DAWN PUTNAM SESSION

Putnam 1975 — Papers A and B

Putnam 120 Archive

Contents

Putnam 1975 — Paper A **5**
1975 A1 5
1975 A2 7
1975 A3 9
1975 A4 12
1975 A5 15
1975 A6 18

Putnam 1975 — Paper B **23**
1975 B1 23
1975 B2 25
1975 B3 27
1975 B4 30
1975 B5 32
1975 B6 35

Putnam 1975 — Paper A

1975 A1

1975 A1. Supposing that an integer n is the sum of two triangular numbers, $n = \frac{a^2+a}{2} + \frac{b^2+b}{2}$, write $4n+1$ as the sum of two squares, $4n+1 = x^2+y^2$, and show how x and y can be expressed in terms of a and b. Show that, conversely, if $4n+1 = x^2+y^2$, then n is the sum of two triangular numbers. [Of course, a, b, x, y are understood to be integers.]

Fields: number theory, sums of two squares, quadratic forms **Position:** A1 (opener) **Answer:** $x = a+b+1$, $y = a-b$; conversely $a = \frac{x+y-1}{2}$, $b = \frac{x-y-1}{2}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The expression to expand is right there: $4n+1 = 4\left(\frac{a^2+a}{2} + \frac{b^2+b}{2}\right) + 1 = 2a^2+2a+2b^2+2b+1$. The phrase "sum of two triangular numbers" primes you to look at a and b as separate objects, and there is a genuinely clean fact about a single triangular number sitting one line away: $4 \cdot \frac{a^2+a}{2} + 1 = 2a^2+2a+1 = a^2+(a+1)^2$, a triangular number's own "$4T+1$" is a sum of two consecutive squares. Finding that in under a minute feels like real progress, and it is real progress — it just isn't the progress the problem asked for. The natural next move is to do the same thing to b and add:

$$4n+2 = \left[a^2+(a+1)^2\right] + \left[b^2+(b+1)^2\right].$$

Two things are wrong with this. First, it's $4n+2$, not $4n+1$ — the two "+1"s from the two separate completions have doubled up. Second, and more seriously, it's a sum of *four* squares, and the problem wants two. At this point you have two perfectly good square-completions in hand and no legal way to merge four squares into two without another identity you may not remember on the spot. Time gets spent trying to force $a^2+(a+1)^2+b^2+(b+1)^2$ into two squares by regrouping terms, with no guarantee the regrouping exists in this form at all.

The model chain. Do not complete the square on a and b separately. Keep the expansion whole and split it differently — into a pure quadratic part and a linear-plus-constant part:

$$4n+1 = \underbrace{2a^2+2b^2}_{\text{quadratic}} + \underbrace{2a+2b+1}_{\text{linear + constant}}.$$

The quadratic part is exactly twice a sum of two squares, and there is a standard identity for turning $2p^2+2q^2$ into a sum of two squares directly:

$$2p^2+2q^2 = (p+q)^2+(p-q)^2.$$

Apply it with $p=a$, $q=b$: $2a^2+2b^2 = (a+b)^2+(a-b)^2$. The remaining piece $2a+2b+1 = 2(a+b)+1$ now has an obvious home: fold it into the $(a+b)^2$ term by completing the square there, since $(a+b)^2+2(a+b)+1 = (a+b+1)^2$. Putting the two pieces back together,

$$4n+1 = (a+b+1)^2+(a-b)^2,$$

so $x = a + b + 1$ and $y = a - b$ — two squares, obtained with one identity applied once, no case-work and no accidental extra squares.

For the converse, don't attack $2a^2 + 2a + 2b^2 + 2b + 1 = x^2 + y^2$ as a fresh Diophantine equation in (a, b); invert the map just found. Since $4n + 1 = x^2 + y^2$ and squares are 0 or 1 $\pmod 4$, a sum of two squares that is $1 \pmod 4$ forces exactly one of x, y to be odd and the other even. So $x + y$ is odd, hence $x + y - 1$ and $x - y - 1$ are both even, and

$$a = \frac{x + y - 1}{2}, \qquad b = \frac{x - y - 1}{2}$$

are integers. These are exactly the algebraic inverse of $x = a + b + 1$, $y = a - b$ (check: $a + b + 1 = \frac{(x+y-1)+(x-y-1)}{2} + 1 = x$, and $a - b = \frac{(x+y-1)-(x-y-1)}{2} = y$). Running the forward identity on this (a, b) therefore reproduces $4\left(\frac{a^2+a}{2} + \frac{b^2+b}{2}\right) + 1 = x^2 + y^2 = 4n + 1$, which forces $\frac{a^2+a}{2} + \frac{b^2+b}{2} = n$: n is a sum of two triangular numbers.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start from the identical expansion $4n + 1 = 2a^2 + 2b^2 + 2a + 2b + 1$ — that much is forced algebra. They split on what to do with it: the natural attempt treats a and b as two separate triangular-number identities and completes the square on each one individually, producing a sum of four squares that needs (and lacks) a further merging step. The model treats $2a^2 + 2b^2$ as a single joint quadratic form in (a, b) and applies the two-squares identity $2p^2 + 2q^2 = (p + q)^2 + (p - q)^2$ before doing anything with the linear terms.

> *I have $2a^2 + 2b^2$ sitting together in one expression — that isn't two separate triangular-number identities waiting to happen, it's twice a sum of two squares, and there's a name for exactly that shape: $(a + b)^2 + (a - b)^2$.*

Layer 3 THE WHY

The identity $2p^2 + 2q^2 = (p + q)^2 + (p - q)^2$ is not an isolated algebra trick; it is the norm-multiplicativity of the Gaussian integers made visible. In $\mathbb{Z}[i]$, the norm $N(w) = w\bar{w} = |w|^2$ is multiplicative: $N(uv) = N(u)N(v)$. Take $u = 1 + i$ (so $N(u) = 2$) and $v = p + qi$ (so $N(v) = p^2 + q^2$). Then $uv = (1 + i)(p + qi) = (p - q) + (p + q)i$, and $N(uv) = (p - q)^2 + (p + q)^2$. Multiplicativity forces $N(uv) = N(u)N(v)$, i.e. $(p - q)^2 + (p + q)^2 = 2(p^2 + q^2)$ — exactly the identity used above. So the coefficient "2" in front of $2a^2 + 2b^2$ is not a nuisance to be divided away or absorbed piecemeal; it is the signature of multiplying a sum-of-two-squares by the specific Gaussian integer $1 + i$, and that multiplication is precisely the operation $(a, b) \mapsto (a + b, a - b)$ (up to the order and sign of the output, which don't matter once everything is squared).

This also explains why the natural attempt's route — complete the square on a alone, then on b alone — cannot land on two squares without extra work: each single completion is really working with the trivial norm-1 identity $p^2 + (\text{shift})^2$ on one variable in isolation, and nothing about that route ever multiplies the two triangular-number contributions together the way the Gaussian-integer product does. The problem's insistence on *exactly two* squares is the tell that the two variables must be combined multiplicatively (in the Gaussian-integer sense) rather than completed independently and then stacked.

The converse direction is the mod-4 fact about squares doing quiet but essential work: $x^2 \equiv 0 \pmod 4$ if x is even and $x^2 \equiv 1 \pmod 4$ if x is odd, so a sum of two squares can only be $\equiv 1 \pmod 4$ if one summand is even and the other odd. That parity fact is exactly what guarantees $x + y - 1$ and $x - y - 1$ are even, i.e. that the inverse formulas for a, b land on integers at all — without it, the converse would only produce rational triangular-number indices.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the coefficient 2 sitting in front of both squared terms once $4n+1$ is expanded: $2a^2+2b^2+2a+2b+1$. On seeing it, write down: "twice a sum of two squares — reach for $2p^2+2q^2=(p+q)^2+(p-q)^2$" before touching the linear terms.
2. That cue suggests completing the square exactly once, on the combined variable $a+b$, after first peeling $(a+b)^2+(a-b)^2$ off the pure quadratic part.
3. Hesitation sets in right after the single-variable identity $4T(a)+1=a^2+(a+1)^2$ is found and looks like it's already solved the problem. Reinforce against it by counting: the problem wants exactly two squares, and the running total after treating a and b separately is four — a mismatch that should stop you before you spend time trying to merge them.
4. The trap is that "sum of two triangular numbers" primes you to keep a and b conceptually separate, when the actual required identity is a joint one in (a,b) together, reached by treating $2a^2+2b^2$ as a single quadratic form rather than as two one-variable triangular-number facts stacked side by side.
5. Recognise this faster next time by treating any expression of the shape $2p^2+2q^2$ (or more generally a coefficient matching a known norm) as a direct cue for the corresponding norm identity, rather than as raw material for completing the square variable-by-variable.
6. Decision rule: to express a quadratic expression as a sum of exactly two squares, first isolate its pure quadratic part and test it against standard two-squares identities (starting with $2p^2+2q^2=(p+q)^2+(p-q)^2$, itself an instance of Gaussian-integer norm multiplicativity) before folding in any linear or constant terms by completing the square.

1975 A2

1975 A2. For which ordered pairs of real numbers b, c do both roots of the quadratic equation $z^2+bz+c=0$ lie inside the unit disk $\{|z|<1\}$ in the complex plane? Draw a reasonably accurate picture (i.e., 'graph') of the region in the real bc-plane for which the above condition holds. Identify precisely the boundary curves of this region.

Fields: complex numbers, quadratics, root location **Position:** A2 **Answer:** The open triangle with vertices $(0,-1)$, $(2,1)$, $(-2,1)$ in the bc-plane, bounded by $L_1: c=1$, $L_2: c-b+1=0$, $L_3: c+b+1=0$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The equation is right there and it's quadratic, so the reflex is to just solve it: $z=\dfrac{-b\pm\sqrt{b^2-4c}}{2}$. Now impose $|z|<1$ on both roots. If $b^2-4c<0$ the roots are a complex-conjugate pair $z=\frac{-b}{2}\pm i\frac{\sqrt{4c-b^2}}{2}$, and $|z|^2=\frac{b^2}{4}+\frac{4c-b^2}{4}=c$, so this case collapses to $c<1$ — clean enough. The real case is where the trouble starts. With $b^2-4c\geq 0$, you need

$$-1<\frac{-b-\sqrt{b^2-4c}}{2} \quad \text{and} \quad \frac{-b+\sqrt{b^2-4c}}{2}<1$$

simultaneously (plus the mirrored lower bounds). Isolating the square root in each inequality, checking the sign of what's left before you're allowed to square it, and then squaring and simplifying — twice, once per inequality, with the discriminant condition $b^2\geq 4c$ tracked alongside throughout — is doable, but it is slow, it is easy to drop a sign when isolating $\sqrt{b^2-4c}$ on one side, and at the end you still have to reconcile whatever region this produces with the separate $c<1$ region from the complex case to get one final picture. The arithmetic is not hard in any single step; there is just a lot of it, and every step invites an error that costs the whole problem.

The model chain. Don't solve for the roots at all. Let r, s be the (possibly complex, possibly equal) roots, so $b = -(r+s)$ and $c = rs$ — symmetric functions, known the moment the equation is written down, no square root required. Split on whether the roots are real or a conjugate pair, but handle each case through the polynomial's *values*, not its roots.

Complex case. If the roots are non-real they are conjugates, $s = \bar{r}$, so $c = rs = r\bar{r} = |r|^2$. The condition $|r| < 1$ (equivalently $|s| < 1$) is then just $c < 1$, together with $b^2 < 4c$ for the roots to be non-real at all.

Real case. Write $f(z) = z^2 + bz + c$, an upward-opening parabola since the leading coefficient is $1 > 0$. An upward parabola is negative only strictly between its two real roots (when they exist) and non-negative everywhere else. So both real roots lie in $(-1, 1)$ exactly when: (i) $f(1) > 0$ and $f(-1) > 0$ (so ± 1 are outside the interval between the roots, hence not "caught" by the dip), (ii) the vertex $z = -\frac{b}{2}$ lies in $(-1, 1)$, i.e. $|b| < 2$ (so the dip itself is centred inside $(-1, 1)$, not straddling one edge), and (iii) $b^2 \geq 4c$ (roots are real at all). Compute

$$f(1) = 1 + b + c > 0 \iff c + b + 1 > 0, \qquad f(-1) = 1 - b + c > 0 \iff c - b + 1 > 0.$$

These are exactly the lines $L_3 : c + b + 1 = 0$ and $L_2 : c - b + 1 = 0$. Condition (ii), $|b| < 2$, turns out to be implied by (i) together with $b^2 \geq 4c$ once the two half-plane conditions from L_2, L_3 are imposed (the two lines meet the parabola boundary $b^2 = 4c$ precisely at $b = \pm 2$), so no separate bookkeeping for the vertex is needed.

Now glue the two cases together. In the complex case the boundary is $c = 1$ (call it L_1) together with $b^2 = 4c$; in the real case the boundary is L_2, L_3 together with $b^2 = 4c$. The curve $b^2 = 4c$ is exactly where the two cases meet, and along it both derivations must agree — which they do, since crossing it just swaps "conjugate pair at $|r|^2 = c$" for "double real root at $z = -b/2$," continuously. Intersecting the three lines: $L_2 \cap L_3$ gives $(0, -1)$, $L_1 \cap L_2$ gives $(2, 1)$, $L_1 \cap L_3$ gives $(-2, 1)$. The full region — real case union complex case — is exactly the open triangle with these three vertices, bounded above by $L_1 : c = 1$ and below by L_2, L_3.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains begin by writing $b = -(r+s)$, $c = rs$ for the roots r, s — that much is forced notation, not a choice. They split on what to do next: the natural attempt solves the quadratic formula explicitly for r, s and bounds $\left|\frac{-b \pm \sqrt{b^2 - 4c}}{2}\right| < 1$ directly with the square root exposed. The model instead reasons about $f(1), f(-1)$, the vertex, and $c = |r|^2$ — properties of the polynomial and its roots' symmetric functions — without ever isolating a square root.

> *I don't need the roots themselves — I need to know where they sit relative to ± 1, and a parabola's shape tells me that from three numbers, $f(1)$, $f(-1)$, and its vertex, without ever solving for a root.*

Layer 3 THE WHY

The whole real-case argument rests on one fact treated as an axiom: an upward-opening parabola $f(z) = z^2 + bz + c$ is negative if and only if z is strictly between its two real roots (when they exist), and non-negative everywhere else, because f has no other sign changes — it's degree two, so it crosses zero at most twice. Once that is accepted, "both roots lie in $(-1, 1)$" translates directly into "f is non-negative at both endpoints of $(-1, 1)$, and its one dip below zero is centred inside that interval" — which is precisely conditions (i)-(iii) above. This is a general-purpose technique, not specific to this problem: any time a question asks whether all the roots of a real polynomial lie inside an interval or region, the values of the polynomial at the boundary of that region, plus symmetric functions of the roots (sum, product, discriminant), answer the question without ever solving for a root. Solving explicitly is the generic, syllabus-taught tool for "find the roots"; it is the wrong tool for "where are the roots,"

because location questions are answered by sign information, and sign information about f at specific points is available for free, directly from the coefficients, with no square root anywhere.

The complex case makes the same point even more starkly: $c = rs = r\bar{r} = |r|^2$ is an immediate consequence of s being the conjugate of r, so the modulus condition $|r| < 1$ becomes the linear condition $c < 1$ with no algebra at all — computing r explicitly and then taking its modulus would only rediscover a fact that was already sitting in the constant term c.

Why does the problem hand you *both* a real-root case and a complex-root case, and expect one unified picture? Because the discriminant $b^2 - 4c$ is itself a continuous function of (b, c), and the boundary curve $b^2 = 4c$ is where the polynomial's root structure changes type without discontinuity — a double real root at that boundary is the limiting case of both a conjugate pair collapsing together and two real roots merging. The three lines found from the two separate cases are required to fit together into a single closed region precisely because the underlying root-location condition, $|r| < 1$ and $|s| < 1$, is itself continuous in (b, c) across that boundary.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the phrase itself: "do both roots... lie inside" — a question about *location*, not about the roots' values. On seeing it, write down: do I need the roots explicitly, or can I answer this from f at specific points plus symmetric functions of the roots?

2. That cue suggests testing the sign of f at the boundary of the target region (here $f(\pm 1)$) together with the vertex location for the real-root case, and using $c = rs = |r|^2$ directly for the complex-conjugate case.

3. Hesitation sets in exactly at the boundary between the two cases, where the discriminant $b^2 - 4c$ changes sign — it's tempting to treat "real roots" and "complex roots" as two unrelated sub-problems with two unrelated final answers. Reinforce against this by checking that the boundary curves derived from each case must meet continuously along $b^2 = 4c$, and confirming they do before trusting the combined picture.

4. The trap is the quadratic formula itself: it is always available and always technically correct, so it's tempting to reach for it on any question about a quadratic's roots, even a pure location question where solving is the expensive route and the coefficients already encode everything needed.

5. Recognise this faster next time by treating "are the roots of this quadratic inside some region" as its own category of question, separate from "find the roots," triggered by any phrasing like "lie in," "inside," or "satisfy" applied to roots rather than "solve for" or "determine" the roots themselves.

6. Decision rule: for a location question about the roots of a real polynomial, work with the polynomial's values at the boundary of the target region and with symmetric functions of the roots (sum, product, discriminant) before ever solving for the roots explicitly; explicit solving is rarely necessary and is almost always more error-prone than the sign-and-symmetric-function route.

1975 A3

1975 A3. Let a, b, c be constants with $0 < a < b < c$. At what points of the set $\{x^b + y^b + z^b = 1,\ x \geq 0,\ y \geq 0,\ z \geq 0\}$ in three-dimensional space $\mathbb{R}^3$ does the function $f(x, y, z) = x^a + y^b + z^c$ assume its maximum and minimum values?

Caveat. The exponents in this problem are illegible in the OCR scan: the constraint prints as "$x \bullet + y \bullet + z \bullet = 1$" and the function as "$f(x, y, z) = x \bullet + y \bullet + z <$", with the exponent characters lost or corrupted. The statement above is reconstructed from the

book's own official solution, which defines $h(x) = x^a - x^b$ and $k(z) = z^c - z^b$ and asserts that f reduces to a function of these on the domain projected to the xz-plane — a reduction that is only algebraically valid if the constraint uses the same exponent b for all three variables and f keeps exponent b specifically on y (so that y^b cancels directly against the constraint). This reconstruction is confirmed independently: the solution's stated critical points $x_0 = (a/b)^{1/(b-a)}$ and $z_0 = (b/c)^{1/(c-b)}$, and its claim that h increases from 0 then decreases back to 0 on $[0, 1]$ using "the hypothesis $0 < a < b$" while k decreases from 0 to a minimum then increases back to 0 using $b < c$, are all exactly reproduced by solving the reconstructed statement below from scratch (see Layer 1). No part of the reconstruction was needed to explain the given numerical/algebraic content of the solution; it was needed only to recover exponents that the OCR dropped.

Fields: multivariable calculus, constrained optimisation, Lagrange multipliers **Position:** A3 **Answer:** Maximum at $\left(x_0,\ (1 - x_0^b)^{1/b},\ 0\right)$ with $x_0 = (a/b)^{1/(b-a)}$; minimum at $\left(0,\ (1 - z_0^b)^{1/b},\ z_0\right)$ with $z_0 = (b/c)^{1/(c-b)}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. This is a constrained optimisation of a function of three variables over the surface $g(x, y, z) = x^b + y^b + z^b = 1$ intersected with the first octant, so the reflex is Lagrange multipliers: solve $\nabla f = \lambda \nabla g$, i.e.

$$ax^{a-1} = \lambda b x^{b-1}, \qquad by^{b-1} = \lambda b y^{b-1}, \qquad cz^{c-1} = \lambda b z^{b-1}.$$

The middle equation looks the most promising to start with, since it's the cleanest: for $y \neq 0$ it forces $\lambda = 1$. Substituting into the other two, for $x \neq 0$: $x^{b-a} = a/b$, giving $x_0 = (a/b)^{1/(b-a)}$; for $z \neq 0$: $z^{c-b} = b/c$, giving $z_0 = (b/c)^{1/(c-b)}$. So far this looks like it's converging on a single interior stationary point (x_0, y, z_0) with y fixed by the constraint.

But the domain $\{x^b + y^b + z^b = 1,\ x, y, z \geq 0\}$ is a curved triangular patch with three straight edges and three corners — it is compact, but it is *not* smooth at those edges and corners, and Lagrange's condition $\nabla f = \lambda \nabla g$ only characterises stationary points in the smooth interior. So the single interior candidate just found is only one of several: you also have to separately check what happens on each of the three coordinate faces ($x = 0$, $y = 0$, $z = 0$ individually, each reducing to a 2-variable problem), on each of the three edges where two coordinates vanish simultaneously (three 1-variable problems), and at the corner where all three vanish (impossible here, since it violates the constraint, but that still has to be checked). That is potentially eight separate sub-problems, each needing its own justification for why it is or is not a genuine extremum, before you can honestly say you've found the global maximum and minimum. The single interior point found above is real and useful, but treating it as the whole answer, or grinding through all eight boundary cases by brute force, is a lot of case-work for a problem that asks a comparatively simple question.

The model chain. Look again at the exponent on y: it is b in f, and it is b for every variable in the constraint. That match is not incidental — it means y^b can be eliminated by direct substitution, with no differentiation and no multiplier at all. From the constraint, $y^b = 1 - x^b - z^b$, so

$$f(x, y, z) = x^a + y^b + z^c = x^a + \left(1 - x^b - z^b\right) + z^c = 1 + \underbrace{\left(x^a - x^b\right)}_{h(x)} + \underbrace{\left(z^c - z^b\right)}_{k(z)}.$$

The three-variable constrained problem collapses to a two-variable unconstrained one, and it is separable: $f = 1 + h(x) + k(z)$ with h a function of x alone and k a function of z alone, on the domain $x, z \geq 0$, $x^b + z^b \leq 1$ (the last inequality is exactly what makes $y = (1 - x^b - z^b)^{1/b}$ real and in $[0, 1]$).

Analyse $h(x) = x^a - x^b$ on $[0,1]$. $h(0) = 0$, $h(1) = 0$. For $x > 0$, $h'(x) = ax^{a-1} - bx^{b-1} = x^{a-1}(a - bx^{b-a})$, which is zero exactly at $x_0 = (a/b)^{1/(b-a)} \in (0,1)$ (using $0 < a < b$, so $a/b \in (0,1)$ and $b - a > 0$), positive for $x < x_0$ and negative for $x > x_0$. So h increases from 0 to a maximum $h(x_0) > 0$ at x_0, then decreases back to 0 at $x = 1$: $h \geq 0$ on $[0,1]$, with $h = 0$ only at the two endpoints.

Analyse $k(z) = z^c - z^b$ on $[0,1]$. By the same computation with roles reversed ($b < c$ now), $k'(z) = 0$ exactly at $z_0 = (b/c)^{1/(c-b)} \in (0,1)$, negative for $z < z_0$ and positive for $z > z_0$. So k decreases from 0 to a minimum $k(z_0) < 0$ at z_0, then increases back to 0 at $z = 1$: $k \leq 0$ on $[0,1]$, with $k = 0$ only at the two endpoints.

Maximise f. Since $h \leq h(x_0)$ and $k \leq 0$ everywhere on the domain, $f = 1 + h + k \leq 1 + h(x_0)$ always. This upper bound is attained exactly when $x = x_0$ and $k(z) = 0$, i.e. $z = 0$ or $z = 1$; but $x = x_0$ forces $z^b \leq 1 - x_0^b < 1$, so $z = 1$ is infeasible and only $z = 0$ works. Hence the bound $1 + h(x_0)$ is genuinely achieved, at $(x, z) = (x_0, 0)$, $y = (1 - x_0^b)^{1/b}$ — this is the global maximum.

Minimise f. Since $h \geq 0$ and $k \geq k(z_0)$ everywhere, $f \geq 1 + k(z_0)$ always. This lower bound is attained when $k(z) = k(z_0)$, i.e. $z = z_0$, and $h(x) = 0$, i.e. $x = 0$ or $x = 1$; here $z = z_0$ forces $x^b \leq 1 - z_0^b < 1$, so $x = 1$ is infeasible and only $x = 0$ works. Hence the bound is achieved at $(x, z) = (0, z_0)$, $y = (1 - z_0^b)^{1/b}$ — the global minimum. No boundary case-work beyond checking these two feasibility conditions was needed.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree up to writing the Lagrangian system $\nabla f = \lambda \nabla g$ and reading off the y-equation, $by^{b-1} = \lambda by^{b-1}$, which forces $\lambda = 1$ whenever $y \neq 0$. They split immediately after that: the natural attempt treats this as one solved sub-equation among three, and carries the multiplier framework forward into the x- and z-equations and then into the full boundary case-work the framework demands. The model instead reads the y-equation as telling it something about the *problem's structure* — that y's exponent in f matches the constraint's exponent for y — and abandons Lagrange multipliers entirely in favour of direct substitution.

> *The exponent on y in f is exactly the exponent the constraint uses for every variable — that's not a coincidence to route around with a multiplier, it's an invitation to substitute $y^b = 1 - x^b - z^b$ directly and collapse three coupled variables into two independent, single-variable calculus problems.*

Layer 3 THE WHY

Lagrange multipliers are the generic tool for constrained optimisation precisely because they work without knowing anything about the specific shape of f and g beyond differentiability — that generality is also their cost here. This problem is not generic: f and g share a term, y^b, with an identical exponent, which means $g = 1$ can be substituted into f to eliminate a whole variable algebraically, exactly, with no approximation and no multiplier. Once that elimination is made, the resulting function $f = 1 + h(x) + k(z)$ is separable — a sum of a function of x alone and a function of z alone — and separable functions on a domain like $\{x, z \geq 0,\ x^b + z^b \leq 1\}$ can be optimised by bounding each summand by its own unconstrained maximum or minimum and then checking, as a single feasibility question, whether those two individual bounds can be hit simultaneously. That is a genuinely smaller task than the Lagrange route's boundary enumeration: instead of separately justifying stationarity or extremality on eight sub-cases of a non-smooth domain, there are exactly two single-variable sign charts (for h and k on $[0,1]$) and one feasibility check per bound.

The reason the boundary case-work disappears rather than merely hiding is worth stating explicitly: h and k are each built to vanish at *both* ends of $[0,1]$ — $h(0) = h(1) = 0$, $k(0) = k(1) = 0$ — because those endpoints are exactly the domain's corners in disguise. $x = 1$ (with $z = 0$) is the corner of the constraint surface where $y = z = 0$; $x = 0, z = 1$ is the corner where

$x = y = 0$; $x = 0, z = 0$ is the corner where $x = z = 0$, $y = 1$. The single-variable functions h, k already encode the values of f at every one of these boundary configurations as their own endpoint values, so checking "h and k on $[0, 1]$" silently is checking the whole boundary of the original domain — the reduction doesn't dodge the boundary, it re-expresses it as something two short sign charts can settle.

The ordering $0 < a < b < c$ given in the problem is not decorative either: $a < b$ is exactly what is needed for $h(x) = x^a - x^b$ to be positive on $(0, 1)$ (a smaller exponent dominates a larger one for a base strictly between 0 and 1), and $b < c$ is exactly what is needed for $k(z) = z^c - z^b$ to be negative on $(0, 1)$ for the same reason with the sign of the exponent gap reversed. The problem's ordering constant is the mechanism, stated up front, that guarantees $h \geq 0$ and $k \leq 0$ throughout — which is exactly what makes "maximise h, maximise k" and "minimise h, minimise k" two genuinely different, non-conflicting optimisation targets rather than the same one.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the exponent match: y carries exponent b in f, and every variable carries exponent b in the constraint. On seeing it, write down: "y^b appears identically in both — substitute the constraint into f to eliminate y, before setting up any multiplier."
2. That cue suggests direct substitution ($y^b = 1 - x^b - z^b$) to reduce a 3-variable constrained problem to a 2-variable unconstrained, separable one — turning it into two single-variable calculus problems.
3. Hesitation sets in right after the Lagrangian y-equation gives $\lambda = 1$ for free — it's tempting to treat that as a small win and keep grinding through the x- and z-equations under the multiplier framework. Reinforce against this by asking: "did that equation just reveal something about the whole problem's structure, not only about y?"
4. The trap is that "at what points does the function assume its maximum and minimum" is exactly the phrasing that triggers a Lagrange-multiplier reflex on a genuinely three-variable constrained problem — the generic, syllabus-taught tool — while the shared exponent quietly turns this into something much smaller that the generic tool would only reach after a long detour through boundary case-work.
5. Recognise this faster next time by checking, the moment a constrained optimisation is set up, whether the objective function and the constraint function share a term in one variable with a matching exponent or coefficient — if so, try eliminating that variable by direct substitution before reaching for multipliers.
6. Decision rule: before invoking Lagrange multipliers, check whether the constraint can be solved for one variable and substituted directly into the objective; if the objective and the constraint share an identical term in that variable, the substitution is exact, collapses the problem's dimension for free, and — if the result separates into single-variable pieces — replaces the constrained optimisation with independent sign-chart analyses plus one feasibility check.

1975 A4

1975 A4. Let $n = 2m$, where m is an odd integer greater than 1. Let $\theta = e^{2\pi i/n}$. Express $(1 - \theta)^{-1}$ explicitly as a polynomial in θ,

$$a_k\theta^k + a_{k-1}\theta^{k-1} + \cdots + a_1\theta + a_0,$$

with integer coefficients a_i.

[Note that θ is a primitive n-th root of unity, and thus it satisfies all of the identities which hold for such roots.]

Caveat. The MAA OCR drops the displayed polynomial $a_k\theta^k + \cdots + a_1\theta + a_0$; it is restored here from the sentence "with integer coefficients a_i". Identity check: the book's solution under 1975 A-4 writes $n = 4k+2$, factors $\theta^n - 1$, and ends with $(1-\theta)^{-1} = \theta + \theta^3 + \cdots + \theta^{2k-1}$, so the label and statement match. Both answers below were also checked numerically for $m = 3, 5, 7, 9, 11, 15$.

Fields: algebra (roots of unity), polynomial identities **Position:** A4 — mid-paper; short once the right identity is chosen, a swamp if you go through cos and sin **Answer:** $(1-\theta)^{-1} = 1 + \theta^2 + \theta^4 + \cdots + \theta^{m-1}$, or equivalently $\theta + \theta^3 + \cdots + \theta^{m-2}$. The representation is not unique: any integer polynomial that agrees with one of these modulo the minimal polynomial of θ also works.

Layer 1 THE WHAT — surface analysis

The natural attempt. You have a complex number of modulus 1 and you are asked for its reciprocal after subtracting from 1. The A-level reflex is to rationalise: multiply top and bottom by the conjugate,

$$\frac{1}{1-\theta} = \frac{1-\bar{\theta}}{|1-\theta|^2} = \frac{1-\theta^{-1}}{2-2\cos(\pi/m)}.$$

The numerator is fine ($\theta^{-1} = \theta^{n-1}$ is a power of θ), but the denominator $2-2\cos(\pi/m)$ is an irrational real number for $m \geq 5$. That is a wall: no integer polynomial is going to come out of dividing by it.

The second reflex is the geometric series, $1/(1-\theta) = 1 + \theta + \theta^2 + \cdots$. It diverges, since $|\theta| = 1$. Truncate it using $\theta^n = 1$: $(1-\theta)(1+\theta+\cdots+\theta^{n-1}) = 1 - \theta^n = 0$. Now you have produced zero, not one, which is useless.

The stronger student reaches for $\theta^m = -1$ (because θ is a half-turn raised to the power $1/m$) and gets

$$(1-\theta)(1+\theta+\cdots+\theta^{m-1}) = 1-\theta^m = 2, \qquad \text{so} \qquad \frac{1}{1-\theta} = \tfrac{1}{2}\big(1+\theta+\theta^2+\cdots+\theta^{m-1}\big).$$

That is correct and a real step forward. But the coefficients are $\frac{1}{2}$, and the problem demands integers. Many students stop here and write it down.

The model chain. Write $m = 2k+1$ with $k \geq 1$. Two facts about $\theta = e^{2\pi i/n}$, $n = 2m$:

- $\theta^m = e^{i\pi} = -1$, and $\theta^{m+1} = -\theta$.
- $\theta \neq -1$, since $\theta = -1$ would need $n = 2$, and $m > 1$.

Set $E = 1 + \theta^2 + \theta^4 + \cdots + \theta^{2k}$ (the even powers up to θ^{m-1}). Then

$$(1-\theta)E = \sum_{j=0}^{k} \theta^{2j} - \sum_{j=0}^{k} \theta^{2j+1} = \sum_{i=0}^{m} (-\theta)^i,$$

because the even-index terms of the alternating sum are $+\theta^{2j}$ and the odd-index terms are $-\theta^{2j+1}$, with i running up to $2k+1 = m$. Sum the geometric progression with ratio $-\theta \neq 1$:

$$\sum_{i=0}^{m} (-\theta)^i = \frac{1-(-\theta)^{m+1}}{1+\theta} = \frac{1-\theta^{m+1}}{1+\theta} = \frac{1+\theta}{1+\theta} = 1,$$

using that $m+1$ is even and $\theta^{m+1} = -\theta$; division is legal since $\theta \neq -1$. Hence

$$(1-\theta)^{-1} = 1 + \theta^2 + \theta^4 + \cdots + \theta^{m-1}.$$

For the second form: $-\theta$ satisfies $(-\theta)^m = -\theta^m = 1$ and $-\theta \neq 1$, so $\sum_{i=0}^{m-1}(-\theta)^i = 0$, i.e. $E = \theta + \theta^3 + \cdots + \theta^{m-2}$. Check at $m = 3$: $\theta = e^{i\pi/3}$, $1 - \theta = e^{-i\pi/3}$, so $(1-\theta)^{-1} = \theta$, and indeed $1 + \theta^2 = \theta$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains reach $1 - \theta^m = 2$. The natural attempt divides by 2 and accepts half-integer coefficients. The model notices that the sign in $\theta^m = -1$ can be moved onto θ: working with $-\theta$ instead of θ turns the troublesome $1 - \theta^m = 2$ into $1 - (-\theta)^{m+1} = 1 + \theta$, which cancels the denominator exactly.

> *Half-integers mean I used the wrong identity, not that the answer has halves. A polynomial in θ is only defined up to relations θ satisfies, and I have not used the relation that makes the 2 disappear.*

> *m is odd and $n = 2m$. So $-\theta$ is an m-th root of unity. The parity condition is telling me to replace θ by $-\theta$ and sum a geometric series in $-\theta$.*

Layer 3 THE WHY

Why is the conjugate route doomed? Because it leaves the world the problem lives in. The problem asks for an element of $\mathbb{Z}[\theta]$, the integer polynomials in θ. Rationalising with $\bar{\theta}$ is legitimate — $\bar{\theta} = \theta^{n-1}$ is in that world — but the denominator $|1-\theta|^2$ is a real number that is not an integer, and a reciprocal of a non-unit is not in $\mathbb{Z}[\theta]$. The question is really: is $1 - \theta$ a *unit* of $\mathbb{Z}[\theta]$, and if so, what is its inverse? Every inverse you can write down with integer coefficients comes from an identity of the shape $(1-\theta) \cdot (\text{integer polynomial}) = 1$, and the only source of such identities is a geometric series, because the geometric series is the one place where $(1 - x)$ appears as a factor of something known.

So the whole problem is: which geometric series? $1 - \theta^n = 0$ is the obvious one and gives zero. $1 - \theta^m = 2$ is the next one and gives 2. The fact that it gives 2 is informative: it says $(1-\theta)$ divides 2, which is true for every primitive $2m$-th root. When n is a prime power, $1 - \theta$ is *not* a unit (for example $n = 4$: $1 - i$ has norm 2), and no integer answer exists. That is why the problem insists that m is odd and greater than 1: then n has two distinct prime factors, 2 and an odd prime, and in that case $1 - \theta$ is a unit. The phrase "m an odd integer greater than 1" is not decoration; it is the precise hypothesis that makes the answer exist.

With that in view, the identity to aim for is one whose right-hand side is $1 -$ (something that equals θ up to a The sign flip does it. Replacing θ by $-\theta$ is natural as soon as you read "$n = 2m$, m odd" as "θ is minus a primitive m-th root of unity". This is the same move as writing $e^{i\pi/3} = -e^{i4\pi/3}$: a primitive 6th root is minus a primitive cube root. Once you are working with $\omega = -\theta$, a primitive m-th root, $1 - \theta = 1 + \omega$, and for odd m

$$(1+\omega)(1-\omega+\omega^2-\cdots+\omega^{m-1}) = 1+\omega^m = 2,$$

while $1 + \omega + \cdots + \omega^{m-1} = 0$. Subtract and halve: $(1+\omega)^{-1} = 1 + \omega^2 + \omega^4 + \cdots + \omega^{m-1}$. The official solution does the same thing by factoring $\theta^{2k+1} + 1 = (\theta+1)(\theta^{2k} - \theta^{2k-1} + \cdots + 1)$; the factor $\theta^{2k} - \theta^{2k-1} + \cdots + 1 = 0$ is exactly $\sum(-\theta)^i = 0$. The book states that factorisation without saying where it comes from. It comes from asking what polynomial relation $-\theta$ satisfies.

The division by 2 in the natural attempt is therefore not wrong, only unfinished: the halves are removed by adding a multiple of a relation that equals zero. That is the general principle: in $\mathbb{Z}[\theta]$ an expression has many representations, and you choose the one with the property you want.

Layer $+\alpha$ BACKTESTING

1. The cue is "$n = 2m$, m odd" next to "$\theta = e^{2\pi i/n}$". Write down $\theta^m = -1$ and $(-\theta)^m = 1$ immediately.

2. Inverse of $1-\theta$ with integer coefficients means a geometric series identity $(1-x)(1+x+\cdots+x^{N-1}) = 1-x^N$ with $1-x^N$ a unit. Try $x=\theta$ and $x=-\theta$ and each useful N.

3. Hesitation comes at $\frac{1}{2}(1+\theta+\cdots+\theta^{m-1})$: it is correct, so it feels finished. Reinforce with: "integer coefficients were demanded; I have an unused relation ($\sum(-\theta)^i = 0$); use it to clear the 2".

4. The trap is rationalising with $\bar{\theta}$, which produces $2-2\cos(\pi/m)$ and drags you into trigonometry. A second trap is dividing by $1+\theta$ without checking $\theta \neq -1$.

5. When a problem hands you a parity condition on the order of a root of unity, rewrite the root as $\pm$ a root of smaller order. When a required representation has non-integer coefficients, look for a zero relation to add.

6. To invert $1-\zeta$ over the integers, find N and a sign with $1-(\pm\zeta)^N$ equal to a unit (or to $1-\zeta$ times a unit); the geometric series then gives the inverse. Representations in $\mathbb{Z}[\zeta]$ are defined modulo the relations ζ satisfies — choose the representative, do not accept the first one.

▷ **Cue:** "$n = 2m$ with m odd" and θ a primitive n-th root of unity $\longrightarrow$ **Move:** write $\theta = -\omega$ with ω a primitive m-th root; $\theta^m = -1$

▷ **Cue:** invert $1-\theta$ (or $1+\omega$) using integer coefficients $\longrightarrow$ **Move:** find a geometric series $(1-x)(1+\cdots+x^{N-1}) = 1-x^N$ whose right side is a unit

▷ **Cue:** an answer in powers of θ with coefficients $\frac{1}{2}$ where integers are required $\longrightarrow$ **Move:** add a multiple of a vanishing relation such as $1+\omega+\cdots+\omega^{m-1}=0$

▷ **Cue:** $1/(1-\theta)$ with $|\theta| = 1$ $\longrightarrow$ **Move:** do not rationalise with $\bar{\theta}$ unless the answer may involve cos; stay inside $\mathbb{Z}[\theta]$

▷ **Cue:** a hypothesis like "m odd and greater than 1" on the order of a root of unity $\longrightarrow$ **Move:** it is there to make an element a unit; find where parity enters the identity

1975 A5

1975 A5. On some interval I of the real line, let $y_1(x)$ and $y_2(x)$ be linearly independent solutions of the differential equation

$$y'' = f(x)y,$$

where $f(x)$ is a continuous real-valued function. Suppose that $y_1(x) > 0$ and $y_2(x) > 0$ on I. Show that there exists a positive constant c such that, on I, the function

$$z(x) = c\sqrt{y_1(x)y_2(x)}$$

satisfies the equation

$$z'' + \frac{1}{z^3} = f(x)z.$$

State clearly the manner in which c depends on $y_1(x)$ and $y_2(x)$.

Caveat. The MAA OCR drops the target equation after "satisfies the equation" and prints "$y_1(x) > 0$ and $y_1(x) > 0$". The equation $z'' + z^{-3} = f(x)z$ is restored from the last line of the book's own solution to 1975 A-5 ("$z'' + z^{-3} = fz$"), and the second positivity hypothesis is read as $y_2 > 0$, which the square root requires. The identity $z^3(z'' + z^{-3} - fz) = 1 - \frac{1}{4}c^4w^2$ used below was also checked symbolically.

Fields: differential equations (second-order linear ODEs, Wronskian) **Position:** A5 — a one-idea verification problem that punishes unplanned differentiation **Answer:** $c = \sqrt{2/|w|}$, where $w =$

$y_1y_2' - y_2y_1'$ is the Wronskian of y_1, y_2; it is a non-zero constant on I, so c depends on y_1, y_2 only through $|w|$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The instruction "show that z satisfies an equation" reads as "substitute and differentiate". So you set $u = y_1y_2$, $z = cu^{1/2}$, and differentiate twice:

$$z' = \frac{c\,u'}{2u^{1/2}}, \qquad z'' = c\Big(\frac{u''}{2u^{1/2}} - \frac{(u')^2}{4u^{3/2}}\Big).$$

Next $u' = y_1'y_2 + y_1y_2'$ and $u'' = y_1''y_2 + 2y_1'y_2' + y_1y_2'' = 2fy_1y_2 + 2y_1'y_2'$, using the equation for each y_i. Substituting,

$$z'' = fz + \frac{c}{4u^{3/2}}\Big(4y_1y_2y_1'y_2' - (y_1'y_2 + y_1y_2')^2\Big).$$

You have fz, which is good, and a leftover term that should be $-1/z^3 = -1/(c^3u^{3/2})$. The powers of u match. What remains is to show that the bracket is a negative constant — and here the natural attempt usually stalls. The bracket is a function of four quantities y_1, y_2, y_1', y_2' that all vary with x; there is no visible reason it should be constant, and the student either expands it into six terms and gets lost, or tries to differentiate it and uses the ODE again, which works but takes another half page with the c still undetermined.

The model chain. *Step 1: the Wronskian is a non-zero constant.* Let $w = y_1y_2' - y_2y_1'$. Then

$$w' = y_1y_2'' - y_2y_1'' = y_1(fy_2) - y_2(fy_1) = 0,$$

so w is constant on the interval I. It is non-zero: if $w = 0$, pick $x_0 \in I$; the vectors $(y_1(x_0), y_1'(x_0))$ and $(y_2(x_0), y_2'(x_0))$ are then linearly dependent, so some $(\alpha, \beta) \neq (0,0)$ has $\alpha y_1 + \beta y_2$ vanishing with its derivative at x_0. By uniqueness of solutions to $y'' = fy$ with given $y(x_0), y'(x_0)$ (f continuous), $\alpha y_1 + \beta y_2 \equiv 0$, contradicting linear independence.

Step 2: the computation. With $u = y_1y_2 > 0$ and $z = cu^{1/2}$, the natural attempt's calculation gives

$$z'' = fz + \frac{c}{4u^{3/2}}\Big(4y_1y_2y_1'y_2' - (y_1'y_2 + y_1y_2')^2\Big).$$

The bracket is $4AB - (A+B)^2 = -(A-B)^2$ with $A = y_1'y_2$, $B = y_1y_2'$, so it equals $-(y_1y_2' - y_2y_1')^2 = -w^2$. Hence

$$z'' - fz = -\frac{c\,w^2}{4u^{3/2}}, \qquad \text{while} \qquad -\frac{1}{z^3} = -\frac{1}{c^3u^{3/2}}.$$

These agree for all $x \in I$ exactly when $c\,w^2/4 = 1/c^3$, that is $c^4 = 4/w^2$. Since $w \neq 0$, the unique positive solution is

$$c = \sqrt{\frac{2}{|w|}} = \Big(\frac{2}{|y_1y_2' - y_2y_1'|}\Big)^{1/2}.$$

Equivalently $z^3\big(z'' + z^{-3} - fz\big) = 1 - \frac{1}{4}c^4w^2$ identically, and c is chosen to make the right side vanish.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains reach the bracket $4y_1y_2y_1'y_2' - (y_1'y_2 + y_1y_2')^2$. The natural attempt sees four varying functions; the model sees two quantities $A = y_1'y_2$ and $B = y_1y_2'$ appearing as $4AB - (A+B)^2$, recognises $-(A-B)^2$, and recognises $A - B$ as (minus) the Wronskian.

> *The answer has to be a constant because the target equation has a bare constant* 1 *in* $1/z^3$*, and* c *is supposed to be a constant. The only constant built out of two solutions of* $y'' = fy$ *that I know is the Wronskian. So this bracket must be a function of* $y_1y_2' - y_2y_1'$ *— let me look for that combination.*

> $4AB - (A+B)^2$*: that is minus a perfect square.* $A - B$ *is the Wronskian.*

Layer 3 THE WHY

Why should the Wronskian appear at all? Read the problem's last sentence: "state clearly the manner in which c depends on y_1 and y_2". That tells you c is a single number manufactured from the pair (y_1, y_2). A number, not a function. Now ask: what numbers can you manufacture from two solutions of $y'' = fy$ without knowing f? Values at a point depend on the point. The one quantity that is automatically independent of x is $w = y_1y_2' - y_2y_1'$, and it is constant for a structural reason: the equation has no y' term, so the "area" spanned by (y_1, y_1') and (y_2, y_2') is conserved (Abel's identity, with zero first-order coefficient). This is the same reason angular momentum is conserved under a central force: $w' = y_1y_2'' - y_2y_1''$ and both second derivatives are the same multiple f of the function. So before computing anything, the answer is predictable in form: c is a power of $|w|$ times a number. Dimensional bookkeeping even fixes the power: if you scale $y_1 \mapsto \lambda y_1$, then $w \mapsto \lambda w$ and $\sqrt{y_1y_2} \mapsto \sqrt{\lambda}\sqrt{y_1y_2}$; the equation $z'' + z^{-3} = fz$ is not scale-invariant, so z must be unchanged, which forces $c \propto |w|^{-1/2}$. That is a thirty-second check that tells you where the computation must land.

Why $4AB - (A+B)^2$? Because z is a square root of a product. Differentiating $\sqrt{u}$ twice always produces $2uu'' - (u')^2$ over $u^{3/2}$; with u a product y_1y_2 the term $(u')^2 = (y_1'y_2 + y_1y_2')^2$ is the square of a sum, and u'' contains the cross term $2y_1'y_2'$, which becomes $4AB$ after multiplying by $2u$. The identity $(A+B)^2 - 4AB = (A-B)^2$ is the discriminant identity from quadratics: the sum-and-product of two numbers determine their difference up to sign. You have met it every time you found $\alpha - \beta$ from $\alpha + \beta$ and $\alpha\beta$. Here $A + B = u'$ and $AB = y_1y_2y_1'y_2'$ are what differentiating the product hands you, and $A - B$ is what you want.

The book's solution avoids square roots by setting $c^2 = 2k$ and differentiating $z^2 = 2ky_1y_2$ instead of z. That choice is sensible housekeeping — z^2 is a polynomial in the y_i, so the product rule replaces the quotient rule — but it hides the same step: the line "$z^3z'' + k^2(y_1'y_2 - y_2y_1')^2 = fz^4$" is the discriminant identity applied silently. Here is where it comes from: $k^2(y_1'y_2 + y_1y_2')^2 - 4k^2y_1y_2y_1'y_2' = k^2(y_1'y_2 - y_1y_2')^2$.

For context: $z'' + z^{-3} = fz$ is the Ermakov–Pinney equation. It describes the amplitude of an oscillator with time-varying frequency, and this problem is its standard solution formula. You do not need the name to solve it; you need the habit of asking what is constant.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "linearly independent solutions of $y'' = f(x)y$" together with "positive constant c … depends on y_1 and y_2". Write down $w = y_1y_2' - y_2y_1'$ and $w' = 0$ before differentiating z.

2. Direct substitution, organised so that everything is expressed through $u = y_1y_2$, u', and w. Squaring first ($z^2 = c^2y_1y_2$) is an equally good way to organise it.

3. Hesitation occurs at the bracket $4y_1y_2y_1'y_2' - (y_1'y_2 + y_1y_2')^2$. It is reinforced away by expecting the Wronskian in advance: if you know the answer must be a constant built from y_1, y_2, you search for $y_1y_2' - y_2y_1'$ inside the bracket.

4. The trap is expanding the bracket into monomials, where the square is no longer visible. A second trap is forgetting to show $w \neq 0$; without it, c would not exist. A third is writing $c = \sqrt{2/w}$ with no absolute value: the sign of w depends on the order of y_1, y_2.

5. Whenever two solutions of the same second-order linear ODE with no first-derivative term appear together, the Wronskian is constant and is the first thing to write. Whenever an expression has the form $4AB - (A+B)^2$, rewrite it as $-(A-B)^2$.

6. Before verifying an identity by differentiation, predict what the leftover must be (here: a constant, hence a function of the conserved quantity), and use a scaling argument to predict the exponent. Then compute toward that target instead of expanding blind.

▷ **Cue:** two solutions y_1, y_2 of $y'' = f(x)y$ appear together $\longrightarrow$ **Move:** write $w = y_1y_2' - y_2y_1'$; $w' = 0$; $w \neq 0$ iff they are independent

▷ **Cue:** "state how the constant c depends on y_1 and y_2" $\longrightarrow$ **Move:** c is a function of the Wronskian, the only x-free quantity available

▷ **Cue:** an expression $4AB - (A+B)^2$ after differentiating a product $\longrightarrow$ **Move:** it is $-(A-B)^2$, the discriminant identity

▷ **Cue:** $z = c\sqrt{\text{product}}$ to be substituted into an ODE $\longrightarrow$ **Move:** differentiate z^2 instead, or write $z'' = c(2uu'' - u'^2)/(4u^{3/2})$ in one line

▷ **Cue:** a nonlinear equation not invariant under $y \mapsto \lambda y$ $\longrightarrow$ **Move:** use scaling to fix the power of the Wronskian before computing

1975 A6

1975 A6. Let P_1, P_2, P_3 be the vertices of an acute-angled triangle situated in three-dimensional space. Show that it is always possible to locate two additional points P_4 and P_5 in such a way that no three of the points are collinear and so that the line through any two of the five points is perpendicular to the plane determined by the other three.

In writing your answer, state clearly the locations at which you place the points P_4 and P_5.

Caveat. Statement read from the MAA OCR with subscripts restored (P_1, P_2, P_3, P_4, P_5). Identity check: the book's solution under 1975 A-6 places P_4, P_5 on the perpendicular to the plane at the orthocentre H with $\vec{HP}_4 \cdot \vec{HP}_5 = -\vec{HP}_1 \cdot \vec{HP}_2$, which solves this statement. The configuration below was verified numerically (all ten perpendicularity conditions) for an acute triangle and, as a check on the remark about the hypothesis, for an obtuse one.

Fields: geometry (3D, vectors, orthocentre) **Position:** A6 — the last problem of the morning paper, but its difficulty is organisational: ten conditions collapse to two once the right coordinates are chosen **Answer:** Let H be the orthocentre of $P_1P_2P_3$ and ℓ the line through H perpendicular to the plane. Put $v_i = \vec{HP}_i$; then $v_1 \cdot v_2 = v_2 \cdot v_3 = v_3 \cdot v_1 =: p < 0$. Place P_4, P_5 on ℓ, on the same side of the plane, at heights s, t with $st = -p$ and $s \neq t$ (for example $s = 1$, $t = -p$ unless $-p = 1$).

Layer 1 THE WHAT — surface analysis

The natural attempt. Five points, ten lines, ten planes. The natural first move is to count: $\binom{5}{2} = 10$ conditions, each "line $\perp$ plane", which is two scalar equations (the line direction is orthogonal to two spanning vectors of the plane). That is twenty equations for the six unknown coordinates of P_4 and P_5. You set P_1, P_2, P_3 in the plane $z = 0$ with general coordinates and $P_4 = (a, b, c)$, $P_5 = (d, e, f)$, write out a few conditions, and they are quadratic in the unknowns: for example "$P_1P_4 \perp$ plane $P_2P_3P_5$" contains $(P_4 - P_1) \cdot (P_5 - P_2) = 0$, which multiplies unknowns together. After three or four equations the student sees a system that is heavily overdetermined and nonlinear, with no way to tell in advance whether it is consistent. This is where most scripts either stop or guess a symmetric configuration (a regular simplex–like shape) that does not fit a general triangle.

The model chain. *Step 1: use the one linear condition first.* The line P_4P_5 must be perpendicular to the plane π of $P_1P_2P_3$. So P_4 and P_5 lie on a common normal line ℓ to π, meeting π at some point H. Take H as origin, π as $\{z = 0\}$, and write $P_i = v_i$ ($i = 1, 2, 3$) with v_i horizontal, $P_4 = (0, 0, s)$, $P_5 = (0, 0, t)$, $s, t \neq 0$, $s \neq t$.

Step 2: the three lines P_iP_j in the triangle. Let $\{i,j,k\} = \{1,2,3\}$. The plane $P_kP_4P_5$ contains the vertical axis and v_k (it is a plane provided $v_k \neq 0$). A horizontal direction is perpendicular to that plane iff it is orthogonal to v_k. So the condition is $(v_j - v_i) \cdot v_k = 0$ for each k: H lies on all three altitudes, i.e. H is the orthocentre. Equivalently

$$v_1 \cdot v_2 = v_2 \cdot v_3 = v_3 \cdot v_1 =: p.$$

Step 3: the six lines P_iP_4, P_iP_5. The plane $P_jP_kP_5$ is spanned by $v_k - v_j$ and $P_5 - P_j = (-v_j, t)$. The line P_iP_4 has direction $(-v_i, s)$. It is perpendicular to the plane iff

$$(-v_i) \cdot (v_k - v_j) = 0 \quad \text{and} \quad v_i \cdot v_j + st = 0.$$

The first holds by Step 2; the second says $st = -p$. The lines P_iP_5 give the same conditions with s, t swapped.

Step 4: the line P_4P_5 is vertical and π is horizontal: done in Step 1. That accounts for all $3+3+3+1 = 10$ lines.

Step 5: the sign of p and non-collinearity. In an acute triangle the orthocentre lies strictly inside, so no v_i is zero, and each angle $\angle P_iHP_j$ equals $\pi - \angle P_k > \pi/2$; hence $p = v_i \cdot v_j < 0$. (In fact $p = -4R^2 \cos P_1 \cos P_2 \cos P_3$ with R the circumradius.) So $st = -p > 0$: choose $s > 0$ with $s^2 \neq -p$ and $t = -p/s$. Then $s \neq t$, both are non-zero, and P_4, P_5 lie on the same side of π. No three points are collinear: P_1, P_2, P_3 form a triangle; a triple containing exactly one of P_4, P_5 has a point off π and two in π that are distinct; a triple P_i, P_4, P_5 would need P_i on ℓ, i.e. $v_i = 0$, which does not happen. Every plane used above is therefore a genuine plane, and the ten conditions hold.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains know the ten conditions. The natural attempt writes them all in general coordinates. The model uses the single condition that involves only P_4P_5 and the given triangle ("$P_4P_5 \perp$ plane $P_1P_2P_3$") to fix the geometry first — P_4, P_5 on a common normal — and chooses the foot H of that normal as the origin. After that, every condition is a dot product of the v_i with each other or a product st.

> *Ten conditions for six unknowns: I will not solve this system head-on. One condition, $P_4P_5 \perp$ the triangle's plane, uses only one of the new lines and the given plane. It tells me P_4, P_5 lie on one perpendicular. Put the origin where that perpendicular meets the plane.*

> *$(v_j - v_i) \cdot v_k = 0$ for all three k — that is "HP_k is perpendicular to the opposite side". H is the orthocentre.*

> *The last condition is $v_i \cdot v_j + st = 0$. So st is minus the common value of $v_i \cdot v_j$, and I only need its sign.*

Layer 3 THE WHY

Why start with P_4P_5? Because of the asymmetry the problem hands you. Three points are given and two are free; among the ten lines, exactly one (P_4P_5) joins two free points and is tested against the plane of the three given points. That condition does not involve any unknown plane, so it is linear, and it pins down a whole family (a common normal). Every other condition involves a plane through at least one unknown point. The general principle is: impose the conditions that involve the fewest unknowns first, and let them choose your coordinates. It is the three-dimensional version of "put the origin at the centre of the circle".

Why the orthocentre? Once P_4, P_5 sit on a vertical line through H, any plane containing P_4 and P_5 is vertical, containing the vertical line and one more point. A horizontal segment is perpendicular to a vertical plane exactly when it is perpendicular to the plane's horizontal trace. So "side $P_iP_j \perp$ plane $P_kP_4P_5$" becomes "side $P_iP_j \perp$ line HP_k". That is the definition of an altitude through P_k passing through H. The orthocentre is not guessed; it is what the words "line perpendicular to a plane containing the vertical" mean when projected onto the triangle's plane.

There is also a structural reason the answer exists, and it explains the whole problem. In the plane, the four points A, B, C and the orthocentre H form an *orthocentric system*: the line through any two is perpendicular to the line through the other two, and each of the four is the orthocentre of the other three. This problem is the three-dimensional analogue: five points such that the line through any two is perpendicular to the plane of the other three. Among the five, any four form a tetrahedron whose altitudes are concurrent, with the fifth point as their common point. For the tetrahedron $P_1P_2P_3P_4$, the altitude from P_4 is ℓ (it hits π at H), and P_5 lies on it. The condition $st = -p$ is where that altitude meets the others.

The book's solution uses the acute hypothesis to conclude $p < 0$, hence P_4, P_5 on the same half-line. Worth knowing: the construction does not actually need acuteness. For an obtuse triangle the orthocentre lies outside, $p > 0$, and $st = -p < 0$ puts P_4, P_5 on opposite sides of the plane; the ten conditions and non-collinearity still hold (the orthocentre is not a vertex, so no v_i is zero). What genuinely fails is a right triangle: there H is the right-angle vertex P_k, $v_k = 0$, and P_k, P_4, P_5 are collinear. The examiners chose "acute" to keep H inside and the picture tidy. Do not spend exam time trying to find where acuteness is essential; it is used for a sign, and the sign only decides which side of the plane P_5 goes.

Layer $+\alpha$ BACKTESTING

1. The first cue is "the line through any two of the five points is perpendicular to the plane determined by the other three", with three points given. Write down the one condition among the given points and the pair of new points: $P_4P_5 \perp \pi$. Then write: P_4, P_5 on a common normal through some $H \in \pi$.

2. Vectors from a well-chosen origin H. Every "line $\perp$ plane" becomes two dot products; almost all of them become dot products of the horizontal vectors v_i.

3. Hesitation comes at the start, facing ten conditions. It is reinforced away by the rule "use the condition with fewest unknowns to choose coordinates". A second hesitation is the sign of $v_i \cdot v_j$; it is removed by the angle fact $\angle P_iHP_j = \pi - \angle P_k$.

4. The trap is writing general coordinates for P_4, P_5 and producing a nonlinear overdetermined system. A subtler trap is forgetting to check non-collinearity and that the planes used are non-degenerate; that is where the right-triangle case fails.

5. When many symmetric conditions constrain a few new points, look for the one condition that does not involve an unknown plane or line; it fixes a locus. Perpendicular-to-plane conditions in 3D reduce, after projection, to perpendicular-to-line conditions in 2D.

6. Order constraints by how many unknowns they involve, satisfy the simplest first, and let it choose the origin and axes. In any problem of the form "every k-subset is perpendicular to its complement", expect an orthocentre, and expect the answer to be an orthocentric system.

▷ **Cue:** "the line through any two points is perpendicular to the plane of the other three", three points given $\longrightarrow$ **Move:** the pair of new points lies on one normal to the given plane; take its foot as origin

▷ **Cue:** a horizontal segment perpendicular to a vertical plane through the origin and a point v $\longrightarrow$ **Move:** equivalent to the segment being orthogonal to v

▷ **Cue:** $(v_j - v_i) \cdot v_k = 0$ for all three k $\longrightarrow$ **Move:** H is the orthocentre; $v_1 \cdot v_2 = v_2 \cdot v_3 = v_3 \cdot v_1$

▷ **Cue:** sign of $\vec{HA} \cdot \vec{HB}$ with H the orthocentre $\longrightarrow$ **Move:** $\angle AHB = \pi - C$; negative iff the triangle is acute (value $-4R^2 \cos A \cos B \cos C$)

▷ **Cue:** ten or more conditions for few unknowns $\longrightarrow$ **Move:** impose the one with fewest unknowns first and let it choose coordinates

Putnam 1975 — Paper B

1975 B1

1975 B1. In the additive group of ordered pairs of integers (m, n) (with addition defined componentwise: $(m, n) + (m', n') = (m + m', n + n')$), consider the subgroup H generated by the three elements $(3, 8)$, $(4, -1)$, $(5, 4)$. Then H has another set of generators of the form $(1, b)$, $(0, a)$ for some integers a, b with $a > 0$. Find a.

Fields: algebra (subgroups of $\mathbb{Z}^2$, lattices) **Position:** B1 – opening problem of the second session
Answer: $a = 7$ (and $b \equiv 5 \pmod 7$).

Layer 1 THE WHAT — surface analysis

The natural attempt. Three generators, and you are told the group has a two-generator form $(1, b), (0, a)$ – so the plan is to combine the three given vectors to kill the first coordinate. You look for integer multiples of $(3, 8)$ and $(4, -1)$ whose first coordinates cancel: $4(3, 8) - 3(4, -1) = (0, 35)$. You do the same with $(5, 4)$ and $(4, -1)$: $4(5, 4) - 5(4, -1) = (0, 21)$. Now $2(0, 21) - (0, 35) = (0, 7)$, so $(0, 7) \in H$. You check that $(1, 5) = (4, -1) - (3, 8) + 2(0, 7)$ is also in H, and you write down $a = 7$. The arithmetic is clean and the answer matches what is expected. But stop and ask what you have actually shown: you have exhibited one element $(0, 7)$ of H with zero first coordinate. You have not shown that 7 is the *smallest* positive such value – nothing yet rules out that some other combination of $(3, 8), (4, -1), (5, 4)$ produces $(0, 1)$ or $(0, 3)$, in which case the true a would be smaller and $(1, 5), (0, 7)$ would not generate H at all (they would generate a proper subgroup of it). You have proved existence, and the problem wants the exact value of a, which is a question about the whole set of combinations, not about one lucky one.

The model chain. Every element of H has the form $n_1(3, 8) + n_2(4, -1) + n_3(5, 4) = (3n_1 + 4n_2 + 5n_3,\ 8n_1 - n_2 + 4n_3)$ for integers n_1, n_2, n_3. The elements of H with first coordinate 0 are exactly those with $3n_1 + 4n_2 + 5n_3 = 0$, and their second coordinates $d = 8n_1 - n_2 + 4n_3$ form a subgroup of $\mathbb{Z}$, hence $d\mathbb{Z} = a\mathbb{Z}$ for a single well-defined $a \geq 0$; that a is the answer, and it is intrinsic to H, not to any particular combination.

To pin it down, look for a scalar c with $(8, -1, 4) \equiv c\,(3, 4, 5) \pmod a$ for the correct a: if such a c exists, then whenever $3n_1 + 4n_2 + 5n_3 = 0$ exactly, reducing mod a gives

$$d = 8n_1 - n_2 + 4n_3 \equiv c(3n_1 + 4n_2 + 5n_3) \equiv 0 \pmod a,$$

so every zero-first-coordinate element of H has second coordinate divisible by a – this is the missing minimality argument. Solve $3c \equiv 1 \pmod a$ candidate-free by working mod 7 first (the value the elimination already produced): $3 \cdot 5 = 15 \equiv 1 \pmod 7$, so $c = 5$. Check the other two coordinates: $4 \cdot 5 = 20 \equiv 6 \equiv -1 \pmod 7$ and $5 \cdot 5 = 25 \equiv 4 \pmod 7$ – both match $(8, -1, 4) \equiv (1, -1, 4) \pmod 7$ exactly (since $8 \equiv 1$). So $(8, -1, 4) \equiv 5\,(3, 4, 5) \pmod 7$, confirming $d \equiv 0 \pmod 7$ for every zero-first-coordinate element of H. Combined with the explicit witness $(0, 7) \in H$ found by elimination, this proves $a = 7$ exactly: not merely that 7 works, but that no smaller positive value can.

Finally $(1,5) = (4,-1) - (3,8) + 2(0,7) \in H$, and since $\gcd(3,4,5) = 1$ the first coordinates of the three generators already certify that some element of H has first coordinate 1, so $(1,5)$ and $(0,7)$ do generate H.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains reach $(0,7) \in H$ by the same elimination. They split immediately after: the natural attempt writes down $a = 7$ and stops, treating the witness as the answer. The model chain asks whether 7 is forced to divide every zero-first-coordinate second coordinate, and answers that with the congruence $(8,-1,4) \equiv 5(3,4,5) \pmod 7$.

> *I have shown $(0,7)$ is in H, but that is existence, not minimality – finding one witness does not rule out that some other combination gives $(0,1)$ or $(0,3)$, and the problem is asking for the unique a that actually generates H, which is a claim about every combination, not one of them.*

> *The two coordinate functionals $3n_1 + 4n_2 + 5n_3$ and $8n_1 - n_2 + 4n_3$ might be forced proportional to each other once I reduce modulo the right integer – if I can find a scalar c with $(8,-1,4) \equiv c(3,4,5) \pmod a$, then every combination that kills the first coordinate automatically kills the second coordinate modulo a, which is exactly the universal statement I am missing.*

Layer 3 THE WHY

Why does elimination alone fail to finish the problem? Because "$(0,7) \in H$" and "$H \cap (\{0\} \times \mathbb{Z}) = 7\mathbb{Z}$" are different logical statements – one is $\exists$, the other is $\forall$ (equivalently, a $\neq$-inequality in the other direction). $H \cap (\{0\} \times \mathbb{Z})$ is the kernel of the first-coordinate projection $H \to \mathbb{Z}$, and the kernel of a homomorphism into $\mathbb{Z}$ is automatically a subgroup of $\mathbb{Z}$, hence automatically of the form $a\mathbb{Z}$ for a single non-negative integer a – this is not extra structure the problem is assuming, it is a fact about subgroups of $\mathbb{Z}$ that holds regardless of which vectors generated H. Once that is recognised, "find a" becomes "find the generator of this specific subgroup of $\mathbb{Z}$," and a generator of a subgroup of $\mathbb{Z}$ is only pinned down by proving divisibility in *both* directions: $(0,7) \in H$ shows $7 \in a\mathbb{Z}$-worth of evidence in one direction ($a \mid 7$, if a is the true generator, is not even the right direction – one needs $a \mid$ every element, so 7 being an element only shows $a \mid 7$), while the congruence argument shows every element is divisible by 7, i.e. $7 \mid a \cdot$(everything) in the sense that 7 already divides every candidate, forcing $a \geq 7$; together with 7 itself lying in the subgroup, $a = 7$ exactly.

Why does the congruence trick work, and why is it not a guess? Because $8n_1 - n_2 + 4n_3$ and $3n_1 + 4n_2 + 5n_3$ are both linear functionals of (n_1, n_2, n_3), and two linear functionals agree up to a scalar multiple modulo m exactly when their coefficient vectors are proportional in $(\mathbb{Z}/m\mathbb{Z})^3$. Testing $m = 7$ (the modulus elimination already handed us) and solving $3c \equiv 1 \pmod 7$ is not a search over many candidates; the multiplicative inverse of 3 mod 7 is unique, so the check is a single verification once 7 is on the table, not a fishing expedition. The deeper reason 7 and only 7 makes this work traces back to the elimination computation itself: $4(3,8) - 3(4,-1) = (0,35)$ and $4(5,4) - 5(4,-1) = (0,21)$ are, in disguise, the 2×2 determinants (cross products) of pairs of the generating vectors – $35 = |3 \cdot (-1) - 8 \cdot 4|$ and $21 = |4 \cdot 4 - (-1) \cdot 5|$ up to sign – and $\gcd(35,21) = 7$. The index of a sublattice of $\mathbb{Z}^2$ generated by several vectors is exactly the gcd of these pairwise determinants; the elimination step was silently computing that gcd all along, it just never said so, which is exactly the gap this layer closes.

Layer $+\alpha$ BACKTESTING

1. The cue is the phrase "generated by" three vectors in $\mathbb{Z}^2$ together with a target normal form $(1,b),(0,a)$. On seeing it, write down: this is asking for the index of the sublattice $H \leq \mathbb{Z}^2$, and the index equals the gcd of the pairwise 2×2 determinants of the generators.

2. That cue suggests integer elimination (cancel first coordinates pairwise) to produce a candidate a, immediately followed by a divisibility/congruence argument to confirm that candidate is forced on every combination, not just the one found.
3. Hesitation occurs right after finding one element $(0,7) \in H$ and being tempted to declare $a = 7$; it is reinforced away by asking explicitly, "have I shown 7 divides every zero-first-coordinate second coordinate, or only that $(0,7)$ happens to be one of them?"
4. The trap is mistaking a witness for a characterisation: exhibiting $(0,7) \in H$ feels conclusive because it matches the expected answer, but by itself it only proves $H \cap (\{0\} \times \mathbb{Z}) \supseteq 7\mathbb{Z}$, not equality.
5. Any time a subgroup of $\mathbb{Z}^2$ is given by several integer generators and a normal form $(1,b),(0,a)$ is wanted, recognise immediately that a is the gcd of the pairwise determinants of the generating vectors; compute that gcd directly rather than hunting for one lucky combination.
6. Reusable rule: to find the unique generator a of a subgroup $K \leq \mathbb{Z}$ that arises as a kernel or a zero-coordinate slice of a finitely generated subgroup of $\mathbb{Z}^2$, do not stop at exhibiting one element of K – prove a divides every element of K (via a proportionality-mod-a argument between the defining linear functionals) and separately exhibit an element equal to a; only the two together pin down the generator.

▷ **Cue:** a subgroup of $\mathbb{Z}^2$ is generated by several integer vectors and a normal form $(1,b),(0,a)$ is wanted $\longrightarrow$ **Move:** compute a as the gcd of the pairwise 2×2 determinants of the generating vectors, since this is the index of the sublattice

▷ **Cue:** one element $(0,d)$ of a subgroup $K \leq \mathbb{Z}$ has been exhibited $\longrightarrow$ **Move:** do not declare d the generator until every element of K is shown divisible by d; existence of one witness is not minimality

▷ **Cue:** two linear functionals of the same integer variables are suspected to agree up to a scalar modulo some integer m $\longrightarrow$ **Move:** solve for the scalar c via a single modular inverse computation ($3c \equiv 1 \pmod{m}$), then verify the remaining coordinates match

1975 B2

1975 B2. In three-dimensional Euclidean space, define a slab to be the open set of points lying between two parallel planes. The distance between the planes is called the thickness of the slab. Given an infinite sequence $S_1, S_2, \ldots$ of slabs of thicknesses $d_1, d_2, \ldots$, respectively, such that $\sum_{i=1}^{\infty} d_i$ converges, prove that there is some point in the space which is not contained in any of the slabs.

Fields: geometry (surface area of the sphere) / real analysis (comparison with a convergent series) **Position:** B2 – second problem of the second session **Answer:** Take any sphere of radius $r > \frac{1}{2}\sum d_i$: the total area it exposes to the slabs is at most $2\pi r \sum d_i < 4\pi r^2$, so uncovered points exist on it (and hence in space).

Layer 1 THE WHAT — surface analysis

The natural attempt. The hypothesis is "$\sum d_i$ converges," which is the standard cue for a series-comparison argument: bound the "amount" covered by each slab by something summable, and compare the total to a fixed finite reference. The cheapest reference object in $\mathbb{R}^3$ is a line – say the z-axis. A slab S_i with normal direction making angle θ_i with the axis meets the axis (when it meets it at all) in a segment of length $d_i/|\cos\theta_i|$, since the axis's projection onto the slab's normal direction must sweep out length d_i. Summing these lengths and comparing to the (infinite) length of the axis looks promising – until you notice the denominator $|\cos\theta_i|$. If S_i's bounding planes happen to contain the axis's direction ($\theta_i = 90°$), the axis either misses S_i entirely or, if the axis happens to lie between the two planes, is *entirely contained* in S_i no matter

how thin d_i is. An adversary choosing the slabs can orient every single one this way around your particular fixed axis, and the whole comparison collapses: a sequence of arbitrarily thin slabs can swallow the entire z-axis. The line was the wrong probe, not because the convergence idea was wrong, but because a 1-dimensional probe's overlap with a slab depends on the angle between them, and that angle is exactly what the adversary controls.

The model chain. Replace the line with a sphere Σ of radius r, centred anywhere. By Archimedes' theorem, the surface area of Σ lying between two parallel planes at distance h apart depends *only* on h, never on the planes' orientation or position: if both planes meet the ball of radius r, the zone they cut has area exactly $2\pi rh$; if the planes are further apart or positioned so that the whole sphere lies between them, the area is at most $4\pi r^2 = 2\pi r \cdot (2r)$, which is still $\leq 2\pi rh$ once $h \geq 2r$. In every case,

$$\text{area}(\Sigma \cap S_i) \leq 2\pi r\, d_i,$$

with no dependence on how S_i is oriented relative to Σ. Summing over all slabs,

$$\text{area}\Big(\Sigma \cap \bigcup_i S_i\Big) \leq \sum_{i=1}^{\infty} \text{area}(\Sigma \cap S_i) \leq 2\pi r \sum_{i=1}^{\infty} d_i.$$

Choose $r > \frac{1}{2}\sum d_i$, so that $2\pi r \sum d_i < 4\pi r^2 = \text{area}(\Sigma)$. Then the covered area on Σ is strictly less than the total area of Σ, so Σ has a point not lying in $\bigcup_i S_i$ – and that point is a point of space not contained in any slab.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start from the same instinct: bound total coverage by a convergent sum and compare it to a fixed finite total. They split on the choice of probe. The natural attempt probes with a 1-dimensional line, where a slab's intersection length is $d_i/|\cos\theta_i|$ – unbounded as $\theta_i \to 90^\circ$. The model probes with a 2-dimensional sphere, where the intersection area is $\leq 2\pi r\, d_i$ for *every* orientation θ_i, with no angle dependence at all.

> *A fixed line is a bad probe here because its overlap with a slab depends on the angle between the line and the slab's planes, not on the slab's thickness alone – I need a probe whose intersection with any slab of thickness d is controlled by d alone, for every possible orientation of the slab, because the slabs' orientations are adversarial and unknown to me in advance.*

> *The sphere's surface area cut off between two parallel planes depends only on how far apart the planes are, never on which direction they face – that is Archimedes' theorem, and it is exactly the orientation-independence the line was missing.*

Layer 3 THE WHY

Why does a line fail structurally, not just in this instance? A line is 1-dimensional; a slab is a region unbounded in two directions and bounded in exactly one (its normal direction). When the line's direction is nearly orthogonal to the slab's normal, the line runs almost entirely *inside* the two bounding planes' common direction, so a line lying in that direction can be swallowed by an arbitrarily thin slab – thinness in one direction offers no protection to a probe that can align itself with the two directions the slab does *not* constrain. This is not a defect of the line probe that a smarter choice of axis fixes; for any fixed line, an adversary who sees the line first can orient every slab's planes to contain that line's direction, so the failure is generic, not accidental.

Why does the sphere avoid this failure? Because the sphere is rotationally symmetric: it has no preferred direction for an adversary to exploit. Concretely, rotate coordinates so a given slab's bounding planes are horizontal, $z = c_1$ and $z = c_2$ with $c_2 - c_1 = d_i$ (taking r centred at

the origin, and $c_1, c_2 \in [-r, r]$ or clipped to that range). Parametrising the sphere by $z = r\cos\theta$, a band of height $|dz|$ has surface area $2\pi r \sin\theta \cdot r\, d\theta = 2\pi r\, |dz|$ – the shrinking circumference ($2\pi r \sin\theta$) is exactly compensated by the steepening of the surface as θ moves away from the equator, so the area element $2\pi r\, dz$ is the same at every latitude. This cancellation is the entire content of Archimedes' theorem, and it holds regardless of which direction was called "z" – i.e. regardless of the slab's orientation – because the sphere itself has no distinguished axis until the slab's normal chooses one for it. That is the structural reason a rotationally symmetric bounded object, and not a line or any other fixed-direction object, is the right probe whenever the adversary is free to choose orientations: the measuring device must itself have no orientation to be exploited.

Layer $+\alpha$ BACKTESTING

1. The cue is "$\sum d_i$ converges." On seeing it, write down: this wants a series-comparison argument – bound each slab's contribution to some fixed finite total by (constant)$\times d_i$, then compare the convergent sum of contributions to that fixed total.

2. That cue suggests choosing a bounded reference object and bounding each slab's overlap with it, rather than attempting a direct point construction or a measure-theoretic argument on all of $\mathbb{R}^3$ at once.

3. Hesitation occurs when choosing which bounded object to measure against – a line is the cheapest first guess, and its failure is not obvious until an adversarial orientation is tested; it is reinforced away by asking explicitly, "does my bound on the overlap hold for *every* orientation of the slab, or only for orientations I happened to picture?"

4. The trap is assuming that because a slab is thin (small d_i), its intersection with any fixed probe must also be small – true only for probes oriented transversally to the slab's normal, false in general, and an adversary who picks the slabs after seeing the probe will exploit exactly this.

5. Any time a countable family of "thin, arbitrarily oriented" regions must be shown not to cover all of space, and each region is controlled by a single scalar (here, thickness), reach immediately for a rotationally symmetric bounded probe rather than a line or fixed plane.

6. Reusable rule: to show a countable union of codimension-1 regions with a single adversarially-orientable scalar parameter cannot cover an ambient space, measure against a rotationally symmetric bounded object on which each region's contribution is bounded by that scalar alone (independent of orientation), then choose the object's size so the convergent sum of contributions is strictly less than the object's fixed total measure.

▷ **Cue:** a countable family of "thin" regions of unconstrained orientation, with convergent total thickness, must be shown not to cover space $\longrightarrow$ **Move:** probe with a rotationally symmetric bounded object (a sphere), not a line or a fixed plane, so that orientation cannot be exploited

▷ **Cue:** need to bound how much of a sphere's surface lies between two parallel planes distance d apart $\longrightarrow$ **Move:** apply Archimedes' theorem: the area is at most $2\pi r\, d$ regardless of the planes' orientation or position

▷ **Cue:** a convergent sum of contributions must be compared against a fixed positive total that itself has a free parameter $\longrightarrow$ **Move:** choose that parameter (here the sphere's radius r) large enough that the convergent sum's bound is strictly less than the total, i.e. $r > \frac{1}{2}\sum d_i$

1975 B3

1975 B3. Let $s_k(a_1, \ldots, a_n)$ denote the k-th elementary symmetric function of $a_1, \ldots, a_n$.

With k held fixed, find the supremum (or least upper bound) M_k of

$$\frac{s_k(a_1,\ldots,a_n)}{\big(s_1(a_1,\ldots,a_n)\big)^k}$$

for arbitrary $n \geq k$ and arbitrary n-tuples $a_1,\ldots,a_n$ of positive real numbers. (Here $s_k(a_1,\ldots,a_n)$ is the sum of all k-fold products of the variables; the supremum is approached but not attained as $n \to \infty$ with the a_i suitably chosen.)

Fields: combinatorics / algebra (symmetric functions, multinomial theorem) **Position:** B3 – third problem, the harder problem of the second session **Answer:** $M_k = \dfrac{1}{k!}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The ratio is homogeneous of degree 0 (numerator and denominator both have degree k in the a_i), and homogeneous extremal problems in symmetric functions are trained reflexes for AM-GM and Lagrange multipliers: fix n, set up

$$\frac{\partial}{\partial a_j}\left(\frac{s_k}{s_1^k}\right) = 0$$

for each j, and try to show the critical point occurs where all a_i are equal. That guess is reasonable – symmetric functions usually extremise at symmetric points – and if you push it through, at $a_1 = \cdots = a_n = 1$ you get $s_1 = n$, $s_k = \binom{n}{k}$, and the ratio is $\binom{n}{k}/n^k$, which does tend to $1/k!$ as $n \to \infty$. But you have not actually finished: you assumed the maximiser (for fixed n) is the all-equal point, and proving that rigorously means either a full Lagrange-multiplier symmetry argument (showing the system of equations forces $a_i = a_j$ for all i, j, and separately ruling out boundary behaviour as some $a_i \to 0^+$) or an inequality like Maclaurin's, none of which is short. You are trying to solve a harder problem – find the exact maximiser for every n – than the one actually being asked, which only wants the supremum over *all* n together.

The model chain. Do not look for the maximiser at all. Expand $s_1^k = (a_1 + \cdots + a_n)^k$ by the multinomial theorem: it is the sum, over every ordered k-tuple of indices $(i_1,\ldots,i_k) \in \{1,\ldots,n\}^k$, of the product $a_{i_1}a_{i_2}\cdots a_{i_k}$. Every one of these n^k terms is positive, since every $a_i > 0$. Group the terms by which *unordered* multiset of indices they use. The terms whose k indices are all distinct correspond exactly to the k-element subsets of $\{1,\ldots,n\}$; each such subset $\{j_1,\ldots,j_k\}$ contributes the product $a_{j_1}\cdots a_{j_k}$ exactly $k!$ times (once for each ordering of the k indices), so this group sums to $k!\,s_k(a_1,\ldots,a_n)$. All the remaining terms (where at least two of the k indices coincide) are still positive numbers, so

$$s_1^k = k!\,s_k + (\text{a sum of further positive terms}) \;\geq\; k!\,s_k,$$

with equality possible only if no such further terms exist, i.e. only when $k \leq 1$ or $n < k$ prevents any repeated-index term from arising. For $n \geq k \geq 2$ such terms always exist (e.g. $a_1^2 a_2 \cdots a_{k-1}$ has index 1 repeated), so the inequality is strict: $s_k/s_1^k < 1/k!$ for every finite $n \geq k \geq 2$.

The bound $s_k \leq s_1^k/k!$ holds for every $n \geq k$ and every choice of positive a_i – no optimisation over configurations was needed. To see $1/k!$ is the supremum, exhibit the approach: set $a_1 = \cdots = a_n = 1$, giving

$$\frac{s_k}{s_1^k} = \frac{\binom{n}{k}}{n^k} = \frac{1}{k!}\prod_{j=0}^{k-1}\left(1 - \frac{j}{n}\right) \;\longrightarrow\; \frac{1}{k!} \quad \text{as } n \to \infty,$$

since each factor $1 - j/n \to 1$. So $M_k = 1/k!$, approached but (for $k \geq 2$) never attained. (For $k = 1$ the ratio s_1/s_1^1 is identically $1 = 1/1!$ for every n, so in that degenerate case the supremum

is trivially attained; the "never attained" remark in the problem's own bracket is about $k \geq 2$, where the multinomial expansion genuinely has extra positive terms to lose.)

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree on the value $1/k!$ and on how to approach it (equal a_i, $n \to \infty$). They split on how to prove the upper bound. The natural attempt tries to identify, for each fixed n, exactly which configuration of a_i maximises s_k/s_1^k – an optimisation problem. The model chain expands s_1^k by the multinomial theorem and reads $k!\, s_k$ off as a sub-sum of manifestly positive terms, bounding the ratio uniformly in n without locating any maximiser.

> *I do not need to know where the ratio is maximised for a fixed n – I only need an upper bound that holds for every n and every choice of positive a_i, and s_1^k's own multinomial expansion already contains $k!\, s_k$ as a sub-sum of positive terms, so the bound falls out of the definition of s_1^k itself, with no optimisation step at all.*

> *Since every one of the extra terms in the expansion of s_1^k (the ones with a repeated index) is strictly positive whenever $n \geq k \geq 2$, the inequality $s_k \leq s_1^k/k!$ is strict for every finite configuration – so the supremum cannot be attained at any finite n, and I should check this directly rather than assume it because the problem's bracket told me so.*

Layer 3 THE WHY

Why does the multinomial expansion settle the bound without any optimisation? Because s_1^k, as an algebraic object, *is* the sum over all ordered k-tuples of indices of the corresponding product – that is what raising a sum to the k-th power and distributing means, not a manipulation performed on it. Once that expansion is written down, the sub-collection of terms using k *distinct* indices is visibly $k!\, s_k$ (each k-subset counted once per ordering), and the hypothesis $a_i > 0$ is exactly what turns "the full sum contains this sub-collection" into "the full sum is at least this sub-collection" – for signed terms the sub-collection could exceed the whole sum, but positivity of every term forbids that. So the entire argument is: read the definition of s_1^k correctly, and use positivity to discard everything except the sub-collection you want, in one step, no search required. This is the structural reason the natural attempt overshoots – optimising over configurations of a_i answers "which configuration is best for this n," a strictly harder question, when the problem only needs "is the ratio always at most $1/k!$."

Why is the bound approached exactly at equal a_i, and why does that not contradict the previous paragraph? The bound $s_k \leq s_1^k/k!$ is tight precisely when the "extra" (repeated-index) terms in the expansion are as small as possible relative to the distinct-index terms. Making all a_i equal does not make the extra terms vanish – they are still present and positive for every finite n – but it makes their total weight, relative to s_1^k, shrink as $n \to \infty$: with $a_i = 1$, the ratio of distinct-index terms to all n^k ordered tuples is $n(n-1)\cdots(n-k+1)/n^k = \prod_{j=0}^{k-1}(1 - j/n)$, and each factor tends to 1 because a uniformly random ordered k-tuple of indices from $\{1, \ldots, n\}$ becomes overwhelmingly likely to have all-distinct entries as n grows (a birthday-paradox-type limit). The two facts are complementary, not competing: positivity gives the bound for free at every n; a population-size limit is what is needed to show the bound's constant $1/k!$ cannot be improved.

Layer $+\alpha$ BACKTESTING

1. The cue is a ratio of a symmetric function s_k (degree k) to a power s_1^k of matching degree k – a scale-invariant, homogeneous-degree-0 quantity built from symmetric functions of positive reals. On seeing it, write down: this is a candidate for a multinomial-expansion comparison, since s_1^k literally expands into a sum containing $k!\, s_k$ as a piece.

2. That cue suggests expanding the denominator combinatorially and comparing term-by-term, rather than differentiating with respect to each a_j to search for a critical point.

3. Hesitation occurs right after guessing (correctly) that the extremal configuration is "all a_i equal" and then trying to prove that guess is the true maximiser for a fixed n; it is reinforced away by asking, "do I actually need to identify the maximiser, or only an inequality that holds everywhere?"

4. The trap is assuming that because the supremum is *approached* at the symmetric configuration, that configuration must also be the exact maximiser for every fixed n – true, but proving it directly via calculus is far more work than proving the bound directly via the multinomial expansion.

5. Any time a ratio's numerator is an elementary or power-sum symmetric function and its denominator is a power of s_1 of matching total degree, reach for expanding the denominator by the multinomial theorem before reaching for Lagrange multipliers.

6. Reusable rule: when s_k (or any "coefficient-selecting" symmetric function) is compared to a power of s_1 of matching degree, expand s_1^k combinatorially, identify $k!\,s_k$ as the sub-sum of all-distinct-index terms, and use positivity of the variables to conclude the sub-sum is at most the whole sum – for free, with no optimisation over configurations required.

▷ **Cue:** ratio of an elementary symmetric function s_k to a power of s_1 of matching degree k $\longrightarrow$ **Move:** expand s_1^k by the multinomial theorem and identify $k!\,s_k$ as the sub-sum of all-distinct-index terms

▷ **Cue:** all variables are positive and one sub-sum of an expansion is being compared to the whole sum $\longrightarrow$ **Move:** sub-sum $\leq$ whole sum immediately; no optimisation over configurations is needed

▷ **Cue:** a claimed bound is approached but never attained, and the extremal configuration is symmetric (all variables equal) $\longrightarrow$ **Move:** let the number of variables $n \to \infty$ at the symmetric point and track the "distinctness" correction factor $\prod_{j=0}^{k-1}(1 - j/n) \to 1$

1975 B4

1975 B4. Does there exist a subset B of the unit circle $x^2 + y^2 = 1$ such that (i) B is topologically closed, and (ii) B contains exactly one point from each pair of diametrically opposite points on the circle?

(A set B is topologically closed if it contains the limit of every convergent sequence of points in B.)

Fields: topology of the line and circle; connectedness; the completeness (least upper bound) property of $\mathbb{R}$ **Position:** B4 — middle of the paper, but short once the right structural fact is named **Answer:** No. If B existed, its antipodal image $-B$ would also be closed, and the circle would split into two disjoint nonempty closed sets B and $-B$. The circle is connected, so this is impossible.

Layer 1 THE WHAT — surface analysis

The natural attempt. The natural first move is to try to build one. Take the upper semicircle. Each antipodal pair $\{P, -P\}$ with P off the x-axis has exactly one member with $y > 0$, so put those in B. That leaves the pair $(1,0), (-1,0)$: choose one, say $(1,0)$. So

$$B_0 = \{(\cos\theta, \sin\theta) : 0 \leq \theta < \pi\}.$$

This satisfies (ii). It fails (i): the points $(\cos\theta, \sin\theta)$ with $\theta \uparrow \pi$ lie in B_0 and converge to $(-1,0) \notin B_0$. Close it up by adding $(-1,0)$ and you have both $(1,0)$ and $(-1,0)$, which breaks (ii).

The student then tries to be cleverer: rotate, use a union of arcs, a Cantor-like set. Every attempt has the same shape. Wherever B "ends" along the circle, the antipodal set $-B$ "begins",

and the boundary point is a limit of one set but belongs to the other. After three or four tries the student suspects the answer is no, but has no argument, only a pile of failed constructions. The wall is that "I could not build one" is not a proof.

The model chain. Write C for the unit circle and $\sigma(P) = -P$ for the antipodal map. Condition (ii) says precisely that

$$B \cap \sigma(B) = \varnothing \qquad \text{and} \qquad B \cup \sigma(B) = C.$$

(Exactly one of $P, -P$ lies in B: if $P \notin B$ then $-P \in B$, so $P \in \sigma(B)$; and P cannot lie in both B and $\sigma(B)$.)

Step 1: $\sigma(B)$ is closed. Let $Q_n \in \sigma(B)$ with $Q_n \to Q$. Then $-Q_n \in B$ and $-Q_n \to -Q$, so $-Q \in B$ because B is closed; hence $Q \in \sigma(B)$.

Step 2: reduce to an interval. Let $P(\theta) = (\cos\theta, \sin\theta)$, a continuous map, and note $P(\theta + \pi) = -P(\theta)$. Put

$$S = \{\theta \in [0, \pi] : P(\theta) \in B\}, \qquad T = \{\theta \in [0, \pi] : P(\theta) \in \sigma(B)\}.$$

By Step 1 and the continuity of P, both S and T are closed subsets of $[0, \pi]$ (if $\theta_n \to \theta$ with $\theta_n \in S$, then $P(\theta_n) \to P(\theta)$, so $P(\theta) \in B$). They are disjoint, and their union is $[0, \pi]$. Since $P(\pi) = -P(0)$, exactly one of $0, \pi$ lies in S; by relabelling ($B \leftrightarrow \sigma(B)$ swaps S and T, and $\sigma(B)$ also satisfies (i) and (ii)) we may assume $0 \in S$ and $\pi \in T$.

Step 3: the crossing point. Let $t = \inf T$. This exists since T is nonempty and bounded below. As T is closed, $t \in T$, so $t \neq 0$ and $t > 0$. Every $\theta \in [0, t)$ lies outside T, hence in S. Choose $\theta_n \in [0, t)$ with $\theta_n \to t$; since S is closed, $t \in S$. So $t \in S \cap T = \varnothing$, a contradiction.

Hence no such B exists. □

The official solution compresses Steps 2 and 3 into one phrase: C is connected, and B, $-B$ would disconnect it.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains begin by looking at what condition (ii) does. The natural attempt reads (ii) as a recipe ("choose one from each pair") and starts choosing. The model reads (ii) as an equation between sets: B and $-B$ are complementary. At that line the paths split. The recipe reading keeps you constructing. The equation reading transfers condition (i) from B to its complement, and a set whose complement is also closed is a very special object.

> *Every construction breaks at a boundary point. Let me stop building and ask what (ii) says about the complement. The complement of B is $-B$, and $-B$ is closed because B is. So B is closed* and *open in the circle. A circle cannot be cut into two closed pieces.*

> *To prove that, walk along the circle from a point of B to its antipode, which is not in B. Somewhere I have to leave B for the last time. Look at the first moment I am in $-B$: it is a limit of points of B.*

- ▷ **Cue:** "exactly one of each pair" or "every point is in exactly one of two sets" ⟶ **Move:** rewrite as a partition equation $X = B \sqcup B'$; any property of B that the pairing map preserves passes to the complement
- ▷ **Cue:** a set and its complement are both closed in a space that looks like one piece (interval, circle, sphere) ⟶ **Move:** contradiction by connectedness; to prove it by hand, take the infimum of the second set along a path and show it lies in both
- ▷ **Cue:** repeated failed constructions that all break at the same kind of point ⟶ **Move:** stop constructing and turn the recurring obstruction into the proof of impossibility

Layer 3 THE WHY

Why is connectedness the thing to reach for, and not a trick? Go back to the definitions. A closed set is one that keeps its limit points. If B and its complement are both closed, then

no point of B is a limit of points outside B, and vice versa. So you can never approach the boundary of B from the other side: B has no boundary at all. In a space you can move through continuously, a set with no boundary must be everything or nothing. That is what "connected" means, and it is exactly what the circle is.

Now ask where the complement came from. Condition (ii) is a pairing: it says that σ swaps B with the rest of C. Any property of B that σ respects passes to the complement free of charge. σ is an isometry, so it preserves convergence, so it preserves closedness. The problem gives you a hypothesis about B and a symmetry, and the symmetry converts the hypothesis about B into a hypothesis about the complement. It is a natural move because the problem has only two conditions, and the second one is about how B sits against its complement.

Why does every construction fail at a boundary point? Step 3 is the answer, written out. If you walk from $P(0) \in B$ to $P(\pi) \notin B$, the set of times you spend outside B has a first moment t. Being first, it is a limit of times inside B, so closedness puts $P(t)$ in B. Being in the closed set T, it is also outside B. The semicircle B_0 failed at $(-1, 0)$ for exactly this reason. The failure was the proof, seen from the wrong side.

Finally, where does the infimum come from? It is the completeness of $\mathbb{R}$: every nonempty set bounded below has a greatest lower bound. That axiom is what makes an interval connected (on $\mathbb{Q}$ the argument fails: $\{q \in \mathbb{Q} : q^2 < 2\}$ is closed and open in $\mathbb{Q}$). So the chain runs back to one axiom: the real line has no gaps, so the circle has no gaps, so it cannot be split into two closed pieces.

It is worth seeing why each hypothesis is needed. Drop closedness, and B_0 works. Replace the circle by two disjoint circles swapped by the pairing, and B can be one of them. The problem is designed so that removing either the topology or the connectedness makes the answer yes.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "exactly one point from each pair". On seeing it, write $C = B \sqcup (-B)$.
2. That equation, together with "B closed", suggests showing that $-B$ is also closed and then invoking connectedness of the circle.
3. Hesitation comes from the question "is it yes or no?", which keeps the student constructing. Reinforce it away by writing the complement explicitly: once you see $C \setminus B = -B$, the closedness of $-B$ is one line.
4. The trap is the half-open semicircle. It satisfies (ii) and fails (i) by one point, which tempts you to believe a small adjustment will work. The one-point failure is structural, not accidental.
5. When a problem pairs points by a symmetry and asks for "exactly one of each pair", ask what the symmetry does to the complement. If the symmetry is continuous, topological properties pass across.
6. Reusable decision rule: *if a set and its complement both inherit a closure property in a space that is in one piece, the answer is "impossible" by connectedness; to prove it by hand, take the first point of the complement along a path and show it is also in the set.*

1975 B5

1975 B5. Let $f_0(x) = e^x$ and $f_{n+1}(x) = xf_n'(x)$ for $n = 0, 1, 2, \ldots$. Show that

$$\sum_{n=0}^{\infty} \frac{f_n(1)}{n!} = e^e.$$

Fields: analysis; power series; generating functions; (in the background) Bell numbers **Position:** B5 — late in the paper, but a short problem once the operator $x\frac{d}{dx}$ is recognised **Answer:** Expand $e^x = \sum_k x^k/k!$. The operator $x\frac{d}{dx}$ multiplies x^k by k, so $f_n(x) = \sum_{k\geq 0} k^n x^k/k!$. Then $\sum_n f_n(1)/n! = \sum_k \frac{1}{k!}\sum_n \frac{k^n}{n!} = \sum_k e^k/k! = e^e$, the swap of sums being legitimate because every term is nonnegative.

Layer 1 THE WHAT — surface analysis

The natural attempt. The recursion is explicit, so compute. Each f_n is a polynomial times e^x:

$$f_1 = xe^x, \quad f_2 = x(e^x + xe^x) = (x + x^2)e^x, \quad f_3 = x\big((1 + 2x) + (x + x^2)\big)e^x = (x + 3x^2 + x^3)e^x,$$

and $f_4 = (x + 7x^2 + 6x^3 + x^4)e^x$. Writing $f_n = p_n(x)e^x$, the recursion becomes $p_{n+1} = x(p_n + p_n')$. At $x = 1$:

$$\frac{f_n(1)}{e} = 1,\ 1,\ 2,\ 5,\ 15,\ 52, \ldots$$

A student who knows the sequence recognises the Bell numbers B_n, and then needs $\sum_n B_n/n! = e^{e-1}$, which is the generating function $\sum B_n t^n/n! = e^{e^t-1}$ at $t = 1$. That is true, but proving it from scratch is harder than the problem. A student who does not know the sequence is left with polynomial coefficients whose pattern is not visible, and a sum over n of $p_n(1)/n!$ with no closed form in sight. The wall is that the recursion was run in the wrong basis: multiplying by e^x and differentiating keeps mixing the coefficients.

The model chain. *Step 1: the action of $x\frac{d}{dx}$ on a power series.* The series $e^x = \sum_{k\geq 0} x^k/k!$ has infinite radius of convergence. A power series may be differentiated term by term inside its radius of convergence, and the result has the same radius. For $g(x) = \sum_k c_k x^k$ this gives

$$x\,g'(x) = \sum_{k\geq 1} kc_k x^k = \sum_{k\geq 0} kc_k x^k.$$

Step 2: closed form for f_n. By induction on n,

$$f_n(x) = \sum_{k=0}^{\infty} \frac{k^n}{k!} x^k \qquad (\text{all } x \in \mathbb{R}),$$

with the convention $0^0 = 1$ so that the case $n = 0$ is e^x. If it holds for n, Step 1 gives $f_{n+1}(x) = xf_n'(x) = \sum_k k \cdot k^n x^k/k!$, which is the claim for $n + 1$.

Step 3: the double sum. Hence

$$\sum_{n=0}^{\infty} \frac{f_n(1)}{n!} = \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} \frac{k^n}{n!\,k!}.$$

Every term is nonnegative. For a double series of nonnegative terms, the sum (finite or $+\infty$) is the same in either order. Summing over n first,

$$\sum_{k=0}^{\infty} \frac{1}{k!} \sum_{n=0}^{\infty} \frac{k^n}{n!} = \sum_{k=0}^{\infty} \frac{e^k}{k!} = e^e,$$

which is finite, so the original series converges and equals e^e. □

The official solution is the same argument in two lines.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains begin with $f_0 = e^x$. The natural attempt keeps e^x as a single function and differentiates it, producing $p_n(x)e^x$. The model rewrites e^x as $\sum x^k/k!$ *before* applying the operator. At that line the paths split: in the power-series basis the operator is diagonal, and the recursion becomes the one-line formula $f_n = \sum k^n x^k/k!$.

> *The recursion is "multiply by x after differentiating". What does $x\frac{d}{dx}$ do to a single power x^k? It gives kx^k. So on monomials it is multiplication by k. I should write e^x as a sum of monomials and let the operator act on each.*

> *Now I have $\sum_n \sum_k k^n/(n!k!)$. The inner structure over n is $\sum k^n/n! = e^k$. Every term is positive, so I may swap the order.*

- ▷ **Cue:** an operator $x\frac{d}{dx}$ (or $\frac{d}{dx}$, or "multiply by x") applied repeatedly to a known series $\longrightarrow$ **Move:** expand in monomials first; $x\frac{d}{dx}x^k = kx^k$, so n applications give $k^n x^k$
- ▷ **Cue:** a sum $\sum_n a_n/n!$ in which a_n is itself a sum $\longrightarrow$ **Move:** write the double sum, swap the order (justified by nonnegativity or absolute convergence), and look for $\sum_n c^n/n! = e^c$ inside
- ▷ **Cue:** a recursion whose terms look like polynomial $\times\, e^x$ with messy coefficients $\longrightarrow$ **Move:** change to the basis in which the operator acts diagonally before computing

Layer 3 THE WHY

Why is expanding e^x the natural move and not a lucky one? Look at the operator. $D = x\frac{d}{dx}$ is defined by what it does to functions, but its simplest description is what it does to x^k: $Dx^k = kx^k$. The monomials are its eigenvectors. Any time you have to apply an operator n times, the first question is which basis makes it diagonal, because in that basis applying it n times is raising the eigenvalue to the nth power. For $\frac{d}{dx}$ the eigenvectors are e^{ax}; for $x\frac{d}{dx}$ they are x^k. The problem gives you e^x, which is the one function most people already know as a power series. The data were chosen to be expanded.

There is a second way to see the whole answer at once, and it shows why the number is e^e. Consider

$$F(x,t) = \sum_{n\geq 0} \frac{t^n}{n!} D^n e^x,$$

the "exponential of tD" applied to e^x. On a monomial, $\sum_n \frac{t^n}{n!}k^n x^k = e^{kt}x^k = (xe^t)^k$. So e^{tD} sends $g(x)$ to $g(xe^t)$: $x\frac{d}{dx}$ is the infinitesimal generator of scaling $x \mapsto xe^t$. Hence $F(x,t) = \exp(xe^t)$, and the requested sum is $F(1,1) = e^e$. The model chain is this identity evaluated at one point. (For the reader: this is also why $f_n(1)/e$ are the Bell numbers. $F(1,t) = e \cdot e^{e^t-1}$, and e^{e^t-1} is the exponential generating function of set partitions. You do not need that fact, and the natural attempt shows what happens if you rely on it.)

Why is the swap of summations allowed? This is the one step where marks go. The rule is a consequence of the definition of an infinite sum as a limit of partial sums: for nonnegative terms, every finite partial sum in one order is bounded by the full sum in the other order, so the two sums bound each other. No convergence needs to be checked in advance; the answer $e^e < \infty$ confirms convergence afterwards. Term-by-term differentiation in Step 1 is the other analytic fact used, and it holds because power series converge uniformly on compact subsets of their disc of convergence.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the operator $f \mapsto xf'$ applied to e^x. On seeing it, write $e^x = \sum x^k/k!$ and $x\frac{d}{dx}x^k = kx^k$.
2. Termwise action of the operator, giving $f_n = \sum k^n x^k/k!$; then swap the double sum.
3. Hesitation occurs after f_3 or f_4, when the polynomial coefficients show no clear pattern. Reinforce it away by asking what the operator does to the simplest possible input, a single monomial.
4. The trap is computing the p_n and hunting for a pattern (Stirling numbers, Bell numbers). It leads to a correct but harder problem. A secondary trap is swapping sums without saying why it is allowed.
5. Whenever an operator is iterated, find its eigenvectors before iterating. For $x\frac{d}{dx}$ they are the monomials; for $\frac{d}{dx}$ they are exponentials; for shifts they are geometric sequences.

6. Reusable decision rule: *to iterate a linear operator, expand the starting object in the operator's eigenbasis, apply the nth power of each eigenvalue, then resum; if a second sum over n appears, swap the order and look for an exponential series.*

1975 B6

1975 B6. Show that if $s_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}$, then

(a) $n(n+1)^{1/n} < n + s_n$ for $n > 1$, and

(b) $(n-1)n^{-1/(n-1)} < n - s_n$ for $n > 2$.

Fields: inequalities; AM–GM; telescoping products **Position:** B6 — the last problem, but among the easiest B6s ever set once the right grouping is seen **Answer:** Split $n + s_n = \sum_{k=1}^{n} \left(1 + \frac{1}{k}\right)$, whose n terms have product $\prod \frac{k+1}{k} = n + 1$; AM–GM gives (a). Split $n - s_n = \sum_{k=2}^{n} \left(1 - \frac{1}{k}\right)$, whose $n - 1$ terms have product $\prod \frac{k-1}{k} = \frac{1}{n}$; AM–GM gives (b). Strictness holds because the terms are not all equal exactly when $n > 1$ in (a) and $n > 2$ in (b).

Layer 1 THE WHAT — surface analysis

The natural attempt. The harmonic sum has no closed form, so the student reaches for estimates. The standard one is $s_n > \ln(n+1)$ (compare with $\int_1^{n+1} dx/x$). For (a) it would be enough to show

$$(n+1)^{1/n} < 1 + \frac{\ln(n+1)}{n}.$$

But $(n+1)^{1/n} = e^{\ln(n+1)/n} > 1 + \frac{\ln(n+1)}{n}$, since $e^u > 1 + u$ for $u \neq 0$. The estimate points the wrong way. The loss in $s_n > \ln(n+1)$ is about $\gamma \approx 0.58$, while the true gap in (a) is small (at $n = 2$ it is $3.5 - 2\sqrt{3} \approx 0.036$). No integral comparison is sharp enough at small n.

The next attempt is induction on n. Going from n to $n+1$ requires comparing $(n+1)(n+2)^{1/(n+1)}$ with $n(n+1)^{1/n} + 1 + \frac{1}{n+1}$, and the fractional powers with different exponents do not combine. The student can check $n = 2, 3, 4$ numerically, sees the inequality is true and tight, and has no mechanism. The wall is that both approaches treat s_n as one number to be estimated, when the inequality is exact in a different way.

The model chain. *The tool.* For positive reals $a_1, \ldots, a_m$,

$$\frac{a_1 + \cdots + a_m}{m} \geq (a_1 a_2 \cdots a_m)^{1/m},$$

with equality if and only if $a_1 = \cdots = a_m$ (AM–GM).

(a) Let $n \geq 2$. Distribute the n ones among the n harmonic terms:

$$n + s_n = \sum_{k=1}^{n} \left(1 + \frac{1}{k}\right) = \sum_{k=1}^{n} \frac{k+1}{k}.$$

The product of these n terms telescopes:

$$\prod_{k=1}^{n} \frac{k+1}{k} = \frac{2}{1} \cdot \frac{3}{2} \cdots \frac{n+1}{n} = n + 1.$$

The terms are not all equal (the first is 2, the second is $\frac{3}{2}$), so AM–GM is strict:

$$\frac{n + s_n}{n} > (n+1)^{1/n}, \qquad \text{i.e.} \qquad n(n+1)^{1/n} < n + s_n.$$

(b) Let $n \geq 3$. The term $k = 1$ of $n - s_n = \sum_{k=1}^{n}(1 - \frac{1}{k})$ is zero, so drop it:

$$n - s_n = \sum_{k=2}^{n}\left(1 - \frac{1}{k}\right) = \sum_{k=2}^{n}\frac{k-1}{k},$$

a sum of $n - 1$ positive terms. Their product telescopes:

$$\prod_{k=2}^{n}\frac{k-1}{k} = \frac{1}{2}\cdot\frac{2}{3}\cdots\frac{n-1}{n} = \frac{1}{n}.$$

Since $n \geq 3$ there are at least two terms, $\frac{1}{2}$ and $\frac{2}{3}$, which differ, so AM–GM is strict:

$$\frac{n - s_n}{n-1} > \left(\frac{1}{n}\right)^{1/(n-1)} = n^{-1/(n-1)}, \qquad \text{i.e.} \qquad (n-1)n^{-1/(n-1)} < n - s_n. \qquad \square$$

The boundary cases. At $n = 1$ in (a), $1 \cdot 2^1 = 2 = 1 + s_1$: equality, one term. At $n = 2$ in (b), $1 \cdot 2^{-1} = \frac{1}{2} = 2 - \frac{3}{2}$: equality, one term. So the ranges $n > 1$ and $n > 2$ are exactly where AM–GM has at least two unequal terms. That the ranges match is a check that the method is the intended one.

The official solution is this argument, stated as "both parts are done easily using the Inequality on the Means".

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start from $n + s_n$. The natural attempt treats it as n plus an unknown quantity s_n and estimates s_n. The model reads $n + s_n$ as a sum of n terms, n ones and n fractions, and pairs them up: $1 + \frac{1}{k}$. At that line the paths split. Estimation loses a constant; pairing turns the inequality into AM–GM, which loses nothing except the strictness.

> *The left side has n multiplied by an nth root. That is n times a geometric mean of n things. The right side should then be n times an arithmetic mean of the same n things. What n things? $n + s_n$ has n ones and n fractions. Pair them: $1 + \frac{1}{k} = \frac{k+1}{k}$. Their product telescopes to $n + 1$. That is the $n + 1$ inside the root.*

> *In (b) the root is an $(n-1)$th root, so I need $n - 1$ terms. $n - s_n$ has n terms $1 - \frac{1}{k}$, and the first one is zero. Throw it away, and there are $n - 1$.*

- ▷ **Cue:** an expression $m \cdot P^{1/m}$ compared with a sum $\longrightarrow$ **Move:** read it as m times a geometric mean; split the sum into m terms whose product is P and apply AM–GM
- ▷ **Cue:** a harmonic sum added to or subtracted from n $\longrightarrow$ **Move:** distribute: $n \pm s_n = \sum(1 \pm \frac{1}{k}) = \sum\frac{k\pm1}{k}$, whose product telescopes
- ▷ **Cue:** an inequality stated for $n > 1$ or $n > 2$, with equality at the excluded value $\longrightarrow$ **Move:** look for an equality condition such as "all terms equal" that holds only when there is one term

Layer 3 THE WHY

Why is AM–GM the natural tool and not a lucky guess? Read the left side of (a) as a shape, not a number. $n(n+1)^{1/n}$ is "n times an nth root". The nth root of a product of n things is, by definition, their geometric mean, and n times a mean is a total. So the left side is the total you would get if n quantities with product $n + 1$ were all equal. The right side, $n + s_n$, is a total of something. AM–GM says the total of n positive quantities with a fixed product is least when they are equal. So the problem is asking: find n quantities with product $n + 1$ and sum $n + s_n$. Once the question is phrased like that, the ones must be distributed across the fractions, because $\frac{k+1}{k}$ is the only natural factor of $n + 1$ involving $\frac{1}{k}$.

Why does the product telescope? Because $1 + \frac{1}{k} = \frac{k+1}{k}$ is a ratio of consecutive integers, and a product of consecutive ratios collapses to last over first. This is the multiplicative version of $\sum(a_{k+1} - a_k) = a_{n+1} - a_1$. Here the harmonic sum turns out to be natural for products: every term of s_n, once shifted by ± 1, becomes a ratio of neighbours. That is the structural fact the problem is built on, and it is why $n + 1$ and $\frac{1}{n}$ appear in the roots.

Why does estimating s_n fail? Because the inequality is tight in a way no size estimate of s_n can see. The gap in (a) comes only from the *spread* of the terms $2, \frac{3}{2}, \frac{4}{3}, \ldots$ around their mean, which is small. An estimate such as $s_n \approx \ln n + \gamma$ carries an error that is not tied to that spread. The only argument precise enough is the one that uses the exact product.

Finally, why are the ranges $n > 1$ and $n > 2$? AM–GM is strict exactly when the terms are not all equal, and here that happens exactly when there are at least two terms. The problem's conditions are the equality case of AM–GM, stated as a range of n. Reading hypotheses as "where does the tool become strict?" is itself a way of identifying the tool.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is $n(n+1)^{1/n}$: n times an nth root. On seeing it, write "$= n \cdot$ GM of n numbers with product $n + 1$".

2. AM–GM, after splitting $n + s_n$ into n terms whose product is $n + 1$.

3. Hesitation occurs because s_n has no closed form, which suggests estimation. Reinforce it away by remembering that an inequality can be exact without a closed form: AM–GM needs only the product of the terms, not their sum.

4. The trap is the logarithmic estimate $s_n > \ln(n+1)$. It is too weak, and in (a) it gives an inequality in the wrong direction. The second trap in (b) is using all n terms of $n - s_n$, including the zero, which gives a geometric mean of 0 and an nth root instead of an $(n-1)$th root.

5. Whenever a root appears multiplied by its own index, think "geometric mean" and look for a sum with that many terms. Whenever $1 \pm \frac{1}{k}$ appears in a sum, check what its product telescopes to.

6. Reusable decision rule: *if one side of an inequality is $m \cdot P^{1/m}$, split the other side into m positive terms with product P and apply AM–GM; the equality case tells you which values of the parameter must be excluded.*