the four nodes:

$$d_i = a_i - a_{i-1}, \qquad e_i = b_i - b_{i-1} \qquad (i = 1, 2, 3).$$

What the hypotheses give. f increasing: $d_i \geq 0$. f convex: slopes of chords increase, and the nodes are equally spaced, so $d_1 \leq d_2 \leq d_3$. Likewise $0 \leq e_1 \leq e_2 \leq e_3$. In these variables $E(f) = (d_3 - d_2) - t(d_2 - d_1)$.

Necessity. Shown above: $x \in S_t$ and $x^2 \notin S_t$ when $t > 1$.

Sufficiency, $0 \leq t \leq 1$. Let $f, g \in S_t$.

(i) fg is nonnegative, increasing and continuous. It is convex: for $u < v$, $s \in [0,1]$, $w = su + (1-s)v$, both $f(w) \leq sf(u) + (1-s)f(v)$ and $g(w) \leq sg(u) + (1-s)g(v)$ have nonnegative sides, so

$$f(w)g(w) \leq \big(sf(u) + (1-s)f(v)\big)\big(sg(u) + (1-s)g(v)\big) = sf(u)g(u) + (1-s)f(v)g(v) - s(1-s)\big(f(v) - f(u)$$

and the last term is ≥ 0 because both functions increase.

(ii) The identity (checked by expanding both sides):

$$E(fg) = g(\tfrac{2}{3})\,E(f) + f(\tfrac{1}{3})\,E(g) + R, \qquad R = (d_2 + d_3)\,e_3 - t\,d_1(e_1 + e_2). \tag{$*$}$$

In the original values, $R = [f(1) - f(\frac{1}{3})][g(1) - g(\frac{2}{3})] - t[f(\frac{1}{3}) - f(0)][g(\frac{2}{3}) - g(0)]$, which is the official solution's form.

(iii) Sign of each term. $g(\frac{2}{3}), f(\frac{1}{3}) \geq 0$ and $E(f), E(g) \geq 0$, so the first two terms are ≥ 0. For R: $d_2 + d_3 \geq 2d_1 \geq 0$ and $e_3 \geq \frac{1}{2}(e_1 + e_2) \geq 0$, so

$$(d_2 + d_3)e_3 \geq 2d_1 \cdot \tfrac{1}{2}(e_1 + e_2) = d_1(e_1 + e_2) \geq t\,d_1(e_1 + e_2),$$

the last step because $d_1(e_1 + e_2) \geq 0$ and $t \leq 1$. So $R \geq 0$, $E(fg) \geq 0$ and $fg \in S_t$.

Answer. S_t is closed under multiplication if and only if $0 \leq t \leq 1$. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains write $E(fg)$ in terms of values. The natural attempt then tries to bound the eight-variable quadratic directly. The model asks a structural question first: *how does E act on a product?* It subtracts off $g(\cdot)E(f) + f(\cdot)E(g)$, a discrete product rule, and looks at what is left, which it writes in differences. The split is at "expand" versus "subtract the Leibniz terms, then expand the remainder".

> *E is a linear operator and I'm applying it to a product. For derivatives that means the product rule: $(fg)'' = gf'' + fg'' + 2f'g'$. I'll imitate it. I'll subtract $g(point)E(f) + f(point)E(g)$ and see whether what is left is only products of first differences, since those I can sign.*

> *Convexity at equally spaced nodes says one thing only: the successive differences $d_1 \leq d_2 \leq d_3$ increase. So I should write everything in d_i and e_i and aim for a remainder where the positive part uses late differences and the negative part uses early ones.*

- ▷ **Cue:** a condition that is a linear combination of function values, and the question is whether it survives multiplication $\longrightarrow$ **Move:** find the discrete Leibniz rule: $E(fg) = g(\cdot)E(f) + f(\cdot)E(g) + (\text{bilinear in differences})$
- ▷ **Cue:** convex function sampled at equally spaced points $\longrightarrow$ **Move:** pass to successive differences; convexity is $d_1 \leq d_2 \leq \cdots$, monotonicity is $d_i \geq 0$

▷ **Cue:** "necessary and sufficient conditions on a parameter" for closure $\longrightarrow$ **Move:** test the simplest element sitting on the boundary ($E(x) = 0$) against its square; that gives the threshold, and the proof only needs to handle that threshold

Layer 3 THE WHY

Where does the identity $(*)$ come from? The official solution states it in one line and does not say how it was found. Here is where it comes from.

Start with the continuous version. If E were h'', the Leibniz rule $(fg)'' = gf'' + fg'' + 2f'g'$ would settle the matter: $f'', g'' \geq 0$, $f, g \geq 0$, and $f'g' \geq 0$ because both increase. Our E is a combination of second differences, so we should expect a discrete version of the same rule: "value times E" twice, plus a term that multiplies a difference of f by a difference of g.

Why is the leftover bilinear in differences? Let $R = E(fg) - \lambda E(f) - \mu E(g)$ with $\lambda = g(\text{some node})$ and $\mu = f(\text{some node})$. If $g \equiv c$ is constant, then $E(fg) = cE(f)$, $\lambda = c$ and $E(g) = 0$ (both brackets of E annihilate constants), so $R = 0$. By symmetry $R = 0$ when f is constant. So R does not depend on the levels of f and g, only on their differences. As a bilinear form it must be a combination of products $d_i e_j$. That is structural, not luck, and it holds whichever nodes are chosen for λ and μ.

Why these nodes? Among all choices, you want the one where every $d_i e_j$ appearing with a plus sign involves *later* differences than those with a minus sign. The only thing convexity supplies is that later differences are larger. With $\lambda = g(\frac{2}{3})$, $\mu = f(\frac{1}{3})$ the remainder is $(d_2 + d_3)e_3 - t\, d_1(e_1 + e_2)$. The positive part pairs late with latest, the negative part pairs earliest with early, and the comparison is two averaging inequalities. The middle nodes are natural anchors because E is centred on them. The first bracket of E is centred at $\frac{2}{3}$ and the second at $\frac{1}{3}$.

Why is the threshold $t = 1$ and not something else? At $t = 1$, $E_1(h) = h(1) - 3h(\frac{2}{3}) + 3h(\frac{1}{3}) - h(0)$ is the third finite difference. The condition becomes "$\Delta^3 h \geq 0$", which is the discrete version of $h''' \geq 0$. A third difference kills every polynomial of degree at most 2, so $E_1(x^2) = 0$. That means x^2 sits exactly on the boundary of S_1 and falls outside S_t for every $t > 1$. For $t \leq 1$, $E_t(h) = \Delta^3 h + (1 - t)\Delta^2 h(0) \geq \Delta^3 h$ for convex h, so the condition gets weaker as t decreases. The proof needs exactly one factor of $t \leq 1$, in the last inequality.

Every hypothesis is used exactly once: nonnegativity for the anchor coefficients, monotonicity for the signs of the differences and for $d_1(e_1 + e_2) \geq 0$, convexity for the ordering of the differences and for the convexity of fg, and continuity for the continuity of fg. When a problem carries four adjectives, check that your proof uses each one. A proof that doesn't should make you suspicious of it.

Layer $+\alpha$ BACKTESTING

1. The first cue is the pattern $h(1) - 2h(\frac{2}{3}) + h(\frac{1}{3})$: second differences on the grid $0, \frac{1}{3}, \frac{2}{3}, 1$. On seeing it, write $d_i = h(\frac{i}{3}) - h(\frac{i-1}{3})$ and note what convexity and monotonicity say about the d_i.

2. Test functions for necessity, and a discrete Leibniz identity with a signed remainder for sufficiency.

3. Hesitation comes right after writing $E(fg)$ out in the eight values: there is no obvious grouping. It is reinforced away by asking "what is the product rule for this operator?" and subtracting $g(\cdot)E(f) + f(\cdot)E(g)$ before expanding anything else.

4. The trap is multiplying inequalities whose sides are not known to be nonnegative, for example multiplying $E(f) \geq 0$ by $E(g) \geq 0$, which says nothing about $E(fg)$. A second trap is forgetting that closure also needs fg to be convex. That needs its own short argument, and it uses monotonicity as well as convexity.

5. Any time a class defined by a linear condition on sampled values must be closed under products, look for the discrete product rule, and write the leftover in successive differences, where convexity is an ordering.

6. Reusable decision rule: *to show a linear condition $E(\cdot) \geq 0$ is preserved by products, write $E(fg) = f(\cdot)E(g) + g(\cdot)E(f) + R$, where R vanishes when either factor is constant. R is then bilinear in differences, so choose the anchor points to make its sign follow from the ordering of those differences.*

1980 B6

1980 B6. An infinite array of rational numbers $G(d, n)$ is defined for integers d and n with $1 \leq d \leq n$ as follows:

$$G(1, n) = \frac{1}{n}, \qquad G(d, n) = \frac{d}{n} \sum_{i=d}^{n} G(d-1,\, i-1) \quad \text{for } d > 1.$$

For $1 < d \leq p$ and p prime, prove that $G(d, p)$ is expressible as a quotient s/t of integers s and t with t not an integral multiple of p. (For example, $G(3, 5) = 7/4$ with the denominator 4 not a multiple of 5.)

Fields: number theory; generating functions; combinatorics **Position:** B6 — the last problem; the recursion hides an explicit formula, and once that formula is found the arithmetic takes one line **Answer:** The generating function $F_d(x) = \sum_n G(d, n)x^n$ satisfies $F_d' = d\, F_{d-1} \cdot \frac{1}{1-x}$, so $F_d = \big(-\log(1-x)\big)^d$ and $G(d, n) = \sum_{k_1+\cdots+k_d=n,\, k_j \geq 1} \frac{1}{k_1 k_2 \cdots k_d}$. For $n = p$ and $d \geq 2$ every $k_j \leq p-1$, so no denominator contains p.

Caveat. The OCR renders both ranges with "<". The recursion uses $G(d-1, d-1)$, so the array must be defined for $d \leq n$. The official solution proves the result for $1 < d \leq p$. $G(p, p) = 1$, so the case $d = p$ holds in either reading. The statement above uses $\leq$ in both places.

Layer 1 THE WHAT — surface analysis

The natural attempt. The recursion has an explicit $\frac{1}{n}$ in front. With $n = p$ it reads

$$G(d, p) = \frac{d}{p} \sum_{j=d-1}^{p-1} G(d-1, j),$$

so the whole problem is to show that the sum $\Sigma_{d-1} = \sum_{j=d-1}^{p-1} G(d-1, j)$ has numerator divisible by p (in the sense of p-adic valuation). The strong student computes a table. Columns $n = 1, \ldots, 5$:

$$G(1, \cdot) = 1, \tfrac{1}{2}, \tfrac{1}{3}, \tfrac{1}{4}, \tfrac{1}{5}; \quad G(2, \cdot) = \cdot, 1, 1, \tfrac{11}{12}, \tfrac{5}{6}; \quad G(3, 5) = \tfrac{7}{4}.$$

Check $d = 2$: $\Sigma_1 = 1 + \frac{1}{2} + \cdots + \frac{1}{p-1}$, the harmonic number H_{p-1}. That its numerator is divisible by p is true: pair $\frac{1}{k} + \frac{1}{p-k} = \frac{p}{k(p-k)}$. So $G(2, p) = \frac{2}{p} H_{p-1}$ works.

Now $d = 3$: $\Sigma_2 = \sum_{j=2}^{p-1} G(2, j)$, and $G(2, j) = \frac{2}{j} H_{j-1}$. The student needs $\sum_j \frac{2}{j} H_{j-1} \equiv 0$, which is a statement about $\sum_{i<j<p} \frac{1}{ij}$. That is true too (it equals $\frac{1}{2}(H_{p-1}^2 - H_{p-1}^{(2)})$, both terms divisible by p), but the student now needs a new identity for every d. The induction hypothesis "$G(d-1, p)$ has no p in the denominator" says nothing about Σ_{d-1}, which is a sum over $n < p$. That is the wall: the statement is not inductive in d at fixed p.

The model chain. *Step 1: the generating function.* For $d \geq 1$ let $F_d(x) = \sum_{n \geq d} G(d, n)x^n$ as a formal power series. Put $L(x) = F_1(x) = \sum_{n \geq 1} \frac{x^n}{n}$, so $L'(x) = \sum_{n \geq 0} x^n = \frac{1}{1-x}$.

For $d > 1$ multiply the recursion by n: $nG(d,n) = d\sum_{j=d-1}^{n-1} G(d-1,j)$. The left side is the coefficient of x^{n-1} in F_d'. The sum is the coefficient of x^{n-1} in $F_{d-1}(x)\cdot\frac{1}{1-x}$, because multiplying by $\frac{1}{1-x}$ replaces each coefficient by a partial sum (and $G(d-1,j)=0$ for $j<d-1$ by convention). Both sides also vanish for $n<d$. Hence

$$F_d' = d\,F_{d-1}\,L', \qquad F_d(0)=0.$$

Step 2: $F_d = L^d$. True for $d=1$. If $F_{d-1}=L^{d-1}$, then $F_d' = dL^{d-1}L' = (L^d)'$, and F_d, L^d both have zero constant term, so $F_d = L^d$.

Step 3: the explicit formula. Reading off the coefficient of x^n in $L^d = \left(\sum_{k\geq 1} x^k/k\right)^d$:

$$G(d,n) = \sum_{\substack{k_1+\cdots+k_d=n\\ k_j\geq 1}} \frac{1}{k_1k_2\cdots k_d}.$$

Step 4: the arithmetic. Let $n=p$ prime and $2\leq d\leq p$. In each term every $k_j\geq 1$, so each $k_j = p-\sum_{i\neq j}k_i \leq p-(d-1)\leq p-1$. Each k_j lies in $\{1,\dots,p-1\}$, so $p\nmid k_1\cdots k_d$. A finite sum of fractions with denominators prime to p can be put over a common denominator $t=\prod$ (or lcm) of those denominators, and t is a product of integers prime to p, hence prime to p. So $G(d,p)=s/t$ with $p\nmid t$. □

Check. $d=3$, $p=5$: compositions of 5 into three parts are permutations of $(1,1,3)$ (three of them) and $(1,2,2)$ (three). So $G(3,5)=3\cdot\frac{1}{3}+3\cdot\frac{1}{4}=\frac{7}{4}$, as stated. The case $d=1$ is excluded, and it has to be: $G(1,p)=\frac{1}{p}$ is the one composition with a part equal to p.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start from the recursion at $n=p$. The natural attempt keeps p fixed and tries to show the partial sum Σ_{d-1} is divisible by p, using the recursion as given. The model does not specialise to p at all at first. It asks for a closed form of $G(d,n)$ for *all* n, encodes the n-direction in a power series, and sees that "$n\,G$" and "$\sum_{i\leq n}$" are, respectively, differentiation and multiplication by $\frac{1}{1-x}$.

> *The recursion has a factor $\frac{1}{n}$ and a running sum $\sum_{i\leq n}$. Multiplying by n is what differentiation does to x^n, and running sums are what $\frac{1}{1-x}$ does. So this is a differential equation for a generating function. Let me stop working at $n=p$ and find F_d.*

> *$F_1=\sum x^n/n$ has derivative $\frac{1}{1-x}$, the same factor that appears in the recursion. So $F_d' = dF_{d-1}F_1'$ is the chain rule for F_1^d. The answer is a d-th power, and the coefficients of a d-th power are sums over compositions.*

- ▷ **Cue:** a recursion of the form $n\,a_n = \sum_{i\leq n} b_i$ (a factor $1/n$ outside a running sum) ⟶ **Move:** pass to generating functions: $na_n \leftrightarrow A'$, running sum $\leftrightarrow B/(1-x)$
- ▷ **Cue:** $F_d' = d\,F_{d-1}\,F_1'$ with $F_d(0)=0$ ⟶ **Move:** recognise the chain rule: $F_d = F_1^d$
- ▷ **Cue:** "the denominator of this rational number is not divisible by p" ⟶ **Move:** find an expression as a sum of unit fractions and check that no denominator can contain p

Layer 3 THE WHY

Why must the thought "find the closed form for all n" come before any arithmetic? Because the property to be proved is *not* inherited along the recursion. p-integrality of $G(d,p)$ depends on the whole column $G(d-1,d-1),\dots,G(d-1,p-1)$ adding up to something divisible by p. That is a cancellation, and an inductive hypothesis about single entries cannot see cancellations. When the target property is a cancellation, you need a formula in which the cancellation has

already happened, or in which no cancellation is needed. The composition formula is the second kind: every term is individually p-integral, so nothing needs to cancel.

Why generating functions? Look at the two operations in the recursion. "Divide by n" and "sum over $i \le n$" are the discrete shadows of integration and of multiplying by $\frac{1}{1-x}$. When a recurrence uses operations that are simple on power series, you should translate it into power series. And $G(1,n) = \frac{1}{n}$ is itself the coefficient sequence of $-\log(1-x)$, whose derivative is $\frac{1}{1-x}$: the same factor the recursion multiplies by. Once you notice that the base case and the recursion use the same object, $F_d = F_1^d$ follows from the chain rule.

The official solution states "$F_d' = dF_{d-1}F_1' \ldots$ an induction gives $F_d = F_1^d$" and then reads off the coefficient. It does not say why the reader should think of F_d in the first place. The reason is the $\frac{d}{n}$ in front of the running sum. It also gives no reason why F_1^d helps. The reason is that the coefficient of x^p in a power of $\sum x^k/k$ with at least two factors never involves $k = p$.

A second view explains the number theory more deeply. Expanding $\frac{(-\log(1-x))^d}{d!} = \sum_n c(n,d)\frac{x^n}{n!}$ defines the unsigned Stirling numbers of the first kind, the number of permutations of n objects with d cycles. So $G(d,n) = d!\,c(n,d)/n!$, and I verified this numerically for $n \le 9$. With $n = p$ the claim becomes $p \mid c(p,d)$ for $1 < d < p$, which is the coefficient form of $x(x+1)\cdots(x+p-1) \equiv x^p - x \pmod p$ (Fermat plus Lagrange). The combinatorial reason is that the cyclic shift $i \mapsto i+1 \bmod p$ acts on permutations with d cycles, and when $1 < d < p$ none of them is fixed, so every orbit has size p. The contest does not require this view. It is the reason the result is true, and it links this problem to the family "p divides the middle coefficients" (binomial $\binom{p}{k}$, Stirling $c(p,d)$, Wolstenholme).

Layer $+\alpha$ BACKTESTING

1. The first cue is the $\frac{d}{n}\sum_{i=d}^{n}$ shape: a $\frac{1}{n}$ outside a running sum. On seeing it, write $F_d(x) = \sum_n G(d,n)x^n$ and translate: $nG(d,n) \to F_d'$, running sum $\to F_{d-1}/(1-x)$.
2. Generating functions, then a closed form via the chain rule, then reading off coefficients as sums over compositions.
3. Hesitation comes after $d = 2$ works by pairing $\frac{1}{k} + \frac{1}{p-k}$ and $d = 3$ needs a new identity. That is the signal that fixed-p induction is the wrong frame. It is reinforced away by asking "what is $G(d,n)$ for general n?" before touching p again.
4. The trap is the visible $\frac{1}{p}$ in the recursion. It invites a proof by cancellation, one d at a time. A second trap is the statement's range: $d = 1$ is excluded because $G(1,p) = 1/p$, and your proof must show exactly where $d \ge 2$ is used (every part $k_j \le p - d + 1 < p$).
5. When a divisibility or integrality claim resists induction because it depends on a cancellation across a whole row of the recursion, look for a closed form as a sum of terms that are each already integral in the required sense.
6. Reusable decision rule: *if a recursion multiplies by n or divides by n and takes running sums, translate it into a differential equation for a generating function. If the base case's series has derivative equal to the recursion's multiplier, the solution is a power of the base series, and its coefficients are sums over compositions.*