DAWN PUTNAM SESSION

Putnam 1979 — Papers A and B

Putnam 120 Archive

Contents

Putnam 1979 — Paper A **5**
1979 A1 5
1979 A2 7
1979 A3 10
1979 A4 13
1979 A5 15
1979 A6 17

Putnam 1979 — Paper B **21**
1979 B1 21
1979 B2 23
1979 B3 26
1979 B4 28
1979 B5 31
1979 B6 35

Putnam 1979 — Paper A

1979 A1

1979 A1. Find positive integers n and $a_1, a_2, \ldots, a_n$ such that $a_1 + a_2 + \cdots + a_n = 1979$ and the product $a_1 a_2 \cdots a_n$ is as large as possible.

Fields: elementary number theory, discrete optimization **Position:** A1 — opener; the whole difficulty is proving the exchange rule, not guessing it **Answer:** $n = 660$, with $a_1 = 2$ and $a_2 = \cdots = a_{660} = 3$; the maximal product is $2 \cdot 3^{659}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You have seen "maximize a product given a fixed sum" before, and the reflex is AM–GM: for a fixed sum, a product of n equal parts S/n is as large as the arithmetic-geometric relationship allows, and the real-valued function $g(x) = (S/x)^x$ (the product if you use x equal parts summing to S) is maximized, by calculus, at $x = S/e$. So the parts should be as close to $e \approx 2.71828$ as the integers allow. Since 3 is the nearer integer to e, you commit to parts of 3.

This gets you to the right neighbourhood fast, but it does not by itself finish the problem. The function g was built for a continuous relaxation where the number of parts is also a free real variable; here n must be a positive integer and every a_i must be a positive integer, and 1979 is a fixed, awkward number that will not divide evenly by 3. You are left checking, by hand and without a general principle, whether the leftover after using as many 3's as possible should be absorbed as one extra piece, split across two pieces, or handled some other way — and whether 3 actually beats every other candidate integer once you account for the remainder. The calculus told you where to aim; it did not tell you that no other combination of integers can do better, and it will not, by itself, rule out a rogue 5 or a pair of 4's sneaking into an optimal solution.

The model chain. The problem lives entirely in $\mathbb{Z}_{>0}^n$ with the linear constraint $\sum a_i = 1979$; it has no derivative to set to zero. Prove three exchange rules, each comparing two ways of using the same amount of "sum" and showing one product is never worse.

No $a_i = 1$. If some $a_i = 1$, merge it into any neighbour a_j: replace a_i, a_j by the single term $a_j + 1$. The sum is unchanged and the product changes from a_j to $a_j + 1 > a_j$. So an optimal partition has no part equal to 1.

No $a_i \geq 5$. If some $a_i = m \geq 5$, split it into 2 and $m - 2$ (both still positive, and $m - 2 \geq 3$). The sum is unchanged and the product changes from m to $2(m-2) = 2m - 4$. Since $m \geq 5 > 4$, $2m - 4 > m$, so the split strictly increases the product. Hence an optimal partition uses only parts equal to 2, 3, or 4 (and we have already excluded 1).

No three 2's. If three parts equal 2 (sum 6, product 8), replace them with two parts equal to 3 (sum 6, product $9 > 8$). So an optimal partition has at most two parts equal to 2; and since $4 = 2 + 2$ contributes the same product either as a single 4 or as two 2's, we may as well write every 4 as two 2's, leaving an optimal partition built entirely from 2's and 3's, with at most two 2's.

Fix the remainder. Since $1979 = 3 \cdot 659 + 2$, use 659 parts equal to 3 and one part equal to 2: this is exactly one 2 (allowed) and no leftover. No partition using only 2's and 3's with at most two 2's can do better, because any other split of the remainder (2 ones of 2, or a shortfall absorbed by fewer 3's) either reduces to this case or is excluded by the rules above. So $n = 660$, $a_1 = 2$, $a_2 = \cdots = a_{660} = 3$, and the maximal product is $2 \cdot 3^{659}$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains land on "use mostly 3's". They split at how that claim gets proved. The natural attempt proves it by locating 3 near the continuous optimum e of $(S/x)^x$ — a fact about a differentiable function of a real variable x. The model chain proves it by comparing two *specific, adjacent integer configurations* with the same sum and showing one product dominates: $2(m-2) \geq m$ for $m \geq 4$, and $3 \cdot 3 > 2 \cdot 2 \cdot 2$. The bifurcating line is exactly this: is the justification "3 is close to e", or is it "here are two inequalities between integers, and I can check both of them by hand"?

> *I don't need real-number optimality at all — I only need to show that any local move which preserves the sum and never decreases the product can be applied until no such move remains, and that terminal state is the answer, because there are only finitely many compositions of* 1979 *to begin with.*

> *Reaching for calculus on an integer partition problem is the expensive way in: it borrows a fact about a continuous relaxation and then still needs a separate argument to round correctly. The two inequalities* $2(m-2) \geq m$ *and* $9 > 8$ *are cheaper, need no derivative, and settle every case, including the remainder, without any appeal to* e*.*

Layer 3 THE WHY

Why must the argument be a pairwise exchange between integers, and not a rounding of the continuous optimum? Because the object being maximized is a function on a finite set: the set of all ordered tuples of positive integers summing to 1979. A finite set always attains its maximum on some element of itself, and the standard way to find that element without inspecting all of them is to exhibit a family of local moves, each of which strictly increases (or never decreases) the objective, and to observe that the process must stop at a configuration admitting no such move — which is then a maximizer, because every move away from it would only make things worse. This is the discrete analogue of setting a derivative to zero; the inequality $2(m-2) \geq m$ *is* the derivative test, transplanted into a world with no derivatives, because $\mathbb{Z}$ has no calculus of its own.

Why does no $a_i \geq 5$ survive, specifically? Splitting m into 2 and $m-2$ changes the product by a factor of $2(m-2)/m$; this exceeds 1 exactly when $m > 4$, which is why 4 is the threshold where the inequality goes from strict to an equality (a 4 can be split into two 2's with no gain or loss, which is why 4's and pairs of 2's are interchangeable and the "at most two 2's" rule can be phrased either way). Nothing about e is invoked in this step, because the step never needed it: the comparison is between two specific, already-integer configurations, which is the only kind of comparison a problem set in $\mathbb{Z}^n_{>0}$ can actually support.

Why is the remainder-handling step ($1979 \equiv 2 \pmod 3$, so use exactly one 2) not an afterthought but a forced consequence of the same rules? Because "at most two 2's" was already proved as a global constraint on any optimal partition, independent of what the remainder happens to be; once that constraint is fixed, the remainder mod 3 determines the only way to hit the exact sum 1979 using 3's and at most two 2's: remainder 2 needs one 2, remainder 1 would need two 2's (since a single leftover 1 is forbidden, and $2+2 = 4 \equiv 1$), and remainder 0 needs no 2's at all. The case here is remainder 2, so exactly one 2.

Layer $+\alpha$ BACKTESTING

1. The cue is "maximize a product subject to a fixed sum, over positive integers." Write down immediately: AM–GM says the product is largest when the parts are as equal as possible; the open question is only how to discretize that fact.

2. The cue suggests an exchange (local-improvement) argument between neighbouring integer configurations with the same sum, not a calculus optimization of a real-valued auxiliary function.

3. Hesitation sits at the step from "3's look best" to a rigorous proof that no other integer, or combination of integers, beats them. Reinforce it away by writing the two comparison inequalities explicitly — $2(m-2) \geq m$ for $m \geq 4$, and $3 \cdot 3 > 2 \cdot 2 \cdot 2$ — and checking that between them they cover every possible part size.

4. The trap is stopping at "use 3's because 3 is near e" and forgetting that the remainder $1979 \bmod 3 = 2$ still needs its own justified handling: a leftover of 1 can never stand alone (it must be absorbed, turning one 3 into two 2's), while a leftover of 2 stands alone as a single extra 2.

5. Recognize the family instantly: "partition N into positive integers to maximize the product" always resolves to threes, with the remainder mod 3 absorbed by zero, one, or two 2's (never a 1, never a part ≥ 5).

6. Reusable decision rule: for any fixed-sum, maximize-product problem over positive integers, prove (i) no part equals 1 (merge it), (ii) no part exceeds 4 (split it, since $2(m-2) \geq m$ for $m \geq 4$), and (iii) three 2's lose to two 3's; then read the answer off N mod 3.

▷ **Cue:** maximize a product of positive integers with a fixed sum $\longrightarrow$ **Move:** rule out parts of 1 (merge), rule out parts ≥ 5 (split into 2 and $m-2$), rule out three 2's (trade for two 3's), then fix the sum mod 3 with zero, one, or two 2's

▷ **Cue:** a finite optimization with no natural derivative $\longrightarrow$ **Move:** find a local move that never decreases the objective; the state admitting no such move is the maximizer, because the underlying set is finite

▷ **Cue:** a continuous relaxation (AM–GM, calculus) points at the right neighbourhood of the answer $\longrightarrow$ **Move:** treat it as a hypothesis to test, not a proof; finish with discrete exchange inequalities between actual integers

1979 A2

1979 A2. Establish necessary and sufficient conditions on the constant k for the existence of a continuous real-valued function $f(x)$ satisfying $f(f(x)) = kx^9$ for all real x.

Fields: real analysis, functional equations **Position:** A2 — second problem, and the trap is at a boundary value most students merge into the wrong case **Answer:** A continuous such f exists exactly for $k \geq 0$: $f(x) = k^{1/4}x^3$ works for every $k > 0$, and $f \equiv 0$ works for $k = 0$; no continuous f exists for $k < 0$.

Caveat. The source scan's official solution states its final answer as "The condition is $k > 0$" and gives, essentially verbatim, the injectivity-and-monotonicity argument reproduced in the model chain below for $k \neq 0$ — but that argument, by its own construction, only ever speaks about $k \neq 0$ (it divides by k at the step $ka^9 = kb^9 \Rightarrow a^9 = b^9$). Direct substitution shows $f \equiv 0$ satisfies $f(f(x)) = 0 = 0 \cdot x^9$ for every real x, so $k = 0$ is also achievable, by a legitimate (if degenerate) continuous witness. This is not an OCR artifact — the surrounding text is intact and unambiguous — so it is recorded here as a substantive correction, verified independently in Layer 1 and Layer 3, rather than assumed to be an error in the transcription.

Layer 1 THE WHAT — surface analysis

The natural attempt. The equation is a "functional square root": find f so that $f \circ f$ is a specific odd power function. The honest first instinct is to try a monomial, $f(x) = cx^a$, and match exponents: $f(f(x)) = c(cx^a)^a = c^{a+1}x^{a^2}$, so you need $a^2 = 9$. That already narrows a to ± 3; because x^9 is odd and must be matched by something well-defined and continuous on the whole real line (ruling out fractional or even exponents, which either aren't real for negative x or produce an even function), $a = 3$ or $a = -3$ is the entire menu.

Where a strong student burns time is in worrying whether a needs to be an integer at all — exploring $f(x) = c\,\mathrm{sgn}(x)|x|^a$ for general real a, checking that this is continuous, checking it composes correctly with itself, essentially reinventing "odd powers of x" from a more general ansatz before arriving back at $a = 3$. That detour is not wasted — it confirms $a = 3$ is forced, not guessed — but it is slower than committing to the integer monomial from the start and only falling back to a more general form if it fails.

The model chain. *Existence for $k > 0$.* Try $f(x) = ax^3$. Then $f(f(x)) = a(ax^3)^3 = a^4x^9$. Choosing $a = k^{1/4}$ (the positive real fourth root, which exists exactly because $k > 0$) gives $f(f(x)) = kx^9$ for every x, and f is continuous. So every $k > 0$ admits a witness.

Existence for $k = 0$. Take $f \equiv 0$. Then $f(f(x)) = f(0) = 0$ for every x, and $0 \cdot x^9 = 0$ for every x, so the identity holds. This f is continuous (constant functions are continuous) and satisfies the equation exactly, though it is not injective.

Nonexistence for $k < 0$. Suppose $k \neq 0$ and $f(f(x)) = kx^9$ for all x. Because $k \neq 0$, the map $x \mapsto kx^9$ is a bijection of $\mathbb{R}$ onto itself (it is continuous, strictly monotonic since 9 is odd, and unbounded in both directions). First, f is surjective: for any $y \in \mathbb{R}$ choose x with $kx^9 = y$; then $y = f(f(x)) \in \mathrm{im}(f)$. Second, f is injective: if $f(a) = f(b)$ then $f(f(a)) = f(f(b))$, so $ka^9 = kb^9$; dividing by k (legal since $k \neq 0$) gives $a^9 = b^9$, and since $x \mapsto x^9$ is injective on $\mathbb{R}$, $a = b$. So f is a continuous bijection of $\mathbb{R}$ onto itself, and a continuous injection $\mathbb{R} \to \mathbb{R}$ is strictly monotonic (an Intermediate Value Theorem argument: if it were neither strictly increasing nor strictly decreasing somewhere it would fail to be injective on some triple of points). Whether f itself is increasing or decreasing, the composition $f \circ f$ is always increasing (increasing-then-increasing is increasing; decreasing-then-decreasing is also increasing). But $x \mapsto kx^9$ is strictly decreasing when $k < 0$. This contradicts $f \circ f = kx^9$. So no continuous f exists when $k < 0$.

Combining the three cases: a continuous f with $f(f(x)) = kx^9$ for all real x exists if and only if $k \geq 0$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. The source volume's own solution states the answer as "$k > 0$" and gives exactly the injectivity-and-monotonicity argument above for $k \neq 0$, correctly excluding $k < 0$. It never separately tests $k = 0$ by direct substitution. The two paths agree completely up to the line "$f(a) = f(b) \Rightarrow ka^9 = kb^9 \Rightarrow a = b$" — and it is precisely that middle step, the division by k, where the argument silently stops covering $k = 0$. The model chain's bifurcation from the source is not a different method; it is the refusal to let the $k \neq 0$ argument's conclusion be read as covering $k = 0$ by default.

> *The instant my proof needs to divide by k, I have to go back and ask what happens at the one value of k where that division is illegal — a case excluded inside an inequality argument is not automatically the same as a case the argument's final answer is entitled to exclude.*

> *$k = 0$ looks too trivial to be a real case, because $kx^9 \equiv 0$ collapses the right-hand side to nothing interesting — but "too trivial to be interesting" is not the same claim as "impossible", and a constant function is a completely legal continuous witness that I should just try before assuming it*

> *is excluded.*

Layer 3 THE WHY

Why does the injectivity argument specifically require $k \neq 0$, rather than being a coincidence of how it happens to be written? Because the step "$ka^9 = kb^9 \Rightarrow a^9 = b^9$" is literally an appeal to the field axiom that every nonzero real number has a multiplicative inverse. At $k = 0$, that inverse does not exist, and the cancellation is not merely unproved for $k = 0$ — it is false for $k = 0$, since $k \cdot a^9 = k \cdot b^9 = 0$ holds for *every* pair a, b, telling you nothing. The entire downstream chain — injectivity of f, hence (with surjectivity) bijectivity, hence monotonicity, hence the always-increasing-composition contradiction with $k < 0$ — is a proof about the multiplicative group $\mathbb{R} \setminus \{0\}$ that was never making a claim about $k = 0$ in the first place. Reading its conclusion as if it also settled $k = 0$ mistakes "this argument does not apply here" for "this argument applies and gives a negative answer".

Why is direct substitution the right way to close the gap, rather than patching the injectivity argument? Because $k = 0$ is exactly the case where injectivity of f is *not* forced — the whole point is that $f \equiv 0$ is allowed to be as degenerate as it likes, since $f(f(x)) = f(0) = 0$ regardless of x. There is no monotonicity contradiction to derive because there is no forced bijectivity to derive it from. The honest content of the problem, once you see this, is that "continuous witness exists" and "continuous witness is forced to be a bijection" are two different claims, and they coincide only when $k \neq 0$; the case $k = 0$ is exactly the case where they come apart, and the only way to resolve it is to check by hand rather than to extend an argument built on a hypothesis ($k \neq 0$) that no longer holds.

Why should the reader treat this as a transferable habit rather than a one-off correction to an old book? Because "prove P is impossible for parameter values where some cancellation fails" is a shape that recurs across functional equations, recurrences, and algebra problems generally: any time a nonexistence proof pivots on dividing by a parameter, the value that zeroes that parameter is exactly the value the proof has not yet spoken about, and it needs its own sentence.

Layer $+\alpha$ BACKTESTING

1. The cue is "$f(f(x)) = k \cdot (\text{odd power of } x)$", a functional square root of a monomial. Write down immediately: try $f(x) = ax^n$ and match exponents by $n^2 = 9$.

2. The cue suggests pure algebraic ansatz-matching (a monomial guess) to build the existence half, and a monotonicity-of-composition argument to build the nonexistence half.

3. Hesitation sits exactly at ruling out $k < 0$ versus $k = 0$: the monotonicity contradiction is airtight for $k \neq 0$, but it is easy to fold $k = 0$ into "the excluded case" without re-deriving injectivity from scratch at $k = 0$. Reinforce it away by writing the cancellation step out explicitly and asking, in words, "does this step still make sense if $k = 0$?"

4. The trap is treating $k = 0$ as obviously degenerate and therefore excluded, when a constant function is a perfectly legal continuous witness and the case needs one line of direct verification, not an appeal to the case that was just ruled out.

5. Whenever a nonexistence proof cancels a parameter, mentally flag the value that annihilates it and test that value by hand, before finalizing the boundary of the answer.

6. Reusable decision rule: for "does $f \circ f = k \cdot (\text{odd bijection})$ have a continuous solution", build the explicit monomial witness for the sign where a real root of the matching equation exists, use the always-increasing-composition argument to rule out the opposite sign, and separately test every value that zeroes a coefficient you divided by, by direct substitution rather than by extension of the excluded case.

▷ **Cue:** $f(f(x)) = k \cdot x^{\text{odd}} \longrightarrow$ **Move:** try the monomial ansatz $f(x) = ax^m$ with m^2 equal to the odd exponent, solve $a^{m+1} = k$ for the sign that admits a real root

▷ **Cue:** f is forced to be a continuous bijection of $\mathbb{R} \longrightarrow$ **Move:** cite that a continuous injection $\mathbb{R} \to \mathbb{R}$ is strictly monotonic, so $f \circ f$ is always increasing, ruling out any target that is strictly decreasing

▷ **Cue:** a nonexistence proof cancels or divides by a parameter k $\longrightarrow$ **Move:** separately test $k = 0$ (or whatever value zeroes the coefficient) by direct substitution; the cancellation-based argument says nothing about that value

1979 A3

1979 A3. Let $x_1, x_2, x_3, \ldots$ be a sequence of nonzero real numbers satisfying

$$x_n = \frac{x_{n-2}x_{n-1}}{2x_{n-2} - x_{n-1}}, \qquad n = 3, 4, 5, \ldots$$

Establish necessary and sufficient conditions on x_1 and x_2 for x_n to be an integer for infinitely many values of n.

Fields: sequences, elementary algebra **Position:** A3 — the whole problem is recognizing one substitution **Answer:** x_n is an integer for infinitely many n if and only if $x_1 = x_2 = m$ for some integer m; in that case $x_n = m$ for every n.

Caveat. The source scan's Problem A-3 statement reads "Let $x_1, x_2, x_3, \ldots$ be a sequence of nonzero real numbers satisfying" followed by a blank line where the recurrence should be: the OCR dropped the formula entirely (almost certainly a typeset display equation the scanner could not render at all, rather than a misread string). The recurrence above was reconstructed, not guessed, from the book's own official solution to this problem, which is intact: it sets $r_n = 1/x_n$ and states outright that "$r_n = (2x_{n-2} - x_{n-1})/x_{n-2}x_{n-1} = 2r_{n-1} - r_{n-2}$". Clearing denominators in $r_n = (2x_{n-2} - x_{n-1})/(x_{n-2}x_{n-1})$ and solving for $x_n = 1/r_n$ gives exactly $x_n = \dfrac{x_{n-2}x_{n-1}}{2x_{n-2} - x_{n-1}}$, which is the unique recurrence consistent with the official solution's opening line. The book also gives an "Alternate solution" whose closed form for x_n was itself OCR-dropped, but whose stated conclusion ("x_n will be an integer for infinitely many values of n if and only if $x_1 = x_2 = m$ for some integer m") matches the answer derived independently below from the reconstructed recurrence, which is the cross-check this guard requires. This is a reconstruction from an intact solution, not an invention; the recurrence recovered is exactly the one the book's own solution solves, term for term.

Layer 1 THE WHAT — surface analysis

The natural attempt. Three-term recurrences with a fraction on the right invite direct computation: pick concrete values, say $x_1 = 1$, $x_2 = 2$, and grind out $x_3, x_4, \ldots$ by substitution, hoping a pattern in the numerators and denominators reveals itself. This does surface *something* — the fractions clearly involve products and differences of the previous two terms, and small examples with $x_1 = x_2$ visibly stay constant — but as a strategy for a "necessary and sufficient condition" it stalls. You can check finitely many pairs (x_1, x_2) and finitely many terms of each, but that never proves a statement about *infinitely many* n; you would need either a closed form for x_n in terms of x_1, x_2, n, or a structural reason the sequence must eventually behave one way or another. Staring at the raw recurrence — a product over a linear combination, both built from the same two variables — does not obviously hand you either.

The model chain. The recurrence's shape is a signature: a product of the two previous terms sits over a linear combination of them. That is exactly what you get if you take a *linear* recurrence in the reciprocals and clear denominators. So substitute $r_n = 1/x_n$, which is well-defined since every x_n is nonzero. Then

$$r_n = \frac{1}{x_n} = \frac{2x_{n-2} - x_{n-1}}{x_{n-2}x_{n-1}} = \frac{2}{x_{n-1}} - \frac{1}{x_{n-2}} = 2r_{n-1} - r_{n-2},$$

which rearranges to $r_n - r_{n-1} = r_{n-1} - r_{n-2}$: the sequence $\{r_n\}$ is an arithmetic progression, with common difference $d = r_2 - r_1$.

So $r_n = r_1 + (n-1)d$ and $x_n = 1/r_n$ for every n where $r_n \neq 0$ (and $r_n = 0$ would mean x_n is undefined, which the problem's hypothesis that every x_n is a nonzero real number already rules out).

If $d \neq 0$, $\{r_n\}$ is a nonconstant arithmetic progression: it is strictly monotonic and unbounded, so $|r_n| \to \infty$. In particular there is some N beyond which $|r_n| > 1$ for all $n > N$ (an unbounded monotonic sequence leaves the interval $[-1, 1]$ permanently once it exits it). For $n > N$, $|x_n| = 1/|r_n| < 1$; since $x_n \neq 0$, an integer with absolute value less than 1 does not exist, so x_n is not an integer for any $n > N$. Hence at most finitely many x_n can be integers.

If $d = 0$, then $r_1 = r_2$, i.e. $x_1 = x_2$; call this common value m. Then $r_n = r_1$ for every n, so $x_n = x_1 = m$ for every n. For infinitely many (indeed every) x_n to be an integer, it is necessary and sufficient that m itself be an integer.

Combining: x_n is an integer for infinitely many n if and only if $x_1 = x_2 = m$ for some integer m, in which case $x_n = m$ for all n.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start the same way — look at the recurrence, notice it mixes a product and a difference of the same two terms — but they split at the very next move. The natural attempt keeps computing with x_n itself, chasing the fraction forward term by term. The model chain stops computing and asks what the fraction's *shape* is: a product over a linear combination, homogeneous, of the previous two terms. That shape is exactly what a cleared-denominator linear recurrence in $1/x_n$ looks like. The bifurcating line is the single substitution $r_n = 1/x_n$; everything downstream — an arithmetic progression, its boundedness, the integer count — follows mechanically once that one move is made.

> *A product on top and a linear combination on the bottom, both built from the same two previous terms and matched in total degree — that is what a linear recurrence looks like after I clear denominators by inverting every term, so I should invert first and see whether the mess turns linear, rather than trying to solve the nonlinear recurrence as it stands.*

> *Once I have $r_n = 2r_{n-1} - r_{n-2}$, the question "for how many n is $x_n = 1/r_n$ an integer" is no longer a question about a complicated recurrence at all — it is a question about how often a specific, explicit arithmetic progression lands inside $[-1, 1]$, and an arithmetic progression with nonzero common difference only ever does that finitely often.*

Layer 3 THE WHY

Why does "invert" have to be the self-generated thought, rather than a fact you would only stumble on by luck? Because a linear recurrence $r_n = 2r_{n-1} - r_{n-2}$, rewritten with $r_n = 1/x_n$, reads $\dfrac{1}{x_n} = \dfrac{2}{x_{n-1}} - \dfrac{1}{x_{n-2}}$; putting the right-hand side over the common denominator $x_{n-2}x_{n-1}$ and inverting both sides mechanically produces exactly $x_n = \dfrac{x_{n-2}x_{n-1}}{2x_{n-2} - x_{n-1}}$. The given recurrence is not an arbitrary nonlinear object dressed up to look hard; it is the disguise that *every* linear recurrence in reciprocals wears once you clear denominators: a product in the numerator, a linear combination in the denominator, matched in degree. Once you know what that disguise looks like, seeing "product over linear combination" and reaching for the reciprocal substitution is not a trick pulled from nowhere — it is running the algebra of "clear denominators" backward, from its own definition.

Why does unboundedness of the reciprocal sequence answer the integrality question, rather than needing some separate number-theoretic argument? Because the problem asks about

$x_n = 1/r_n$ specifically, and a nonzero integer always has absolute value at least 1; so "x_n is an integer" forces $|r_n| \leq 1$. An arithmetic progression with nonzero common difference is, by the definition of arithmetic progression, strictly monotonic and hence eventually and permanently leaves any fixed bounded interval such as $[-1, 1]$. There is no coincidence to explain here: monotonic and unbounded is what "nonzero common difference" *means* for a sequence defined by repeated addition of a fixed nonzero quantity, and "visits $[-1, 1]$ only finitely often" is the direct translation of that fact into the language the problem is asking about.

Why is the necessary condition exactly $x_1 = x_2$, and not some other relationship between the two seeds? Because $d = r_2 - r_1 = 0$ is the unique condition under which the arithmetic progression $\{r_n\}$ fails to run away to infinity, and $r_2 = r_1$ means precisely $1/x_2 = 1/x_1$, i.e. $x_1 = x_2$ (both nonzero, so this is a genuine equality, not a vacuous one). There is no weaker condition to hope for: any nonzero difference between the two seeds, however small, is amplified without bound by the recurrence, because that is what adding a fixed nonzero quantity repeatedly does.

Layer $+\alpha$ BACKTESTING

1. The cue is a three-term recurrence with a product of the two previous terms on top and a linear combination of them on the bottom. Write down immediately: this is the shape a linear recurrence in the reciprocals takes after you clear denominators.

2. The cue suggests the substitution $r_n = 1/x_n$, converting the given nonlinear recurrence into a linear one, and then analysing that linear recurrence (here, an arithmetic progression) instead of the original.

3. Hesitation sits at "why would I invert an already complicated fraction instead of just computing terms". Reinforce it away by checking the recurrence's total degree on both sides: numerator and denominator match in degree, which is exactly the fingerprint clearing denominators leaves behind, and that fingerprint is the cue to invert.

4. The trap is grinding out $x_3, x_4, \ldots$ numerically from a guessed pair x_1, x_2 and trying to spot integrality by inspection; this only ever tests finitely many cases and can never close a "necessary and sufficient" claim about infinitely many n.

5. Memorize the signature "product over linear combination, same degree" as the standing cue for "invert to linearize" — the same family as harmonic-mean-style recursions such as $x_{n+1} = \dfrac{2x_n x_{n-1}}{x_n + x_{n-1}}$.

6. Reusable decision rule: whenever a recurrence has the shape $x_n = \dfrac{P(x_{n-2}, x_{n-1})}{L(x_{n-2}, x_{n-1})}$ with P a product and L a linear combination of matching total degree, try $r_n = 1/x_n$ before anything else; if the recurrence in r_n is linear, the sequence's long-run behaviour (boundedness, integrality, limits) reduces to the behaviour of that linear recurrence's solutions, here an arithmetic progression that is unbounded unless its common difference is zero.

▷ **Cue:** a recurrence with a product of the two previous terms on top and a linear combination of them on the bottom, matched in degree $\longrightarrow$ **Move:** substitute $r_n = 1/x_n$ to linearize; analyse the resulting linear recurrence instead of the original

▷ **Cue:** "an integer term occurs infinitely often" in a sequence whose reciprocal is a known linear recurrence $\longrightarrow$ **Move:** translate integrality into a boundedness statement about the reciprocal sequence, and use that an arithmetic progression with nonzero common difference is unbounded

▷ **Cue:** two seed values x_1, x_2 must satisfy a condition for a property to hold for infinitely many n $\longrightarrow$ **Move:** look for the condition that makes the governing linear recurrence's "growth mode" vanish (here, common difference $d = 0$), since any nonzero growth mode is eventually amplified without bound

1979 A4

1979 A4. Let A be a set of $2n$ points in the plane, no three of which are collinear. Suppose that n of them are colored red and the remaining n blue. Prove or disprove: there are n closed straight line segments, no two with a point in common, such that the endpoints of each segment are points of A having different colors.

Fields: combinatorial geometry, extremal principle **Position:** A4 — mid-paper; one idea, and the idea is a choice of what to minimise **Answer:** True. Among the $n!$ red–blue pairings choose one of minimal total segment length; if two of its segments met, swapping partners would strictly shorten the total by the triangle inequality, so the minimal pairing has pairwise disjoint segments.

Layer 1 THE WHAT — surface analysis

The natural attempt. "Prove or disprove" with a clean geometric picture: you draw a few examples, pair things up by eye, and find it always works. So you believe it and try to prove it.

The first plan is local repair. Start from any red–blue pairing. If two segments RB and $R'B'$ cross, uncross them: re-pair as RB' and $R'B$. Draw it — the crossing is gone. Repeat until nothing crosses.

Then the doubt arrives, and it is the right doubt. The new segment RB' may cut through three other segments that RB never touched. You fixed one crossing and created three. Does the process stop? You try to count crossings as a measure of progress, and it does not decrease: a single swap can raise the number of crossings. You try an ordering — always uncross the leftmost crossing — and cannot prove that terminates either. This is the wall: a repair move that is locally good, with no evidence that it is globally going anywhere.

The second plan is induction. Find one red–blue pair R, B such that the line RB has equally many reds as blues on each side; pair R with B and recurse on the two sides, which are separated by a line, so their segments cannot meet. This works, but proving that such a balanced pair exists needs a rotating-line continuity argument (sweep a line about a point of the convex hull and track the red-minus-blue count on one side). It is a correct route and twice the length.

The model chain. The statement is true.

A pairing is a bijection between the n red points and the n blue points; there are $n!$ of them, finitely many. For each pairing, let L be the sum of the lengths of its n segments. Choose a pairing π with L minimal. We show no two segments of π share a point.

Let RB and $R'B'$ be two segments of π (R, R' red, B, B' blue, $R \neq R'$, $B \neq B'$), and suppose they share a point O.

Where can O be? The four endpoints are distinct points of A, so O is not an endpoint of both. If O were an endpoint of one segment, say $O = R$, then R lies on segment $R'B'$ and R, R', B' are collinear — forbidden. Likewise for the other three endpoints. If the two segments overlapped along a stretch, all four points would be collinear — forbidden. So O is a single point lying strictly inside both segments.

Swap. Replace $RB, R'B'$ by $RB', R'B$; this is still a red–blue pairing. By the triangle inequality in triangles ROB' and $R'OB$,

$$|RB'| \leq |RO| + |OB'|, \qquad |R'B| \leq |R'O| + |OB|.$$

Adding, and using $|RO| + |OB| = |RB|$ and $|R'O| + |OB'| = |R'B'|$ (because O lies on both segments),

$$|RB'| + |R'B| \leq |RB| + |R'B'|.$$

Strictness. Equality in the first inequality means O lies on segment RB', so R, O, B' are collinear. But R, O, B are collinear with $O \neq R$, so the line RO contains B and B', and R, B, B' are three collinear points of A — forbidden. So the inequality is strict, and the new pairing has total length less than L, contradicting minimality.

Hence in π no two segments share a point, and π supplies the n required segments.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains contain the same move: uncross $RB, R'B'$ into $RB', R'B$. The natural attempt applies it as a *procedure* and then needs to prove the procedure halts. The model never runs the procedure. It asks what quantity the move decreases, finds that it is total length, and goes straight to the configuration where that quantity is smallest.

> *Every time I uncross, the total length goes down — that is the triangle inequality. I don't need the process to finish. There are only $n!$ pairings, so one of them has the least total length, and that one can't have a crossing, because uncrossing it would make it shorter.*

> *Before I trust the swap, where exactly can the two segments meet? At an endpoint means three points collinear. Overlapping means four collinear. So they cross at an interior point — and that is the only case the triangle inequality has to handle.*

Layer 3 THE WHY

Why is length the quantity to minimise, and not the number of crossings? Look at what the swap does, measured by everything you could measure. The number of crossings can go up. The positions of the other segments do not change. What changes is two lengths replaced by two other lengths — and a crossing is, by definition, a point O on both segments, which is precisely the point that lets you write $|RB| = |RO| + |OB|$ and compare with the triangle inequality. The crossing is not an obstacle to measure; it is the hypothesis of a triangle inequality. So the move has a monotone quantity attached to it the moment you draw O.

Why is "take the minimum" legitimate and not a trick? Because the space of pairings is finite. On a finite set every real-valued function attains its minimum; that is the whole content. Once a local move strictly decreases a function, the minimiser is automatically a configuration where the move is impossible. That is the extremal principle, and it is the same fact as "a process that strictly decreases a quantity on a finite set halts" — read backwards, without running the process. The official solution opens with "there are a finite number (actually $n!$) of ways of pairing". Here is where that sentence comes from: it is the finiteness that converts "the swap helps" into "the best pairing has no crossings".

Why does the problem say "no three collinear"? Because without it the claim is false. Put R, R', B, B' in that order on a line, with no other points: the pairing $RB, R'B'$ overlaps on the stretch $R'B$, and the pairing $RB', R'B$ nests one segment inside the other, so both fail, and the strict inequality in the swap fails exactly in the collinear case. The hypothesis is the strictness of the triangle inequality, written in the language of the problem. When a hypothesis looks like housekeeping, ask which inequality it makes strict.

Why "prove or disprove"? Examiners use it to stop you reading the answer off the question. Drawing three small examples settles which way to go; nothing in the wording should be taken as a hint that it is false.

Layer $+\alpha$ BACKTESTING

1. The cue is "there exist n segments, no two with a point in common" over a finite set of choices. Write down: the set of all red–blue pairings is finite, and a crossing pair can be swapped.

2. The extremal principle: attach a quantity (total length) that the repair move strictly decreases, and take the configuration minimising it.

3. Hesitation comes after the first uncrossing, when you notice new crossings can appear. Reinforce it away by asking "what does the swap decrease?" instead of "does the process stop?". If something decreases, you never need to run the process.

4. Two traps. Counting crossings as the potential — it is not monotone. And forgetting the degenerate meetings (touching at an endpoint, collinear overlap): the triangle inequality is only strict because no three points are collinear, and the proof must say so.
5. Any "show a configuration with no bad pairs exists" where a local swap fixes one bad pair but may create others: stop simulating and find the monotone quantity.
6. Reusable rule: if a local move strictly improves a numerical quantity and the configurations are finitely many, the optimal configuration admits no such move. Prove existence by choosing the optimum, not by running the algorithm.

▷ **Cue:** "no two segments with a point in common" for a matching of points ⟶ **Move:** take the matching of minimal total length; a crossing pair can be swapped to shorten it
▷ **Cue:** a repair move that fixes one defect but may create new ones ⟶ **Move:** find a quantity the move strictly decreases; take the extremal configuration instead of iterating
▷ **Cue:** two segments meet at a point O interior to both ⟶ **Move:** split each length at O and apply the triangle inequality to the swapped pairs
▷ **Cue:** "no three collinear" in a hypothesis ⟶ **Move:** it makes a triangle inequality strict, or rules out degenerate intersections; find which and cite it there
▷ **Cue:** "prove or disprove" on a combinatorial-geometry statement ⟶ **Move:** test three small drawings to decide the direction before proving anything

1979 A5

1979 A5. Denote by $[x]$ the greatest integer less than or equal to x and by $S(x)$ the sequence $[x], [2x], [3x], \ldots$. Prove that there are distinct real solutions α and β of the equation $x^3 - 10x^2 + 29x - 25 = 0$ such that infinitely many positive integers appear both in $S(\alpha)$ and in $S(\beta)$.

Fields: number theory (Beatty sequences), algebra (Vieta), combinatorics (counting / pigeonhole) **Position:** A5 — late-paper; short once the right statistic is chosen, and nothing in the statement names that statistic **Answer:** The roots $a < b < c$ are real, lie in $(1,2)$, $(2,3)$, $(5,6)$, and satisfy $1/a + 1/b + 1/c = 29/25 > 1$. $S(x)$ has about n/x terms up to n, so the three sequences put about $\frac{29}{25}n$ terms into $\{1, \ldots, n\}$; the excess over n tends to infinity, so some pair of sequences shares infinitely many values.

Layer 1 THE WHAT — surface analysis

The natural attempt. The cubic is the concrete object, so you start there. You look for rational roots: candidates $\pm 1, \pm 5, \pm 25$; $f(1) = -5$, $f(5) = -5$, none work. So the roots are irrational and you cannot write them down. You locate them: $f(1) = -5$, $f(2) = 1$, $f(3) = -1$, $f(5) = -5$, $f(6) = 5$, so there are roots $a \in (1,2)$, $b \in (2,3)$, $c \in (5,6)$.

Now you pick a pair and try to show $S(a)$ and $S(b)$ collide infinitely often. You need integers j, k with $[ja] = [kb]$, i.e. ja and kb in the same unit interval. You refine the roots: $a \approx 1.623$, $b \approx 2.726$. You try to find a pattern in $j/k \approx b/a$ and get tangled in continued fractions of numbers you only know to three decimal places.

Then a better idea: density. $S(a)$ hits about a fraction $1/a$ of the integers, $S(b)$ about $1/b$. If $1/a + 1/b > 1$ they must overlap. But $1/a + 1/b \approx 0.616 + 0.367 = 0.983 < 1$. It fails, and narrowly. $1/a + 1/c$ is worse. For two sequences whose densities add to less than 1, nothing forces an overlap — Beatty's theorem says two sequences with $1/\alpha + 1/\beta = 1$ can be exactly disjoint. This is the wall: every pair you can name is too thin to force a collision.

The model chain. Let $f(x) = x^3 - 10x^2 + 29x - 25$. From $f(1) = -5$, $f(2) = 1$, $f(3) = -1$, $f(5) = -5$, $f(6) = 5$ and the intermediate value theorem, f has roots $a \in (1,2)$, $b \in (2,3)$, $c \in (5,6)$; these are three distinct real roots, hence all of them. By Vieta,

$$ab + bc + ca = 29, \qquad abc = 25, \qquad \text{so} \qquad \frac{1}{a} + \frac{1}{b} + \frac{1}{c} = \frac{ab + bc + ca}{abc} = \frac{29}{25}.$$

Counting one sequence. Let $x > 1$ and $T_x(n) = \{[kx] : k \geq 1\} \cap \{1, \ldots, n\}$. Since $(k+1)x - kx = x > 1$, the values $[kx]$ are strictly increasing in k, hence distinct, and all are $\geq [x] \geq 1$. Also $[kx] \leq n \iff kx < n+1 \iff k < (n+1)/x$. So $|T_x(n)|$ equals the number of positive integers $k < (n+1)/x$, which is at least $(n+1)/x - 1$.

Counting three. Write T_a, T_b, T_c for $T_a(n), T_b(n), T_c(n)$. All three lie in $\{1, \ldots, n\}$, so $|T_a \cup T_b \cup T_c| \leq n$. An element lying in exactly m of the three sets ($m = 1, 2, 3$) contributes m to $|T_a| + |T_b| + |T_c|$, and contributes 1 to the union plus $\binom{m}{2}$ to the sum of pairwise intersections; since $m \leq 1 + \binom{m}{2}$ for $m = 1, 2, 3$,

$$|T_a| + |T_b| + |T_c| \leq |T_a \cup T_b \cup T_c| + |T_a \cap T_b| + |T_b \cap T_c| + |T_c \cap T_a|.$$

Therefore

$$|T_a \cap T_b| + |T_b \cap T_c| + |T_c \cap T_a| \;\geq\; \frac{29}{25}(n+1) - 3 - n \;=\; \frac{4n}{25} - \frac{46}{25} \;\xrightarrow[n\to\infty]{}\; \infty.$$

If each of the three sets $S(a) \cap S(b)$, $S(b) \cap S(c)$, $S(c) \cap S(a)$ of positive integers were finite, the left side would be bounded independently of n. So at least one is infinite, and the corresponding two roots are the required $\alpha \neq \beta$.

Caveat. The argument does not say which pair works. A numerical check (not part of the proof) with $n = 2 \times 10^5$ gives $|T_a \cap T_b| = 45204$, $|T_a \cap T_c| = 21806$, $|T_b \cap T_c| = 12983$, suggesting every pair shares infinitely many values; proving that for a specific pair needs finer information (equidistribution of ja/b modulo 1) and is not asked.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains locate the roots and both think about how often each $S(x)$ hits the integers. The natural attempt then chooses a pair and compares two densities. The model refuses to choose: it adds all three densities, because the problem only asks for *some* pair.

> *The problem says "there are distinct solutions α and β" — it doesn't tell me which. That's a pigeonhole shape. I shouldn't pick a pair; I should throw all three sequences into $\{1, \ldots, n\}$ and count the overflow.*

> *The density of $S(x)$ is $1/x$, and a sum of reciprocals of roots is a Vieta quantity: $e_2/e_3 = 29/25$. That's bigger than 1. So the coefficients were chosen for exactly this.*

Layer 3 THE WHY

Why density at all? $S(x)$ is defined by floors, and the only thing floors reliably tell you in bulk is how many terms fall below a given size: $[kx] \leq n$ exactly when $k < (n+1)/x$. Any question of the form "infinitely many common values" becomes, under that lens, a question about how many values each sequence plants in $\{1, \ldots, n\}$ — and there is only room for n of them. Overflow forces collisions. That is pigeonhole, and it is the only general tool that proves collisions without finding them.

Why all three roots and not two? Because the numbers force it. Once the density idea is on the table, you compute $1/a + 1/b$ and find it below 1. At that point the natural reaction is to abandon density; the correct reaction is to re-read the question. "There are distinct solutions α, β" is an existence claim over the three pairs. Existence over a finite menu is exactly where you can pool: if three sets overflow a box of size n by an unbounded amount, the overflow is carried by pairwise intersections, and some one of the three must carry an unbounded share. You were never required to know which.

Why is $1/a + 1/b + 1/c$ the natural quantity to compute, rather than something you notice? Because you cannot compute any single $1/a$ exactly — the roots are irrational cubic numbers. The only functions of the roots you *can* compute exactly from the coefficients are the symmetric ones. So the moment density suggests "reciprocals of roots", and the moment exact values are unavailable, the symmetric sum is the one exact statement in reach: $\sum 1/r = e_2/e_3 = 29/25$. The official solution states "$1/a + 1/b + 1/c = 29/25 > 1$" without motivation. Here is where it comes from: it is the only exact information about the roots the cubic can give you, and it is precisely the density bookkeeping of the three sequences.

The benchmark behind the number 1 is Beatty's (Rayleigh's) theorem: for irrational $\alpha, \beta > 1$ with $1/\alpha + 1/\beta = 1$, the sequences $S(\alpha)$ and $S(\beta)$ partition the positive integers. So density sum exactly 1 is compatible with no collisions at all; strictly above 1, linearly growing overlap is forced. The examiners set the threshold at $29/25$ so that the three-way pooling works and no two-way comparison does.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "infinitely many positive integers appear both in $S(\alpha)$ and in $S(\beta)$" with S defined by floors. Write down: $|S(x) \cap [1, n]| \approx n/x$. The second cue is that α, β are roots of a given cubic: write down Vieta.

2. Counting (densities) plus pigeonhole: total count of the three sequences up to n versus the n available slots.

3. Hesitation occurs when $1/a + 1/b < 1$ kills the two-sequence argument. Reinforce it away by re-reading the quantifier: the problem lets you choose the pair after the fact, so pool all three.

4. The trap is to fix a pair (the two smallest roots look most promising) and try to prove collisions for it directly, which needs Diophantine information about b/a that the problem does not supply.

5. When the objects are roots of an explicit polynomial and the problem needs a numerical fact about them, the fact is almost always a symmetric function; compute e_1, e_2, e_3 and the ratios before anything else.

6. Reusable rule: to show that some two of several sets intersect infinitely, add their counting functions up to n and compare with the size of the ambient box; an unbounded excess is carried by pairwise intersections, and one of them must be infinite.

▷ **Cue:** "infinitely many integers in both $S(\alpha)$ and $S(\beta)$" with $S(x) = \{[kx]\}$ $\longrightarrow$ **Move:** count terms up to n: about n/x each; compare total with n

▷ **Cue:** "there exist two of these objects such that…" with a small finite menu $\longrightarrow$ **Move:** pool all candidates and pigeonhole; do not choose the pair in advance

▷ **Cue:** a needed quantity is a sum of reciprocals of roots of an explicit polynomial $\longrightarrow$ **Move:** Vieta: $\sum 1/r = e_{d-1}/e_d$

▷ **Cue:** density comparison of Beatty sequences $\longrightarrow$ **Move:** sum of $1/\alpha$ equal to 1 allows disjointness (Beatty); strictly greater forces unbounded overlap

▷ **Cue:** a cubic with integer coefficients whose roots are needed $\longrightarrow$ **Move:** rational-root test, then sign table at integers to bracket the roots

1979 A6

1979 A6. Let $0 \le p_i \le 1$ for $i = 1, 2, \ldots, n$. Show that

$$\sum_{i=1}^{n} \frac{1}{|x - p_i|} \le 8n \left(1 + \frac{1}{3} + \frac{1}{5} + \cdots + \frac{1}{2n-1}\right)$$

for some x satisfying $0 \le x \le 1$.

Caveat. The OCR of the MAA volume drops the displayed inequality from the problem statement and prints it, garbled, beside Problem B-1. The statement above is reconstructed from those fragments and confirmed by the book's solution, which proves exactly $\sum_i 1/|x_j - p_i| \le B$ with $B = 8n(1 + \frac{1}{3} + \cdots + \frac{1}{2n-1})$.

Fields: analysis (inequalities), combinatorics (averaging / pigeonhole) **Position:** A6 — last of the paper; hard because the obvious averaging diverges, and the fix is a choice of discrete test points that the constant encodes **Answer:** Cut $[0,1]$ into $2n$ open intervals of length $1/(2n)$; at least n contain no p_i. At their midpoints $x_1, \ldots, x_n$, each p_i sees at most 2 midpoints closer than $3/(4n)$, at most 4 closer than $5/(4n)$, and so on, so $\sum_j 1/|x_j - p_i| \le B$; summing over i and averaging over j, some x_j has $\sum_i 1/|x_j - p_i| \le B$.

Layer 1 THE WHAT — surface analysis

The natural attempt. "For some x" — an existence statement with a sum. The reflex is averaging: if the average of $g(x) = \sum_i 1/|x - p_i|$ over $[0,1]$ is at most B, some x does at least as well. So you integrate:

$$\int_0^1 \frac{dx}{|x-p|} = \infty$$

for every $p \in [0,1]$, because of the log singularity at $x = p$. The average is infinite. The reflex is dead on arrival.

You patch it: avoid the singularities. Remove a small interval of radius δ around each p_i and average over what remains. Each removed interval costs length at most 2δ, so at least $1 - 2n\delta$ remains; each p_i contributes at most $\int_{\delta \le |x-p_i| \le 1} dx/|x-p_i| \le 2\log(1/\delta)$. Then the average of g over the remaining set is at most $2n\log(1/\delta)/(1-2n\delta)$. Take $\delta = 1/(4n)$: the bound is $4n\log(4n)$. That is the right order ($B \approx 4n\log n$), and it is a genuine proof of a bound of the right shape. But now you must compare $4n\log(4n)$ with $8n\sum_{h<n} 1/(2h+1)$ for every n, the continuous estimate has lost the odd reciprocals the problem is written in, and the constant bookkeeping for small n is fiddly. You have a proof of *a* theorem, not yet of *this* one.

The model chain. Write $H = \sum_{h=0}^{n-1} \frac{1}{2h+1}$ and $B = 8nH$.

Test points. For $k = 0, 1, \ldots, 2n-1$ let $I_k = \left(\frac{k}{2n}, \frac{k+1}{2n}\right)$, open. These $2n$ intervals are pairwise disjoint, so each p_i lies in at most one of them, and at most n of them contain some p_i. Choose n intervals containing no p_i and let $x_1, \ldots, x_n$ be their midpoints. Each $x_j \in [0,1]$, and every midpoint has the form $\frac{2k+1}{4n}$, so distinct midpoints differ by a positive multiple of $\frac{1}{2n}$.

One p_i against all test points. Fix i and write $d_{ij} = |x_j - p_i|$. Since p_i is not in the open interval of half-length $\frac{1}{4n}$ centred at x_j, $d_{ij} \ge \frac{1}{4n} > 0$; in particular every $1/d_{ij}$ is defined. For $m \ge 1$, the midpoints with $d_{ij} < \frac{2m+1}{4n}$ lie in

$$\left[p_i + \tfrac{1}{4n},\ p_i + \tfrac{2m+1}{4n}\right) \cup \left(p_i - \tfrac{2m+1}{4n},\ p_i - \tfrac{1}{4n}\right],$$

two half-open intervals of length $\frac{m}{2n}$ each. Points spaced by at least $\frac{1}{2n}$ can occupy at most m positions in a half-open interval of length $\frac{m}{2n}$. So at most $2m$ indices j have $d_{ij} < \frac{2m+1}{4n}$.

List $d_{i1}, \ldots, d_{in}$ in increasing order as $\delta_1 \le \delta_2 \le \cdots \le \delta_n$. For $r \ge 1$ write $r = 2m+1$ or $r = 2m+2$ with $m \ge 0$. At most $2m$ distances are below $\frac{2m+1}{4n}$ (for $m = 0$ this is the bound $d_{ij} \ge \frac{1}{4n}$), so $\delta_r \ge \frac{2m+1}{4n}$. Hence

$$\sum_{j=1}^{n} \frac{1}{d_{ij}} = \sum_{r=1}^{n} \frac{1}{\delta_r} \le \sum_{m=0}^{\lceil n/2 \rceil - 1} 2 \cdot \frac{4n}{2m+1} \le 2 \sum_{m=0}^{n-1} \frac{4n}{2m+1} = 8nH = B.$$

Average over test points. Summing over i,

$$\sum_{j=1}^{n}\Big(\sum_{i=1}^{n}\frac{1}{d_{ij}}\Big)=\sum_{i=1}^{n}\Big(\sum_{j=1}^{n}\frac{1}{d_{ij}}\Big)\le nB.$$

A sum of n non-negative terms at most nB has a term at most B: there is j with $\sum_i 1/|x_j-p_i|\le B$. Take $x=x_j$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains average. The natural attempt averages over the continuum $[0,1]$ (and then over $[0,1]$ minus neighbourhoods of the p_i). The model averages over a finite set of n points chosen in advance to sit on a grid, away from every p_i. The split is the line "$\int_0^1 dx/|x-p|=\infty$": the natural attempt patches the integral; the model replaces it with a sum.

> *The average over all x is infinite, so I need to average over points I've chosen to be far from every p_i. The right-hand side has $1,\frac{1}{3},\frac{1}{5},\ldots$ — odd numbers. Odd multiples of $\frac{1}{4n}$ are midpoints of a grid of spacing $\frac{1}{2n}$. So: cut $[0,1]$ into $2n$ pieces, and use the midpoints of the empty ones.*

> *I can't control the sum for one x directly, because I don't know where the p_i are. But I can swap the order: for each fixed p_i, the grid midpoints are spread out, so p_i's total over all of them is bounded no matter where p_i is. Then some midpoint gets at most the average.*

Layer 3 THE WHY

Why average at all? Because the claim is "for some x", and the p_i are arbitrary. You cannot locate a good x without knowing the p_i, and the only way to prove existence without locating is to show the average is small. That much the natural attempt has right.

Why does the average have to be *discrete*? Look at the definition of the function. Each term $1/|x-p_i|$ has a non-integrable pole: near p, $\int dx/|x-p|$ behaves like log, which diverges. A continuous average is dominated by the poles. The remedy is to average only over points that you know stay at distance $\gtrsim 1/n$ from every p_i — and such points exist because n points can only spoil n of $2n$ cells. That is pigeonhole: twice as many cells as points guarantees half the cells are empty.

Why swap the order of summation? The double sum $\sum_j\sum_i 1/d_{ij}$ is hard to read row by row (a row is one test point against all the unknown p_i) and easy column by column (a column is one p_i against a known, evenly spaced set of test points). An evenly spaced set is the one thing whose distances to an arbitrary point you can bound: at most two within one spacing, at most two more within the next, and so on. This is the discrete analogue of $\int_{1/(4n)}^1 dx/x$, and it produces exactly the harmonic-type sum. Fubini is the idea in the continuous version; interchanging a finite double sum is its honest discrete form.

Why the grid $\frac{1}{2n}$ and the odd reciprocals? Read the constant backwards. $\frac{4n}{2h+1}$ is $1/\text{distance}$ for the distance $\frac{2h+1}{4n}$ — an odd multiple of half a cell of width $\frac{1}{2n}$. The factor 2 in $8n=2\cdot 4n$ is "one neighbour on each side". So the right-hand side is the sum a point p would see if it sat at the centre of an empty cell with midpoints on both sides at every odd distance. The examiners wrote the proof into the constant. The official solution's opening — "let I_k be the open interval $(k/2n,(k+1)/2n)$" — is not a guess. It is read off B.

Why do $2n$ cells and n test points suffice with room to spare? The official solution notes "this inequality can be improved": the model chain uses only $\lceil n/2\rceil$ of the n pairs of terms available in B. The factor-two slack is the price of not caring where the empty cells are; it is what lets a crude count work.

Layer $+\alpha$ BACKTESTING

1. The cue is "for some x" in front of a sum of $1/|x-p_i|$ with arbitrary p_i. Write down: averaging, and the warning that $\int dx/|x-p|$ diverges. Then write down what the constant looks like: $4n/(2h+1)$ is a reciprocal distance on a grid of step $1/(2n)$.

2. Discrete averaging over a pigeonhole-chosen set of test points, with the order of summation reversed so each p_i is bounded against a spread-out set.

3. Hesitation occurs when the integral diverges. Reinforce it away: a divergent average means the averaging set is wrong, not that averaging is wrong. Choose a set that avoids the singularities.

4. The trap is to push the continuous average through with excised neighbourhoods and then fight to compare log with the odd harmonic sum for all n. A second trap: taking the x_j as midpoints of *all* $2n$ cells, some of which contain a p_i and give an unbounded term.

5. Any "there exists a point where a sum of singular terms is small": choose a grid of about twice as many cells as singularities, keep the empty cells, and bound each singularity's contribution against the grid.

6. Reusable rule: to show some choice makes a double-indexed sum small, sum over the choices, interchange the order, and bound each fixed term against all choices; if a continuous average diverges, replace it by a finite set of choices kept away from the singularities by pigeonhole. When a bound has an unusual constant, decode it into the grid or configuration that generates it.

▷ **Cue:** "for some x, a sum of $1/|x-p_i|$ is at most B" with arbitrary p_i $\longrightarrow$ **Move:** average over test points, not over $[0,1]$; the integral of $1/|x-p|$ diverges

▷ **Cue:** n points to avoid in an interval $\longrightarrow$ **Move:** cut into $2n$ cells; at least n cells are empty; use their midpoints

▷ **Cue:** a double sum $\sum_j \sum_i$ where rows are hard and columns easy $\longrightarrow$ **Move:** interchange the order and bound each column

▷ **Cue:** distances from an arbitrary point to an evenly spaced set $\longrightarrow$ **Move:** at most two within each successive half-step band; the reciprocal sum is harmonic-type

▷ **Cue:** a constant built from $1, \frac{1}{3}, \frac{1}{5}, \ldots$ times $4n$ $\longrightarrow$ **Move:** odd multiples of $\frac{1}{4n}$ are midpoints of a $\frac{1}{2n}$-grid; the constant names the grid

Putnam 1979 — Paper B

1979 B1

1979 B1. Prove or disprove: there is at least one straight line normal to the graph of $y = \cosh x$ at a point $(a, \cosh a)$ and also normal to the graph of $y = \sinh x$ at a point $(c, \sinh c)$. [At a point on a graph, the normal line is the perpendicular to the tangent at that point. Also, $\cosh x = (e^x + e^{-x})/2$ and $\sinh x = (e^x - e^{-x})/2$.]

Fields: analysis (hyperbolic functions, tangent/normal lines) **Position:** B1 – opening problem of the second session **Answer:** Disprove: no such common normal line exists.

Layer 1 THE WHAT — surface analysis

The natural attempt. You write down what a common normal requires. The tangent to $y = \cosh x$ at a has slope $\sinh a$, so the normal there has slope $-1/\sinh a$ (which forces $a \neq 0$, since at $a = 0$ the normal is vertical and cannot equal a normal to $\sinh x$, whose tangent slope $\cosh c \geq 1$ never gives a vertical normal). The tangent to $y = \sinh x$ at c has slope $\cosh c$, so its normal has slope $-1/\cosh c$. For one line to be normal to both curves you need the two normal lines to coincide, which needs (i) equal slopes, $\sinh a = \cosh c$, and (ii) the line through $(a, \cosh a)$ with that slope to actually pass through $(c, \sinh c)$. You write both conditions down and now have two equations in a, c. Since $\cosh c$ ranges over $[1, \infty)$ as c ranges over $\mathbb{R}$, and $\sinh a$ also ranges over all of $\mathbb{R}$ (in particular over $[1, \infty)$ for $a \geq \operatorname{arcsinh} 1$), condition (i) alone looks easily satisfiable – for any c there is an $a > 0$ with $\sinh a = \cosh c$. This makes the problem feel like it should have a "yes" answer, and you start trying to satisfy condition (ii) as well, substituting $\sinh a = \cosh c$ into the point-equation and attempting to solve the resulting transcendental relation between a and c directly – picking a value of c, computing the required a, checking whether the line happens to hit $(c, \sinh c)$. This is a search with no termination guarantee: you cannot try every c, and nothing in the algebra tells you when to stop looking.

The model chain. Impose both conditions symbolically instead of searching numerically. Equal slopes give

$$\sinh a = \cosh c. \tag{$*$}$$

The line through $(a, \cosh a)$ with slope $-1/\sinh a$ passes through $(c, \sinh c)$ exactly when

$$\sinh c - \cosh a = -\frac{1}{\sinh a}(c - a) \iff \sinh a\big(\sinh c - \cosh a\big) = a - c.$$

Substitute $(*)$ into the left side:

$$\cosh c\big(\sinh c - \cosh a\big) = a - c, \qquad \text{i.e.} \qquad \cosh c\big(\cosh a - \sinh c\big) = c - a. \tag{$**$}$$

Now bound both sides using one fact: $\cosh x - \sinh x = e^{-x} > 0$ for every real x, so $\sinh x < \cosh x$ always, with no exceptions and no case split on the sign of x.

From $(*)$, $\sinh a = \cosh c \geq 1 > 0$, and since sinh is increasing with $\sinh 0 = 0$, this forces $a > 0$.

Chain the inequality $\sinh x < \cosh x$ twice: $\sinh c < \cosh c$ (at $x = c$) and $\cosh c = \sinh a < \cosh a$ (at $x = a$, using $(*)$ again). So

$$\sinh c < \cosh c = \sinh a < \cosh a.$$

In particular $\cosh c < \cosh a$ with $a > 0$; since cosh is even and strictly increasing on $[0, \infty)$, $\cosh c < \cosh a$ forces $|c| < a$, hence $c < a$, hence $c - a < 0$.

But the same chain gives $\cosh a - \sinh c > 0$ (indeed $\cosh a > \sinh a = \cosh c > \sinh c$), and $\cosh c > 0$ always, so the left side of $(**)$, $\cosh c(\cosh a - \sinh c)$, is a product of two strictly positive numbers – strictly positive. The right side, $c - a$, is strictly negative. A positive number cannot equal a negative number, so $(**)$ has no solution. No common normal line exists.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains reach the same two conditions – equal slopes $(*)$ and the point-collinearity relation that becomes $(**)$ after substitution. They split right after that: the natural attempt treats $(**)$ as an equation to be solved for a pair (a, c), and starts searching numerically. The model treats $(**)$ as an equation whose two sides can be bounded in sign using only $\sinh x < \cosh x$, without ever solving for a or c individually.

> *I do not need to find a and c – I need to know whether the system is consistent, and consistency of an equation between two real expressions can be refuted by showing the two sides always disagree in sign, which is a strictly weaker (and cheaper) claim than solving it.*

> *Every quantity in this problem is built from* cosh *and* sinh *at real arguments, and there is exactly one universal inequality relating them,* $\sinh x < \cosh x$ *for all* x *– so before doing anything else I should see whether applying that one fact, twice, to the two arguments a and c already pins down the sign of both sides of my equation.*

Layer 3 THE WHY

Why does a single inequality, applied twice, settle a two-variable transcendental system that the natural attempt could not solve in closed form? Because the problem never actually asked for a and c – it asked whether a certain equation in a, c has *any* solution, and an equation between two real quantities has no solution whenever those quantities can be shown to lie in disjoint sign classes for every (a, c) satisfying the other constraint. The moment $(*)$ pins $a > 0$, the entire rest of the problem lives inside the single inequality $\sinh x < \cosh x$: this inequality is not an estimate that happens to be true here, it is the algebraic content of the definitions themselves, since $\cosh x - \sinh x = e^{-x}$ is not merely positive but is manifestly an exponential, hence positive for literally every real x with no case analysis required. Once you know one universal fact about the two functions in play, applying it is not a trick – it is the only structural handle the problem offers, and reaching for it before reaching for algebraic elimination is what separates "prove or disprove" problems (which reward finding an obstruction) from "solve" problems (which reward finding a value). The phrase "prove or disprove" is itself the cue: it tells you before you start that a clean nonexistence argument is being invited, not a construction, so the correct first question is "what quantity, forced to be simultaneously positive and negative, would kill this?" rather than "what are a and c?"

Layer $+\alpha$ BACKTESTING

1. The cue is the phrase "prove or disprove" applied to the existence of a line normal to two named curves at unspecified points. On seeing it, write down the two conditions a common normal line requires (equal slopes; the line through one point hits the other) as symbolic equations in the two free parameters, before trying to solve for them.

2. That cue suggests looking for an obstruction (a forced sign contradiction, a parity clash, a range mismatch) rather than an explicit construction, since the problem statement itself signals that non-existence is on the table.
3. Hesitation occurs right after deriving the slope condition $\sinh a = \cosh c$, because both ranges overlap and the system looks satisfiable at a glance; it is reinforced away by asking "have I used the *point* condition, or only the slope condition?" – range overlap of one condition says nothing about the conjunction of two conditions.
4. The trap is mistaking "the two conditions look individually satisfiable" for "the system is satisfiable," and responding by hunting for explicit values instead of testing the conjunction algebraically.
5. Whenever a hyperbolic (or otherwise paired) function pair appears with a fixed identity between them ($\cosh^2 - \sinh^2 = 1$, or here $\cosh x - \sinh x = e^{-x} > 0$), recognise instantly that a universal inequality is available and try applying it to every argument in the problem before attempting to solve anything explicitly.
6. Reusable rule: to refute the existence of a solution to a coupled system of two real equations, do not attempt to solve the system; substitute one equation into the other to get a single equation, then try to show its two sides are forced into strictly disjoint sign classes using whatever universal inequality the problem's functions supply.

▷ **Cue:** "prove or disprove" that a geometric object (line, point, curve) exists satisfying two simultaneous conditions on named functions $\longrightarrow$ **Move:** write both conditions as symbolic equations in the free parameters, substitute one into the other, and look for a forced sign contradiction rather than solving explicitly

▷ **Cue:** cosh and sinh of possibly different real arguments appear together in an equation $\longrightarrow$ **Move:** use $\cosh x - \sinh x = e^{-x} > 0$ for all real x as the universal comparison tool, applied at each argument separately

▷ **Cue:** one condition (e.g. equal slopes) looks satisfiable because two ranges overlap $\longrightarrow$ **Move:** do not stop there – substitute that condition into the remaining condition and test the conjunction, since range overlap of a single equation is not evidence for the full system

▷ **Cue:** need to compare $\cosh x$ at two different arguments, say $\cosh c < \cosh a$ with $a > 0$ $\longrightarrow$ **Move:** use that cosh is even and strictly increasing on $[0, \infty)$ to conclude $|c| < a$

1979 B2

1979 B2. Let $0 < a < b$. Evaluate

$$\lim_{t \to 0} \left[\int_0^1 \left(bx + a(1-x) \right)^t dx \right]^{1/t}.$$

[The final answer should not involve any operations other than addition, subtraction, multiplication, division, and exponentiation.]

Fields: analysis (definite integrals, L'Hôpital's rule, indeterminate limits) **Position:** B2 – second problem of the second session **Answer:** $\frac{1}{e} \left(\frac{b^b}{a^a} \right)^{1/(b-a)}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You see a limit of the form $[\,\cdot\,]^{1/t}$ as $t \to 0$ and immediately recognise the shape "1^∞": at $t = 0$ the integral is $\int_0^1 1\,dx = 1$, and you're raising something tending to 1 to a power tending to ∞. The standard move is to take logs: let $I(t) = \int_0^1 (bx + a(1-x))^t\,dx$ and compute $\lim_{t \to 0} \frac{\ln I(t)}{t}$ by L'Hôpital, which needs $I'(t)$. Differentiating under the integral

sign, $I'(t) = \int_0^1 (bx + a(1-x))^t \ln(bx + a(1-x))\,dx$, and you need this at $t = 0$: $I'(0) = \int_0^1 \ln(bx + a(1-x))\,dx$. This is a genuine integral you now have to evaluate – substitute $u = bx + a(1-x) = a + (b-a)x$, so $du = (b-a)\,dx$, and

$$I'(0) = \frac{1}{b-a}\int_a^b \ln u\,du = \frac{1}{b-a}\Big[u\ln u - u\Big]_a^b = \frac{b\ln b - a\ln a - (b-a)}{b-a}.$$

This works, and gets you to the right place – but only because you happened to differentiate under the integral sign correctly and then still had a second integral, $\int \ln u\,du$, to do by parts. That is two nontrivial calculus operations (differentiating a parametrised integral, then integrating by parts) stacked on top of the L'Hôpital bookkeeping, and it is easy to make a sign slip in the fundamental-theorem-of-calculus evaluation or lose track of which variable is being integrated against which.

The model chain. Do the elementary substitution *before* taking any limit or derivative, not after. Set $u = bx + a(1-x) = a + (b-a)x$; as x runs over $[0,1]$, u runs over $[a,b]$, and $du = (b-a)\,dx$. So

$$I(t) = \int_0^1 \big(bx + a(1-x)\big)^t\,dx = \frac{1}{b-a}\int_a^b u^t\,du = \frac{b^{t+1} - a^{t+1}}{(t+1)(b-a)}.$$

Now $I(t)$ is an explicit elementary function of t – a ratio of two smooth functions of t, no integral sign left. Write $I(t) = N(t)/D(t)$ with $N(t) = b^{t+1} - a^{t+1}$ and $D(t) = (t+1)(b-a)$, both trivial to differentiate: $N'(t) = b^{t+1}\ln b - a^{t+1}\ln a$, $D'(t) = b - a$.

We want $L = \lim_{t\to 0} I(t)^{1/t}$. Since $I(0) = \dfrac{b-a}{b-a} = 1$ (an explicit check, not an assumption), this is 1^∞, and $\ln L = \lim_{t\to 0} \dfrac{\ln I(t)}{t}$, which by L'Hôpital (both numerator and denominator vanish at $t = 0$) equals $\dfrac{I'(0)}{I(0)} = I'(0)$.

Using the quotient rule at $t = 0$, where $N(0) = b - a = D(0)$:

$$I'(0) = \frac{N'(0)D(0) - N(0)D'(0)}{D(0)^2} = \frac{(b\ln b - a\ln a)(b-a) - (b-a)(b-a)}{(b-a)^2} = \frac{b\ln b - a\ln a - (b-a)}{b-a}.$$

So

$$\ln L = \frac{b\ln b - a\ln a}{b-a} - 1, \qquad L = \exp\left(\frac{b\ln b - a\ln a}{b-a}\right)\cdot e^{-1}.$$

Since $b\ln b - a\ln a = \ln(b^b) - \ln(a^a) = \ln(b^b/a^a)$, the first factor is $(b^b/a^a)^{1/(b-a)}$. Hence

$$L = \frac{1}{e}\left(\frac{b^b}{a^a}\right)^{1/(b-a)}.$$

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree that the limit is 1^∞ and that the resolution is $L = \exp\big(\lim_{t\to 0} \ln I(t)/t\big)$, computed by L'Hôpital, which needs $I'(0)$. They split at the very first step – what to do with the integral defining $I(t)$ before differentiating it. The natural attempt keeps $I(t)$ as an integral and differentiates it with respect to t under the integral sign, producing a second integral, $\int_0^1 \ln(bx + a(1-x))\,dx$, that must then itself be evaluated by substitution and integration by parts. The model performs the substitution $u = a + (b-a)x$ *first*, evaluating the x-integral in closed form for general t, so that $I(t)$ is already an elementary function of t before any derivative is taken – differentiating it is then bookkeeping, not integration.

> *The integrand $(bx + a(1-x))^t$ is a power of a* linear *function of x – I can substitute that linear function itself, $u = bx + a(1-x)$, and turn the whole integral into $\int u^t\,du$, which I can antidifferentiate for a general exponent t in one line. I should do this substitution before touching t at all, because it removes the integral entirely and leaves ordinary single-variable calculus in t.*

> *Differentiating under the integral sign is legitimate here, but it is not free – it trades "differentiate a function of t" for "evaluate a new integral in x," and I only need $I'(0)$, a single number, not $I'(t)$ as a function. If I can get a closed form for $I(t)$ itself first, I never have to integrate again; I only differentiate an explicit formula, which is strictly cheaper than integrating* $\ln(\cdot)$ *by parts.*

Layer 3 THE WHY

Why is substituting first so clearly the right order, rather than a matter of taste? Because the integrand's dependence on x and its dependence on t are not entangled in the way they look – the expression $bx + a(1-x)$ is a fixed linear function of x that does not involve t at all, and t only ever acts on it from outside, as an exponent. Substituting $u = bx + a(1-x)$ is therefore not a trick discovered by trial; it is what "change of variables" means whenever an integral's integrand is $f(\text{linear in } x)$ for some function f depending on a separate parameter – the substitution separates the roles of x (which only parametrises which value of $u \in [a,b]$ you are at) from t (which only parametrises which power of u you are raising). Once separated, $\int_a^b u^t\,du = \dfrac{b^{t+1} - a^{t+1}}{t+1}$ is not an estimate or an approximation, it is the exact antiderivative of a power function, valid for every real $t \neq -1$ – the single most basic integral in single-variable calculus. The natural attempt's route works, but it works by *postponing* the linear substitution until after differentiating in t, at which point the substitution still has to be done (it is exactly the step that evaluates $\int_0^1 \ln(bx + a(1-x))\,dx$), so no work is saved and a second nontrivial technique (integration by parts) is now needed as well. The general principle: when an operation (here, substitution) will have to be performed eventually regardless of order, perform it as early as possible, before any other operation (here, differentiation in t) has a chance to multiply the amount of calculus required downstream.

Layer $+\alpha$ BACKTESTING

1. The cue is an integral $\int_0^1 f(x)^t\,dx$ where f is linear in x and t appears only as an exponent, evaluated in the limit $t \to 0$ giving a 1^∞ form. On seeing it, the useful first move to write down is: substitute $u = f(x)$ to turn the integral into $\int u^t\,du$ over the range of f, before touching t.

2. That cue suggests evaluating $I(t)$ in closed form as an elementary function of t first, then applying $L = \exp(\lim \ln I(t)/t)$ and ordinary (not under-the-integral-sign) L'Hôpital/differentiation to that closed form.

3. Hesitation occurs right at the start, deciding whether to differentiate under the integral sign immediately (since that is the textbook route to $\lim \ln I(t)/t$) or to simplify the integral first; it is reinforced away by asking "is the integrand a fixed function of x composed with something that depends only on t? If so, can I do the x-integral for general t before differentiating?"

4. The trap is applying differentiation-under-the-integral-sign reflexively because the limit has the classic 1^∞/L'Hôpital shape, without first checking whether the integral itself has an elementary closed form in t that would make differentiating under the integral sign unnecessary.

5. Whenever the integrand is (a function depending only on the exponent t) applied to (a fixed expression in x that does not involve t), recognise instantly that a substitution turning x

into that fixed expression will produce a closed form in t, and do the substitution before any t-differentiation.

6. Reusable rule: for $\lim_{t\to 0}\left[\int f(x)^t\,dx\right]^{1/t}$ type limits where f does not depend on t, first evaluate the integral in closed form as a function of t by substituting $u = f(x)$; only then take logs and apply L'Hôpital to the resulting elementary function – never differentiate under the integral sign before checking whether the integral can be closed first.

▷ **Cue:** integral $\int f(x)^t\,dx$ with t appearing only as an exponent and f not depending on t $\longrightarrow$ **Move:** substitute $u = f(x)$ to reduce to $\int u^t\,du$, obtaining a closed form in t before differentiating

▷ **Cue:** a limit of the form $[\,\cdot\,]^{1/t}$ as $t \to 0$ where the bracket $\to 1$ $\longrightarrow$ **Move:** recognise 1^∞, take $L = \exp(\lim_{t\to 0} \ln(\cdot)/t)$, and apply L'Hôpital to the resulting $0/0$

▷ **Cue:** need to differentiate a function of t that is currently an unevaluated integral in x $\longrightarrow$ **Move:** check first whether the x-integral has an elementary closed form for general t; if so, differentiate that closed form rather than differentiating under the integral sign

▷ **Cue:** expression of the form $b\ln b - a\ln a$ appears in an exponent $\longrightarrow$ **Move:** rewrite as $\ln(b^b/a^a)$ so the exponential becomes $(b^b/a^a)^{\text{power}}$, matching the required answer format of only $+, -, \times, \div, ^\wedge$

1979 B3

1979 B3. Let F be a finite field having an odd number m of elements. Let $p(x)$ be an irreducible (i.e., nonfactorable) polynomial over F of the form

$$p(x) = x^2 + bx + c, \qquad b, c \in F.$$

For how many elements k in F is $p(x) + k$ irreducible over F?

Fields: abstract algebra (finite fields, irreducible polynomials, quadratic residues) **Position:** B3 – third problem of the second session **Answer:** $(m-1)/2$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You know the discriminant test for quadratics from ordinary algebra: $x^2 + bx + d$ factors over a field (of characteristic $\neq 2$) exactly when $b^2 - 4d$ is a square in that field. You try to import this wholesale, and it is in fact valid – m odd forces the characteristic of F to be an odd prime, so $2 \neq 0$ in F and completing the square is legitimate. But now you have to count, for $d = c + k$ ranging as k ranges over F, how many values of d make $b^2 - 4d$ a *non*-square in the abstract finite field F – and "how many elements of a finite field are squares" is not something you can answer by picturing a number line. You reach for small examples: take $F = \mathbb{Z}/5\mathbb{Z}$, list the squares ($0, 1, 4, 4, 1$ for $x = 0, 1, 2, 3, 4$, i.e. the square values are $\{0, 1, 4\}$, three of the five elements), and guess a pattern – roughly half the elements are squares. That guess is correct, but turning "roughly half, for one example field" into a proof for a general odd m means either (a) proving the general fact that a finite field of odd order m has exactly $(m+1)/2$ squares (itself a small argument, via the 2-to-1 map $x \mapsto x^2$ on F^* off of 0), or (b) trying to track, for each of the m values of k, whether $b^2 - 4(c+k)$ is a square directly, which just restates the problem without simplifying it. You can feel that the answer should be $(m-1)/2$ from the pattern, but assembling a rigorous count of "how many of m shifted values are non-squares" without an extra idea feels like it needs its own lemma, proved from scratch, before you can even start on the actual counting.

The model chain. Complete the square, exactly as over $\mathbb{R}$: since m is odd, the characteristic of F is odd, so 2 is invertible in F. Set $h = b/2 \in F$. Then for any $d \in F$,

$$x^2 + bx + d = (x + h)^2 - (h^2 - d),$$

and this factors over F (has a root, hence a linear factor, since it is quadratic) exactly when $h^2 - d$ is a square in F – if $h^2 - d = s^2$ then $(x+h)^2 - s^2 = (x+h-s)(x+h+s)$; conversely a root r gives $(r+h)^2 = h^2 - d$. So $x^2 + bx + d$ is irreducible over F if and only if $h^2 - d$ is *not* a square in F.

Apply this with $d = c$: $p(x)$ is irreducible, so $h^2 - c$ is not a square in F. Apply it with $d = c + k$: $p(x) + k = x^2 + bx + (c+k)$ is irreducible exactly when $h^2 - (c+k) = (h^2 - c) - k$ is not a square in F.

Now the key move: as k ranges over all m elements of F, the quantity $y = (h^2 - c) - k$ also ranges over all m elements of F, hitting each one exactly once – the map $k \mapsto (h^2 - c) - k$ is a bijection of F with itself (it is affine with invertible "slope" -1). So counting "how many k make y a non-square" is exactly counting "how many elements y of F are non-squares," full stop, with no reference to b, c, h left at all.

Count the non-squares in F: the squaring map $s \mapsto s^2$ on $F^* = F \setminus \{0\}$ (which has $m-1$ elements) is exactly 2-to-1, since $s^2 = t^2 \iff s = \pm t$ and $s \neq -s$ (as $\operatorname{char} F \neq 2$ and $s \neq 0$). So the number of nonzero squares is $(m-1)/2$, the number of squares including 0 is $1 + (m-1)/2 = (m+1)/2$, and the number of non-squares is $m - (m+1)/2 = (m-1)/2$.

Hence the number of $k \in F$ for which $p(x) + k$ is irreducible is $(m-1)/2$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains reach the same reduction: irreducibility of $x^2 + bx + d$ over F (odd characteristic) is equivalent to $h^2 - d$ being a non-square. They split immediately after that, at the point of actually counting how many of the m shifted values $d = c + k$ give a non-square. The natural attempt tries to compute this count example-by-example or value-by-value, in effect re-deriving the problem's own casework. The model observes that $k \mapsto (h^2 - c) - k$ is a bijection $F \to F$, which converts "count k such that $(h^2 - c) - k$ is a non-square" into "count non-squares in F" – a fact about the field alone, independent of b, c, k.

> *I don't need to track which specific k works – I need a* count*, and the map from k to the quantity I'm testing, $(h^2 - c) - k$, is a bijection of F with itself, so the count of k's with a given property equals the count of field elements with that property, period. The problem "how many k" silently converts into "how many elements of F" the instant I notice the map is a bijection.*

> *The fact that must be true here is a fact about the field F in isolation – "how many elements are non-squares" – not a fact entangled with b, c, or h; once I see the bijection, every trace of b, c, h can be discarded, which tells me I am now solving a strictly easier, purely structural sub-problem.*

Layer 3 THE WHY

Why does completing the square work here at all, and why is the bijection not a coincidence but forced? Completing the square requires dividing by 2, which is legitimate precisely when $2 \neq 0$ in the field – and the hypothesis "F has an *odd* number m of elements" is not decoration, it is exactly the condition that forces the characteristic of F to be an odd prime (the characteristic must divide $m = |F|$, and if it were 2, m would be a power of 2, hence even). So the very first structural fact the problem hands you, "m is odd," is the license for the one algebraic manipulation the whole solution rests on; noticing that is the difference between treating "m odd" as flavour text and treating it as an instruction.

Why is the bijection the right tool for the counting step, rather than a lucky shortcut? Because a bijection between two finite sets is, by definition, the statement that they have the same size – and more specifically, that a subset defined by a property on one side corresponds exactly to the subset defined by the same property on the other side. The map $k \mapsto (h^2 - c) - k$ is affine with coefficient $-1 \neq 0$, hence invertible, hence a bijection of the finite set F; this is not special to this problem, it is true of every map $y = \alpha - k$ on a field with α fixed. Once you recognise

the quantity you are testing ("is y a square?") depends on k only through this one bijective substitution, the count you want is unconditionally the same as the count of non-squares in F itself – a fact that does not depend on b, c, or the specific irreducible $p(x)$ you started with. This is why the answer $(m-1)/2$ has no b, c in it: the problem is engineered so that after completing the square, the entire remaining question is a property of F as a set with a squaring map, and the count of non-squares in a field of odd order m is a 2-to-1-map argument that would answer the same question for *any* irreducible quadratic, not just this one.

Layer $+\alpha$ BACKTESTING

1. The cue is a quadratic polynomial over a field of odd characteristic (signalled here by "m odd") whose irreducibility is being tested. On seeing it, write down the completed-square form $(x+h)^2-(h^2-d)$ immediately, since 2 being invertible is guaranteed and the discriminant/square test transfers unchanged from $\mathbb{R}$.
2. That cue suggests translating "is x^2+bx+d irreducible?" into "is h^2-d a non-square in F?", then treating the varying parameter (k here) as acting on the quantity being tested for squareness, not on the polynomial.
3. Hesitation occurs right after the reduction to a square/non-square question, when it is tempting to check cases (specific small fields, specific k) instead of asking whether the map from k to the tested quantity is itself invertible; it is reinforced away by asking explicitly "does the parameter I'm counting over act as a bijection on the thing I'm testing?"
4. The trap is treating "count how many k satisfy a condition on $(h^2-c)-k$" as a problem tied to the specific values b, c, and trying to compute case by case, instead of noticing the substitution is a bijection that erases b, c from the counting question entirely.
5. Whenever a counting problem over a finite field or finite set asks "for how many values of a parameter does [property depending affinely on that parameter] hold," check immediately whether the map from the parameter to the tested quantity is a bijection; if so, the count reduces to counting the property directly on the whole set.
6. Reusable rule: to count solutions to a shifted or translated condition over a finite algebraic structure, first check whether the shift/translation map is a bijection of the structure with itself; if it is, replace "count parameter values giving the property" with "count elements of the structure with the property," which is usually a pure structural fact (here, the standard 2-to-1 squaring-map count of quadratic residues) independent of the specific instance.

▷ **Cue:** quadratic x^2+bx+d over a field of odd characteristic, irreducibility being tested $\longrightarrow$ **Move:** complete the square, $(x+h)^2-(h^2-d)$ with $h=b/2$; irreducible iff h^2-d is a non-square

▷ **Cue:** "F has an odd number m of elements" as a hypothesis $\longrightarrow$ **Move:** read this as "char $F \neq 2$," licensing division by 2 and the usual discriminant/completing-the-square manipulations

▷ **Cue:** counting how many values of a parameter k make an affine expression in k have a given property $\longrightarrow$ **Move:** check whether $k \mapsto$ (expression) is a bijection of the ambient set; if so, the count equals the count of elements of the ambient set with that property, independent of the other constants involved

▷ **Cue:** need the number of squares (or non-squares) in a finite field of odd order m $\longrightarrow$ **Move:** use that $s \mapsto s^2$ is exactly 2-to-1 on F^* (since $s^2=t^2 \iff s=\pm t$ and $s \neq -s$ when char $\neq 2, s \neq 0$), giving $(m-1)/2$ nonzero squares, $(m+1)/2$ squares including 0, and $(m-1)/2$ non-squares

1979 B4

1979 B4. (a) Find a solution that is not identically zero, of the homogeneous linear differ-

ential equation

$$(3x^2+x-1)y''-(9x^2+9x-2)y'+(18x+3)y=0.$$

Intelligent guessing of the form of a solution may be helpful.

(b) Let $y=f(x)$ be the solution of the nonhomogeneous differential equation

$$(3x^2+x-1)y''-(9x^2+9x-2)y'+(18x+3)y=6(6x+1)$$

that has $f(0)=1$ and $(f(-1)-2)(f(1)-6)=1$. Find integers a, b, c such that $(f(-2)-a)(f(2)-b)=c$.

Fields: differential equations; linear ODEs with polynomial coefficients; undetermined coefficients
Position: B4 — middle of the paper; almost pure computation once the right trial functions are chosen
Answer: (a) $y=e^{3x}$ and $y=x^2+x$ both work, so any nonzero $he^{3x}+k(x^2+x)$. (b) $f(x)=2-e^{3x}+2x^2+2x$, so $f(-2)=6-e^{-6}$, $f(2)=14-e^6$, and $(a,b,c)=(6,14,1)$.

Caveat. The OCR of the book renders the right-hand side of (b) as "$6(6x+1)$" with some garbling and the conditions as "$f(0)-1$", "$(f(-1)-2)(f(1)-6)-1$"; the $-$ signs are OCR renderings of $=$. The reading above was checked against the book's own solution, which verifies $f(0)=1$, $(f(-1)-2)(f(1)-6)=1$ and the particular solution $y=2$, and against the index entry "Differential equations — homogeneous/nonhomogeneous linear 1979 B-4". All three trial solutions were re-verified by direct substitution below.

Layer 1 THE WHAT — surface analysis

The natural attempt. A strong student sees a second-order linear ODE with variable coefficients and reaches for the syllabus. The syllabus says: constant coefficients means an auxiliary equation; variable coefficients means power series or a substitution. So the natural first move is a power series $y=\sum c_n x^n$. Substitute and collect x^n:

$$3n(n-1)c_n+(n+1)n\,c_{n+1}-(n+2)(n+1)c_{n+2}-9(n-1)c_{n-1}-9n\,c_n+2(n+1)c_{n+1}+18c_{n-1}+3c_n=0.$$

That is a three-term recurrence for c_{n+2} in terms of c_{n+1}, c_n, c_{n-1}. It is correct, and it is a swamp. You can run it for a few terms from $c_0=1, c_1=0$ and see nothing you recognise, because the true solutions mix e^{3x} (infinitely many terms) with a quadratic, and a generic starting pair produces a generic mixture. After twenty minutes you have a recurrence, no closed form, and part (b) — which needs exact values at $x=\pm 2$ — is out of reach.

The second natural attempt is the textbook one for part (b): variation of parameters. But that needs two independent homogeneous solutions first, which is exactly what you do not yet have. The syllabus machinery has no entry point.

The model chain. Write $L[y]=(3x^2+x-1)y''-(9x^2+9x-2)y'+(18x+3)y$. The problem says to guess, so guess from the two families the operator treats best.

Exponential trial. For $y=e^{mx}$, $L[y]=e^{mx}\big(m^2(3x^2+x-1)-m(9x^2+9x-2)+18x+3\big)$. The x^2 coefficient is $3m^2-9m$, so $m\in\{0,3\}$. For $m=0$ the bracket is $18x+3\neq 0$. For $m=3$:

$$9(3x^2+x-1)-3(9x^2+9x-2)+18x+3=(27-27)x^2+(9-27+18)x+(-9+6+3)=0.$$

So $y_1=e^{3x}$.

Polynomial trial. For $y=x^d+\cdots$ the top power in $L[y]$ is x^{d+1}, from $-9x^2\cdot dx^{d-1}+18x\cdot x^d$, with coefficient $18-9d$. So $d=2$. Try $y=x^2+px+q$. The x^3 terms cancel automatically; the x^2 coefficient is $9p-9$, the x coefficient is $6-6p+18q$ and the constant is $-2+2p+3q$, so $p=1$, $q=0$. Check $y_2=x^2+x$ in full:

$$2(3x^2+x-1)-(9x^2+9x-2)(2x+1)+(18x+3)(x^2+x)$$

$$= (6x^2 + 2x - 2) - (18x^3 + 27x^2 + 5x - 2) + (18x^3 + 21x^2 + 3x) = 0.$$

That answers (a).

Particular solution. The right-hand side is $6(6x + 1) = 2(18x + 3)$, twice the coefficient of y. A constant $y = C$ gives $L[C] = C(18x + 3)$, so $y_p = 2$.

Every solution has this form. The Wronskian is

$$W(e^{3x}, x^2 + x) = e^{3x}(2x + 1) - 3e^{3x}(x^2 + x) = -e^{3x}(3x^2 + x - 1).$$

On each interval where $3x^2 + x - 1 \neq 0$ the equation is regular and $W \neq 0$, so there every solution is $2 + he^{3x} + k(x^2 + x)$. The leading coefficient vanishes at $x_\pm = (-1 \pm \sqrt{13})/6 \approx 0.434, -0.768$, and the points $-2, -1, 0, 1, 2$ straddle both. Could f use different constants on different intervals? The difference of two pieces glued at $x_0 \in \{x_\pm\}$ would be a homogeneous solution d with $d(x_0) = d'(x_0) = 0$; since $W(x_0) = 0$ there is exactly one such d up to scale, namely $d = (x_0^2 + x_0)e^{3x} - e^{3x_0}(x^2 + x)$, and $d''(x_0) = e^{3x_0}(9x_0^2 + 9x_0 - 2) = e^{3x_0}(1 + 6x_0) \neq 0$ (using $3x_0^2 = 1 - x_0$). So a glued function has a jump in y'' at x_0 and is not a solution. Hence a solution on $\mathbb{R}$ is a single global $f = 2 + he^{3x} + k(x^2 + x)$.

Constants. $f(0) = 2 + h = 1$ gives $h = -1$. Then $f(-1) - 2 = -e^{-3}$ and $f(1) - 6 = 2k - 4 - e^3$, so

$$-e^{-3}(2k - 4 - e^3) = 1 \iff (2k - 4)e^{-3} = 0 \iff k = 2.$$

So $f(x) = 2 - e^{3x} + 2x^2 + 2x$, $f(-2) = 6 - e^{-6}$, $f(2) = 14 - e^6$, and $(f(-2) - 6)(f(2) - 14) = (-e^{-6})(-e^6) = 1$. Take $a = 6$, $b = 14$, $c = 1$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start from the same equation. They split at the first line: the natural attempt writes $y = \sum c_n x^n$, a trial that fits *every* analytic solution and therefore identifies none. The model writes $y = e^{mx}$ and $y = x^d$, two trials each of which turns $L[y]$ into a polynomial whose *leading* coefficient alone pins down the parameter.

> *The problem tells me to guess. A guess is only intelligent if one coefficient decides it. For e^{mx} everything is e^{mx} times a quadratic in x, so the x^2 coefficient $3m^2 - 9m$ must vanish: $m = 3$. For x^d the top power is x^{d+1} with coefficient $18 - 9d$: $d = 2$.*

> *The right-hand side $6(6x + 1)$ is twice $18x + 3$, the coefficient of y. So a constant does it: $y = 2$. I should look at the forcing term against the coefficients before I reach for variation of parameters.*

> *The final expression pairs $f(-2)$ with $f(2)$, and the answer must be integers. The transcendental parts $e^{\pm 6}$ must cancel in a product, which tells me a and b are the polynomial parts and $c = e^{-6}e^6 = 1$.*

- ▷ **Cue:** a linear ODE whose coefficients are polynomials in x of degree at most 2, with the instruction "guess" ⟶ **Move:** try e^{mx} and read m off the highest power of x; try x^d and read d off the highest power of x
- ▷ **Cue:** the forcing term is a constant multiple of the coefficient of y ⟶ **Move:** a constant is a particular solution
- ▷ **Cue:** a condition like $(f(p) - u)(f(q) - v) = 1$ with $e^{\pm}$ terms ⟶ **Move:** the exponentials must cancel as $e^s e^{-s}$; subtract the polynomial parts

Layer 3 THE WHY

Why is the exponential-plus-polynomial guess not a trick? Look at what L does to each family. Differentiation maps e^{mx} to a multiple of itself and lowers the degree of x^d by one. Multiplication by a coefficient of degree j raises degree by j. So L sends e^{mx} to $e^{mx} \cdot$ (polynomial of degree $\leq$

2) and sends a degree-d polynomial to a polynomial of degree at most $d+1$. In both cases the output lives in a family where "is zero" means "every coefficient is zero", and the highest coefficient is the cheapest one to compute. That is exactly the auxiliary-equation idea from the constant-coefficient case, with one change: the auxiliary equation is now read off the top coefficient instead of being the whole story. The other coefficients become checks. The guess is intelligent because it has one free parameter and one equation to fix it.

The power series fails for the opposite reason. It has infinitely many free parameters and produces a recurrence that encodes the whole two-dimensional solution space at once. It is a tool for proving that solutions exist, not for finding the ones you can name.

Why did the setters choose these coefficients? Because e^{3x} and x^2+x are the solutions and the equation was built backwards from them. The Wronskian gives it away: $W=-e^{3x}(3x^2+x-1)$, so the leading coefficient *is* (minus) the Wronskian divided by e^{3x}. Abel's formula says $W = C\exp\big(\int(9x^2+9x-2)/(3x^2+x-1)\,dx\big)$, and $(9x^2+9x-2) = 3(3x^2+x-1)+(6x+1)$, whose integral is $3x+\ln|3x^2+x-1|$. So Abel's formula alone predicts $W \propto e^{3x}(3x^2+x-1)$, and an e^{3x} in the Wronskian is a strong hint that e^{3x} is a solution. If the guess had failed, this is the calculation that would have rescued you.

The official solution says "trial of e^{mx} shows" and "trial of a polynomial $x^d+\cdots$ shows that d must be 2" without saying which coefficient decides. The deciding coefficient is the top one in each case, as above. It also asserts that f has the form $2+he^{3x}+k(x^2+x)$ without comment, although the equation is singular at two points between -2 and 2. The gluing argument in the model chain is what makes "*the* solution" well defined on $\mathbb{R}$: the only homogeneous solution vanishing to first order at a singular point has nonzero second derivative there, so no piecewise solution exists.

Finally, part (b)'s strange condition $(f(-1)-2)(f(1)-6)=1$ is designed so that x^2+x vanishes at $x=-1$, isolating h in one factor and k in the other. The target $(f(-2)-a)(f(2)-b)=c$ repeats the design: x^2+x is 2 at -2 and 6 at 2, so $2x^2+2x$ contributes 4 and 12, and $a=2+4$, $b=2+12$.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "Intelligent guessing of the form of a solution may be helpful" next to polynomial coefficients of degree 2. Write down $L[e^{mx}]$ and $L[x^d]$ and circle their top coefficients.
2. Undetermined coefficients with exponential and polynomial trials; then a constant particular solution; then two conditions for two constants.
3. Hesitation comes when the problem has variable coefficients and the student feels that exponentials "belong" to constant-coefficient equations. Reinforce it away with the observation that e^{mx} is an eigenfunction of d/dx whatever the coefficients are, so $L[e^{mx}]/e^{mx}$ is always a polynomial.
4. The trap is the power series: it is correct, it is what the syllabus suggests, and it never ends. A second trap is ignoring the singular points; the official solution does, and so will most contestants, but a careful answer should say why the global form is forced.
5. When an operator maps a family of functions into a family where vanishing is checked coefficient by coefficient, the leading coefficient chooses the parameter and the rest are checks. Look for families the operator preserves before looking for series.
6. Reusable decision rule: *before solving a linear equation in general, test the functions the operator acts on predictably (eigenfunctions of the differentiations, or polynomials whose degree is controlled), and fix each trial's parameter from the top coefficient alone. Compare the forcing term with the coefficients before any general method.*

1979 B5

1979 B5. In the plane, let C be a closed convex set that contains $(0,0)$ but no other point with integer coordinates. Suppose that $A(C)$, the area of C, is equally distributed among the four quadrants. Prove that $A(C) \leq 4$.

Fields: geometry; convex sets; lattice points; support lines **Position:** B5 — second hardest on the paper; a lattice-point area bound in the spirit of Minkowski, but one that needs no theorem, only one support line per neighbour of the origin **Answer:** Through each of $(\pm 1, 0), (0, \pm 1)$ draw a line that has C on one side. If one of these lines is steep, the quadrant beside it contains at most a triangle of area ≤ 1. Otherwise the four lines cut off, in some quadrant, a quadrilateral with vertices $(0,0), (1,0), (h,k), (0,1)$ whose angle at (h,k) is at least 90°; that forces $h + k \leq 2$ and area ≤ 1. So $A(C)/4 \leq 1$. In fact $A(C) < 4$.

Caveat. The OCR prints the conclusion as "$A(C) < 4$" in both the statement and the book's solution, and the same OCR renders $\leq$ as $<$ elsewhere on the page (for instance in Cauchy–Schwarz in 1979 B-6). The standard wording is "Prove that $A(C) \leq 4$", used above. The model chain proves $A(C) \leq 4$ and then shows the inequality is in fact strict, so the analysis is correct under either reading. The book's solution also says "the part of the area of C in the fourth quadrant" in the steep case; with the support line through $(0,1)$ of slope $m \geq \frac{1}{2}$ the quadrant that is controlled is the second, and the text below uses the corrected quadrant.

Layer 1 THE WHAT — surface analysis

The natural attempt. A strong student sees "convex", "no lattice points" and "area" and thinks of Minkowski's theorem: a convex set symmetric about the origin with area > 4 contains a nonzero lattice point. The number 4 in the problem matches. So the natural attempt is to make C symmetric. Take $D = \frac{1}{2}(C - C) = \{\frac{1}{2}(p - q) : p, q \in C\}$, which is convex and symmetric. If D contained a nonzero lattice point $u = \frac{1}{2}(p - q)$ then… nothing: $p - q = 2u$ does not put a lattice point in C. The standard trick gives lattice points in $C - C$ only, and $C - C$ has area up to $6A(C)$ (Rogers–Shephard), not $4A(C)$. The quadrant hypothesis is never used. That is the wall: the problem has an extra hypothesis, and a proof that does not use it is proving a false statement (the triangle with vertices $(0,0)$, $(0, \frac{1}{2})$, $(N, \frac{1}{2})$ meets the lattice only at the origin and has area $N/4$).

The second natural attempt is to fix a quadrant and ask directly how large $C \cap Q_1$ can be. Since $(1,0), (0,1), (1,1) \notin C$ one tries to fence $C \cap Q_1$ inside the unit square. It does not fit: C can poke out past $x = 1$ along a shallow slope. The student realises that one quadrant alone is *not* bounded by 1, and gets stuck on the fact that some quadrant has to be.

The model chain. Write $Q_1, \ldots, Q_4$ for the closed quadrants, so that $A(C \cap Q_i) = A(C)/4$ for each i. It suffices to find one i with $A(C \cap Q_i) \leq 1$.

Support lines. The point $(0,1)$ is not in the closed convex set C, so there is a line strictly separating them; translate it to pass through $(0,1)$. It cannot be the vertical line $x = 0$, since that line contains $(0,0) \in C$. So it is $y = 1 + m_N x$, and $C \subseteq \{y \leq 1 + m_N x\}$ (the side containing the origin). In the same way:

$$C \subseteq \{x \leq 1 + m_E y\}, \qquad C \subseteq \{y \geq -1 + m_S x\}, \qquad C \subseteq \{x \geq -1 + m_W y\}.$$

Steep case. If $m_N \geq \frac{1}{2}$, then $C \cap Q_2$ lies in $\{x \leq 0,\ y \geq 0,\ y \leq 1 + m_N x\}$, the triangle with vertices $(0,0), (0,1), (-1/m_N, 0)$, of area $1/(2m_N) \leq 1$. If $m_N \leq -\frac{1}{2}$ the same holds in Q_1. The other three lines are handled identically. So assume from now on that all four slopes m_N, m_E, m_S, m_W lie in $(-\frac{1}{2}, \frac{1}{2})$.

The quadrant pieces. In Q_1, C lies in $R_1 = \{x, y \geq 0,\ y \leq 1 + m_N x,\ x \leq 1 + m_E y\}$. The two lines meet at $V_1 = (h,k)$ with

$$h = \frac{1 + m_E}{D}, \qquad k = \frac{1 + m_N}{D}, \qquad D = 1 - m_E m_N > 0,$$

and $h, k > 0$. R_1 is bounded (a direction $(a, b) \geq 0$ of recession would need $b \leq \frac{1}{2}a$ and $a \leq \frac{1}{2}b$, so $a = b = 0$), so R_1 is the quadrilateral $(0,0), (1,0), (h,k), (0,1)$, whose area by the shoelace formula is $\frac{1}{2}(h+k)$. The same holds in each quadrant.

The right angle. A direct computation gives

$$(1,0) - V_1 = -\tfrac{1+m_N}{D}(m_E, 1), \qquad (0,1) - V_1 = -\tfrac{1+m_E}{D}(1, m_N),$$

so

$$\big((1,0) - V_1\big) \cdot \big((0,1) - V_1\big) = \frac{(1+m_N)(1+m_E)}{D^2}\,(m_N + m_E).$$

The prefactor is positive. Doing the same at the other three vertices, the dot products have the signs of

$$s_1 = m_N + m_E, \quad s_2 = -(m_N + m_W), \quad s_3 = m_W + m_S, \quad s_4 = -(m_S + m_E),$$

and $s_1 + s_2 + s_3 + s_4 = 0$. So some $s_i \leq 0$; by the rotation $(x, y) \mapsto (-y, x)$, which preserves the lattice, the quadrants and the hypotheses, take $i = 1$. (This is the book's "at least one angle of the quadrilateral is not acute".)

The bound. $s_1 \leq 0$ says $(1-h)(-h) + (-k)(1-k) \leq 0$, that is $h^2 + k^2 \leq h + k$: V_1 lies in the disc with diameter from $(1,0)$ to $(0,1)$. Then

$$(h+k)^2 \leq 2(h^2 + k^2) \leq 2(h+k), \qquad \text{so} \qquad h + k \leq 2,$$

and $A(C \cap Q_1) \leq A(R_1) = \frac{1}{2}(h+k) \leq 1$. Hence $A(C) = 4A(C \cap Q_1) \leq 4$. □

Strictness. Equality would need $A(C \cap Q_i) = 1$ for the chosen i. In the steep case that forces $m_N = \frac{1}{2}$ and $C \cap Q_2$ equal to the closed triangle $(0,0), (0,1), (-2,0)$ (a closed convex subset of a convex region with the same positive area is the whole region), which contains $(-1,0)$. In the main case it forces $h + k = 2$ with $h^2 + k^2 \leq h + k$, so $h = k = 1$, and $C \cap Q_1 = R_1 \ni (1,1)$. Both contradict the hypothesis, so $A(C) < 4$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains know that the four points $(\pm 1, 0), (0, \pm 1)$ are excluded. They split on what to do with that. The natural attempt tries to turn "excluded" into a fence around one quadrant, and finds the fence leaks. The model turns each excluded point into a *line* (a support line), and then asks not "is this quadrant small?" but "which quadrant is small?" — a question whose answer comes from four quantities summing to zero.

> *C is convex and misses $(0,1)$. For a convex set, missing a point means there's a whole half-plane it misses. So I get a line through $(0,1)$, and four such lines, one per neighbour of the origin.*

> *I can't bound every quadrant, only one. But I'm told all four have the same area, so one is enough. I need a pigeonhole over the four quadrants: something that adds up to a fixed total and is good when it's small.*

> *The corner of $C \cap Q_1$ is at (h,k) and the area is $(h+k)/2$. When is $h + k \leq 2$? When (h,k) is inside the circle on $(1,0)$–$(0,1)$ as diameter, which is when the angle at (h,k) is at least a right angle. The four angles of a quadrilateral add up to $360°$, so one of them is.*

- ▷ **Cue:** a convex set is known to miss certain points ⟶ **Move:** draw a support (separating) line through each missed point; the set lies in the intersection of the half-planes
- ▷ **Cue:** a hypothesis that several parts have *equal* size, and you need to bound the total ⟶ **Move:** bound only the smallest, or any one of your choosing; set up a pigeonhole over the parts

▷ **Cue:** a vertex whose position must be controlled, with two fixed points on its two edges ⟶ **Move:** the angle at the vertex is $\geq 90^\circ$ iff the vertex lies in the disc on the segment between the two fixed points as diameter

Layer 3 THE WHY

Why support lines? Because that is what convexity *is*, from the outside. A closed convex set is exactly the intersection of the closed half-planes that contain it. So the only usable information in "C is convex and misses $(0,1)$" is a half-plane through $(0,1)$ that contains C. There is nothing else to extract. Once you see that, drawing four lines is not a trick. It is writing down the hypothesis.

Why is Minkowski the wrong door, even though the 4 looks like his? Minkowski's theorem is about *symmetric* sets, and its proof uses the symmetry to turn a lattice point of $C - C$ into one of C. Here C is not symmetric. The quadrant hypothesis is the replacement for symmetry: it says the mass of C is balanced around the origin in a weaker, four-fold sense. Without it the statement is false (the long triangle $(0,0), (0,\frac{1}{2}), (N,\frac{1}{2})$), so any proof must use it, and the cheapest way to use "all four equal" is to bound one of them.

Why does one quadrant have to be small? Think of the four support lines as nearly horizontal and nearly vertical. They form a quadrilateral around the origin, and its four angles add to 360°. If every corner angle were acute the sum would be below 360°. So some corner is at least a right angle. The algebraic version of this is the identity $s_1 + s_2 + s_3 + s_4 = 0$: the slopes cancel in pairs, so their signed sums cannot all be positive. The official solution states the angle fact in one sentence; the identity is where it comes from.

Why does a right angle give area 1? This is the circle example from the Blueprint, almost literally. The corner (h,k) sees the segment from $(1,0)$ to $(0,1)$ at an angle $\geq 90^\circ$, and by Thales that means (h,k) is inside the circle with that segment as diameter. The farthest that disc reaches in the direction $(1,1)$ is the point $(1,1)$ itself, where $h + k = 2$. So the circle is not decoration: the region's area is linear in $h + k$, the angle condition is a disc, and the maximum of a linear function on a disc is at the tangent point. That point is the lattice point $(1,1)$, which is exactly why the bound is 1 per quadrant and why it is never attained.

Why the steep case first? Because the quadrilateral argument needs the four lines to actually form a quadrilateral with one corner per quadrant, and that needs the slopes to be small. A steep line, on the other hand, cuts a quadrant down to a triangle of height 1 and base $1/|m| \leq 2$. The threshold $\frac{1}{2}$ is chosen so that the triangle has area ≤ 1; it is the only number that makes both cases close.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "closed convex set" next to "no other point with integer coordinates". Write down the four excluded neighbours $(\pm 1, 0), (0, \pm 1)$ and a support line through each.

2. Support lines, then a two-case split on their slopes, then a pigeonhole over four angles (or four signed slope sums).

3. Hesitation occurs when a single quadrant turns out not to be bounded by 1. Reinforce it away with the reminder that "equally distributed" means bounding *any one* quadrant suffices; you get to choose which.

4. The trap is Minkowski's theorem. The constant 4 invites it, and it cannot be used because C is not symmetric. A second trap is using only one support line and trying to confine a fixed quadrant.

5. When a problem says several pieces are equal, expect to bound only the best piece, and look for a quantity that sums to a constant across the pieces. When a problem gives convexity and excluded points, expect support lines through those points.

6. Reusable decision rule: *translate "a convex set misses a point" into "it lies in a half-plane through that point"; if you need a bound on one of several symmetric pieces, set up a sum over the pieces that is fixed (angles of a polygon, signed slopes) and take the piece where the summand has the good sign.*

1979 B6

1979 B6. For $k = 1, 2, \ldots, n$ let $z_k = x_k + iy_k$, where the x_k and y_k are real and $i = \sqrt{-1}$. Let r be the absolute value of the real part of

$$\pm\sqrt{z_1^2 + z_2^2 + \cdots + z_n^2}.$$

Prove that $r \leq |x_1| + |x_2| + \cdots + |x_n|$.

Fields: complex numbers; inequalities; Cauchy–Schwarz **Position:** B6 — last on the paper, though short once the right quantity is computed exactly; the difficulty is entirely in not throwing information away **Answer:** With $W = \sum z_k^2$ and $w = a + bi$ a square root, $a^2 = \frac{1}{2}(|W| + \operatorname{Re} W) \leq \frac{1}{2}\left(\sum |z_k|^2 + \sum (x_k^2 - y_k^2)\right) = \sum x_k^2$. So $r \leq \sqrt{x_1^2 + \cdots + x_n^2} \leq |x_1| + \cdots + |x_n|$.

Caveat. The OCR of the book loses the displayed expression and the conclusion: it breaks off at "Let r be the absolute value of the real part of". The statement above is the standard wording of 1979 B-6. It was checked against the book's own solution, which sets $a + bi$ to be "either square root of $z_1^2 + \cdots + z_n^2$" and ends with "the desired $r \leq |x_1| + \cdots + |x_n|$", and against the book's index ("Complex numbers 1979 B-6", "Cauchy–Schwarz 1979 B-6"). The mathematics below was verified independently of that reconstruction.

Layer 1 THE WHAT — surface analysis

The natural attempt. The student reads "absolute value of the real part" and uses the first inequality they know about real parts: $|\operatorname{Re} w| \leq |w|$. With $w = \pm\sqrt{W}$, $W = \sum z_k^2$,

$$r \leq |w| = \sqrt{|W|} \leq \sqrt{\sum |z_k|^2} = \sqrt{\sum (x_k^2 + y_k^2)}.$$

That is a true bound, and it is useless: it contains the y_k, and the target does not. Take $n = 1$, $z_1 = iy$: the bound says $r \leq |y|$, while the truth is $r = 0$. The inequality $|\operatorname{Re} w| \leq |w|$ threw away exactly the information the problem is about.

The second natural attempt is induction on n or a triangle inequality for square roots: "$\sqrt{u^2 + v^2}$ should be at most $|u| + |v|$ in real part." But complex square roots are not subadditive in any usable sense, and there is no branch of $\sqrt{\ }$ that is continuous on all of $\mathbb{C}$. The induction step needs to compare $\operatorname{Re}\sqrt{W + z^2}$ with $\operatorname{Re}\sqrt{W} + |x|$, which is the original problem in disguise. After this the student starts writing z_k in polar form, and the page fills with half-angles.

The model chain. Let $W = z_1^2 + \cdots + z_n^2$ and let $w = a + bi$ (a, b real) be either square root of W; then $r = |a|$, and the two choices $\pm w$ give the same $|a|$. Write $X = (x_1, \ldots, x_n)$, $Y = (y_1, \ldots, y_n)$ and $\|\cdot\|$ for the Euclidean norm. Expanding,

$$W = \sum (x_k^2 - y_k^2) + 2i \sum x_k y_k, \qquad \text{so} \qquad \operatorname{Re} W = \|X\|^2 - \|Y\|^2.$$

Compute a^2 exactly. From $w^2 = W$: $a^2 - b^2 = \operatorname{Re} W$, and $a^2 + b^2 = |w|^2 = |W|$. Adding,

$$a^2 = \tfrac{1}{2}\left(|W| + \operatorname{Re} W\right).$$

Estimate only the modulus. By the triangle inequality, $|W| \leq \sum |z_k|^2 = \|X\|^2 + \|Y\|^2$. Hence

$$a^2 \leq \tfrac{1}{2}\left(\|X\|^2 + \|Y\|^2 + \|X\|^2 - \|Y\|^2\right) = \|X\|^2.$$

So $r = |a| \leq \|X\| = \sqrt{x_1^2 + \cdots + x_n^2}$. Finally $\|X\|^2 = \sum x_k^2 \leq \left(\sum |x_k|\right)^2$, since the cross terms $2|x_j||x_k|$ are nonnegative. So $r \leq |x_1| + \cdots + |x_n|$. $\square$

The book's route. The official solution reaches $|a| \le \|X\|$ by contradiction. From $w^2 = W$, $ab = X \cdot Y$ and $a^2 - b^2 = \|X\|^2 - \|Y\|^2$. If $|a| > \|X\|$, then $b^2 = a^2 - \|X\|^2 + \|Y\|^2 > \|Y\|^2$, so $|b| > \|Y\|$, and $|X \cdot Y| = |a||b| > \|X\|\|Y\|$, contradicting Cauchy–Schwarz. The two proofs are the same inequality: $|W| \le \|X\|^2 + \|Y\|^2$ is equivalent to $4(X \cdot Y)^2 \le 4\|X\|^2\|Y\|^2$ after squaring $|W|^2 = (\|X\|^2 - \|Y\|^2)^2 + 4(X \cdot Y)^2$.

Sharpness. The intermediate bound $r \le \|X\|$ is attained whenever all z_k are real. The final bound is attained when, in addition, at most one x_k is nonzero.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start with $w = \pm\sqrt{W}$ and want $|\operatorname{Re} w|$. The natural chain's first line is $|\operatorname{Re} w| \le |w|$, an estimate. The model's first line is $(\operatorname{Re} w)^2 = \frac{1}{2}(|W| + \operatorname{Re} W)$, an identity. Everything after that is determined: the natural chain has already lost the sign information that separates x from y, and the model keeps $\operatorname{Re} W = \sum x_k^2 - \sum y_k^2$ exact, so the y's cancel.

> *$|\operatorname{Re} w| \le |w|$ can't work: at $z = iy$ it gives $|y|$ and the answer is 0. I have to keep the real part. I can't take a square root of a sum nicely, but I can square: the real part of w is determined by $|W|$ and $\operatorname{Re} W$. So compute a^2 exactly and put the inequality somewhere later.*

> *Where do I spend the one inequality I'm allowed?* $\operatorname{Re} W$ *I know exactly, it's* $\sum x^2 - \sum y^2$. $|W|$ *I don't, so that is the only place to estimate.* $|W| \le \sum |z_k|^2$ *brings in* $+\sum y^2$*, and it cancels the* $-\sum y^2$.

- ▷ **Cue:** a bound on the real (or imaginary) part of a complex square root $\longrightarrow$ **Move:** write $(\operatorname{Re}\sqrt{W})^2 = \frac{1}{2}(|W| + \operatorname{Re} W)$ and $(\operatorname{Im}\sqrt{W})^2 = \frac{1}{2}(|W| - \operatorname{Re} W)$
- ▷ **Cue:** an expression mixes a quantity you know exactly with one you can only bound $\longrightarrow$ **Move:** isolate the known part as an identity and apply the inequality only to the unknown part
- ▷ **Cue:** an estimate fails on a one-term test case where the truth is 0 $\longrightarrow$ **Move:** the estimate threw away a sign; go back one step and keep the quantity signed

Layer 3 THE WHY

Why is the half-angle identity the natural first line and not a trick? Because a square root is defined by squaring. The problem hands you w only through $w^2 = W$, and squaring $w = a + bi$ gives two real equations, $a^2 - b^2 = \operatorname{Re} W$ and $2ab = \operatorname{Im} W$, plus the modulus equation $a^2 + b^2 = |W|$ that follows from them. Those are all the facts about w there are. Solving them for a^2 is the most direct thing you can do with the definition. It is the same as the A-level method for finding $\sqrt{3 + 4i}$ by equating real and imaginary parts, and the same as the half-angle formula $\cos^2(\theta/2) = \frac{1}{2}(1 + \cos\theta)$ scaled by $|W|$.

Why is $|\operatorname{Re} w| \le |w|$ wrong here, when it is the reflex? That inequality is sharp only when w is real. It forgets the direction of w. But the direction is the whole problem: the hypothesis separates z_k into x_k and y_k, the conclusion keeps only x_k, and Re is the operation that tells them apart. Any first step that treats w through its modulus alone cannot produce a bound free of y. Testing the step on $n = 1$, $z_1 = iy$ exposes this in five seconds. That test is the cheapest thing in the problem, and it should be run before the step is written, not after twenty minutes.

Why does the cancellation happen? $\operatorname{Re} W$ carries $-\sum y_k^2$ exactly. The only inequality used, $|W| \le \sum |z_k|^2$, overestimates by bringing in $+\sum y_k^2$. They cancel because the triangle inequality for $|\sum z_k^2|$ is tight in the "real" direction: it is an equality when all the z_k^2 point the same way, and the real part is where x and y enter with opposite signs. Structurally, $a^2 = \frac{1}{2}(|W| + \operatorname{Re} W)$ is $|W|$ times $\cos^2$ of half the argument, and the proof bounds the length while keeping the angle.

The official solution hides this behind Cauchy–Schwarz. It writes $ab = X \cdot Y$ and argues by contradiction. Here is where that comes from: squaring out $|W|^2 = (\|X\|^2 - \|Y\|^2)^2 + 4(X \cdot Y)^2$

shows that $|W| \le \|X\|^2 + \|Y\|^2$ *is* Cauchy–Schwarz, $(X \cdot Y)^2 \le \|X\|^2\|Y\|^2$, in disguise. So both proofs use one inequality in one place. The direct version tells you which place, and why the y's disappear.

Finally, the problem's $|x_1| + \cdots + |x_n|$ is weaker than what is true. The proof gives $r \le \|X\|$, the Euclidean length, and then drops to the ℓ^1 sum using $\sum x_k^2 \le (\sum |x_k|)^2$. When a problem's bound is weaker than the natural one, prove the natural one; it is usually easier, because it has the right shape.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "the real part of $\pm\sqrt{\cdots}$". On seeing it, write $w = a + bi$, $w^2 = W$, and the three real equations $a^2 - b^2 = \operatorname{Re} W$, $2ab = \operatorname{Im} W$, $a^2 + b^2 = |W|$.

2. Exact computation of a^2, then one application of the triangle inequality (equivalently Cauchy–Schwarz) to $|W|$ only.

3. Hesitation occurs at the square root: students distrust $\sqrt{\ }$ on $\mathbb{C}$ and reach for polar form or branch cuts. Reinforce it away by noting that the $\pm$ in the statement is there precisely so that no branch has to be chosen, and $|a|$ is the same for both roots.

4. The trap is $|\operatorname{Re} w| \le |w|$, which gives a bound involving the y_k. A second trap is induction on n, which reproduces the problem at every step.

5. Before applying an inequality, test it on the smallest case where the answer is 0 or degenerate. If it fails there, it has discarded a sign or a direction, and you should keep that quantity exact.

6. Reusable decision rule: *compute every quantity you can compute exactly, and spend inequalities only on the parts you cannot; for complex square roots, square first and use* $a^2 = \frac{1}{2}(|W| + \operatorname{Re} W)$. *If the stated bound is weaker than a natural one (here* ℓ^1 *versus* ℓ^2*), prove the natural one.*