DAWN PUTNAM SESSION

Putnam 1986 — Papers A and B

Putnam 120 Archive

Contents

Putnam 1986 — Paper A **5**
1986 A1 5
1986 A2 7
1986 A3 10
1986 A4 12
1986 A5 15
1986 A6 18

Putnam 1986 — Paper B **23**
1986 B1 23
1986 B2 25
1986 B3 28
1986 B4 30
1986 B5 33
1986 B6 35

Putnam 1986 — Paper A

1986 A1

1986 A1. Find, with explanation, the maximum value of $f(x) = x^3 - 3x$ on the set of all real numbers x satisfying $x^4 + 36 \leq 13x^2$.

Fields: algebra, single-variable calculus (constrained extrema) **Position:** A1 — the opener **Answer:** The maximum is 18, attained at $x = 3$. The constraint is $(x^2-4)(x^2-9) \leq 0$, i.e. $x \in [-3,-2] \cup [2,3]$, and $f'(x) = 3(x^2-1) > 0$ on both intervals, so f is increasing on each and the candidates are the right-hand endpoints: $f(-2) = -2$, $f(3) = 18$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You see "maximum value of f" and your hand reaches for the derivative, because that is what "maximum" has meant in every calculus course you have taken. So you write

$$f'(x) = 3x^2 - 3 = 0 \quad \Longrightarrow \quad x = \pm 1, \qquad f''(x) = 6x,$$

so $x = -1$ is a local maximum with $f(-1) = -1 + 3 = 2$, and $x = 1$ is a local minimum with $f(1) = -2$. You now have a candidate answer, 2, and you have spent your first minutes on it.

Then you remember the constraint. You test $x = -1$: $x^4 + 36 = 37$ and $13x^2 = 13$, and $37 \leq 13$ is false. The stationary point is not even in the set. Nor is $x = 1$. So everything you have computed so far is irrelevant, and you go back and work out the set. You treat $x^4 - 13x^2 + 36 \leq 0$ as a quadratic in $u = x^2$: $u^2 - 13u + 36 = (u-4)(u-9) \leq 0$, so $4 \leq x^2 \leq 9$. At this point a common slip is to write "$2 \leq x \leq 3$", taking only the positive square root. On $[2,3]$ you evaluate $f(2) = 2$, $f(3) = 18$, and answer 18 — correct, but only because the negative branch happens not to matter here. On a different f (try $f(x) = -x^3 + 3x$) the same slip returns a wrong answer.

The attempt works in the end. What it cost was the wrong order: you optimised first and located the domain second, so the first block of work was thrown away, and the domain was then found carelessly because you were already in a hurry.

The model chain. *Step 1: identify the set.* Put $u = x^2 \geq 0$. The condition $x^4 + 36 \leq 13x^2$ reads $u^2 - 13u + 36 \leq 0$, i.e. $(u-4)(u-9) \leq 0$, i.e. $4 \leq u \leq 9$. Hence

$$x^4 + 36 \leq 13x^2 \iff (x-3)(x-2)(x+2)(x+3) \leq 0 \iff x \in [-3,-2] \cup [2,3].$$

Step 2: the shape of f on the set. $f'(x) = 3(x^2-1)$, and on the set $x^2 \geq 4$, so $f'(x) \geq 9 > 0$. Thus f is strictly increasing on $[-3,-2]$ and on $[2,3]$. There are no interior stationary points, and on each interval the maximum is at the right-hand endpoint.

Step 3: compare. $f(-2) = -8 + 6 = -2$ and $f(3) = 27 - 9 = 18$. The maximum on the whole set is $\max\{-2, 18\} = 18$, attained at $x = 3$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains contain the same two computations — the derivative and the factorisation of the constraint. They split at the very first line, over which one comes first. The

natural attempt writes $f'(x) = 0$ first; the model writes $(x^2 - 4)(x^2 - 9) \leq 0$ first. Once the set is on the page, the derivative is only needed to answer "is f monotone here?", and the answer takes one line.

> *A maximum is a statement about f* on a set. *I do not know the set yet, so I do not know where to look. Draw the set first.*

> *x^4 and x^2 and a constant, no odd powers: this is a quadratic in x^2. Solve for x^2, then take both square roots — the set is symmetric under $x \mapsto -x$, so it must come in a positive and a negative piece.*

> *The set is a union of closed intervals. On each one the maximum sits at an endpoint or at a stationary point inside. Is there a stationary point inside? $x^2 \geq 4$ there, so $f' = 3(x^2 - 1) > 0$: no. So only endpoints matter, and only right-hand ones.*

Layer 3 THE WHY

Why is "domain first" the natural move and not a matter of taste? Go back to what the theorem you are using actually says. The Extreme Value Theorem guarantees that a continuous function on a closed bounded set attains a maximum, and Fermat's theorem says that at an *interior* maximum of a differentiable function the derivative vanishes. Put together: the maximum of f on $[a, b]$ is found among (i) the endpoints a, b and (ii) the interior points with $f' = 0$. Both lists are defined *relative to the set*. "Endpoints" means nothing until you know the set; "interior stationary point" means a stationary point that lies inside the set. So the set is logically prior to the derivative. The natural attempt runs the second half of the procedure before the first half exists. That is why the work at $x = \pm 1$ was wasted: the stationary points of f on $\mathbb{R}$ are not the stationary points of f on the set.

Why read the constraint as a quadratic in x^2? Because the polynomial $x^4 - 13x^2 + 36$ is even: it contains only even powers, so it is a polynomial in x^2. That is structural, not a trick — an even function of x is a function of x^2. And evenness tells you something before you solve anything: if x is in the set, so is $-x$. That is the reason the positive-root-only slip is a genuine error, not a small one. The set must be symmetric, so $[2, 3]$ alone cannot be it.

Why is monotonicity the right question on the set? Because $f'(x) = 3(x^2 - 1)$ changes sign only at $x^2 = 1$, and the set lives at $x^2 \geq 4$, away from that. The same variable x^2 that organised the constraint also organises f'. Once you see that the set is "x^2 between 4 and 9" and the sign of f' is "x^2 versus 1", the comparison is immediate. That is why the problem-setter chose 4 and 9 as the roots: to put the whole set on the increasing side of f.

Finally, the comparison across the two pieces. f is odd, so $f(-2) = -f(2) = -2$; the negative piece contributes only negative values when f is increasing there and ends at -2. The answer 18 comes from the positive piece. You must still check the negative piece — the problem says "the set of all real numbers", and a reader marking your work wants to see that you considered the whole set.

Layer $+\alpha$ BACKTESTING

1. The first cue is the phrase "on the set of all real numbers x satisfying" followed by an inequality. Write down the set, as a union of intervals, before touching f: $x \in [-3, -2] \cup [2, 3]$.

2. A quadratic in x^2 (factor $(x^2 - 4)(x^2 - 9)$), then the closed-interval method: endpoints plus interior stationary points, with monotonicity settling the second list at once.

3. Hesitation occurs after finding $x = \pm 1$ as stationary points — "is the answer $f(-1) = 2$?" It is reinforced away by testing membership in the set before evaluating: $x = -1$ gives $37 \leq 13$, false.

4. Two traps. First, reporting the unconstrained local maximum $f(-1) = 2$. Second, solving $4 \le x^2 \le 9$ as $2 \le x \le 3$ and losing the negative interval; here it does not change the answer, but it makes the proof incomplete and would be fatal for a decreasing or non-odd f.

5. Any optimisation phrased "on the set of … satisfying …" is a two-stage problem, and the stage that is easier to forget is the one that must be done first. An even polynomial constraint is a quadratic in the square.

6. Reusable rule: to optimise a continuous function on a constrained set, (1) describe the set explicitly as closed intervals or a region with boundary, (2) list boundary points and interior stationary points *of the set*, (3) evaluate and compare. If the sign of f' is constant on a piece, only that piece's endpoints matter.

▷ **Cue:** "maximum/minimum of f on the set of x satisfying [inequality]" $\longrightarrow$ **Move:** solve the inequality first and write the set as a union of closed intervals; only then differentiate

▷ **Cue:** a polynomial with only even powers, e.g. $x^4 - 13x^2 + 36$ $\longrightarrow$ **Move:** substitute $u = x^2$, factor the quadratic, and take both signs of the square root

▷ **Cue:** a stationary point of f has been found $\longrightarrow$ **Move:** check that it lies in the constrained set before evaluating f there

▷ **Cue:** f' has constant sign on each piece of the domain $\longrightarrow$ **Move:** the extremum is at an endpoint; compare the relevant endpoint on each piece

1986 A2

1986 A2. What is the units (i.e., rightmost) digit of

$$\left\lfloor \frac{10^{20000}}{10^{100}+3} \right\rfloor ?$$

Fields: number theory, algebra (polynomial division) **Position:** A2 — early, but with a sign trap **Answer:** The units digit is 3. With $N = 10^{100}$, polynomial division gives $N^{200} = (N+3)Q + 3^{200}$ with $0 < 3^{200} < N+3$, so the floor is Q; reducing $(N+3)Q = N^{200} - 3^{200}$ modulo 10 gives $3Q \equiv -1 \equiv 9$, so $Q \equiv 3 \pmod{10}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You see a huge power of ten divided by "a huge power of ten plus a little bit", and the reflex is to pull out the big part and expand the little bit as a series:

$$\frac{10^{20000}}{10^{100}+3} = 10^{19900} \cdot \frac{1}{1+3\cdot 10^{-100}} = 10^{19900} \sum_{k\ge 0} (-3)^k 10^{-100k} = \sum_{k\ge 0} (-3)^k 10^{19900-100k}.$$

The terms with $k \le 199$ are integers; the terms with $k \ge 200$ carry negative powers of ten. So you say: the integer part is $\sum_{k=0}^{199} (-3)^k 10^{19900-100k}$, and its units digit comes from the last term, $k = 199$, which is $(-3)^{199} \cdot 10^0 = -3^{199}$. You compute 3^{199} modulo 10: the cycle of 3^k mod 10 is $3, 9, 7, 1$ with period 4, and $199 \equiv 3 \pmod 4$, so $3^{199} \equiv 7$. Many students write 7 here and stop.

Those who pause notice two problems. First, the last term is *negative*: -3^{199} ends in … what, exactly? A negative number does not have a units digit in the usual sense; you need the units digit of a large positive integer that has -3^{199} added to it, so $-7 \equiv 3 \pmod{10}$ — the answer is 3, not 7. Second, and more seriously, "the integer part is the sum of the integer terms" is not a theorem. The tail $\sum_{k\ge 200} (-3)^k 10^{19900-100k}$ is an alternating series of non-integers; if it were negative, the floor would be one *less* than the integer sum and the digit would change. The series approach can be repaired (the tail sums to $3^{200}/(10^{100}+3)$, which is positive and less than 1), but you have to know to repair it, and in the arena the two sign issues are where the marks go.

The model chain. Let $N = 10^{100}$, so the number in question is $\lfloor N^{200}/(N+3) \rfloor$.

Step 1: divide exactly. Since 200 is even, $N^{200} - 3^{200} = N^{200} - (-3)^{200}$, and $x - y$ divides $x^{200} - y^{200}$ as polynomials; with $x = N$, $y = -3$,

$$N^{200} - 3^{200} = (N+3)\,Q, \qquad Q = \sum_{k=0}^{199} N^{199-k}(-3)^k \in \mathbb{Z}.$$

Equivalently, dividing the polynomial x^{200} by $x + 3$ leaves remainder $(-3)^{200} = 3^{200}$ by the Remainder Theorem. So

$$N^{200} = (N+3)Q + 3^{200}.$$

Step 2: identify the floor. We have $0 < 3^{200} = 9^{100} < 10^{100} < N + 3$. So 3^{200} is a legitimate remainder on division by $N + 3$, and

$$\frac{N^{200}}{N+3} = Q + \frac{3^{200}}{N+3}, \qquad 0 < \frac{3^{200}}{N+3} < 1,$$

hence $\lfloor N^{200}/(N+3) \rfloor = Q$. (Also $Q > 0$, since $N^{200} > 3^{200}$.)

Step 3: Q modulo 10. Reduce $(N+3)Q = N^{200} - 3^{200}$ modulo 10. As $10 \mid N$, $N + 3 \equiv 3$ and $N^{200} \equiv 0$. As $3^4 = 81 \equiv 1$, $3^{200} = (3^4)^{50} \equiv 1$. So

$$3Q \equiv -1 \equiv 9 \pmod{10}.$$

Since $3 \cdot 3 = 9$ and 3 is invertible modulo 10 (its inverse is 7), $Q \equiv 3 \pmod{10}$. The units digit is **3**.

(Check against the explicit sum: every term of Q except $k = 199$ is a multiple of N, so $Q \equiv (-3)^{199} = -3^{199} \equiv -7 \equiv 3 \pmod{10}$. The two routes agree.)

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains begin by seeing $10^{20000} = (10^{100})^{200}$. They split at the next line. The natural attempt divides by $N + 3$ *approximately*, as a real number, via the expansion of $1/(1 + 3N^{-1})$. The model divides *exactly*, as integers, writing $N^{200} = (N+3)Q + r$ with $0 \le r < N + 3$. The first produces an infinite series whose integer part has to be argued for; the second produces the quotient and remainder, and the floor is the quotient by definition.

> *A floor of a fraction of integers is a quotient in long division: $a = qb + r$ with $0 \le r < b$ means $\lfloor a/b \rfloor = q$. So I want the quotient and remainder of N^{200} divided by $N + 3$, and I must check the remainder lies in $[0, N + 3)$.*

> *$N = 10^{100}$ is enormous next to 3, so it behaves like the variable x. Divide x^{200} by $x + 3$ as polynomials: the remainder is $(-3)^{200} = 3^{200}$, and $3^{200} = 9^{100} < 10^{100}$, so it is a genuine remainder.*

> *I do not need Q, only Q mod 10. I know $(N+3)Q$ exactly mod 10, and $N + 3 \equiv 3$ is invertible mod 10. Divide out.*

Layer 3 THE WHY

Why is exact division the natural move? Go back to the definition of the floor of a quotient. For positive integers a, b, the Division Algorithm says there is a unique pair (q, r) with $a = qb + r$ and $0 \le r < b$, and then $a/b = q + r/b$ with $0 \le r/b < 1$, so $\lfloor a/b \rfloor = q$. The floor of a fraction of integers *is* the integer quotient. The problem's phrasing — a floor bracket around a fraction whose top and bottom are integers — is therefore a request for a quotient, and the tool for

computing a quotient is division, not approximation. The series approach replaces an exact integer question by an analytic one, and then has to climb back to the integers by controlling the tail; that climb is where both sign errors live.

Why polynomial division? Because the only numbers in sight are powers of $N = 10^{100}$ and the small constant 3. An integer written in base N is a polynomial in N; when the "digits" are small compared with N, arithmetic in base N behaves exactly like arithmetic with polynomials in x. Dividing x^{200} by $x + 3$ is a school computation — the Remainder Theorem says the remainder is the value at $x = -3$ — and the only thing that could go wrong when substituting $x = N$ is that the polynomial remainder might be too large, or negative, to be an integer remainder. That is the one check the problem is testing: 3^{200} against 10^{100}. It passes because $3^2 = 9 < 10$, so $3^{200} = 9^{100} < 10^{100}$. Notice the problem-setter's choices here. The exponent $20000 = 200 \cdot 100$ was chosen so that 3^{200} and 10^{100} compete on equal footing and the comparison is 9 versus 10; the evenness of 200 was chosen so that the remainder $(-3)^{200}$ is positive. Had the exponent been odd, the polynomial remainder would be -3^{odd}, negative, and you would have to borrow one from the quotient — exactly the case where the series approach silently goes wrong.

Why is modular arithmetic the last step and not a direct attack? Because Q is a 19900-digit number and you need one digit of it. A units digit is a residue modulo 10, and residues are computed from any exact relation involving Q. The relation $(N + 3)Q = N^{200} - 3^{200}$ is exact, and everything in it except Q has an easy residue. Dividing by 3 modulo 10 is legitimate because $\gcd(3, 10) = 1$; with $+2$ or $+5$ in the denominator this final cancellation would not be available, and you would have to read the digit off the explicit sum instead.

Layer $+\alpha$ BACKTESTING

1. The cue is a floor bracket around a quotient of integers with a power-of-10^{100} structure. Write down $N = 10^{100}$ and $N^{200} = (N + 3)Q + r$.

2. Exact polynomial division with the Remainder Theorem (or the factorisation $x^{200} - y^{200} = (x - y)(\cdots)$ with $y = -3$), followed by arithmetic modulo 10.

3. Hesitation occurs at the last term $(-3)^{199}$, or at the remainder: "is the floor Q or $Q - 1$?" It is reinforced away by checking $0 \le r < N + 3$ explicitly ($3^{200} = 9^{100} < 10^{100}$) and by working with residues of an exact identity instead of reading digits off a signed sum.

4. Two traps. First, the sign: computing $3^{199} \equiv 7$ and answering 7, when the term is $-3^{199} \equiv 3$. Second, assuming the floor of a series equals the sum of its integer terms without checking that the non-integer tail lies in $[0, 1)$.

5. Whenever a problem contains a floor of a/b with a, b integers, the problem is about the Division Algorithm. Whenever the numbers are powers of one huge base plus small constants, treat the base as a variable and divide polynomials.

6. Reusable rule: to find $\lfloor a/b \rfloor$ modulo m, write $a = bq + r$ exactly, verify $0 \le r < b$, then reduce $bq = a - r$ modulo m and cancel b if $\gcd(b, m) = 1$. Never read a digit off an infinite expansion without controlling the tail.

▷ **Cue:** $\lfloor a/b \rfloor$ with a, b integers $\longrightarrow$ **Move:** write $a = bq + r$ with $0 \le r < b$; the floor is q

▷ **Cue:** expressions built from powers of a huge base N and small constants, e.g. $10^{20000}/(10^{100} + 3)$ $\longrightarrow$ **Move:** set N as a variable and divide polynomials; the Remainder Theorem gives the remainder, then check it is in range

▷ **Cue:** need one digit (a residue) of a gigantic integer defined implicitly $\longrightarrow$ **Move:** find an exact identity it satisfies and reduce the identity modulo 10, cancelling any factor coprime to 10

▷ **Cue:** an alternating expansion whose integer part you want $\longrightarrow$ **Move:** check the sign and size of the discarded tail before taking the floor

1986 A3

1986 A3. Evaluate $\sum_{n=0}^{\infty} \operatorname{Arccot}(n^2+n+1)$, where $\operatorname{Arccot} t$ for $t \geq 0$ denotes the number θ in the interval $0 < \theta \leq \pi/2$ with $\cot\theta = t$.

Fields: analysis (series), trigonometry **Position:** A3 — mid-paper, one idea **Answer:** The sum is $\pi/2$. For $n \geq 0$, $\operatorname{Arccot}(n^2+n+1) = \arctan(n+1) - \arctan n$, so the partial sum of the first m terms telescopes to $\arctan m$, which tends to $\pi/2$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You compute the first few terms. $n=0$: $\operatorname{Arccot}1 = \pi/4$. $n=1$: $\operatorname{Arccot}3$. $n=2$: $\operatorname{Arccot}7$. $n=3$: $\operatorname{Arccot}13$. None of these after the first is a standard angle, so you cannot add them numerically by hand in any useful way. You note that $\operatorname{Arccot} t \approx 1/t$ for large t, so the terms behave like $1/n^2$ and the series converges; that is worth a line but it does not evaluate anything.

So you reach for the addition formula, because two angles are the thing it adds:

$$\cot(\alpha+\beta) = \frac{\cot\alpha\cot\beta - 1}{\cot\alpha + \cot\beta}.$$

With $\cot\alpha = 1$, $\cot\beta = 3$: $\cot(\alpha+\beta) = \frac{3-1}{4} = \frac{1}{2}$. Then add $\operatorname{Arccot}7$: $\cot = \frac{\frac{1}{2}\cdot 7 - 1}{\frac{1}{2}+7} = \frac{5/2}{15/2} = \frac{1}{3}$. The pattern $1, \frac{1}{2}, \frac{1}{3}$ appears, so you guess that the first m terms sum to $\operatorname{Arccot}(1/m)$ and prove it by induction; the inductive step is

$$\frac{\frac{1}{m}(m^2+m+1) - 1}{\frac{1}{m} + m^2 + m + 1} = \frac{(m^2+1)/m}{(m^2+1)(m+1)/m} = \frac{1}{m+1},$$

using $m^3 + m^2 + m + 1 = (m^2+1)(m+1)$. Then $\operatorname{Arccot}(1/m) \to \operatorname{Arccot}0 = \pi/2$.

This is Kedlaya's Solution 1, and it works. It costs a guess, an induction, and a branch check you might forget: the cot addition formula only returns Arccot of the right thing when the sum of angles stays in $(0, \pi/2]$, which needs the product of the two cotangents to be at least 1 at every step. Without that check the identity is not proved.

The model chain. For $t > 0$, $\operatorname{Arccot} t = \arctan(1/t)$: both lie in $(0, \pi/2)$ and have cotangent t.

Step 1: rewrite each term as a difference. Fix $n \geq 0$ and let $A = \arctan(n+1)$, $B = \arctan n$. Then $0 \leq B < A < \pi/2$, so $0 < A - B < \pi/2$, and

$$\tan(A-B) = \frac{(n+1) - n}{1 + n(n+1)} = \frac{1}{n^2+n+1}.$$

Hence $\cot(A-B) = n^2+n+1$ with $A - B \in (0, \pi/2)$, which is exactly the defining property of $\operatorname{Arccot}(n^2+n+1)$. So

$$\operatorname{Arccot}(n^2+n+1) = \arctan(n+1) - \arctan n.$$

Step 2: telescope. For $m \geq 1$,

$$\sum_{n=0}^{m-1} \operatorname{Arccot}(n^2+n+1) = \sum_{n=0}^{m-1} \big(\arctan(n+1) - \arctan n\big) = \arctan m - \arctan 0 = \arctan m.$$

Step 3: take the limit. $\arctan m \to \pi/2$ as $m \to \infty$. So the series converges and

$$\sum_{n=0}^{\infty} \operatorname{Arccot}(n^2+n+1) = \frac{\pi}{2}.$$

(In Arccot language, Step 1 reads $\operatorname{Arccot}(n^2+n+1) = \operatorname{Arccot} n - \operatorname{Arccot}(n+1)$, using $\operatorname{Arccot} 0 = \pi/2$ — which is why the problem defines Arccot at $t = 0$.)

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start from the same fact: the trigonometric addition formula is the only tool that combines two inverse-trig values. They split on which way to use it. The natural attempt uses it forwards, to *add* consecutive terms and watch the partial sums. The model uses it backwards, to *split* one term into a difference: it looks at $\frac{1}{n^2+n+1}$ and asks which a, b make it $\frac{a-b}{1+ab}$. The answer $a = n+1$, $b = n$ turns the series into a telescoping sum with no guessing and no induction.

> *An infinite sum of non-geometric terms that the problem expects me to evaluate exactly: that means it telescopes. So I want each term as $g(n+1) - g(n)$.*

> *Differences of inverse tangents come from $\tan(A-B) = \frac{a-b}{1+ab}$. Is n^2+n+1 of the form $1+ab$ with $a-b=1$? Yes: $1+n(n+1)$, and $(n+1)-n=1$.*

> *Before I write "=", check the branch: $\arctan(n+1) - \arctan n$ lies in $(0, \pi/2)$, the same interval the problem uses to define* Arccot*. So the identity is exact, not "up to π".*

Layer 3 THE WHY

Kedlaya's Solution 2 remarks that the identity $\operatorname{Arccot}(n^2+n+1) = \operatorname{Arccot} n - \operatorname{Arccot}(n+1)$ "can also be proved from the difference formula for cot, but then one needs to guess the identity in advance." Here is where the guess comes from, so that it is no longer a guess.

Start from what you are asked to do: evaluate an infinite series exactly. At this level there are very few ways that can happen. A geometric series has a formula; a power series might be a known function evaluated at a point; and otherwise the series telescopes. The terms here are angles, not powers, so the first two are out. That leaves telescoping. So the question is not "is there a trick?" but "which g makes $\operatorname{Arccot}(n^2+n+1) = g(n+1) - g(n)$?" — a definite question with a definite search.

Now: where do differences of inverse-trig values come from? From the subtraction formula, and only from there. For arctan it reads

$$\arctan a - \arctan b = \arctan \frac{a-b}{1+ab} \qquad \text{(when the left side lies in } (-\tfrac{\pi}{2}, \tfrac{\pi}{2})\text{)}.$$

So the *fingerprint* of a difference of arctangents is an argument of the shape $\frac{a-b}{1+ab}$: a small numerator over "one plus a product". Convert the term: $\operatorname{Arccot}(n^2+n+1) = \arctan \frac{1}{n^2+n+1}$. The denominator n^2+n+1 is $1+n(n+1)$, a product of consecutive integers plus one; the numerator 1 is the difference of those same consecutive integers. The fingerprint matches, and it matches with $a = n+1$, $b = n$ — consecutive values of one function, which is precisely what telescoping needed. The problem-setter wrote n^2+n+1 rather than, say, n^2+1 because $n^2+n+1 = 1+n(n+1)$ is the form that telescopes.

There is a second, deeper reason, which is Kedlaya's Solution 2. The argument of a complex number $x+iy$ with $x > 0$ is $\arctan(y/x)$, and arguments *add* when complex numbers multiply. The identity

$$(n+i)(n+1-i) = n^2+n+1+i$$

therefore says $\arg(n^2+n+1+i) = \arg(n+i) + \arg(n+1-i)$, i.e. $\arctan \frac{1}{n^2+n+1} = \arctan \frac{1}{n} - \arctan \frac{1}{n+1}$ for $n \geq 1$ — the same telescoping identity in cotangent form. So "every arctan

identity is a factorisation of Gaussian integers" is the axiom-level fact underneath the addition formula: the addition formula is the statement $\arg(zw) = \arg z + \arg w$ written in coordinates. If you ever have to find an arctan decomposition cold, factor $t + i$.

Finally, why must the branch be checked? Because tan is not injective on $\mathbb{R}$: $\tan\theta = \tan\phi$ gives only $\theta \equiv \phi \pmod{\pi}$. The problem takes care to define Arccot with values in $(0, \pi/2]$. An identity between angles is only proved when both sides are shown to lie in an interval on which tan (or cot) is one-to-one. Here both sides lie in $(0, \pi/2)$, so the identity holds exactly. This is the step that the forward-adding attempt must check afresh at every stage of its induction, and the telescoping approach checks once.

Layer $+\alpha$ BACKTESTING

1. The first cue is "Evaluate $\sum_{n=0}^{\infty}$" of inverse-trig terms: an exact value for a non-geometric infinite sum. Write down "it telescopes: find g with term $= g(n+1) - g(n)$", and convert to arctan $\frac{1}{n^2+n+1}$.

2. The arctan subtraction formula, used in reverse to split each term: match $\frac{1}{n^2+n+1}$ to $\frac{a-b}{1+ab}$ with $a = n+1, b = n$.

3. Hesitation occurs when the terms $\operatorname{Arccot} 3, \operatorname{Arccot} 7, \ldots$ refuse to be standard angles. It is reinforced away by not trying to evaluate terms at all: an exact sum of unevaluable terms can only come from cancellation.

4. Two traps. First, grinding partial sums forward and then needing a guess plus an induction, with a branch condition ($xy \geq 1$) to verify at each step. Second, writing $\arctan a - \arctan b = \arctan\frac{a-b}{1+ab}$ without checking that the left side lies in the principal range; the identity is only true modulo π in general.

5. Any sum of arctan or Arccot of a rational function of n is a telescoping candidate; test whether the argument can be put in the form $\frac{a-b}{1+ab}$ with a, b consecutive values of one function, or factor $t + i$ over the Gaussian integers.

6. Reusable rule: when asked for the exact value of an infinite sum whose terms are not geometric and not individually computable, look for a telescoping decomposition, and find it by inverting the identity that produces differences of the relevant function (partial fractions for rationals, $\log\frac{a}{b}$ for logs, $\frac{a-b}{1+ab}$ for arctan).

▷ **Cue:** "evaluate" an infinite sum whose individual terms cannot be computed $\longrightarrow$ **Move:** it telescopes; look for g with term $= g(n+1) - g(n)$

▷ **Cue:** $\arctan\frac{p}{q}$ or $\operatorname{Arccot}\frac{q}{p}$ where q can be written as $1 + ab$ and $p = a - b$ $\longrightarrow$ **Move:** write the term as $\arctan a - \arctan b$

▷ **Cue:** a quadratic like $n^2 + n + 1$ inside an inverse trig function $\longrightarrow$ **Move:** rewrite as $1 + n(n+1)$ and test the arctan difference formula with $a = n+1, b = n$

▷ **Cue:** an identity between angles derived from tan or cot values $\longrightarrow$ **Move:** verify both sides lie in the same interval of length π on which the function is injective

▷ **Cue:** an arctan identity is needed but cannot be guessed $\longrightarrow$ **Move:** factor $t + i$ over the Gaussian integers; arguments add under multiplication

1986 A4

1986 A4. A *transversal* of an $n \times n$ matrix A consists of n entries of A, no two in the same row or column. Let $f(n)$ be the number of $n \times n$ matrices A satisfying the following two conditions:

(a) Each entry $a_{i,j}$ of A is in the set $\{-1, 0, 1\}$.

(b) The sum of the n entries of a transversal is the same for all transversals of A.

An example of such a matrix A is

$$A = \begin{pmatrix} -1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Determine with proof a formula for $f(n)$ of the form

$$f(n) = a_1 b_1^n + a_2 b_2^n + a_3 b_3^n + a_4,$$

where the a_i's and b_i's are rational numbers.

Fields: combinatorics, linear algebra (structure of matrices) **Position:** A4 — mid-paper; one structural lemma turns a condition on $n!$ transversals into a condition on pairs of rows, after which it is a four-case count **Answer:** $f(n) = 4^n + 2 \cdot 3^n - 4 \cdot 2^n + 1$. Condition (b) holds exactly when any two rows differ by a constant vector $(c, c, \ldots, c)$, equivalently $\alpha_{i,j} = r_i + s_j$; counting by what values the first row takes gives the formula. Check: $f(1) = 3$, $f(2) = 19$, $f(3) = 87$ (all three confirmed by exhaustive search).

Layer 1 THE WHAT — surface analysis

The natural attempt. You start small, as you should. For $n = 1$ every one of the 3 matrices works. For $n = 2$ there are two transversals, and (b) says $\alpha_{11} + \alpha_{22} = \alpha_{12} + \alpha_{21}$. You count solutions of $p + q = r + s$ over $\{-1, 0, 1\}$: the common sum k ranges over $-2, \ldots, 2$ with $1, 2, 3, 2, 1$ representations, so $1 + 4 + 9 + 4 + 1 = 19$. Good.

For $n = 3$ there are 6 transversals and $3^9 = 19683$ matrices. You write out the six sums

$$\alpha_{11} + \alpha_{22} + \alpha_{33},\ \alpha_{11} + \alpha_{23} + \alpha_{32},\ \alpha_{12} + \alpha_{21} + \alpha_{33},\ \ldots$$

and set them all equal: five linear equations in nine unknowns. You try to solve the system, find it has rank 4, parametrise the solution space with five free parameters, and then try to count the lattice points of that five-dimensional space that land in $\{-1, 0, 1\}^9$. The parametrisation is ugly and the box constraint cuts it in an irregular way. Somewhere around here you start guessing: $3, 19, \ldots$ and a desired shape $a_1 b_1^n + \cdots + a_4$, with no reliable third or fourth data point and four unknown bases. The attempt stalls because you are treating (b) as $n! - 1$ separate equations rather than asking what they say collectively.

The model chain. *Lemma.* For an $n \times n$ matrix, (b) holds if and only if any two rows differ by a constant vector, i.e. $\alpha_{i,k} - \alpha_{j,k}$ does not depend on k.

Proof. ($\Leftarrow$) If every row is row 1 plus a constant, $\alpha_{i,k} = \alpha_{1,k} + d_i$, then a transversal $\{(i, \sigma(i))\}$ has sum $\sum_i \alpha_{1,\sigma(i)} + \sum_i d_i = \sum_k \alpha_{1,k} + \sum_i d_i$, since σ is a permutation. That is independent of σ.

($\Rightarrow$) Take rows $i \neq j$ and columns $k \neq l$. Choose a transversal σ with $\sigma(i) = k$, $\sigma(j) = l$ (complete it by any bijection of the remaining rows onto the remaining columns). Swapping gives σ' with $\sigma'(i) = l$, $\sigma'(j) = k$, equal to σ elsewhere. The two sums are equal by (b), and they share all other entries, so

$$\alpha_{i,k} + \alpha_{j,l} = \alpha_{i,l} + \alpha_{j,k}, \qquad \text{i.e.} \qquad \alpha_{i,k} - \alpha_{j,k} = \alpha_{i,l} - \alpha_{j,l}.$$

(When $k = l$ or $i = j$ both sides of the identity are equal term by term.) So $\alpha_{i,\cdot} - \alpha_{j,\cdot}$ is constant. □

Counting. By the lemma a valid matrix is determined by its first row $v \in \{-1, 0, 1\}^n$ and, for each $i \geq 2$, a constant d_i with $v + d_i(1, \ldots, 1) \in \{-1, 0, 1\}^n$. The allowed d_i depend only on the set of values that v takes.

- *v constant* (3 choices). Then every row is constant, and each of rows $2, \ldots, n$ is independently one of $(0, \ldots, 0)$, $(1, \ldots, 1)$, $(-1, \ldots, -1)$. Total: $3 \cdot 3^{n-1} = 3^n$.

- v *uses exactly* $\{0,1\}$ (2^n-2 choices). A shift must keep both 0 and 1 inside $\{-1,0,1\}$, so $d_i \in \{0,-1\}$. Total: $(2^n-2)2^{n-1}$.
- v *uses exactly* $\{-1,0\}$. The negatives of the previous case: $(2^n-2)2^{n-1}$.
- v *uses both* 1 *and* -1. Then $d_i = 0$ only, so the matrix is n copies of v. The number of such v is $3^n - 3 - 2(2^n-2) = 3^n - 2\cdot 2^n + 1$ (all vectors, minus the three constant ones, minus the two two-valued families of the previous cases).

Adding,

$$f(n) = 3^n + 2(2^n-2)2^{n-1} + 3^n - 2\cdot 2^n + 1 = 4^n + 2\cdot 3^n - 4\cdot 2^n + 1.$$

So $(a_1,b_1) = (1,4)$, $(a_2,b_2) = (2,3)$, $(a_3,b_3) = (-4,2)$, $a_4 = 1$. For $n = 1$ the lemma is vacuous and the formula gives $4+6-8+1 = 3$; for $n = 2$ it gives 19, matching the hand count.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains write down the condition "all transversal sums are equal". The natural attempt then writes all $n!$ of them as a linear system. The model instead compares *two transversals that differ as little as possible* — a single swap of two entries — and reads off a 2×2 condition.

> *There are $n!$ transversals and I cannot handle them all at once. What is the smallest change I can make to a transversal and still have a transversal? Swap the columns of two of its rows. Then everything cancels except four entries, and I get $\alpha_{ik}+\alpha_{jl} = \alpha_{il}+\alpha_{jk}$.*

> *$\alpha_{ik}-\alpha_{jk} = \alpha_{il}-\alpha_{jl}$ says row i minus row j is the same in every column. So the matrix is "one row, shifted up and down". Now the count is about which shifts keep the entries in $\{-1,0,1\}$, and that depends only on which values the first row uses.*

Layer 3 THE WHY

Why the swap is natural and not a trick. The condition in (b) is about a set of objects — permutations — and the permutations of $\{1,\dots,n\}$ are *generated by transpositions*: any permutation can be reached from any other by swapping two images at a time. So "$S(\sigma)$ is the same for all σ" is equivalent to "$S(\sigma) = S(\sigma\circ\tau)$ for every σ and every transposition τ". That is the whole content of the lemma's proof. Whenever a problem says "the same for all [objects built from a group]", the cheapest test is invariance under the generators, and the transpositions are the generators everyone knows.

Why the resulting condition has to be "rows differ by constants". The equation $\alpha_{ik}+\alpha_{jl} = \alpha_{il}+\alpha_{jk}$ for all i,j,k,l is the discrete version of $\partial^2 F/\partial i\,\partial k = 0$, whose solutions are $F = R(i)+S(k)$. Here: set $r_i = \alpha_{i1}$ and $s_k = \alpha_{1k}-\alpha_{11}$; then $\alpha_{ik} = \alpha_{i1}+\alpha_{1k}-\alpha_{11} = r_i+s_k$. So a matrix satisfies (b) exactly when it is a "sum table" r_i+s_k, and then every transversal sum is $\sum r_i + \sum s_k$ because a transversal hits each row once and each column once. Seen this way the lemma is almost a definition: the transversal sum can only be independent of the transversal if it only ever reads row totals and column totals.

Why the count splits by the *set of values* in the first row. Once the matrix is "row v plus row-shifts d_i", the only question is which d_i keep $v+d_i$ in $\{-1,0,1\}^n$, and that depends on $\min v$ and $\max v$ alone: $d_i \in [-1-\min v,\ 1-\max v]$. The spread $\max v - \min v \in \{0,1,2\}$ gives 3, 2, 1 admissible shifts. The four cases are exactly spread 0, spread 1 (two flavours), spread 2. So the bases $4,3,2$ come from: $2^n\cdot 2^{n-1}$ contributes $4^n/2$ twice; 3^n appears once as "vectors of spread 2" and once as "constant rows, each with 3 shifts"; 2^n counts two-valued vectors. The form the problem asks for is the problem telling you the answer is a sum of such exponential counts.

Kedlaya–Poonen–Vakil (solution by Doug Jungreis) open with the lemma "any two rows differ by a constant vector" and prove it with the swap. Here is where it comes from: equal sums over all permutations means equal sums under every transposition, and one transposition touches only four entries.

A note on the natural attempt: the linear-system approach is not wrong. The rank computation, done properly, would reveal that the solution space is $\{r_i + s_k\}$, of dimension $2n - 1$. But it reaches that conclusion by elimination on a system of size $n!$, where the swap reaches it in one line.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "the same for all transversals": an invariance statement over all permutations. On seeing it, write down two transversals differing by one swap and subtract: $\alpha_{ik} + \alpha_{jl} = \alpha_{il} + \alpha_{jk}$.

2. Test invariance on generators (transpositions), then solve the resulting local equation: it forces $\alpha_{ik} = r_i + s_k$, i.e. rows differ by constant vectors.

3. Hesitation occurs at the count: how do the row-shifts interact with the box $\{-1, 0, 1\}$? Reinforce it away by noting the admissible shifts depend only on min and max of the first row, so classify first rows by their set of values.

4. The trap is to treat (b) as $n! - 1$ independent linear equations and try to count lattice points in their solution space by elimination. A second, smaller trap: in the last case, forgetting to subtract the constant vectors when counting first rows that use both 1 and -1 (they are counted only if both values appear).

5. Any condition of the form "the same value for every permutation / arrangement / path" should immediately be reduced to invariance under the smallest move (a swap, an adjacent transposition, a single-edge flip), which produces a local equation.

6. Reusable rule: to exploit "$S(x)$ is constant over a family", find the minimal move connecting members of the family and write $S(x) = S(x')$ for one move; the resulting local identity is usually a clean structure theorem. Then count by the few statistics (here, min and max) that the structure actually depends on.

▷ **Cue:** a quantity is the same for all permutations or transversals $\longrightarrow$ **Move:** compare two that differ by a single transposition; everything cancels except four terms

▷ **Cue:** $a_{ik} + a_{jl} = a_{il} + a_{jk}$ for all i, j, k, l $\longrightarrow$ **Move:** the matrix is a sum table $a_{ik} = r_i + s_k$; rows differ by constant vectors

▷ **Cue:** counting objects of the form "base vector plus admissible shifts" inside a box $\longrightarrow$ **Move:** the admissible shifts depend only on the min and max of the base; split cases by its range

▷ **Cue:** answer requested in the form $\sum a_i b_i^n + a_4$ $\longrightarrow$ **Move:** expect a case split whose cases count b_i^n-type families (vectors over b_i-letter alphabets); check small n by hand

1986 A5

1986 A5. Suppose $f_1(x), f_2(x), \ldots, f_n(x)$ are functions of n real variables $x = (x_1, \ldots, x_n)$ with continuous second-order partial derivatives everywhere on $\mathbb{R}^n$. Suppose further that there are constants c_{ij} such that

$$\frac{\partial f_i}{\partial x_j} - \frac{\partial f_j}{\partial x_i} = c_{ij}$$

for all i and j, $1 \leq i \leq n$, $1 \leq j \leq n$. Prove that there is a function $g(x)$ on $\mathbb{R}^n$ such that $f_i + \partial g / \partial x_i$ is linear for all i, $1 \leq i \leq n$. (A linear function is one of the form

$$a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_n x_n.)$$

Fields: multivariable calculus, vector fields **Position:** A5 — late-paper, but short once you see it; the difficulty is recognising the "curl" and knowing that curl-free fields on $\mathbb{R}^n$ are gradients **Answer:** Such g exists. The c_{ij} are antisymmetric; the linear field $h_i = \frac{1}{2}\sum_j c_{ij}x_j$ has the same "curl" c_{ij} as f, so $u = h - f$ is curl-free on $\mathbb{R}^n$, hence $u = \nabla g$ with $g(x) = \int_0^1 \sum_i x_i u_i(tx)\,dt$, and then $f_i + \partial g/\partial x_i = h_i$ is linear.

Layer 1 THE WHAT — surface analysis

The natural attempt. You read "find g with $f_i + \partial_i g$ prescribed" as a system of PDEs and try to solve it by integrating. Start with $n = 2$, writing $c = c_{12}$. You want $\partial_1 g = L_1 - f_1$ and $\partial_2 g = L_2 - f_2$ for some linear L_1, L_2 still to be chosen. Integrate the first in x_1:

$$g(x_1, x_2) = \int_0^{x_1} (L_1 - f_1)(s, x_2)\,ds + \varphi(x_2).$$

Differentiate in x_2, use $\partial_2 f_1 = \partial_1 f_2 + c$, and you find

$$\partial_2 g = \int_0^{x_1} \partial_2 L_1\,ds - \big(f_2(x_1, x_2) - f_2(0, x_2)\big) - cx_1 + \varphi'(x_2).$$

You need this to equal $L_2 - f_2$. The $f_2(x_1, x_2)$ cancels. What is left forces $L_2 = x_1\partial_2 L_1 - cx_1 + f_2(0, x_2) + \varphi'(x_2)$, and choosing $\varphi' = -f_2(0, \cdot)$ and $L_1 = 0$ gives $L_2 = -cx_1$. So $n = 2$ works: $L = (0, -cx_1)$.

Then you try $n = 3$. The same scheme integrates along x_1, then along x_2 with $x_1 = 0$, then along x_3 with $x_1 = x_2 = 0$, and at each stage you must choose the linear corrections so that three separate compatibility conditions close. It can be pushed through, but the bookkeeping grows with n, and it is easy to lose a sign or forget that c_{ij} and c_{ji} are the same data. The attempt works but does not tell you *why* a linear correction is always enough, and a general-n writeup becomes a page of nested integrals.

The model chain. *Step 1: the c_{ij} are antisymmetric.* Putting $j = i$ gives $c_{ii} = 0$; swapping i and j gives $c_{ji} = -c_{ij}$.

Step 2: a linear field with the same curl. Define

$$h_i(x) = \tfrac{1}{2}\sum_{j=1}^{n} c_{ij}x_j.$$

Each h_i is linear, $\partial h_i/\partial x_j = \frac{1}{2}c_{ij}$, and

$$\frac{\partial h_i}{\partial x_j} - \frac{\partial h_j}{\partial x_i} = \tfrac{1}{2}c_{ij} - \tfrac{1}{2}c_{ji} = c_{ij}.$$

Step 3: the difference is curl-free. Put $u_i = h_i - f_i$. Each u_i is C^1 (indeed C^2), and

$$\frac{\partial u_i}{\partial x_j} - \frac{\partial u_j}{\partial x_i} = c_{ij} - c_{ij} = 0 \qquad \text{for all } i, j.$$

Step 4: curl-free on $\mathbb{R}^n$ means gradient. Define

$$g(x) = \int_0^1 \sum_{i=1}^{n} x_i\, u_i(tx)\,dt.$$

The integrand is C^1 in (t, x), so we may differentiate under the integral sign:

$$\frac{\partial g}{\partial x_k} = \int_0^1 \Big(u_k(tx) + \sum_i x_i\, t\,(\partial_k u_i)(tx)\Big)dt = \int_0^1 \Big(u_k(tx) + t\sum_i x_i\,(\partial_i u_k)(tx)\Big)dt,$$

using $\partial_k u_i = \partial_i u_k$ from Step 3. By the chain rule the integrand is $\frac{d}{dt}\big[t\, u_k(tx)\big]$, so

$$\frac{\partial g}{\partial x_k} = \Big[t\, u_k(tx)\Big]_{t=0}^{t=1} = u_k(x).$$

Step 5: conclude. $f_i + \partial g/\partial x_i = f_i + u_i = h_i$, which is linear (with $a_0 = 0$). □

Checked symbolically for $n = 3$ with a non-trivial f (a gradient of $\sin x_1\, x_2^2 + x_1 e^{x_3}$ plus a linear field with prescribed antisymmetric c_{ij}): the constructed g gives $f_i + \partial_i g = h_i$ exactly.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains read the goal as "make $f + \nabla g$ linear". The natural attempt solves for g by integrating coordinate by coordinate. The model first asks what the hypothesis says about f as a vector field: $\partial_j f_i - \partial_i f_j$ is the curl, and the problem says the curl is constant. Then it removes the constant curl with a linear field and uses the one theorem that says curl-free fields are gradients.

> *$\partial f_i/\partial x_j - \partial f_j/\partial x_i$ is the curl of the vector field f. Adding a gradient never changes the curl, because mixed partials of g commute. So the curl is the only thing I cannot change, and the problem says it is constant. I need a linear field with that same constant curl.*

> *A linear field $h = Mx$ has curl $M - M^T$. I want $M - M^T = C$ with C antisymmetric, and $M = \frac{1}{2}C$ does it. Then $h - f$ has zero curl, and on all of $\mathbb{R}^n$ zero curl means gradient.*

Layer 3 THE WHY

Why the curl is the object of interest. The problem asks for g such that $f + \nabla g$ lies in a certain class. Adding ∇g to f changes $\partial_j f_i$ by $\partial_j \partial_i g$, which is symmetric in i, j by equality of mixed partials (Clairaut–Schwarz, valid because everything is C^2). So the antisymmetric part $\partial_j f_i - \partial_i f_j$ is invariant under the moves you are allowed to make. Whenever a problem lets you modify an object by some family of moves, the first question is "what do the moves not change?" Here that invariant is exactly the expression the problem hands you, and the problem says it is constant. That is not a coincidence of phrasing; it is the setter telling you what the invariant is.

Why a linear correction is always enough. The curl operator is linear, and on linear fields $h = Mx$ it is $M - M^T$, whose image is every antisymmetric matrix. So every constant curl is the curl of some linear field. Once you have that, the question reduces to the case of zero curl. The factor $\frac{1}{2}$ is not decoration: $h_i = \sum_j c_{ij} x_j$ would give curl $2c_{ij}$. (Taking only the strictly upper-triangular part of C works equally well; any M with $M - M^T = C$ does.)

Why zero curl means gradient on $\mathbb{R}^n$. This is the Poincaré lemma, and the proof in Step 4 is the fundamental theorem of calculus along straight lines from the origin: $g(x)$ is the line integral of u along the segment from 0 to x. Its derivative in the direction e_k differs from u_k by an error that, after swapping $\partial_k u_i$ for $\partial_i u_k$, is a total t-derivative, so it integrates to the boundary term. Zero curl is used exactly once, at that swap. The domain matters: $\mathbb{R}^n$ is star-shaped about 0, so the segment from 0 to tx stays inside for every $t \in [0, 1]$. On a punctured plane the lemma fails (the angle field $(-y, x)/(x^2 + y^2)$ is curl-free but not a gradient). That is why the problem says "everywhere on $\mathbb{R}^n$".

Kedlaya–Poonen–Vakil's solution sets $h_i = \frac{1}{2}\sum_j c_{ij} x_j$ and then says "hence $(h_1 - f_1, \ldots, h_n - f_n)$ is a gradient". Here is where both come from: h is the simplest field with the given invariant, and "hence" is the Poincaré lemma, which in a written solution you should either cite by name or prove as in Step 4.

Why the natural attempt worked for $n = 2$ but felt unmotivated. Integrating along x_1 and then along x_2 is the Poincaré lemma along a staircase path instead of a straight segment. It does the same job, but because it never names the curl, it cannot see in advance that a linear

correction with $L = (0, -cx_1)$ was inevitable: $(0, -cx_1)$ is precisely a linear field with curl c, the lower-triangular choice $M = \begin{pmatrix} 0 & 0 \\ -c & 0 \end{pmatrix}$, for which $M - M^T = C$.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the expression $\partial f_i/\partial x_j - \partial f_j/\partial x_i$. On seeing it, write "this is the curl of f, adding ∇g does not change it" and note $c_{ji} = -c_{ij}$, $c_{ii} = 0$.

2. Reduce to the zero-curl case by subtracting a linear field with the same constant curl, $h = \frac{1}{2}Cx$; then apply the Poincaré lemma on $\mathbb{R}^n$.

3. Hesitation occurs at "curl-free implies gradient", which many students half-remember only in $\mathbb{R}^3$. Reinforce it by having the one-line proof ready: $g(x) = \int_0^1 x \cdot u(tx)\,dt$, differentiate, swap $\partial_k u_i = \partial_i u_k$, recognise $\frac{d}{dt}[t\,u_k(tx)]$.

4. The trap is the coordinate-by-coordinate PDE solve, which succeeds for $n = 2$ and turns into unmanageable bookkeeping for general n. A smaller trap: writing $h_i = \sum_j c_{ij}x_j$ without the $\frac{1}{2}$, which doubles the curl.

5. When a problem allows you to modify an object by a family of moves and asks you to reach a target class, compute the invariant of the moves first; the target is reachable exactly when some member of the class shares that invariant.

6. Reusable rule: "the obstruction is an invariant; kill it with the simplest object that has the same invariant, then use the theorem for the zero-invariant case". Here: constant curl, linear field, Poincaré lemma. The same shape solves exact-differential ODEs (integrating factors), cohomology questions, and "make this sequence arithmetic by adding a telescoping term" problems.

▷ **Cue:** an expression $\partial f_i/\partial x_j - \partial f_j/\partial x_i$ $\longrightarrow$ **Move:** recognise the curl; it is unchanged by adding a gradient, since mixed partials commute

▷ **Cue:** a constant antisymmetric matrix C that must be realised as a curl $\longrightarrow$ **Move:** use the linear field $h = \frac{1}{2}Cx$, whose curl is C

▷ **Cue:** a field with $\partial_j u_i = \partial_i u_j$ on all of $\mathbb{R}^n$ $\longrightarrow$ **Move:** it is a gradient: $g(x) = \int_0^1 x \cdot u(tx)\,dt$ (Poincaré lemma)

▷ **Cue:** "find g such that $f + \nabla g$ lies in a class" $\longrightarrow$ **Move:** first find what adding ∇g cannot change; match that invariant with an explicit member of the class

1986 A6

1986 A6. Let $a_1, a_2, \ldots, a_n$ be real numbers, and let $b_1, b_2, \ldots, b_n$ be distinct positive integers. Suppose that there is a polynomial $f(x)$ satisfying the identity

$$(1-x)^n f(x) = 1 + \sum_{i=1}^{n} a_i x^{b_i}.$$

Find a simple expression (not involving any sums) for $f(1)$ in terms of $b_1, b_2, \ldots, b_n$ and n (but independent of $a_1, a_2, \ldots, a_n$).

Fields: algebra, polynomials, differential operators **Position:** A6 — the hardest slot on paper A, and the scores show it (one top mark in the official tally); short once the right operator is chosen, a determinant slog otherwise **Answer:** $f(1) = \dfrac{b_1 b_2 \cdots b_n}{n!}$. Apply $L = \prod_{i=1}^{n}\left(x\frac{d}{dx} - b_i\right)$ to both sides and set $x = 1$: it annihilates every $a_i x^{b_i}$ and sends 1 to $(-1)^n \prod b_i$, while on the left only the n-th derivative of $(1-x)^n$ survives, giving $(-1)^n n!\, f(1)$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You test small cases. For $n = 1$, $(1-x)f(x) = 1 + a_1 x^{b_1}$ forces $a_1 = -1$

and $f = 1 + x + \cdots + x^{b_1-1}$, so $f(1) = b_1$. For $n = 2$ with $(b_1, b_2) = (1, 2)$, $(1-x)^2 = 1 - 2x + x^2$ gives $f = 1$, and $b_1 b_2 / 2! = 1$. With $(2, 5)$ a short computation gives $f(1) = 5 = 10/2$. The guess $f(1) = \prod b_i / n!$ appears quickly.

To prove it you do what the factor $(1-x)^n$ suggests: $x = 1$ is a root of multiplicity n of $P(x) = 1 + \sum a_i x^{b_i}$. So $P^{(j)}(1) = 0$ for $0 \le j \le n-1$, and by Leibniz on the left, $P^{(n)}(1) = (-1)^n n!\, f(1)$. Writing $(b)_j = b(b-1)\cdots(b-j+1)$, this is the linear system

$$1 + \sum_i a_i = 0, \quad \sum_i a_i (b_i)_1 = 0, \ \ldots, \ \sum_i a_i (b_i)_{n-1} = 0, \quad \sum_i a_i (b_i)_n = (-1)^n n!\, f(1).$$

Now you have $n+1$ equations in the n unknowns a_i and the one quantity $f(1)$ you want. You can solve the first n equations for the a_i by Cramer's rule and substitute into the last, or observe that $(-1, a_1, \ldots, a_n)$ is a nonzero kernel vector of an $(n+1) \times (n+1)$ matrix, so its determinant vanishes. Either way you must evaluate determinants whose rows are falling factorials $(b_i)_j$. Row-reduce them to powers b_i^j (each $(b)_j$ is monic of degree j with no constant term), expand along the first column, recognise two Vandermonde determinants, factor b_i from each column of one of them, and cancel. That is Kedlaya–Poonen–Vakil's Solution 1. It is correct and complete, but it demands that you know the Vandermonde determinant, carry an $(n+1) \times (n+1)$ matrix through row operations, and track signs; under time pressure most people did not finish.

The model chain. Write $\theta = x \dfrac{d}{dx}$. Two facts about θ:

- $\theta\, x^b = b x^b$ for every integer $b \ge 0$; in particular $\theta(1) = 0$. So $(\theta - b)\, x^b = 0$ and $(\theta - b)\, 1 = -b$.
- $\theta^m = x^m \dfrac{d^m}{dx^m} + \sum_{k<m} s_{m,k}\, x^k \dfrac{d^k}{dx^k}$ for some constants $s_{m,k}$ (induction on m: $\theta(x^k D^k) = k x^k D^k + x^{k+1} D^{k+1}$, where $D = d/dx$).

Let $L = (\theta - b_1)(\theta - b_2)\cdots(\theta - b_n)$. The factors are polynomials in the single operator θ, so they commute. By the second fact,

$$L = x^n D^n + \sum_{k<n} p_k\, x^k D^k \qquad (p_k \text{ constants}).$$

Right-hand side. Each x^{b_i} is killed by its own factor $(\theta - b_i)$, which may be applied first. And $L(1) = \prod_i (0 - b_i) = (-1)^n\, b_1 \cdots b_n$. Hence

$$L\Big(1 + \sum_i a_i x^{b_i}\Big) = (-1)^n\, b_1 b_2 \cdots b_n \quad \text{(a constant, whatever the } a_i \text{ are)}.$$

Left-hand side. Let $Q(x) = (1-x)^n f(x)$. For $k < n$, every term of the Leibniz expansion of $Q^{(k)}$ contains a derivative of $(1-x)^n$ of order at most $k < n$, which vanishes at $x = 1$; so $Q^{(k)}(1) = 0$. For $k = n$ the only surviving term is $\big[D^n (1-x)^n\big] f(1) = (-1)^n n!\, f(1)$. Therefore

$$(LQ)(1) = 1^n\, Q^{(n)}(1) + \sum_{k<n} p_k\, Q^{(k)}(1) = (-1)^n n!\, f(1).$$

Equate. $(-1)^n n!\, f(1) = (-1)^n b_1 \cdots b_n$, so

$$f(1) = \frac{b_1 b_2 \cdots b_n}{n!}.$$

Checks. $n = 1$ gives b_1; $(b_1, b_2) = (1, 2)$ gives 1. I also verified by computer algebra, for $(1, 2)$, $(2, 5)$, $(1, 3, 4)$, $(2, 3, 7)$, $(1, 4, 6, 9)$, that solving for the a_i, dividing by $(1-x)^n$ and evaluating at 1 gives exactly $\prod b_i / n!$.

Remark on the hypotheses. The computation never used that the b_i are distinct; distinctness is what guarantees that some choice of a_i exists (the $n \times n$ falling-factorial system $\sum_i a_i (b_i)_j = -[j = 0]$, $0 \le j < n$, is then invertible), so the "suppose" is not vacuous.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree that $x = 1$ is an n-fold root and that the information sits in the derivatives at $x = 1$. They split at the next line. The natural attempt differentiates P with d/dx and gets falling factorials $(b_i)_j$, which leaves the a_i entangled in a linear system. The model differentiates with $x\,d/dx$, for which each monomial x^{b_i} is an eigenvector, and builds the operator that deletes every $a_i x^{b_i}$ at once.

> *The answer is supposed to be independent of $a_1, \dots, a_n$. So I want an operation that makes the a_i disappear without my ever solving for them. Something that kills x^{b_i} regardless of its coefficient. $x\,d/dx$ multiplies x^b by b, so $(x\,d/dx - b_i)$ kills x^{b_i}, and the product over i kills all of them.*

> *On the other side, $(1-x)^n$ vanishes to order n at 1, so any differential operator of order n evaluated at $x = 1$ only sees the top derivative. The leading term of $\prod(x\,d/dx - b_i)$ is $x^n d^n/dx^n$, which at $x = 1$ is d^n/dx^n.*

Layer 3 THE WHY

Why $x\,d/dx$ and not d/dx. The right-hand side is a sum of monomials x^{b_i} with unknown coefficients and known exponents. The operator whose eigenfunctions are exactly the monomials is $\theta = x\,d/dx$: $\theta x^b = b x^b$. (Equivalently, put $x = e^t$; then $\theta = d/dt$ and $x^b = e^{bt}$ are exponentials. This is Kedlaya–Poonen–Vakil's Solution 2: the right side minus 1 becomes $\sum a_i e^{b_i t}$, which solves the constant-coefficient ODE with characteristic polynomial $\prod(z - b_i)$.) Once you have an operator with the monomials as eigenvectors, killing a given monomial is the definition of an eigenvalue: $(\theta - b_i)x^{b_i} = 0$. The ordinary derivative, by contrast, sends x^b to $b x^{b-1}$, changing the monomial, so the a_i never separate cleanly and you are forced into a determinant.

Why the phrase "independent of $a_1, \dots, a_n$" is the cue. The setters are telling you the answer does not see the a_i. That is an instruction: find a procedure that eliminates them. In linear algebra language, you want a linear functional that vanishes on $\operatorname{span}\{x^{b_1}, \dots, x^{b_n}\}$; the functional $P \mapsto (LP)(1)$ is one, and it is the natural one because the monomials are eigenvectors of θ. Solution 1's determinant is the same elimination done by brute force: Cramer's rule is a way of finding a combination of the equations in which the a_i cancel, and the Vandermonde determinant is the price of not knowing that combination in advance.

Why the left side collapses to one term. The factor $(1-x)^n$ is a statement about vanishing order at $x = 1$: Q and its first $n-1$ derivatives vanish there. Any linear differential operator of order n is "top derivative plus lower derivatives", and evaluated at $x = 1$ the lower ones see only zeros. So you do not need to know the lower-order coefficients p_k of L at all; only the leading coefficient x^n, which is 1 at $x = 1$. That is why the computation is short: every piece of information you do not have is multiplied by zero.

Why the answer has the form it does. The identity $(-1)^n n! f(1) = L(1) = \prod(-b_i)$ says that $n!\,f(1)$ is the characteristic polynomial $\prod(z - b_i)$ evaluated at $z = 0$, up to sign: the constant term 1 on the right is x^0, the eigenvector with eigenvalue 0. That is the structural reason $f(1)$ is a product of the b_i.

Kedlaya–Poonen–Vakil's Solution 1 begins by differentiating j times and setting $x = 1$. Here is where that comes from: $(1-x)^n$ means an n-fold root at 1, and derivatives at 1 are the way to read vanishing order. The archive's model keeps that idea and changes only the choice of derivative.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "independent of $a_1, \ldots, a_n$" together with the monomials x^{b_i}. On seeing it, write $\theta = x\,d/dx$, $\theta x^b = bx^b$, and $L = \prod(\theta - b_i)$. The second cue is $(1-x)^n$: write "n-fold root at $x = 1$, only the n-th derivative survives there".

2. Build the annihilating operator of the unwanted terms and apply it to both sides of the identity, then evaluate at the point where the other side simplifies.

3. Hesitation occurs when you have the $n+1$ equations $\sum a_i (b_i)_j = \ldots$ in hand and must decide how to eliminate the a_i. Reinforce it away: do not solve for unknowns the answer does not depend on; find the operator (or linear combination) that kills them.

4. The trap is the ordinary derivative, which produces falling factorials and forces a Vandermonde computation that is correct but long and sign-prone. A secondary trap: guessing $\prod b_i/n!$ from small cases and not proving it.

5. When a sum $\sum a_i x^{b_i}$ has unknown coefficients and known exponents (or $\sum a_i e^{\lambda_i t}$, or $\sum a_i r_i^k$ in a sequence), think of the operator that has those terms as eigenvectors: $x\,d/dx$, d/dt, or the shift E. Its polynomial $\prod(\text{operator} - \text{eigenvalue})$ annihilates the whole sum.

6. Reusable rule: if the answer is claimed independent of some parameters, construct the linear operation that annihilates the terms carrying those parameters, and apply it to the whole identity. It is the same principle as the characteristic polynomial of a linear recurrence and as Newton's identities.

▷ **Cue:** answer stated to be independent of the coefficients a_i $\longrightarrow$ **Move:** do not solve for the a_i; build an operation that annihilates every term carrying them

▷ **Cue:** a sum $\sum a_i x^{b_i}$ with known exponents $\longrightarrow$ **Move:** use $\theta = x\,d/dx$, which has x^b as eigenvector with eigenvalue b; $\prod(\theta - b_i)$ kills the sum

▷ **Cue:** a factor $(1-x)^n$ multiplying an unknown polynomial $\longrightarrow$ **Move:** at $x = 1$ every derivative of order below n vanishes; an order-n operator sees only $(-1)^n n!\, f(1)$ times its leading coefficient

▷ **Cue:** a linear system whose rows are $(b_i)_j$ or b_i^j $\longrightarrow$ **Move:** Vandermonde structure; either compute the determinant $\prod_{i>j}(b_i - b_j)$ or look for the operator that makes the system unnecessary

Putnam 1986 — Paper B

1986 B1

1986 B1. Inscribe a rectangle of base b and height h in a circle of radius one, and inscribe an isosceles triangle in the region of the circle cut off by one base of the rectangle (with that side as the base of the triangle). For what value of h do the rectangle and triangle have the same area?

Fields: plane geometry, elementary algebra **Position:** B1 – opening problem of the second session, the easiest problem on the paper **Answer:** $h = \dfrac{2}{5}$ (and then $b = \dfrac{4\sqrt{6}}{5}$); this is the only value.

Layer 1 THE WHAT — surface analysis

The natural attempt. You draw the circle $x^2 + y^2 = 1$ and the rectangle centred at the origin with corners $(\pm b/2, \pm h/2)$. Inscribed means the corners lie on the circle, so

$$\frac{b^2}{4} + \frac{h^2}{4} = 1, \qquad b = \sqrt{4 - h^2}.$$

The triangle sits on the top side, from $(-b/2, h/2)$ to $(b/2, h/2)$, and its apex is on the arc. Isosceles with that side as base puts the apex on the perpendicular bisector of the base, i.e. on the y-axis, so the apex is $(0, 1)$ and the triangle's height is $1 - h/2$. Now you set up the equation with b written out:

$$h\sqrt{4 - h^2} = \frac{1}{2}\sqrt{4 - h^2}\Big(1 - \frac{h}{2}\Big).$$

A student in "solve the equation" mode squares both sides here, because there is a square root in the way:

$$h^2(4 - h^2) = \frac{1}{4}(4 - h^2)\Big(1 - \frac{h}{2}\Big)^2,$$

expands to a quartic, and starts hunting for roots. The quartic factors as $(4 - h^2)\big(h^2 - \frac{1}{4}(1 - \frac{h}{2})^2\big)$, and the second factor gives $h = \frac{2}{5}$ or $h = -\frac{2}{3}$, while the first gives $h = \pm 2$. Now you have four candidates, two introduced by squaring, one degenerate ($h = 2$ gives $b = 0$, no rectangle at all), and you must argue each away. It works. It costs five minutes and creates three chances to keep a spurious root.

The model chain. Let O be the centre and let X be the point where the radius perpendicular to the top base meets the circle. The base of the rectangle is a chord at distance $h/2$ from O (the rectangle is symmetric about O, because the perpendicular bisectors of its sides are diameters of the circle). The isosceles triangle on that chord has its apex on the chord's perpendicular bisector, which passes through O; the apex is on the arc cut off by the chord, so it is X. Along the radius OX, of length 1, the first $h/2$ lies inside the rectangle and the rest is the triangle's altitude, so

$$\text{altitude of triangle} = 1 - \frac{h}{2}.$$

Equal areas means

$$bh = \frac{1}{2}\, b\Big(1 - \frac{h}{2}\Big).$$

Since $b > 0$ (a rectangle has positive base), divide by b:

$$h = \frac{1}{2} - \frac{h}{4} \quad \Longrightarrow \quad \frac{5h}{4} = \frac{1}{2} \quad \Longrightarrow \quad h = \frac{2}{5}.$$

This is admissible: $0 < \frac{2}{5} < 2$, and then $b = \sqrt{4 - \frac{4}{25}} = \frac{4\sqrt{6}}{5}$, with both areas equal to $\frac{8\sqrt{6}}{25}$. Because the equation after cancelling b is linear, $h = \frac{2}{5}$ is the only value.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree up to the equal-area equation $bh = \frac{1}{2}b(1 - h/2)$. At that line the natural attempt replaces b by $\sqrt{4 - h^2}$ – the substitution it has been trained to make, because "one equation, one unknown" – and then squares. The model notices that b appears as a factor on both sides and cancels it. The circle constraint $b^2 + h^2 = 4$ is never used at all.

> *Before I substitute for b, is b even in the answer? Both areas are b times something. It cancels. The circle only enters through the radius $OX = 1$.*

> *The triangle's height is the part of a radius that sticks out beyond the rectangle: $1 - h/2$. I do not need the apex's coordinates, I need the length of one radius.*

Layer 3 THE WHY

Why is cancelling b the natural move and not luck? Look at what an area is here. Both figures stand on the same segment: the rectangle's top side *is* the triangle's base. Area of a rectangle is base times height; area of a triangle is half base times height. When two figures share a base, their area ratio is the ratio of their heights. That is not a trick, it is the definition of area for these two shapes read side by side. So the moment you see "with that side as the base of the triangle", the problem has told you that the comparison is a comparison of heights: h against $\frac{1}{2}(1 - h/2)$. The base is common and drops out, exactly as the common base drops out when you compare two triangles between the same parallels.

The phrasing of the question confirms it. It asks "for what value of h", not for b and h. If the answer needed the circle equation linking b to h, the problem would be a genuine two-variable system. The fact that it names only h is a hint that h is determined by a relation not involving b.

Where does the circle enter, then? Through the single fact that a circle is the set of points at distance one from O. The apex X is on the circle, so $OX = 1$. The rectangle is inscribed and therefore centred at O (its diagonals are chords through its centre that bisect each other and are equal, so they are diameters), so its top side is at height $h/2$. The triangle's altitude is the difference of those two lengths along one line. That is the Blueprint circle principle in its plainest form: the centre is the object of interest, so measure along a radius.

Kedlaya, Poonen and Vakil's solution is two lines: "The radius OX has length equal to $h/2$ plus the altitude of the triangle". Here is where that comes from: the altitude of an isosceles triangle lies on the perpendicular bisector of its base; the perpendicular bisector of a chord passes through the centre; so the altitude, extended, is a radius. Once you see that the altitude and the rectangle's half-height are collinear pieces of one radius, the rest is subtraction.

The squaring route is a clean case of zombie maths. It is not wrong, but it manufactures roots ($h = -\frac{2}{3}$, $h = \pm 2$) that the geometry never had, and then it charges you to remove them. The rule is: never square an equation while a common factor is still sitting on both sides of it.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "with that side as the base of the triangle": the two figures share a base. On seeing it, write down both areas with the shared base factored out, $b \cdot h$ and $b \cdot \frac{1}{2}(1 - h/2)$.

2. It suggests comparing heights, not areas: cancel the common base and solve a linear equation in h.

3. Hesitation comes at the equation $bh = \frac{1}{2}b(1 - h/2)$, where the reflex is to eliminate b using $b^2 + h^2 = 4$. What reinforces it away is asking, before any substitution, whether the variable you are about to eliminate appears as a common factor.

4. The trap is using the inscribed-rectangle constraint at all. It is a true fact that is irrelevant to the answer, and invoking it leads to squaring, a quartic, and spurious roots. A secondary trap is placing the apex anywhere other than the midpoint of the arc; "isosceles with that side as base" forces the apex onto the chord's perpendicular bisector.

5. When two figures are built on the same segment, their areas are proportional to their heights. When a problem asks for one variable out of two, suspect that the other cancels before you eliminate it.

6. Cancel common factors before substituting or squaring; and in a circle, turn every length you need into a piece of a radius.

▷ **Cue:** two figures share a base ("with that side as the base") $\longrightarrow$ **Move:** compare heights: the area ratio is the height ratio, and the base cancels

▷ **Cue:** the question asks for one variable although the configuration has two $\longrightarrow$ **Move:** look for the unasked variable to cancel before eliminating it by a constraint

▷ **Cue:** an isosceles triangle stands on a chord of a circle $\longrightarrow$ **Move:** its altitude lies on a diameter; measure it as radius minus the chord's distance from the centre

▷ **Cue:** an equation has a square root and a common factor on both sides $\longrightarrow$ **Move:** divide out the common factor first; square only if a root is still in the way

1986 B2

1986 B2. Prove that there are only a finite number of possibilities for the ordered triple $T = (x - y, y - z, z - x)$, where x, y, z are complex numbers satisfying the simultaneous equations

$$x(x-1) + 2yz = y(y-1) + 2zx = z(z-1) + 2xy,$$

and list all such triples T.

Fields: algebra (polynomial systems, factorisation) **Position:** B2 – early second-session problem, routine once the right subtraction is made **Answer:** Exactly four triples: $T = (0,0,0)$, $(0,-1,1)$, $(1,0,-1)$, $(-1,1,0)$.

Layer 1 THE WHAT — surface analysis

The natural attempt. Three unknowns, and the chain of equalities gives two equations. You call the common value k and write

$$x^2 - x + 2yz = k, \qquad y^2 - y + 2zx = k, \qquad z^2 - z + 2xy = k,$$

now three equations in four unknowns x, y, z, k. You try to solve: from the first, $yz = (k - x^2 + x)/2$; you substitute into the others and the system turns into a tangle of quadratics. Or you try adding all three: $x^2 + y^2 + z^2 + 2(xy + yz + zx) - (x + y + z) = 3k$, i.e. $(x + y + z)^2 - (x + y + z) = 3k$ – a pleasant identity which pins down k in terms of $s = x + y + z$ but says nothing about the differences. At this point the solution set is visibly infinite (one free parameter at least: if (x, y, z) works, does $(x + c, y + c, z + c)$?), and you are not sure how to get a finite list out of

an infinite set. The attempt stalls not because the algebra is hard, but because it is aimed at the wrong unknowns: the question asks about T, and every line so far is about x, y, z, k.

The model chain. Subtract the second expression from the first:

$$(x^2 - x + 2yz) - (y^2 - y + 2zx) = (x^2 - y^2) - (x - y) - 2z(x - y) = (x - y)(x + y - 1 - 2z).$$

By the cyclic symmetry $x \to y \to z \to x$ of the system, the same computation gives

$$(y^2 - y + 2zx) - (z^2 - z + 2xy) = (y - z)(y + z - 1 - 2x),$$

$$(z^2 - z + 2xy) - (x^2 - x + 2yz) = (z - x)(z + x - 1 - 2y).$$

The system says the three expressions are equal, which is equivalent to the vanishing of all three differences (any two of them suffice, and imply the third, since the three differences sum to zero). So the system is equivalent to

$$(x - y)(x + y - 1 - 2z) = 0, \quad (y - z)(y + z - 1 - 2x) = 0, \quad (z - x)(z + x - 1 - 2y) = 0. \ (*)$$

Case 1: x, y, z pairwise distinct. Then all three second factors vanish. Adding them: $(x + y - 2z) + (y + z - 2x) + (z + x - 2y) - 3 = 0$, i.e. $-3 = 0$. Impossible. So at least two of x, y, z are equal.

Case 2: all three equal. Then $T = (0, 0, 0)$, and $(*)$ holds (e.g. $x = y = z = 0$).

Case 3: exactly two equal. If $x = y \neq z$: the first equation of $(*)$ holds; the second requires $y + z - 1 - 2x = 0$, i.e. $z = x + 1$; the third then reads $(z - x)(z + x - 1 - 2y) = 1 \cdot (x + 1 + x - 1 - 2x) = 0$, which holds. So $T = (0,\ x - z,\ z - x) = (0, -1, 1)$. If $y = z \neq x$: the third equation forces $z + x - 1 - 2y = 0$, i.e. $x = y + 1$, and then the first reads $(x - y)(x + y - 1 - 2z) = 1 \cdot (y + 1 + y - 1 - 2y) = 0$. So $T = (1, 0, -1)$. If $z = x \neq y$: the first equation forces $x + y - 1 - 2z = 0$, i.e. $y = x + 1$, and the second reads $(y - z)(y + z - 1 - 2x) = 1 \cdot 0 = 0$. So $T = (-1, 1, 0)$.

Every triple in the list occurs: $(x, y, z) = (0, 0, 0), (0, 0, 1), (1, 0, 0), (0, 1, 0)$ give $T = (0, 0, 0), (0, -1, 1), (1, 0, -$ respectively; for instance at $(0, 0, 1)$ the three expressions are $0, 0, 1 \cdot 0 + 0 = 0$. Hence there are exactly four possible triples.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start from the same two equations. The natural attempt introduces a common value k and tries to solve for x, y, z. The model subtracts the expressions pairwise, and the split happens at the very first line: "call the common value k" versus "subtract adjacent members of the chain".

> *The problem asks about $x - y$, $y - z$, $z - x$, not about x, y, z. So I want equations whose unknowns are the differences. Subtracting two of the expressions will throw away whatever the three have in common and leave me something about the differences.*

> *Each difference vanishes when $x = y$, so it must have $(x - y)$ as a factor. I will pull that factor out before doing anything else.*

> *If none of the variables coincide, three linear equations must all hold. Let me add them and see whether they are even consistent.*

Layer 3 THE WHY

Why is subtraction the natural move? Read what the problem is asking for. It asks for the set of possible $T = (x - y, y - z, z - x)$, and it says that set is finite. But the solution set

in (x, y, z) is plainly not finite: the equations are equalities between three expressions, and a chain of equalities with three unknowns leaves at least one degree of freedom. The only way an infinite set of (x, y, z) can produce a finite set of differences is if the infinite freedom lies in a direction that does not change the differences – adding the same amount to each variable – or if the equations only constrain the differences in the first place. Either way, the problem is telling you to find equations in the differences. And the only operation that manufactures differences from a chain $A = B = C$ is subtraction: $A - B = 0$, $B - C = 0$.

(An aside you should check rather than trust: is the solution set invariant under $(x, y, z) \mapsto (x + c, y + c, z + c)$? The three expressions themselves are not – each changes by a different amount, since the $-x$ terms and the products yz shift differently. But the conditions $(*)$ involve only $x - y$ and $x + y - 2z = (x - z) + (y - z)$, which are both differences, so $(*)$, and hence the solution set, *is* invariant under the shift. For example $x = y = z = t$ is a solution for every t. Subtraction has done exactly what we wanted: it has projected away the common-value information, including k, and left a system purely in differences.)

The second decision, factoring out $(x - y)$, is forced by a structural fact about polynomials: a polynomial in x that vanishes when $x = y$ is divisible by $x - y$ (the factor theorem, with y treated as a constant). The expression $(x^2 - x + 2yz) - (y^2 - y + 2zx)$ is antisymmetric under swapping x and y – swapping them negates it – so it vanishes on $x = y$, so $x - y$ divides it. You do not need to spot the factorisation; you need to know it must be there and then divide.

The third decision, adding the second factors, is the standard test of consistency for an overdetermined linear system: three equations, and the linear parts $x + y - 2z$, $y + z - 2x$, $z + x - 2y$ sum to zero because each variable appears with coefficients $1, 1, -2$. Any time you have cyclic linear forms, add them; the cyclic sum of $x + y - 2z$ is the first thing to compute because cyclic symmetry makes it collapse.

Kedlaya, Poonen and Vakil's solution opens with "Subtracting ... we find that the given system is equivalent to". The motivation is the one above: the question is about differences, and a factorisation into (difference)(linear) turns a quadratic system into a finite case split. Every $(x - y)(\ldots) = 0$ is a fork with two branches, three forks give eight branches, and most die at once.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that the unknowns of the question are $x - y$, $y - z$, $z - x$, and the equations are a chain $A = B = C$. On seeing it, write down $A - B = 0$ and $B - C = 0$.

2. Factor each difference by the factor theorem (it vanishes when two variables are equal), then split into cases according to which factor is zero.

3. Hesitation comes in the "all distinct" case, where you have three linear equations in three unknowns and may start solving them. Reinforce it away by adding them first: cyclic linear forms usually sum to a constant, and here the sum is $-3 = 0$.

4. The trap is naming the common value k and solving for x, y, z, which keeps you in a space where the solution set is infinite. A second trap is forgetting to exhibit a solution for each triple: "the only possibilities are ..." needs every listed triple to be attained.

5. When the quantity asked for is a set of differences, the equations should be subtracted; when an expression is antisymmetric in two variables, their difference is a factor.

6. Rewrite the system in the variables the question asks about before solving anything; and when a product of factors is zero, branch on the factors and kill branches by summing.

▷ **Cue:** a chain of equalities $A = B = C$ and the question asks about differences of the variables $\longrightarrow$ **Move:** subtract adjacent members; each difference is an equation in the differences

▷ **Cue:** a polynomial expression changes sign when two variables are swapped $\longrightarrow$ **Move:** the difference of those variables is a factor; divide it out

▷ **Cue:** a product $(x - y)(\text{linear}) = 0$ for each pair $\longrightarrow$ **Move:** split into the cases "some variables equal" and "all distinct"; in the distinct case the linear factors all vanish

▷ **Cue:** several cyclic linear equations such as $x+y-2z-1=0$ and its rotations $\longrightarrow$ **Move:** add them first; the variable parts often cancel and expose an inconsistency

▷ **Cue:** "list all possibilities" in a problem $\longrightarrow$ **Move:** after the necessity argument, exhibit one explicit solution for each item in the list

1986 B3

1986 B3. Let Γ consist of all polynomials in x with integer coefficients. For f and g in Γ and m a positive integer, let $f \equiv g \pmod{m}$ mean that every coefficient of $f-g$ is an integral multiple of m. Let n and p be positive integers with p prime. Given that f, g, h, r and s are in Γ with $rf+sg \equiv 1 \pmod{p}$ and $fg \equiv h \pmod{p}$, prove that there exist F and G in Γ with $F \equiv f \pmod{p}$, $G \equiv g \pmod{p}$, and $FG \equiv h \pmod{p^n}$.

Fields: number theory, polynomial algebra (Hensel lifting) **Position:** B3 – middle of the second session; short once the induction is set up, but only 26 of the roughly 200 top-ranked contestants received full marks **Answer:** Proof by induction on the exponent: lift F_k, G_k to $F_k + p^k st$, $G_k + p^k rt$ where $h - F_k G_k = p^k t$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You read "prove there exist F and G" and try to write them down. The obvious candidates are $F=f$, $G=g$, which work for $n=1$ but not in general: $h-fg$ is divisible by p and nothing more. So you try to correct f and g all at once: set $F=f+pu$, $G=g+pv$ and demand $FG \equiv h \pmod{p^n}$, i.e.

$$p(uf+vg)+p^2uv \equiv h-fg \pmod{p^n}.$$

Writing $h-fg=pt$, you need $uf+vg+puv \equiv t \pmod{p^{n-1}}$. That is a *nonlinear* congruence in the unknown polynomials u, v, modulo a high power of p, and the only tool you are given, $rf+sg \equiv 1$, is a statement modulo p alone. You can solve the linear part mod p (take $u=st$, $v=rt$), but then the error puv and the error in $rf+sg \equiv 1$ reappear at the next power of p, and the one-shot attempt has nowhere to go. The tempting escape here is to write "continuing in this way" – but that phrase is the whole proof, left unwritten.

The model chain. We prove by induction on $k \geq 1$: there exist $F_k, G_k \in \Gamma$ with

$$F_k \equiv f, \quad G_k \equiv g \pmod{p}, \qquad F_k G_k \equiv h \pmod{p^k}.$$

First note the congruence notation behaves like ordinary congruence: $f \equiv g \pmod{m}$ means $f-g \in m\Gamma$, and $m\Gamma$ is closed under addition and under multiplication by any element of Γ, so congruences mod m may be added and multiplied.

Base case $k=1$. Take $F_1=f$, $G_1=g$; then $F_1G_1 = fg \equiv h \pmod{p}$ by hypothesis.

Inductive step. Suppose F_k, G_k exist. Since $h-F_kG_k$ has every coefficient divisible by p^k, write

$$h - F_k G_k = p^k t, \qquad t \in \Gamma.$$

Look for the next pair as corrections of size p^k:

$$F_{k+1} = F_k + p^k \Delta_1, \qquad G_{k+1} = G_k + p^k \Delta_2, \qquad \Delta_1, \Delta_2 \in \Gamma.$$

Since $k \geq 1$, $F_{k+1} \equiv F_k \equiv f$ and $G_{k+1} \equiv G_k \equiv g \pmod{p}$. Expanding,

$$F_{k+1}G_{k+1} = F_kG_k + p^k(\Delta_2 F_k + \Delta_1 G_k) + p^{2k}\Delta_1\Delta_2.$$

As $2k \geq k+1$, the last term is $\equiv 0 \pmod{p^{k+1}}$. So we need

$$p^k(\Delta_2 F_k + \Delta_1 G_k) \equiv p^k t \pmod{p^{k+1}}, \quad \text{i.e.} \quad \Delta_2 F_k + \Delta_1 G_k \equiv t \pmod{p}.$$

(The equivalence holds because $p^k w$ has all coefficients divisible by p^{k+1} exactly when w has all coefficients divisible by p.) Choose

$$\Delta_1 = st, \qquad \Delta_2 = rt.$$

Then, using $F_k \equiv f$, $G_k \equiv g \pmod{p}$,

$$\Delta_2 F_k + \Delta_1 G_k = t(rF_k + sG_k) \equiv t(rf + sg) \equiv t \pmod{p}.$$

Hence $F_{k+1}G_{k+1} \equiv F_k G_k + p^k t = h \pmod{p^{k+1}}$, completing the induction. Taking $k = n$ gives the required $F = F_n$, $G = G_n$.

(I checked the construction by machine for $p = 3$, $f = x + 1$, $g = x + 2$, $h = fg + 3(3x^2 + x + 5)$, $r = -1$, $s = 1$: the lifted pairs satisfy $FG \equiv h \pmod{3^k}$ for $k = 2, \ldots, 6$.)

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start by perturbing f, g by multiples of p and expanding the product. They split at the size of the step. The natural attempt tries to reach p^n in one jump, $F = f + pu$, and meets a nonlinear congruence mod p^{n-1}. The model reaches only p^{k+1} from p^k, with correction $p^k\Delta$: then the cross term $p^{2k}\Delta_1\Delta_2$ is already invisible mod p^{k+1}, and what remains is a *linear* congruence mod p – exactly the kind the hypothesis $rf + sg \equiv 1 \pmod{p}$ can solve.

> *I only know something modulo p. So each step must only need information modulo p. If I go from p^k to p^{k+1}, the new unknown is one digit in base p, and one digit is a mod-p question.*

> *If the correction is $p^k\Delta$, the product of two corrections is p^{2k}, which is beyond p^{k+1}. The quadratic term dies and the equation is linear in Δ_1, Δ_2.*

> *I need $\Delta_2 f + \Delta_1 g \equiv t$, and I am handed $rf + sg \equiv 1$. Multiply the given identity by t.*

Layer 3 THE WHY

Why is a one-power-at-a-time induction the natural move, rather than a device? Look at the shape of the hypotheses and the conclusion. The hypotheses are all mod p; the conclusion is mod p^n, with n arbitrary. Whenever a statement is to be proved for every n and the data live at level 1, the proof has to climb from level 1 to level n, and the ladder's rungs must each be something the level-1 data can pay for. That is what induction on the exponent is.

Now drive to the axiom that makes each rung payable. Think of integers written in base p: knowing a number mod p^{k+1} is knowing it mod p^k plus one more digit. A polynomial mod p^{k+1} is, coefficientwise, the same. So the passage from p^k to p^{k+1} asks for one new "digit" Δ (a polynomial read mod p), and the equation that digit must satisfy is the p^k-coefficient of the product. When you expand $(F_k + p^k\Delta_1)(G_k + p^k\Delta_2)$, the p^k-part is $\Delta_2 F_k + \Delta_1 G_k$ – linear, like the derivative in the product rule – and the p^{2k}-part is too high to see. This is Newton's method with p playing the role of a small number: squared errors are negligible, so first-order corrections suffice.

The linear equation $\Delta_2 f + \Delta_1 g \equiv t \pmod{p}$ is solvable for every t precisely because f and g are "coprime mod p" in the strong sense $rf + sg \equiv 1$: the map $(\Delta_1, \Delta_2) \mapsto \Delta_1 g + \Delta_2 f$ hits 1, hence hits every multiple of 1. That is why the hypothesis is phrased as a Bézout identity rather than as "f and g have no common factor": it is the form you can multiply by t.

Kedlaya, Poonen and Vakil's solution says "We will try $F_{k+1} = F_k + p^k\Delta_1$". Here is where the p^k comes from: the correction must not disturb what is already right mod p^k (so it is a

multiple of p^k) and must be allowed to change things mod p^{k+1} (so it need not be a multiple of p^{k+1}). The step size is dictated by the definition of the congruence, not chosen.

Two remarks on what the proof does *not* use. It never uses that p is prime: the argument works verbatim for any integer $p \geq 2$. And it never controls degrees, which is fine because the problem does not ask for $\deg F = \deg f$. The Remark in the book identifies the statement as a case of Hensel's lemma; you do not need the name, only the mechanism.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the mismatch between the hypothesis modulus (p) and the conclusion modulus (p^n, n arbitrary). On seeing it, write down the inductive statement: $F_k \equiv f$, $G_k \equiv g \pmod{p}$, $F_k G_k \equiv h \pmod{p^k}$.

2. Induction on the exponent, with correction terms of the form $p^k\Delta$ and expansion of the product to first order.

3. Hesitation occurs after expanding the product, when the term $p^{2k}\Delta_1\Delta_2$ looks like a nuisance. Reinforce it away with $2k \geq k+1$: for $k \geq 1$ the quadratic term is invisible at the level being worked on.

4. The trap is trying to jump from p to p^n in one step, which gives a nonlinear congruence that the mod-p hypothesis cannot solve. A subtler trap is writing "and so on" instead of an explicit induction: the inductive hypothesis must carry $F_k \equiv f \pmod{p}$ along, since the step uses it.

5. Whenever data are given modulo p and the conclusion is wanted modulo p^n (or to precision ε^n), climb one power at a time; the equation at each rung is linear, because products of corrections are of higher order.

6. Lift by successive first-order corrections: write the current error as $p^k t$, correct by $p^k\Delta$, discard p^{2k} terms, and solve the linear congruence mod p using the given Bézout identity multiplied by t.

▷ **Cue:** hypotheses hold mod p, conclusion is required mod p^n for all n $\longrightarrow$ **Move:** induct on the exponent: go from p^k to p^{k+1}

▷ **Cue:** the current approximation has error $p^k t$ $\longrightarrow$ **Move:** correct by $p^k\Delta$; the cross term $p^{2k}\Delta_1\Delta_2$ vanishes mod p^{k+1}, leaving a linear congruence mod p

▷ **Cue:** a Bézout identity $rf + sg \equiv 1$ is given $\longrightarrow$ **Move:** to solve $\Delta_2 f + \Delta_1 g \equiv t$, multiply it by t: $\Delta_2 = rt$, $\Delta_1 = st$

▷ **Cue:** a statement "there exist objects satisfying a congruence to arbitrary precision" $\longrightarrow$ **Move:** think Newton's method: successive linear corrections, squared errors negligible

1986 B4

1986 B4. For a positive real number r, let $G(r)$ be the minimum value of $|r - \sqrt{m^2 + 2n^2}|$ for all integers m and n. Prove or disprove the assertion that $\lim_{r\to\infty} G(r)$ exists and equals 0.

Fields: number theory; analysis; approximation **Position:** B4 — middle of the paper; the decision (prove or disprove) is the hard part, the proof is a few lines **Answer:** The assertion is true. Greedily take the largest m with $m^2 \leq r^2$, then the largest n with $2n^2 \leq r^2 - m^2$; the leftover $r^2 - m^2 - 2n^2$ is $O(\sqrt{r})$, and dividing by r to pass back to $\sqrt{m^2 + 2n^2}$ gives $G(r) = O(r^{-1/2}) \to 0$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The strong student reads "$m^2 + 2n^2$" and thinks number theory: these are the norms $N(m + n\sqrt{-2})$ from $\mathbb{Z}[\sqrt{-2}]$, and a prime p is of this form exactly when $p = 2$ or $p \equiv 1, 3 \pmod{8}$. So which integers k are of the form $m^2 + 2n^2$? Those in whose factorisation each prime $\equiv 5, 7 \pmod{8}$ appears to an even power. That is a thin set — like the sums of

two squares, the representable integers up to N number only about $N/\sqrt{\log N}$ (a theorem of Landau–Bernays type, which I quote and do not need). So the representable integers have density zero, their gaps are unbounded, and the student starts to believe the assertion is *false*: surely r can be placed in the middle of a long gap and kept a fixed distance from every $\sqrt{k}$.

Carry this to the wall. Suppose k and $k+g$ are consecutive representable integers with g large. The corresponding values $\sqrt{k}$ and $\sqrt{k+g}$ differ by

$$\sqrt{k+g}-\sqrt{k}=\frac{g}{\sqrt{k+g}+\sqrt{k}}\approx\frac{g}{2\sqrt{k}}.$$

A gap of g in k is only a gap of $g/(2\sqrt{k})$ in $\sqrt{k}$. To disprove the assertion you would need gaps g of size comparable to $\sqrt{k}$ infinitely often. "Density zero" says nothing about that — it only says gaps are unbounded on average, which is far weaker. The student now has no proof in either direction and is holding a heavy theory of which primes are norms, none of which touches the question.

The model chain. *Step 1: square the target.* For integers m,n put $k=m^2+2n^2\geq 0$. Then

$$|r-\sqrt{k}|=\frac{|r^2-k|}{r+\sqrt{k}}\leq\frac{|r^2-k|}{r}.$$

So it is enough to find, for each large r, a value $k=m^2+2n^2$ with $|r^2-k|$ small compared with r.

Step 2: greedy choice of m. Let $m\geq 0$ be the largest integer with $m^2\leq r^2$, i.e. $m=\lfloor r\rfloor$. Then $m^2\leq r^2<(m+1)^2$, so

$$0\leq r^2-m^2<2m+1\leq 2r+1.$$

Step 3: greedy choice of n. Put $X=r^2-m^2\in[0,2r+1)$ and let $n\geq 0$ be the largest integer with $2n^2\leq X$. Then $2n^2\leq X<2(n+1)^2$, so

$$0\leq X-2n^2<2(n+1)^2-2n^2=4n+2,\qquad n\leq\sqrt{X/2}<\sqrt{r+\tfrac{1}{2}}.$$

Step 4: conclude. With $k=m^2+2n^2$ we have $0\leq r^2-k<4n+2<2+4\sqrt{r+\frac{1}{2}}$, so by Step 1

$$0\leq G(r)\leq\frac{2+4\sqrt{r+\frac{1}{2}}}{r}\xrightarrow[r\to\infty]{}0.$$

The squeeze gives $\lim_{r\to\infty}G(r)=0$; the assertion is **true**. □

Caveat. The published solution (Jungreis, in Kedlaya–Poonen–Vakil) runs the same greedy argument. In the archive's copy of the book, the chain $2+4\sqrt{(r^2-m^2)/2}\leq 2+\sqrt{2m+1}$ appears; as printed this inequality is off by a constant factor (the left side can be about $2\sqrt{2}\sqrt{2m+1}$). It may be an extraction artefact. It is harmless: the bound above, $G(r)\leq(2+4\sqrt{r+1/2})/r$, is verified line by line and gives the same limit.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start from the same object, the set of values $\sqrt{m^2+2n^2}$. They split at the first question asked about it. (N) asks *which integers are of the form m^2+2n^2?* — an arithmetic question about k. (M) asks *how far is r^2 from the nearest such integer, measured against r?* — a size question. Once the question is about size, the arithmetic of $\mathbb{Z}[\sqrt{-2}]$ is irrelevant, and the greedy choice (largest square, then largest twice-square) answers it.

> *The square root compresses everything: a gap of g between integers near r^2 becomes a gap of about $g/2r$ between their square roots. So I do not need every integer to be representable; I only need a representable integer within $o(r)$ of r^2.*

> *How close can I get to r^2 with one square? Within $2r+1$. Now spend the second, smaller square on the leftover: the leftover is only about $2r$, and twice-squares up to $2r$ are spaced at most about $4\sqrt{r}$ apart. $\sqrt{r}$ is $o(r)$. Done.*

- ▷ **Cue:** a limit of $\min|r - \sqrt{Q(m,n)}|$ over integer points of a quadratic form $\longrightarrow$ **Move:** square it: $|r - \sqrt{k}| \le |r^2 - k|/r$, and ask whether Q comes within $o(r)$ of r^2
- ▷ **Cue:** approximate a large number by a sum of squares (or square-like terms) $\longrightarrow$ **Move:** greedy: take the largest first square, then approximate the remainder, which is much smaller, with the next term
- ▷ **Cue:** a thin set of integers (density zero) under a square root $\longrightarrow$ **Move:** do not conclude the square roots are sparse; gaps shrink by a factor of about $2\sqrt{k}$

Layer 3 THE WHY

Why is squaring natural rather than a trick? Look at what the problem gives you. The object is $\sqrt{m^2+2n^2}$, and square roots of integers are awkward to compare directly; integers are not. The identity $\sqrt{a} - \sqrt{b} = (a-b)/(\sqrt{a}+\sqrt{b})$ is the definition of rationalising, and it converts a question about square roots into a question about integers with an explicit exchange rate: divide by roughly $2r$. The moment you write that exchange rate down, the scale of the problem is visible — you need an error $o(r)$ in the integer world, which is a very generous target. That is the real content of B4: the tolerance is enormous once you measure it in the right units.

Why is the greedy choice natural? Because of that generosity. If the target had been $|r^2 - k|$ bounded, you would need the representable integers to have bounded gaps, and they do not (that is where the natural attempt's number theory was correct and relevant). But the target is $o(r)$, and a two-stage greedy approximation is the cheapest machine that produces sublinear error. The first square does the heavy lifting — it gets you within $2r$, because consecutive squares near r^2 differ by $2r+1$. The second term only has to approximate a number of size $2r$, and twice-squares near X are spaced about $2\sqrt{2X} \approx 4\sqrt{r}$ apart. Each stage takes a square root of the remaining error. Two stages take r^2 to an error of order $\sqrt{r}$, which is what the proof shows.

Kedlaya's book presents Jungreis's solution beginning "select the largest integer $m \ge 0$ such that $r^2 - m^2 \ge 0$". Here is where that comes from: it is the first step of writing a number in any "base" built from squares, exactly as the first step of a decimal expansion is the largest multiple of the leading power of ten that fits. You spend the biggest coin first. The book does not say why two coins suffice; the answer is the exchange rate from Step 1 — $\sqrt{r}$ in the integer world is $\sqrt{r}/r \to 0$ in the root world.

Why does the natural attempt fail? Not because the number theory is wrong, but because it answers a question with the wrong resolution. Density zero is a statement about averages over $[1, N]$; the question needs a statement about the largest gap near N relative to $\sqrt{N}$. Those are different quantities, and the greedy argument gives the second one directly: gaps between representable integers near N are at most about $4N^{1/4}$, far below $\sqrt{N}$.

Note the coefficient 2 in $2n^2$ plays no role. The same proof works for $m^2 + cn^2$ with any fixed $c > 0$, and even for $m^2 + n^2$. The "2" is there to tempt you into $\mathbb{Z}[\sqrt{-2}]$.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the square root around an integer-valued quadratic, with $r \to \infty$. On seeing it write $|r - \sqrt{k}| = |r^2 - k|/(r + \sqrt{k})$ and ask what error in k is affordable: anything $o(r)$.

2. Greedy approximation: the largest square below r^2, then the largest admissible second term below the remainder. Each stage replaces an error E by roughly $\sqrt{E}$.
3. Hesitation comes at the prove-or-disprove decision, driven by the knowledge that numbers of the form $m^2 + 2n^2$ are rare. Reinforce it away by computing what rarity means after a square root: a gap g at height k becomes $g/(2\sqrt{k})$.
4. The trap is the specific form $m^2 + 2n^2$, which invites the arithmetic of $\mathbb{Z}[\sqrt{-2}]$ and a density-zero intuition pointing towards "disprove". Neither is needed; the constant 2 is decoration.
5. Any time a problem asks how well an unfamiliar set approximates a real number, first change units to where the set is easy (here: square everything) and read off the allowed error in those units. A generous tolerance means a crude construction will do.
6. Reusable decision rule: *before deciding whether a thin set is dense enough, compute the tolerance in the set's own units; if the tolerance grows, build the approximation greedily, largest piece first, and check that the error shrinks faster than the tolerance grows.*

1986 B5

1986 B5. Let $f(x,y,z) = x^2 + y^2 + z^2 + xyz$. Let $p(x,y,z), q(x,y,z), r(x,y,z)$ be polynomials with real coefficients satisfying

$$f(p(x,y,z), q(x,y,z), r(x,y,z)) = f(x,y,z).$$

Prove or disprove the assertion that the sequence p, q, r consists of some permutation of $\pm x, \pm y, \pm z$, where the number of minus signs is 0 or 2.

Fields: algebra; polynomials; Vieta jumping **Position:** B5 — late in the paper, but a one-line counterexample once you look at f as a quadratic in one variable **Answer:** The assertion is false: $(p,q,r) = (x,\ y,\ -xy - z)$ satisfies $f(p,q,r) = f(x,y,z)$ and is not a signed permutation of x, y, z.

Layer 1 THE WHAT — surface analysis

The natural attempt. The phrasing "prove or disprove" plus a very specific assertion (signed permutations with an even number of minus signs) reads like a classification theorem, and the strong student sets out to prove it. First the easy direction: signed permutations preserve $x^2 + y^2 + z^2$, and the product xyz is preserved exactly when the number of minus signs is even. So the listed triples do work. Now the converse, by degrees.

Write $a = \deg p$, $b = \deg q$, $c = \deg r$ and order them $a \geq b \geq c$. The identity is

$$p^2 + q^2 + r^2 + pqr = x^2 + y^2 + z^2 + xyz, \tag{$*$}$$

whose right side has degree 3. If $a = b = c = d \geq 2$, then pqr has degree $3d$, and p^2, q^2, r^2 have degree $2d < 3d$, so nothing can cancel the top part of pqr and the left side has degree $3d > 3$. Contradiction. Good; so the degrees are not all equal and large. Next try $a \geq 2$ with b, c smaller. Then p^2 has degree $2a$ and pqr has degree $a + b + c$. If $2a > a + b + c$, i.e. $a > b + c$, then p^2 dominates and the left side has degree $2a \geq 4$, contradiction. If $a < b + c$, then pqr dominates with degree $a + b + c$; is that > 3? Since $a \geq 2$ and $b + c > a \geq 2$, $a + b + c \geq 5$. Contradiction again.

That leaves the case $a = b + c$. Here the top-degree parts of p^2 and of pqr both have degree $2a$, and they *can* cancel, since the top part of $p^2 + pqr$ is $P(P + QR)$ where P, Q, R are the leading forms. The simplest instance is $a = 2$, $b = c = 1$. The degree argument has run into a wall, and the wall has a precise shape: degrees $(2, 1, 1)$ with $P = -QR$.

The model chain. Look at $(*)$ with $p = x$, $q = y$ fixed and r unknown. It says

$$r^2 + (xy)\, r + (x^2 + y^2) = z^2 + (xy)\, z + (x^2 + y^2),$$

that is, $r^2 + xy\,r = z^2 + xy\,z$, or

$$(r - z)(r + z + xy) = 0.$$

One root is $r = z$. By Vieta's formulas (the two roots of $t^2 + xy\,t +$ const sum to $-xy$), the other root is $r = -xy - z$.

Check. With $(p, q, r) = (x, y, -xy - z)$:

$$r^2 + pqr = (xy + z)^2 - xy(xy + z) = (xy + z)\big((xy + z) - xy\big) = z(xy + z) = z^2 + xyz,$$

so $f(p, q, r) = x^2 + y^2 + z^2 + xyz = f(x, y, z)$.

Not of the asserted form. $r = -xy - z$ has degree 2, while each of $\pm x, \pm y, \pm z$ has degree 1. So the triple is not a permutation of $\pm x, \pm y, \pm z$ with any signs.

The assertion is **false**. □

Caveat. Kedlaya–Poonen–Vakil go further and, in a remark, describe all solutions: iterate the moves "permute", "change an even number of signs", and the jump $(p, q, r) \mapsto (-qr - p, q, r)$, starting from (x, y, z) and from a second family with one of p, q, r a constant of absolute value greater than 2; they justify it by a degree-descent argument. That classification is not needed for the problem and has not been independently checked for this archive; only the counterexample above is verified.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains write down $(*)$. They split at how they read it. (N) reads $(*)$ as a polynomial identity to be analysed *by total degree*, in all three unknowns at once, aiming at a proof. (M) freezes two of the unknowns ($p = x$, $q = y$) and reads $(*)$ as a *quadratic equation in the third*, aiming at a second root. The natural attempt's own wall — degrees $(2, 1, 1)$ with the top parts of r^2 and pqr cancelling — is exactly the case where r satisfies a quadratic with pq as its linear coefficient.

> *Every variable in f appears to degree at most 2. So if I fix x and y, f is a quadratic in z, and a quadratic with one known root has a second one — the sum of the roots is $-xy$. That second root is a polynomial, so it is a new solution.*

> *My degree argument breaks only when $\deg p = \deg q + \deg r$, because then p^2 and pqr can cancel. When a proof breaks in exactly one case, I should try to build a counterexample in that case before trying to patch it.*

- ▷ **Cue:** a polynomial (or Diophantine) equation that is quadratic in each variable separately ⟶ **Move:** fix all but one variable and use Vieta: from a root t of $t^2 + Bt + C$, jump to $-B - t$
- ▷ **Cue:** "prove or disprove" with a suspiciously neat classification ⟶ **Move:** spend two minutes hunting for a counterexample first — start from the identity solution and look for a move that preserves the equation
- ▷ **Cue:** a degree-counting proof that fails in one boundary case ⟶ **Move:** the boundary case is a blueprint: construct an object with exactly those degrees and cancelling leading terms

Layer 3 THE WHY

Why is freezing two variables the natural move and not a trick? Because of the shape of f. A polynomial in three variables in general has no special structure in one variable, but $x^2 + y^2 + z^2 + xyz$ has degree exactly 2 in each variable separately — z appears only as z^2 and xyz. That is a structural fact visible on the page before any computation. A quadratic in z has

two roots, and Vieta's relation (the sum of the roots is minus the linear coefficient) gives the second root as a polynomial in the others, with no square roots needed. That is the whole of "Vieta jumping": whenever an equation is quadratic in one unknown and you hold one solution, the other root is free.

The official solution states the counterexample and then, under "Motivation", factors $(r - z)(r + xy + z) = 0$. Here is where the idea to look at r alone comes from: the assertion to be tested lists solutions that are *symmetries* — permutations and sign changes — and the question is whether f has other symmetries. The cheapest way to find a new symmetry is to change one coordinate and keep the others, and changing one coordinate while preserving a quadratic in that coordinate means moving to the other root. You are not guessing; you are listing the degrees of freedom of the cheapest possible modification.

Why should you suspect "disprove" at all? Two reasons, both from the phrasing. First, the insistence on "0 or 2" minus signs shows the setters have already computed which *linear* substitutions work; nothing in the statement forces p, q, r to be linear, so the assertion is really "every solution is linear", and that is a strong claim about polynomials of arbitrary degree. Second, the degree argument you try in order to prove it fails in exactly one configuration, and failure localised to one case is the signature of a counterexample hiding there. The case is $\deg r = \deg p + \deg q$ with $R = -PQ$ at top degree; with $p = x$, $q = y$ that forces $r = -xy + (\text{lower})$, and the lower part is found by solving the equation: $-z$.

The family this belongs to is recognisable. Markov's equation $x^2 + y^2 + z^2 = 3xyz$ has the same shape, and its solutions are generated from $(1, 1, 1)$ by exactly this jump $z \mapsto 3xy - z$ together with permutations. Here the coefficient is $+1$ instead of -3, and the jump is $z \mapsto -xy - z$. Once you have seen one member of the family, "quadratic in each variable, symmetric, with a product term" should trigger the jump immediately.

Cheapness: the counterexample costs three lines. A student who spends the time proving a classification by degree arguments is doing the zombie maths this problem punishes — and worse, trying to prove something false.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that f has degree 2 in each variable separately. On seeing it write: *as a quadratic in z: $z^2 + (xy)z + (x^2 + y^2)$; roots sum to $-xy$.*

2. Vieta jumping: keep $p = x$, $q = y$ and replace z by the other root $-xy - z$.

3. Hesitation comes at the prove-or-disprove decision, because the assertion looks like a carefully checked theorem. Reinforce it away by asking what the assertion quietly claims — that every solution is linear — and testing that claim against the cheapest non-linear modification.

4. The trap is the precision of the assertion: "0 or 2" minus signs is correct for linear substitutions, which lends the whole statement authority. The second trap is a degree-counting proof that almost works and absorbs the time.

5. Whenever an identity to be preserved is quadratic in one variable at a time, the symmetry group contains the root-swapping moves, and they usually raise degree. Look for a second root before looking for a proof of rigidity.

6. Reusable decision rule: *when a "prove or disprove" classification lists only the obvious symmetries, first try to build a new one by changing a single coordinate; if the equation is quadratic in that coordinate, the other root is the new symmetry.*

1986 B6

1986 B6. Suppose A, B, C, D are $n \times n$ matrices with entries in a field F, satisfying the conditions that AB^T and CD^T are symmetric and $AD^T - BC^T = I$. Here I is the $n \times n$ identity matrix, and if M is an $n \times n$ matrix, M^T is its transpose. Prove that $A^T D - C^T B = I$.

Fields: linear algebra; block matrices **Position:** B6 — nominally the hardest problem; one idea, after which the proof is four lines **Answer:** Assemble the $2n \times 2n$ matrices $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$ and $N = \begin{pmatrix} D^T & -B^T \\ -C^T & A^T \end{pmatrix}$. The hypotheses say exactly $MN = I_{2n}$; hence $NM = I_{2n}$, whose lower-right block is $A^TD - C^TB = I$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The student starts manipulating. The hypotheses are

$$AB^T = BA^T, \qquad CD^T = DC^T, \qquad AD^T - BC^T = I,$$

and the target $A^TD - C^TB = I$ has the transposes on the *other* side. First instinct: assume A is invertible and solve. From $AB^T = BA^T$, the matrix $S = A^{-1}B$ satisfies $S^T = B^TA^{-T} = A^{-1}(AB^T)A^{-T} = A^{-1}(BA^T)A^{-T} = A^{-1}B = S$, so S is symmetric. From $AD^T = I + BC^T$,

$$D^T = A^{-1} + SC^T, \qquad D = A^{-T} + CS.$$

Then $A^TD = I + A^TCS$ and $C^TB = C^TAS$, so

$$A^TD - C^TB = I + (A^TC - C^TA)S.$$

Now use CD^T symmetric: $CD^T = CA^{-1} + CSC^T$, and CSC^T is symmetric, so CA^{-1} is symmetric, i.e. $CA^{-1} = A^{-T}C^T$, i.e. $A^TC = C^TA$. The bracket vanishes and $A^TD - C^TB = I$.

That is a correct proof *when A is invertible*. Now the wall. Over $\mathbb{R}$ you might hope to perturb A to an invertible matrix, but the perturbation must keep all three hypotheses true, and it is not clear how to do that. Worse, the problem says "a field F" — possibly finite, where there is no perturbation and no limit at all. The student has a proof of a special case and no way to reach the general case.

The model chain. Write the hypotheses as four equations. Symmetry of AB^T and CD^T gives

$$AB^T = BA^T, \qquad CD^T = DC^T. \tag{1, 2}$$

The third hypothesis and its transpose give

$$AD^T - BC^T = I, \qquad DA^T - CB^T = I. \tag{3, 4}$$

Form the $2n \times 2n$ block matrices

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}, \qquad N = \begin{pmatrix} D^T & -B^T \\ -C^T & A^T \end{pmatrix}.$$

Block multiplication gives

$$MN = \begin{pmatrix} AD^T - BC^T & -AB^T + BA^T \\ CD^T - DC^T & -CB^T + DA^T \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix}$$

by (3), (1), (2), (4) respectively. For square matrices over a field, $MN = I$ implies $NM = I$. (Proof: $MN = I$ gives $\det M \det N = 1$, so $\det M \neq 0$ and M is invertible; then $N = M^{-1}MN = M^{-1}$, so $NM = M^{-1}M = I$.) Now compute the lower-right block of

$$NM = \begin{pmatrix} D^TA - B^TC & D^TB - B^TD \\ -C^TA + A^TC & -C^TB + A^TD \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix}:$$

it reads $A^TD - C^TB = I$. □

Bonus. The other three blocks of $NM = I$ give $D^TA - B^TC = I$, $A^TC = C^TA$ and $B^TD = D^TB$ — the "column" versions of all three hypotheses. The problem asks for one of four equivalent facts.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start from the same three equations. They split at the first action. (N) treats A, B, C, D as four separate unknowns and tries to solve for one in terms of the others, which needs an inverse and so a special case. (M) observes that the hypotheses are all of the form "(row of blocks)×(column of blocks) = something" and packages all four $n \times n$ matrices into *one* $2n \times 2n$ matrix, so that the only inverse needed is the inverse of the big matrix, which exists automatically.

> *Each hypothesis has the shape $XY^T - ZW^T$: a sum of two products, each a matrix from $\{A, B\}$ or $\{C, D\}$ times a transpose. That is exactly what an entry of a 2×2 block product looks like. So let me write all of them as blocks of one product MN and see what N has to be.*

> *The conclusion has the transposes on the left, the hypotheses on the right. Swapping the order of a product is the step "$MN = I \Rightarrow NM = I$", and that step is free for square matrices. So I want the hypotheses to say $MN = I$ and the conclusion to be a block of NM.*

- ▷ **Cue:** several matrix identities, each a sum of two products like $AD^T - BC^T$ $\longrightarrow$ **Move:** assemble a 2×2 block matrix and read the identities as block entries of one product
- ▷ **Cue:** the conclusion is the hypothesis with the order of products reversed (transposes moved from right to left) $\longrightarrow$ **Move:** show $MN = I$ and use $MN = I \Rightarrow NM = I$ for square matrices
- ▷ **Cue:** a matrix proof that needs A invertible, over an arbitrary field $\longrightarrow$ **Move:** do not try to perturb; find a larger matrix whose invertibility is forced by the hypotheses

Layer 3 THE WHY

Why is the block matrix the natural move and not a rabbit from a hat? Go to the definition of block multiplication:

$$\begin{pmatrix} P & Q \\ R & S \end{pmatrix} \begin{pmatrix} W & X \\ Y & Z \end{pmatrix} = \begin{pmatrix} PW + QY & PX + QZ \\ RW + SY & RX + SZ \end{pmatrix}.$$

Every entry is a sum of two products. The hypothesis $AD^T - BC^T = I$ is a sum of two products with A, B on the left: it is the $(1,1)$ block of $\begin{pmatrix} A & B \\ \cdot & \cdot \end{pmatrix} \begin{pmatrix} D^T & \cdot \\ -C^T & \cdot \end{pmatrix}$. That fixes the first row of M and the first column of N. The symmetry $AB^T = BA^T$, rewritten as $A(-B^T) + B(A^T) = 0$, is another sum of two products with A, B on the left; it fixes the second column of N as $\begin{pmatrix} -B^T \\ A^T \end{pmatrix}$. Now N is determined, and the remaining two blocks of MN are $CD^T - DC^T$ and $DA^T - CB^T$, which are hypotheses (2) and (4). Nothing was guessed: the four hypotheses *are* the four blocks, once you write each as a two-term sum of products. That is the circle-and-centre level of "of course" for this problem.

Kedlaya's book says "these four equations are the entries in the block matrix identity". Here is what the book leaves unsaid: the fourth equation is not given — it is the transpose of the third, and you are meant to notice that "$= I$" is transpose-invariant. A student who writes only the three given equations finds one block of MN missing and stalls. Always transpose every equation whose right side is symmetric; it is a free equation.

Why does $MN = I$ imply $NM = I$? This is the structural fact the whole problem rests on, and it fails for infinite matrices (the shift operator on sequences has a one-sided inverse only).

For $m \times m$ matrices over a field it is a consequence of determinants being multiplicative, or of rank–nullity: $MN = I$ forces M surjective, hence bijective in finite dimension. This is why the problem says "a field": it is all the structure needed, and it defeats the natural attempt's perturbation strategy while leaving the determinant argument untouched.

Why does the natural attempt get stuck, and what does its special-case proof actually show? It shows the conclusion holds on the set where $\det A \neq 0$. Over $\mathbb{R}$ or $\mathbb{C}$ a density argument *might* be arranged, but the hypotheses cut out a curved set of quadruples and you would have to show the invertible-A points are dense in it — real work with no guarantee. The block matrix sidesteps all of this: you need an inverse, so find something whose inverse is guaranteed. A is not guaranteed invertible; M is, because the hypotheses hand you its inverse.

What M is. With $J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$ one checks $N = -JM^TJ$, so the hypotheses say $MJM^T = J$: M is a *symplectic* matrix. The conclusion is part of $M^TJM = J$, the statement that M^T is symplectic too. For $n = 1$ with A, B, C, D scalars a, b, c, d, the symmetry conditions are empty and $ad - bc = 1$ is the determinant condition; the conclusion is $ad - cb = 1$, the same equation. The problem is the $n \times n$ version of "$\det M^T = \det M$" for 2×2 matrices.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that every hypothesis is a difference of two products of the form XY^T.
On seeing it write: 2×2 *block matrix* $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$*; find* N *with* $MN = I$. Also write the transpose of $AD^T - BC^T = I$.

2. Block-matrix packaging, then the finite-dimensional fact that a one-sided inverse of a square matrix is two-sided.

3. Hesitation comes after the special case A invertible works: the student tries to extend it and loses time. Reinforce it away with the question: *whose invertibility is guaranteed here?* Not A's; the hypotheses exhibit an inverse for the big matrix.

4. The trap is solving for D in terms of A, which needs A^{-1}, and then the general field F blocks every density or perturbation rescue. A secondary trap is writing only three of the four block equations.

5. When the conclusion looks like the hypothesis with every product reversed, think "$MN = I \Rightarrow NM = I$". When several matrix equations each involve two products, think "they are entries of one block product".

6. Reusable decision rule: *if an argument needs an inverse that may not exist, enlarge the object — bundle the unknowns into a bigger matrix (or bigger structure) whose inverse the hypotheses provide, and read the conclusion off a block of the reversed product.*