DAWN PUTNAM SESSION

Putnam 1992 — Papers A and B

Putnam 120 Archive

Contents

Putnam 1992 — Paper A **5**
1992 A1 5
1992 A2 7
1992 A3 10
1992 A4 12
1992 A5 15
1992 A6 18

Putnam 1992 — Paper B **23**
1992 B1 23
1992 B2 25
1992 B3 27
1992 B4 30
1992 B5 32
1992 B6 35

Putnam 1992 — Paper A

1992 A1

1992 A1. Prove that $f(n) = 1 - n$ is the only integer-valued function defined on the integers that satisfies the following conditions.

(i) $f(f(n)) = n$, for all integers n;

(ii) $f(f(n+2)+2) = n$ for all integers n;

(iii) $f(0) = 1$.

Fields: functional equations, algebra on $\mathbb{Z}$ **Position:** A1 — the opener **Answer:** Applying f to both sides of (ii) and using (i) gives $f(n) = f(n+2)+2$; with $f(0) = 1$ and $f(1) = f(f(0)) = 0$ this pins down f on the evens and on the odds, forcing $f(n) = 1 - n$, which does satisfy (i)–(iii).

Layer 1 THE WHAT — surface analysis

The natural attempt. A strong student meets a functional equation on the integers and starts computing values. From (iii), $f(0) = 1$. From (i) with $n = 0$, $f(1) = f(f(0)) = 0$. Now (ii) with $n = -2$ gives $f(f(0)+2) = -2$, i.e. $f(3) = -2$. Then (i) with $n = 3$ gives $f(-2) = 3$. Then (ii) with $n = -1$: $f(f(1)+2) = -1$, so $f(2) = -1$, and (i) gives $f(-1) = 2$. So far:

$$f(-2) = 3,\ f(-1) = 2,\ f(0) = 1,\ f(1) = 0,\ f(2) = -1,\ f(3) = -2,$$

and every one agrees with $1 - n$. The pattern looks right, and the student now tries to turn the chase into an induction.

This is where it gets sticky. To use (ii) at n you need to know $f(n+2)$ *before* you learn anything, and what you learn is the value of f at the point $f(n+2)+2$ — a point you did not choose. The natural induction hypothesis "$f(k) = 1 - k$ for $|k| \le K$" has to be fed into (ii) at the right n, and then (i) has to be used to flip the conclusion round, and the bookkeeping of which n produces which new value (sometimes one step out on the right, sometimes on the left) becomes a two-sided induction with four cases. It can be done. It costs ten minutes and invites an off-by-one error, and the written proof reads as a chase rather than an argument.

The model chain. Let $f : \mathbb{Z} \to \mathbb{Z}$ satisfy (i)–(iii). Fix an integer n and put $m = f(n+2)+2$. Condition (ii) says $f(m) = n$. Apply f to both sides and use (i) at m:

$$f(n) \;=\; f(f(m)) \;=\; m \;=\; f(n+2)+2.$$

So for every integer n,

$$f(n+2) \;=\; f(n) - 2. \tag{$*$}$$

Condition $(*)$ runs in both directions: read forwards it gives f at $n+2$ from f at n; read as $f(n) = f(n+2)+2$ it gives f at n from f at $n+2$. So f is determined on all of $2\mathbb{Z}$ by $f(0)$ and on all of $2\mathbb{Z}+1$ by $f(1)$, and a two-sided induction on $k \in \mathbb{Z}$ gives

$$f(2k) = f(0) - 2k, \qquad f(2k+1) = f(1) - 2k.$$

Now $f(0) = 1$ by (iii), and $f(1) = f(f(0)) = 0$ by (i) at $n = 0$. Hence $f(2k) = 1 - 2k$ and $f(2k+1) = -2k = 1 - (2k+1)$; that is, $f(n) = 1 - n$ for every integer n.

Finally, $f(n) = 1 - n$ does satisfy the conditions, so "the only" is not vacuous: $f(f(n)) = 1 - (1 - n) = n$; $f(n+2) + 2 = (1 - n - 2) + 2 = 1 - n$, so $f(f(n+2)+2) = f(1-n) = n$; and $f(0) = 1$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start with the same four symbols in front of them, (i) and (ii). The natural attempt *substitutes numbers into* (ii). The model *applies f to both sides of* (ii) as an identity in n, and lets (i) cancel the outer f. That single move converts the tangled condition (ii) into the clean shift law $f(n+2) = f(n) - 2$, after which there is nothing left to chase.

> *Condition (i) says f is its own inverse. Condition (ii) has an f wrapped round the outside. If I apply f to both sides, the outer f disappears and I get a statement with only one f on each side.*

> *Once I have $f(n+2) = f(n) - 2$, this is a recursion of step 2, so I need two starting values, one for the evens and one for the odds. (iii) gives me $f(0)$; I have to find $f(1)$ somewhere else, and (i) at $n = 0$ gives it.*

Layer 3 THE WHY

Why is "apply f to both sides" natural rather than lucky? Because of what (i) *is*. The statement $f(f(n)) = n$ for all n says $f \circ f = \mathrm{id}$: f is a bijection and $f^{-1} = f$. Whenever you are told a function is invertible, and you have an equation of the form $f(\text{stuff}) = n$, the reflex is to invert: $\text{stuff} = f^{-1}(n)$. Here $f^{-1} = f$, so inverting is applying f. There is no trick; this is the definition of an inverse used once.

Rewrite (ii) with that lens. Let $T(n) = n + 2$ be the shift. Then $f(f(n+2)+2) = f(T(f(T(n))))$, so (ii) says $f \circ T \circ f \circ T = \mathrm{id}$. Since $f = f^{-1}$, compose with f on the left: $T \circ f \circ T = f$, i.e. $f(n+2) + 2 = f(n)$. In words: *conjugating f by the shift gives back f up to a shift*. That is exactly the statement that f interacts with translation in a rigid way, and the only thing a rigid interaction with translation by 2 can produce on $\mathbb{Z}$ is an affine law on each residue class mod 2. So the problem was always "find the step-2 recursion hidden in (ii)", and the recursion is found by undoing the involution.

The second critical thought is forced by the step size. A recursion $f(n+2) = f(n) - 2$ has no link between even and odd arguments — it sees $\mathbb{Z}$ as two separate chains $\ldots, -2, 0, 2, \ldots$ and $\ldots, -1, 1, 3, \ldots$. Each chain needs its own anchor. The problem hands you exactly one anchor, $f(0) = 1$, on the even chain. A reader who notices this asks: "what connects the two chains?" Only (i) mixes arguments freely, and (i) at $n = 0$ sends the anchor across: $f(1) = f(f(0)) = 0$. If you forget this, you prove only $f(2k) = 1 - 2k$ and have silently assumed the odd values.

A structural remark tells you what shape the answer must have before you finish. The law $(*)$ alone says $f(n) + n$ is constant on each chain: $f(n) = c_0 - n$ for even n and $f(n) = c_1 - n$ for odd n. Condition (iii) gives $c_0 = 1$, and (i) at $n = 0$ gives $f(1) = 0$, so $c_1 = 1$ as well. Thus f is the reflection $n \mapsto 1 - n$, whose mirror sits at $\frac{1}{2}$. So $1 - n$ is the reflection of the integers in the point $\frac{1}{2}$, and conditions (i)–(iii) are the three facts "it is a reflection", "it commutes suitably with the shift", "its mirror is at $\frac{1}{2}$".

Kedlaya's archive has no official solution for 1992 in the project's sources, so the chain above was constructed here and checked in both directions: the derivation of $(*)$ uses (i) only at the point $m = f(n+2) + 2$, which is an integer because f is integer-valued, and the verification that $1 - n$ works is written out explicitly.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is (i), $f(f(n)) = n$. On seeing it, write down "f is a bijection and $f^{-1} = f$".

Then look at (ii) and see an outer f you now know how to remove.

2. Apply f to both sides of (ii) and use (i) to cancel: $f(n) = f(n+2) + 2$. Then solve the step-2 recursion on each residue class mod 2.

3. Hesitation occurs after computing half a dozen values that all fit $1 - n$, when the student cannot see how to make the value-chase into an induction. What reinforces it away is to stop plugging in numbers and treat (ii) as an identity between functions: an identity can have operations applied to both sides.

4. Two traps. First, deriving $f(2k) = 1 - 2k$ and forgetting that the odd chain needs its own anchor $f(1) = 0$. Second, forgetting that "the only function that satisfies" has two halves: uniqueness (the derivation) and existence (the check that $1 - n$ works). Graders expect both.

5. Any condition of the form $f(f(\cdot)) = \mathrm{id}$, or any stated invertibility, is a licence to apply f (or f^{-1}) to both sides of every other condition. Look for the outermost f in each hypothesis and ask whether it can be peeled off.

6. Reusable rule: in a functional equation, before substituting values, try applying known operations (inverses, the function itself, shifts) to both sides of each hypothesis to simplify its shape. A recursion of step d on $\mathbb{Z}$ needs d anchors, one per residue class; count them before claiming the function is determined.

▷ **Cue:** a hypothesis $f(f(n)) = n$ sits alongside another equation with an outer f ⟶ **Move:** apply f to both sides of the other equation and cancel with $f \circ f = \mathrm{id}$

▷ **Cue:** a functional equation reduces to $f(n+d) = f(n) + c$ on $\mathbb{Z}$ ⟶ **Move:** split $\mathbb{Z}$ into d residue classes and find one anchor value in each; the function is affine on each class

▷ **Cue:** the problem says "prove this is the only function satisfying …" ⟶ **Move:** prove uniqueness by derivation and existence by direct verification; write both

▷ **Cue:** you have computed several values and all fit a guess, but the induction is getting tangled ⟶ **Move:** stop plugging numbers; treat the hypothesis as an identity and transform it algebraically

1992 A2

1992 A2. Define $C(\alpha)$ to be the coefficient of x^{1992} in the power series about $x = 0$ of $(1+x)^\alpha$. Evaluate

$$\int_0^1 \left(C(-y-1) \sum_{k=1}^{1992} \frac{1}{y+k} \right) dy.$$

Fields: calculus, binomial series, logarithmic differentiation **Position:** A2 — routine once the structure is seen **Answer:** 1992. Writing $P(y) = \prod_{k=1}^{1992}(y+k)$, one has $C(-y-1) = P(y)/1992!$ and $\sum_k 1/(y+k) = P'(y)/P(y)$, so the integrand is $P'(y)/1992!$ and the integral is $\big(P(1) - P(0)\big)/1992! = (1993! - 1992!)/1992! = 1992$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The student knows the generalised binomial series: the coefficient of x^r in $(1+x)^\alpha$ is

$$\binom{\alpha}{r} = \frac{\alpha(\alpha-1)\cdots(\alpha-r+1)}{r!}.$$

So $C(\alpha) = \binom{\alpha}{1992}$, and $C(-y-1) = \dfrac{(-y-1)(-y-2)\cdots(-y-1992)}{1992!}$. So far so good.

Then the student looks at the integrand as a product of two separate things — a degree-1992 polynomial in y and a sum of 1992 reciprocals — and starts multiplying out. Each term $\frac{1}{y+k}\prod_j (y+j)$ is a polynomial of degree 1991 (the factor $y+k$ cancels), so the integrand is a sum of 1992 polynomials, each of which would have to be integrated from 0 to 1. At this point

the student either tries to expand $\prod_{j\neq k}(y+j)$ in elementary symmetric functions (hopeless) or starts evaluating $\int_0^1 \prod_{j\neq k}(y+j)\,dy$ for each k separately (also hopeless). The wall is visible: 1992 integrals of degree-1991 polynomials, with no pattern forthcoming. This is zombie maths in its purest form: the problem has been read as "integrate this expression" instead of "recognise this expression".

The model chain. *Step 1: identify C.* For any real α, the Maclaurin series of $(1+x)^\alpha$ is $\sum_{r\geq 0}\binom{\alpha}{r}x^r$ with $\binom{\alpha}{r} = \alpha(\alpha-1)\cdots(\alpha-r+1)/r!$. So

$$C(\alpha) = \frac{\alpha(\alpha-1)\cdots(\alpha-1991)}{1992!}.$$

Step 2: substitute $\alpha = -y-1$. The factors $\alpha - j$ for $j = 0,\ldots,1991$ become $-y-1-j = -(y+j+1)$, so with $k = j+1$,

$$C(-y-1) = \frac{\prod_{k=1}^{1992}(-(y+k))}{1992!} = \frac{(-1)^{1992}\prod_{k=1}^{1992}(y+k)}{1992!} = \frac{P(y)}{1992!}, \qquad P(y) := \prod_{k=1}^{1992}(y+k).$$

The sign disappears because 1992 is even.

Step 3: recognise the sum. For $y \in [0,1]$ every factor $y+k$ is positive, so $\log P(y) = \sum_{k=1}^{1992}\log(y+k)$ is defined and differentiable, and

$$\frac{P'(y)}{P(y)} = \frac{d}{dy}\log P(y) = \sum_{k=1}^{1992}\frac{1}{y+k}.$$

(Equivalently, by the product rule, $P'(y) = \sum_k \prod_{j\neq k}(y+j) = P(y)\sum_k \frac{1}{y+k}$, which holds for every y not equal to $-1,\ldots,-1992$.)

Step 4: integrate. The integrand is therefore

$$C(-y-1)\sum_{k=1}^{1992}\frac{1}{y+k} = \frac{P(y)}{1992!}\cdot\frac{P'(y)}{P(y)} = \frac{P'(y)}{1992!},$$

and by the fundamental theorem of calculus

$$\int_0^1 \frac{P'(y)}{1992!}\,dy = \frac{P(1)-P(0)}{1992!} = \frac{1993!-1992!}{1992!} = 1993-1 = 1992.$$

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains compute $C(-y-1)$ as a product of 1992 linear factors. They split at the next line. The natural attempt treats $\sum_k \frac{1}{y+k}$ as a separate object to be multiplied in; the model sees that $\sum_k \frac{1}{y+k}$ is built from exactly the same 1992 numbers $y+1,\ldots,y+1992$ as the product, and asks what operation turns a product of $(y+k)$'s into a sum of $1/(y+k)$'s.

> *The product is $\prod(y+k)$ and the sum is $\sum 1/(y+k)$ over the same k. A sum of reciprocals of the factors of a product is what you get when you take the log of the product and differentiate. So the sum is P'/P, and the integrand is P' up to a constant.*

> *An integral over $[0,1]$ of a derivative is the difference of two values. $P(1)$ and $P(0)$ are factorials. I do not need to expand anything.*

Layer 3 THE WHY

Why is logarithmic differentiation the move you are *supposed* to see, not one you might happen to see? Start from what the logarithm is for: it is the homomorphism that turns products into sums. Differentiation is linear, so it respects sums. Put them together and you get the rule

$$\frac{d}{dy}\log\prod_k g_k(y) = \sum_k \frac{g_k'(y)}{g_k(y)}.$$

For linear factors $g_k(y) = y + k$ the numerators are all 1, so the right-hand side is exactly $\sum_k \frac{1}{y+k}$. That is not a coincidence the examiner was lucky to find; it is the only everyday operation that produces a sum of reciprocals of the factors of a product. So the moment you see both a product of linear factors and a sum of their reciprocals in the same expression, the pairing "$P \cdot (P'/P) = P'$" is the structure of the problem, and the question is essentially: "integrate a derivative".

The second half of the WHY is about the phrasing. Why does the problem go to the trouble of saying "the coefficient of x^{1992} in the power series of $(1+x)^\alpha$" and then evaluate it at $\alpha = -y-1$, instead of printing the product? Because the examiners are hiding the product. The disguise has two layers: the binomial coefficient $\binom{\alpha}{1992}$ hides a falling factorial, and the substitution $\alpha = -y-1$ turns that falling factorial $\alpha(\alpha-1)\cdots$ into the rising factorial $(y+1)(y+2)\cdots(y+1992)$, up to a sign $(-1)^{1992} = 1$. When a problem wraps something in a definition you have to unpack, unpack it fully first and *then* look at the whole integrand afresh; the unpacked object is usually the thing that makes the rest of the expression make sense.

Finally, why does the answer come out as the year? $P(1)/1992! = 1993!/1992! = 1993$ and $P(0)/1992! = 1$. In general, with N in place of 1992, one has $\binom{-y-1}{N} = (-1)^N\binom{y+N}{N}$, and the integral is $(-1)^N\left[\binom{N+1}{N} - \binom{N}{N}\right] = (-1)^N N$. For even N, such as 1992, the problem is therefore: "$C(-y-1) = \binom{y+N}{N}$, its derivative in y is $C(-y-1)\sum_k 1/(y+k)$, and the answer N is the change in $\binom{y+N}{N}$ as y goes from 0 to 1."

There is no official solution in the archive's sources for 1992; the chain above was built independently and checked numerically (the integral evaluates to 1992.000… by quadrature).

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the pairing of $C(-y-1)$, once unpacked into $\prod_k(y+k)/1992!$, with $\sum_k \frac{1}{y+k}$ over the same index range $k = 1, \ldots, 1992$. On seeing it, write down $P(y) = \prod_{k=1}^{1992}(y+k)$ and $P'(y)/P(y) = \sum_k \frac{1}{y+k}$.

2. Logarithmic differentiation, followed by the fundamental theorem of calculus: the integrand is an exact derivative.

3. Hesitation occurs after computing $C(-y-1)$, when the student has a big product and a big sum and no plan. It is reinforced away by the habit: before multiplying two expressions together, ask whether one is the derivative (or log-derivative) of the other.

4. The trap is expanding: distributing the sum over the product and integrating 1992 polynomials. A secondary trap is a sign slip in $C(-y-1)$: forgetting that $(-1)^{1992} = +1$, or getting the product as $\prod_{k=0}^{1991}$ instead of $\prod_{k=1}^{1992}$.

5. A product $\prod(x-a_k)$ and a sum $\sum 1/(x-a_k)$ appearing together, or a sum $\sum g_k'/g_k$ anywhere, is the fingerprint of logarithmic differentiation. Also, any integral whose integrand looks too complicated to integrate term by term should be tested for being an exact derivative before anything else.

6. Reusable rule: when a problem defines an object indirectly (a coefficient, a value of a special function), unpack the definition completely, then re-read the whole expression with the unpacked object in hand; the rest of the expression is usually designed to interact with it. And: look for an antiderivative before looking for a method of integration.

- ▷ **Cue:** a product of linear factors $\prod(y+k)$ appears with the sum $\sum 1/(y+k)$ over the same k $\longrightarrow$ **Move:** recognise $\sum 1/(y+k) = P'/P$ (logarithmic derivative), so the product times the sum is P'
- ▷ **Cue:** an integrand is $F \cdot G$ where G could be F'/F $\longrightarrow$ **Move:** the integrand is F'; the integral is $F(b) - F(a)$
- ▷ **Cue:** the coefficient of x^r in $(1+x)^\alpha$ is evaluated at a negative argument like $\alpha = -y-1$ $\longrightarrow$ **Move:** use $\binom{-y-1}{r} = (-1)^r\binom{y+r}{r} = (-1)^r\,(y+1)\cdots(y+r)/r!$, a rising factorial
- ▷ **Cue:** a definite integral whose integrand would take pages to expand $\longrightarrow$ **Move:** first test whether it is an exact derivative of something visible in the problem

1992 A3

1992 A3. For a given positive integer m, find all triples (n, x, y) of positive integers, with n relatively prime to m, which satisfy

$$(x^2+y^2)^m = (xy)^n.$$

Fields: number theory, Diophantine equations **Position:** A3 — mid-paper **Answer:** If m is odd there are no solutions. If m is even the only solution is $(n, x, y) = \left(m+1,\, 2^{m/2},\, 2^{m/2}\right)$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The natural first move is to try small cases. With $m = 1$ the equation is $x^2+y^2 = (xy)^n$; the student tries $x = y$ and gets $2x^2 = x^{2n}$, i.e. $2 = x^{2n-2}$, impossible. Unequal small values ($x = 1, y = 2$: 5 versus 2^n) fail too. With $m = 2$ and $x = y$: $4x^4 = x^{2n}$, so $4 = x^{2n-4}$, and $x = 2$, $n = 3$ works: $(4+4)^2 = 64 = 4^3$. So solutions exist sometimes.

The student then tries a prime-by-prime attack directly on x and y: "let p be a prime dividing x; then $p \mid (xy)^n$, so $p \mid (x^2+y^2)^m$, so $p \mid x^2+y^2$, so $p \mid y^2$, so $p \mid y$." This shows x and y have the same prime divisors. But then the argument stalls. The same primes, fine — with what exponents? Comparing p-adic valuations on the two sides requires knowing $v_p(x^2+y^2)$, and that is not determined by $v_p(x)$ and $v_p(y)$ when they are equal (for example $v_2(2^2+2^2) = 3$, not 2). The student is now inside a case analysis on valuations that does not close, and the coprimality condition on n and m has not yet been used anywhere, which is a sign that the route is not the one intended.

The model chain. Suppose (n, x, y) is a solution.

Step 1: $n > m$. By AM–GM, $x^2+y^2 \geq 2xy > xy$. If $xy = 1$ then $x = y = 1$ and the equation reads $2^m = 1$, which is false; so $xy \geq 2$. Now $(xy)^n = (x^2+y^2)^m > (xy)^m$, and since $xy \geq 2$ this forces $n > m$.

Step 2: strip the common factor. Let $d = \gcd(x, y)$ and write $x = da$, $y = db$ with $\gcd(a, b) = 1$. Substituting,

$$d^{2m}(a^2+b^2)^m = d^{2n}(ab)^n, \qquad \text{so} \qquad (a^2+b^2)^m = d^{2(n-m)}(ab)^n,$$

using $n > m$ to divide by d^{2m} and keep a non-negative power of d.

Step 3: $a = b = 1$. Suppose a prime p divides a. Then p divides the right-hand side (since $n \geq 1$), so $p \mid (a^2+b^2)^m$, so $p \mid a^2+b^2$ (p is prime), so $p \mid b^2$, so $p \mid b$, contradicting $\gcd(a, b) = 1$. Hence a has no prime divisor, i.e. $a = 1$. By symmetry of the equation in a and b (the same argument with the roles swapped), $b = 1$. So $x = y = d$.

Step 4: solve for d. With $a = b = 1$ the equation is $2^m = d^{2(n-m)}$. By unique factorisation d is a power of 2, say $d = 2^r$, and comparing exponents,

$$m = 2r(n-m). \qquad (*)$$

Since $m \geq 1$, $r \geq 1$ and $n - m \geq 1$.

Step 5: use $\gcd(n, m) = 1$. From $(*)$, $n - m$ divides m. Also $\gcd(n - m, m) = \gcd(n, m) = 1$. A positive integer that divides m and is coprime to m is 1. So $n - m = 1$, and $(*)$ becomes $m = 2r$.

Therefore m must be even, $r = m/2$, $n = m + 1$ and $x = y = 2^{m/2}$. If m is odd, $(*)$ with $n - m = 1$ is impossible, so there is no solution.

Step 6: check. For even m, $\gcd(m + 1, m) = 1$, and with $x = y = 2^{m/2}$: $x^2 + y^2 = 2^{m+1}$, so $(x^2 + y^2)^m = 2^{m(m+1)}$, while $(xy)^n = (2^m)^{m+1} = 2^{m(m+1)}$. The triple works.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains observe that a prime dividing x must divide y. They split on what to do with that observation. The natural attempt keeps x and y and tries to track exponents prime by prime. The model *first divides out* $\gcd(x, y)$, so that the same observation becomes a contradiction instead of a constraint: in coprime a, b there is no room for a shared prime, so $a = b = 1$ at once.

> *Every prime of x is a prime of y and vice versa. That is exactly the kind of fact that becomes a contradiction if I first make x and y coprime. So write $x = da$, $y = db$ with $\gcd(a, b) = 1$ and run the same argument on a and b.*

> *The hypothesis $\gcd(n, m) = 1$ has not been used yet, so it must be what kills the extra solutions. Where does n meet m in a divisibility? In $m = 2r(n - m)$: $n - m$ divides m, and $\gcd(n - m, m) = \gcd(n, m) = 1$.*

> *Before dividing by d^{2m}, I need to know which of m and n is bigger. $x^2 + y^2 > xy$ by AM–GM settles it.*

Layer 3 THE WHY

Why is "divide by the gcd" the natural move? Look at the shape of the equation: it is *homogeneous* in (x, y) — both sides are homogeneous polynomials, of degrees $2m$ and $2n$. For a homogeneous equation, scaling $(x, y) \mapsto (dx, dy)$ multiplies each side by a fixed power of d. That means the equation factors cleanly into a "scale" part (d) and a "shape" part (the primitive pair (a, b)). Separating scale from shape is the standard first move for homogeneous Diophantine equations for the same reason it is for Pythagorean triples: the primitive pair carries all the arithmetic constraints, and the scale is left to a single exponent comparison. It is not a trick; it is what homogeneity is for.

Once (a, b) is primitive, the argument the natural attempt already had — "a prime of one divides the other" — goes through without any valuations, because $a^2 + b^2$ and ab share no prime when $\gcd(a, b) = 1$. That is the real structural fact: *ab and $a^2 + b^2$ are coprime whenever a and b are*. In the reduced equation $(a^2 + b^2)^m = d^{2(n-m)}(ab)^n$, the factor $(ab)^n$ divides $(a^2 + b^2)^m$; but a number that divides a power of something coprime to it must be 1. So $ab = 1$ is forced, and all the remaining information sits in the scale d.

The coprimality of n and m deserves its own "of course". Without it the problem has more solutions: from $2^m = d^{2(n-m)}$, any factorisation $m = 2rt$ with $t = n - m$ gives a solution $x = y = 2^r$, $n = m + t$. For instance $m = 4$, $t = 2$, $r = 1$ gives $(n, x, y) = (6, 2, 2)$: $(4 + 4)^4 = 4096 = 4^6$. Here $\gcd(6, 4) = 2$. So the hypothesis $\gcd(n, m) = 1$ is exactly the condition that selects $t = 1$. When a hypothesis is stated but unused late in a solution, find the line where it bites; here it is the observation that $t \mid m$ and $\gcd(t, m) = \gcd(n, m)$, so t is a divisor of m coprime to m.

Finally, the parity condition on m comes from the "2" in $2^m = d^{2(n-m)}$: the right-hand side is a perfect square, so 2^m must be a perfect square, so m is even. That is a quicker way to see the odd case dies, independent of Step 5.

There is no official solution in the archive's sources for 1992. The chain above was constructed independently, each step checked, and the answer confirmed by a computer search over $1 \le m \le 8$, $n < 30$, $x, y < 70$, which found exactly the triples $(m+1, 2^{m/2}, 2^{m/2})$ for even m and nothing for odd m.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that both sides are homogeneous in (x, y). On seeing it, write $d = \gcd(x, y)$, $x = da$, $y = db$, $\gcd(a, b) = 1$.

2. Reduce to the primitive pair and use the coprimality of ab and $a^2 + b^2$ to force $a = b = 1$; then compare exponents of 2.

3. Hesitation occurs when the student knows x and y have the same prime divisors but cannot finish. It is reinforced away by the reflex: "same primes" plus "make them coprime" equals "no primes". A second hesitation is at the sign of $n - m$ before dividing; AM–GM removes it.

4. The trap is overlooking the role of $\gcd(n, m) = 1$ and reporting a family of solutions $x = y = 2^r$ with $m = 2r(n - m)$, or conversely forgetting that for odd m the answer is "none". A smaller trap: omitting the verification that the found triple satisfies the equation and the coprimality condition.

5. Homogeneous Diophantine equations in two or more variables; any equation where a prime dividing one variable is forced to divide another. Also, any stated hypothesis not yet used: it marks the line of the proof still to be written.

6. Reusable rule: for a homogeneous equation, factor out the gcd first and work with a primitive solution; coprime variables turn "shares a prime" arguments into immediate contradictions. And audit your hypotheses: an unused condition tells you where the remaining work is.

▷ **Cue:** a Diophantine equation is homogeneous in its variables $\longrightarrow$ **Move:** write the variables as d times a primitive (coprime) tuple, and separate the scale d from the shape

▷ **Cue:** a prime dividing x is forced to divide y $\longrightarrow$ **Move:** if x, y are coprime this is a contradiction, so the variable is 1; reduce to coprime variables first

▷ **Cue:** a, b coprime and both ab and $a^2 + b^2$ appear $\longrightarrow$ **Move:** use $\gcd(ab, a^2 + b^2) = 1$

▷ **Cue:** a hypothesis such as $\gcd(n, m) = 1$ has not been used by the end of the derivation $\longrightarrow$ **Move:** find where n and m meet in a divisibility; use $\gcd(n - m, m) = \gcd(n, m)$

▷ **Cue:** an equation $2^m = d^{2k}$ in positive integers $\longrightarrow$ **Move:** d is a power of 2 and m must be even

1992 A4

1992 A4. Let f be an infinitely differentiable real-valued function defined on the real numbers. If

$$f\left(\frac{1}{n}\right) = \frac{n^2}{n^2+1}, \qquad n = 1, 2, 3, \ldots,$$

compute the values of the derivatives $f^{(k)}(0), k = 1, 2, 3, \ldots$.

Fields: real analysis (Rolle's theorem, smooth functions); power series **Position:** A4 — middle of the paper. The computation is short; the content is justifying that the data at the points $1/n$ determine every derivative at 0. **Answer:** $f^{(k)}(0) = 0$ for k odd and $f^{(k)}(0) = (-1)^{k/2}\, k!$ for k even.

Layer 1 THE WHAT — surface analysis

The natural attempt. You rewrite the data: $\dfrac{n^2}{n^2+1} = \dfrac{1}{1+1/n^2}$, so with $x = 1/n$ the condition

reads $f(x) = \dfrac{1}{1+x^2}$ at $x = 1, \frac{1}{2}, \frac{1}{3}, \ldots$. The temptation is now to declare $f(x) = \dfrac{1}{1+x^2}$ and differentiate. Expand $\dfrac{1}{1+x^2} = 1 - x^2 + x^4 - \cdots$ for $|x| < 1$, read off $f^{(2m)}(0) = (-1)^m (2m)!$, odd derivatives zero, and write down the answer.

The answer is right. The argument is not. Nothing says $f = \dfrac{1}{1+x^2}$: the function

$$h(x) = \begin{cases} e^{-1/x^2} \sin(\pi/x), & x \neq 0, \\ 0, & x = 0, \end{cases}$$

is infinitely differentiable, vanishes at every $1/n$ (because $\sin(\pi n) = 0$), and is not zero. So $f = \dfrac{1}{1+x^2} + h$ satisfies every hypothesis. A student who wrote "so $f(x) = 1/(1+x^2)$" has written a false sentence and would lose most of the marks. The honest alternative that many then try — difference quotients $\dfrac{f(1/n) - f(0)}{1/n}$ for $f'(0)$, then second differences for $f''(0)$ — works for $k = 1$ and becomes an unreadable nest of limits by $k = 3$.

The model chain. Let $g(x) = \dfrac{1}{1+x^2}$ and $h = f - g$. Both f and g are infinitely differentiable on $\mathbb{R}$, so h is too, and $h(1/n) = 0$ for every $n \geq 1$.

Claim: $h^{(k)}(0) = 0$ for every $k \geq 0$. We prove, by induction on k, the stronger statement: *$h^{(k)}$ has zeros $z_{k,1} > z_{k,2} > \cdots > 0$ with $z_{k,j} \to 0$ as $j \to \infty$.*

For $k = 0$ take $z_{0,j} = 1/j$.

Suppose the statement holds for k. For each j, $h^{(k)}$ is differentiable on $\mathbb{R}$ and $h^{(k)}(z_{k,j+1}) = h^{(k)}(z_{k,j}) = 0$, so by Rolle's theorem there is $z_{k+1,j}$ with $z_{k,j+1} < z_{k+1,j} < z_{k,j}$ and $h^{(k+1)}(z_{k+1,j}) = 0$. These points interlace the previous ones, so they are strictly decreasing, positive, and squeezed to 0: $0 < z_{k+1,j} < z_{k,j} \to 0$. That is the statement for $k+1$.

Now $h^{(k)}$ is continuous (it is differentiable), so

$$h^{(k)}(0) = \lim_{j \to \infty} h^{(k)}(z_{k,j}) = 0.$$

This proves the claim, and hence $f^{(k)}(0) = g^{(k)}(0)$ for every k.

Computing $g^{(k)}(0)$. For $|x| < 1$,

$$g(x) = \frac{1}{1+x^2} = \sum_{m \geq 0} (-1)^m x^{2m},$$

a power series with radius of convergence 1. A power series may be differentiated term by term inside its interval of convergence, so its k-th derivative at 0 is $k!$ times the coefficient of x^k. The coefficient of x^k is 0 for odd k and $(-1)^{k/2}$ for even k. Therefore

$$f^{(k)}(0) = g^{(k)}(0) = \begin{cases} 0, & k \text{ odd}, \\ (-1)^{k/2}\, k!, & k \text{ even}. \end{cases}$$

So $f'(0) = 0$, $f''(0) = -2$, $f'''(0) = 0$, $f^{(4)}(0) = 24$, $f^{(6)}(0) = -720$, and so on. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains spot that the data are $\dfrac{1}{1+x^2}$ sampled at $x = 1/n$. At that line the natural attempt writes "so $f(x) = \dfrac{1}{1+x^2}$". The model writes "so $f - \dfrac{1}{1+x^2}$ vanishes at every $1/n$", and asks what a smooth function vanishing on a sequence accumulating at 0 must do at 0.

> *I've found a nice function that fits the data, but the data are only a sequence of points. f isn't determined by them — so the question can only be asking about the part of f that* is *determined. Subtract my candidate. The difference has zeros piling up at* 0*. What does that force?*

> *Zeros piling up at* 0 *from one side: Rolle puts a zero of the derivative between each consecutive pair, and those pile up at* 0 *too. Continuity then gives value* 0 *at* 0*. Repeat, level by level.*

Layer 3 THE WHY

Why subtract rather than identify? Read what the question asks for. It does not ask for f; it asks for $f^{(k)}(0)$. A problem that could determine f would ask for f. The word choice is a signal that f is not pinned down, and the right reaction is to separate f into a part you know and a part you don't, then show the unknown part is invisible to the question. Writing $f = g + h$ with $h(1/n) = 0$ is that separation, and it costs one line.

Why Rolle? Go back to what a derivative at 0 is: a limit built from values near 0. The only values we have near 0 are at $1/n$, and there $h = 0$. So $h(0) = 0$ by continuity, and

$$h'(0) = \lim_{n\to\infty} \frac{h(1/n) - h(0)}{1/n} = 0.$$

That handles $k = 1$ from the definition. For $k = 2$ the definition wants values of h' near 0, which we do not have. Rolle's theorem is precisely the device that converts "two zeros of h" into "one zero of h' between them". So it is not a trick; it is the only standard theorem that produces information about h' from information about the values of h. Once you have zeros of h' accumulating at 0, you are in the same situation one level up, and induction is the bookkeeping.

Why "infinitely differentiable" and not "analytic"? Because the problem is precisely poised on the difference. If f were analytic, the identity theorem would force $f = g$ near 0, and the naive attempt would be correct. For merely smooth f, the example $h(x) = e^{-1/x^2}\sin(\pi/x)$ shows f is not determined, yet all derivatives at 0 still are: h is flat at 0, all its derivatives vanish there, and the Rolle argument proves that *every* smooth h vanishing at the $1/n$ is flat at 0. The hypotheses give you exactly enough for the Taylor coefficients at 0 and not an inch more. Seeing that is seeing why the question is phrased as it is.

Why the power series for g? Because it is the cheapest route to all derivatives at once. Differentiating $\dfrac{1}{1+x^2}$ by the quotient rule eight times is zombie maths. The geometric series gives every derivative at 0 in one line, using the fact that a convergent power series's k-th derivative at the centre is $k!$ times its k-th coefficient.

There is no official solution in the archive's sources for 1992. The argument above is constructed here and checked: the interlacing in the induction is written out, continuity of $h^{(k)}$ follows from differentiability of $h^{(k)}$, and the values $f''(0) = -2$, $f^{(4)}(0) = 24$, $f^{(6)}(0) = -720$, $f^{(8)}(0) = 40320$ were confirmed by direct symbolic differentiation of $1/(1+x^2)$.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that $\dfrac{n^2}{n^2+1}$ is a function of $1/n$: write down $\dfrac{1}{1+(1/n)^2}$ and the candidate $g(x) = \dfrac{1}{1+x^2}$. The second cue is "compute the derivatives at 0" rather than "find f": write $h = f - g$ and note $h(1/n) = 0$.

2. Show the difference h has all derivatives zero at 0 by iterated Rolle plus continuity; then read the derivatives of g from its power series.

3. Hesitation occurs at "is f really $1/(1+x^2)$?". The weak student suppresses the doubt and asserts it; the strong student acts on the doubt. Reinforce with the flat function e^{-1/x^2}:

smooth functions can hide anything in a region where all derivatives vanish, so a sequence of values never determines a smooth function.

4. The trap is identifying f with the obvious interpolant. It gives the right numbers with a false proof. A second trap is differentiating $1/(1+x^2)$ repeatedly by hand instead of using the geometric series.

5. Slogan: *if you are asked for less than the whole object, the whole object is probably not determined — subtract a candidate and show the remainder is invisible.* And: *zeros accumulating at a point force every derivative to vanish there, by Rolle and continuity.*

6. Decision rule: when data at a sequence of points $x_n \to a$ are given for a smooth function and derivatives at a are asked for, find a nice function matching the data, subtract, and apply iterated Rolle to the difference. More generally, when the question asks for less than the hypotheses seem to determine, split the object into known part plus unknown part and prove the unknown part does not affect the answer.

▷ **Cue:** a smooth f given at the points $1/n$ only, derivatives at 0 requested $\longrightarrow$ **Move:** subtract the obvious interpolant; the difference vanishes on a sequence tending to 0

▷ **Cue:** a differentiable function with zeros $z_1 > z_2 > \cdots \to a$ $\longrightarrow$ **Move:** Rolle gives zeros of the derivative interlacing them, also tending to a; continuity gives value 0 at a

▷ **Cue:** all derivatives at 0 of a rational function like $1/(1+x^2)$ $\longrightarrow$ **Move:** expand as a power series; $f^{(k)}(0) = k! \cdot [x^k]$

▷ **Cue:** a statement that seems to identify a smooth function from countably many values $\longrightarrow$ **Move:** test against the flat function e^{-1/x^2} times something vanishing on the data

1992 A5

1992 A5. For each positive integer n, let $a_n = 0$ (or 1) if the number of 1's in the binary representation of n is even (or odd), respectively. Show that there do not exist positive integers k and m such that

$$a_{k+j} = a_{k+m+j} = a_{k+2m+j},$$

for $0 \le j \le m-1$.

Fields: combinatorics on words; number theory (binary representation); infinite descent **Position:** A5 — hard end of the paper. The statement says the Thue–Morse sequence contains no cube XXX; the proof is a minimal-counterexample descent that splits on the parity of m. **Answer:** No such k, m exist: the sequence (a_n) contains no three consecutive identical blocks.

Layer 1 THE WHAT — surface analysis

The natural attempt. You write out the sequence. Extending the definition to $n = 0$ (zero 1's, so $a_0 = 0$):

$$a_0a_1a_2\cdots = 0\,1\,1\,0\;1\,0\,0\,1\;1\,0\,0\,1\;0\,1\,1\,0\;\cdots$$

You notice the block structure: each block of length 2^r is followed by its complement. You try to use that directly. A repeated block XXX of length m would have to sit somewhere inside this nested pattern, and you start casing on where k falls relative to the powers of 2 and on how m compares with 2^r. Small m go by inspection: $m = 1$ would need 000 or 111, and you can check none occurs in the first few dozen terms. But "I checked the first few dozen" is not a proof, and the case analysis on block positions for general m has no end. You have the pattern; you have no mechanism.

The model chain. Write t_n for a_n with the convention $t_0 = 0$. Two facts follow from reading binary digits. Appending a 0 to the binary expansion of n gives $2n$, and appending a 1 gives $2n+1$, so

$$t_{2n} = t_n, \qquad t_{2n+1} = 1 - t_n \qquad (n \ge 0). \tag{$*$}$$

In particular $t_{2n} \neq t_{2n+1}$ for every $n \geq 0$.

Call a pair (k, m) with $k \geq 0$, $m \geq 1$ a *cube* if $t_{k+j} = t_{k+m+j} = t_{k+2m+j}$ for $0 \leq j \leq m-1$. We show no cube exists; the problem is the case $k \geq 1$. Suppose cubes exist and choose one, (k, m), with m as small as possible. Write $x_i = t_{k+i}$ for $0 \leq i \leq 3m-1$. The cube condition says exactly

$$x_i = x_{i+m} \qquad (0 \leq i \leq 2m-1). \tag{P}$$

(For $i < m$ this is the first equality with $j = i$; for $m \leq i \leq 2m-1$ it is the second with $j = i-m$.)

Case 1: m odd. We claim $x_i \neq x_{i+1}$ for every $0 \leq i \leq 2m-2$. If $k+i$ is even, this is $t_{2n} \neq t_{2n+1}$ with $2n = k+i$. If $k+i$ is odd, then $k+i+m$ is even because m is odd, and $i+m+1 \leq 3m-1$, so $x_{i+m} \neq x_{i+m+1}$ by the same fact; by (P) applied at i and at $i+1$ (both $\leq 2m-1$), $x_i = x_{i+m}$ and $x_{i+1} = x_{i+m+1}$, hence $x_i \neq x_{i+1}$.

So $x_0, x_1, \ldots, x_{2m-1}$ alternate, and $x_i = x_0$ exactly when i is even. Since $m \leq 2m-1$ and m is odd, $x_m \neq x_0$. But (P) at $i = 0$ says $x_m = x_0$. Contradiction.

Case 2: m even, $m = 2m'$ with $m' \geq 1$. We produce a cube of length $m' < m$, contradicting minimality.

If k is even, write $k = 2k'$ and take $j = 2i$ with $0 \leq i \leq m'-1$ (so $0 \leq j \leq m-1$). By $(*)$, $t_{2(k'+i)} = t_{k'+i}$, and likewise for the other two terms, so the cube condition at $j = 2i$ reads

$$t_{k'+i} = t_{k'+m'+i} = t_{k'+2m'+i} \qquad (0 \leq i \leq m'-1),$$

i.e. (k', m') is a cube.

If k is odd, write $k+1 = 2k'$ and take $j = 2i+1$ with $0 \leq i \leq m'-1$ (so $1 \leq j \leq m-1$). Then $k+j = 2(k'+i)$, $k+m+j = 2(k'+m'+i)$, $k+2m+j = 2(k'+2m'+i)$, and $(*)$ again gives that (k', m') is a cube.

Either way minimality fails. So no cube exists, and in particular none with $k \geq 1$. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains see the self-similar block structure. The natural attempt tries to use it at the level of blocks of length 2^r — a global picture. The model uses it at the level of single indices, as the recurrence $(*)$: halving an even index keeps the value. That local rule immediately suggests the move "take every other term of a cube and get a half-size cube", and descent follows. The split is at the line "the sequence is built by doubling"; the model reads it as a rule about $n \mapsto 2n$, $2n+1$.

> *The sequence is defined digit by digit, so what I actually know is how t_{2n} and t_{2n+1} relate to t_n. If I take every second term of a cube, I get the sequence at half the index — another cube, half as long. That's descent. It only works if m is even, so the odd case must be killed some other way.*

> *For odd m, what does $(*)$ give me locally? Each pair $(2n, 2n+1)$ is 01 or 10. Shifting by an odd m swaps the parity of the index, so the "pairs differ" fact transfers across the period to every adjacent pair. Then the block alternates — and an alternating sequence cannot have odd period.*

Layer 3 THE WHY

Why descent? The object is defined by binary digits, and the only operation that interacts cleanly with binary digits is multiplication by 2 (append a digit) and its inverse (delete the last digit). So the natural question about any pattern in (t_n) is: what happens to it under $n \mapsto \lfloor n/2 \rfloor$? A cube of length m, sampled at every other index, becomes a cube of length $m/2$ in the original sequence. When a structure reproduces itself at half size, you look at the smallest instance and derive a contradiction. That is minimal counterexample, which is induction in its most economical form.

Why does parity of m appear? Because sampling every other term only lines up with the doubling structure when m is even: the three copies start at k, $k+m$, $k+2m$, and for the even-indexed samples in each copy to correspond, m must be even. So descent handles even m and fails for odd m, and odd m needs a separate argument. This is the moment to ask what else $(*)$ gives. It gives a local fact: $t_{2n} \neq t_{2n+1}$, so the sequence never has three equal terms in a row — every run of length 3 contains an aligned pair $(2n, 2n+1)$. That settles $m = 1$. For general odd m, periodicity with odd period moves each index to one of the opposite parity, so the "differ" fact, known only at even starting positions, is copied to odd starting positions too. Two constraints — alternate, and have odd period — are incompatible, because alternation has period 2 and an odd period m combined with period 2 forces period $\gcd(2, m) = 1$, a constant sequence, which does not alternate.

Why include $k = 0$? Because the descent from odd k lands at $k' = (k+1)/2 \geq 1$, but from even k it lands at $k/2$, which can be 0. Setting up the minimal counterexample over all $k \geq 0$ makes the argument closed under the descent. This is a standard hygiene move: prove the statement for a class closed under your reduction, even if the problem asks for a subclass.

This is the theorem of Thue (1912) that the Thue–Morse sequence is cube-free; in fact it is overlap-free (no $aXaXa$ with a a letter), which is stronger. There is no official solution in the archive's sources for 1992. The proof above is constructed and checked here: every index range in both cases is verified in the text ($i + m + 1 \leq 3m - 1$ in Case 1, $j \leq m - 1$ in both subcases of Case 2), and a computer search over $k < 3000$, $m < 300$ found no cube in the sequence, as the theorem requires.

Layer $+\alpha$ BACKTESTING

1. The first cue is "number of 1's in the binary representation". On seeing it, write down $t_{2n} = t_n$ and $t_{2n+1} = 1 - t_n$ immediately — the definition converted into a recurrence.

2. A recurrence under $n \mapsto 2n$ suggests descent: take a minimal cube and halve it. The parity split falls out of when halving works.

3. Hesitation occurs when descent fails for odd m. It is reinforced away by asking what the recurrence says *locally* — adjacent aligned terms differ — and combining that with the periodicity the cube gives.

4. The trap is working with the block picture (complements of blocks of length 2^r) and casing on positions relative to powers of 2. A second trap is restricting the minimal counterexample to $k \geq 1$, which the descent can leave.

5. Slogan: *a sequence defined by binary digits is a sequence with a doubling recurrence; patterns in it descend under halving.* And: *odd period plus alternation is a contradiction.*

6. Decision rule: when a structure is self-similar under a map (here $n \mapsto 2n$), assume a minimal bad configuration and push it through the map. Where the map does not apply cleanly (a parity obstruction), the obstruction usually carries a local constraint strong enough to finish by hand. Set up the minimal counterexample in a class closed under the reduction.

▷ **Cue:** a sequence defined by the digits of n in base b $\longrightarrow$ **Move:** write the recurrence relating t_{bn+r} to t_n

▷ **Cue:** a forbidden pattern in a sequence with a doubling recurrence $\longrightarrow$ **Move:** take a minimal occurrence, sample every other term, obtain a half-size occurrence

▷ **Cue:** descent works only when a parameter is even $\longrightarrow$ **Move:** handle the odd case with a local constraint from the recurrence plus the periodicity the pattern imposes

▷ **Cue:** a finite sequence that both alternates and has odd period $\longrightarrow$ **Move:** contradiction: compare x_0 with x_m

▷ **Cue:** a reduction that can leave the problem's stated range (here $k \geq 1 \to k = 0$) $\longrightarrow$ **Move:** prove the statement on the larger range closed under the reduction

1992 A6

1992 A6. Four points are chosen at random on the surface of a sphere. What is the probability that the center of the sphere lies inside the tetrahedron whose vertices are at the four points? (It is understood that each point is independently chosen relative to a uniform distribution on the sphere.)

Fields: geometric probability; linear algebra (convex combinations, linear dependence); symmetry arguments **Position:** A6 — the last problem of the morning paper, and it was scored accordingly: very few full solutions in 1992. It is hard because the natural route needs spherical geometry most students do not have, and the short route needs a change of viewpoint that nothing in the statement announces.
Answer: $\frac{1}{8}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. Write O for the centre and P_1, P_2, P_3, P_4 for the four points. The honest first move is to condition on the first three points and ask where the fourth must land.

Fix P_1, P_2, P_3. The centre lies inside the tetrahedron exactly when it is a combination $O = \sum r_i P_i$ (vectors from O) with all $r_i > 0$. Rearranging, $-r_4 P_4 = r_1 P_1 + r_2 P_2 + r_3 P_3$, so $-P_4$ must lie in the open cone spanned by P_1, P_2, P_3; on the sphere, $-P_4$ must lie in the spherical triangle T with vertices P_1, P_2, P_3. Equivalently P_4 lies in the antipodal triangle $-T$, which has the same area. Since P_4 is uniform,

$$\Pr(\text{success} \mid P_1, P_2, P_3) = \frac{\text{area}(T)}{4\pi} \quad (\text{unit sphere}),$$

and the answer is $\mathbb{E}[\text{area}(T)]/4\pi$.

So far this is good mathematics. Now the wall. You need the expected area of a random spherical triangle. The student with only A-level tools reaches for coordinates: put P_1 at the north pole, parametrise P_2, P_3 by spherical angles, and try to write the area of T as an integral. The area of a spherical triangle in terms of its vertex coordinates is not something you can produce in the exam room from first principles, and the resulting four-dimensional integral over the positions of two points is where most scripts in 1992 stopped.

The student who knows Girard's theorem — $\text{area}(T) = \alpha + \beta + \gamma - \pi$, the angles of T — can finish. Given P_1, rotational symmetry about the axis OP_1 makes the tangent directions from P_1 towards P_2 and towards P_3 independent and uniform on a circle, so the angle between them is uniform on $[0, \pi]$ and has mean $\pi/2$. By linearity of expectation $\mathbb{E}[\text{area}(T)] = 3 \cdot \frac{\pi}{2} - \pi = \frac{\pi}{2}$, and the probability is $\frac{\pi/2}{4\pi} = \frac{1}{8}$. This works. It costs a named theorem from spherical geometry, a symmetry argument about angle distributions, and a conditioning step — three things to get right under time pressure, one of which most candidates have never met.

The model chain. Identify points of the unit sphere with unit vectors in $\mathbb{R}^3$, the centre being 0.

Step 1: the event in linear-algebra terms. Almost surely every three of $P_1, \ldots, P_4$ are linearly independent (given two independent points, the third lies on the great circle through them with probability 0). Then the 3×4 matrix $M = [P_1\, P_2\, P_3\, P_4]$ has rank 3, so its kernel is one-dimensional, spanned by some $w = (w_1, w_2, w_3, w_4)$ with

$$w_1 P_1 + w_2 P_2 + w_3 P_3 + w_4 P_4 = 0. \qquad (*)$$

Every $w_i \neq 0$: if, say, $w_4 = 0$, then $(*)$ is a nontrivial relation among P_1, P_2, P_3, which are independent.

Claim: 0 lies inside the tetrahedron if and only if w_1, w_2, w_3, w_4 all have the same sign. If 0 is inside, then $0 = \sum r_i P_i$ with $r_i > 0$, $\sum r_i = 1$ (interior points of a nondegenerate tetrahedron have all barycentric coordinates positive). So $(r_1, \ldots, r_4)$ is in the kernel, hence a multiple of w, and the

w_i share a sign. Conversely, if all w_i share a sign then $s = \sum w_i \neq 0$ and $r_i = w_i/s > 0$ with $\sum r_i = 1$, so $0 = \sum r_i P_i$ is an interior point. Points on the boundary would need some $r_i = 0$, i.e. some $w_i = 0$, which we have excluded; so "inside" and "in the interior" agree almost surely.

Step 2: split each point into a line and a sign. Generate each point in two stages: choose Q_i uniformly on the sphere, choose $\epsilon_i \in \{+1, -1\}$ by a fair coin, and set $P_i = \epsilon_i Q_i$, all eight choices independent. Because the uniform distribution is invariant under $x \mapsto -x$, each P_i is uniform, and the P_i are independent: they have exactly the distribution in the problem.

Step 3: condition on the lines. Fix $Q_1, \ldots, Q_4$ (generic, as in Step 1) and let w span the kernel of $[Q_1\, Q_2\, Q_3\, Q_4]$. Then

$$\sum_i (\epsilon_i w_i)(\epsilon_i Q_i) = \sum_i w_i Q_i = 0,$$

so $(\epsilon_1 w_1, \ldots, \epsilon_4 w_4)$ spans the kernel for the P_i. By the Claim, the centre is inside exactly when $\epsilon_1 w_1, \ldots, \epsilon_4 w_4$ all share a sign. With w fixed and every $w_i \neq 0$, the sign of $\epsilon_i w_i$ is determined by ϵ_i, and each assignment of four signs to the $\epsilon_i w_i$ comes from exactly one $(\epsilon_1, \ldots, \epsilon_4)$. Of the $2^4 = 16$ equally likely sign vectors, exactly 2 (all positive, all negative) succeed. So

$$\Pr(\text{success} \mid Q_1, \ldots, Q_4) = \frac{2}{16} = \frac{1}{8}.$$

Step 4. This conditional probability does not depend on the Q_i, so averaging over them gives the answer $\frac{1}{8}$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree on the first sentence: *condition on part of the randomness and make the rest tractable.* They split on *what* to condition on. The natural attempt conditions on three of the points and leaves the fourth point free — a continuous variable ranging over a region whose area is hard. The model conditions on all four *lines* $\{\pm P_i\}$ and leaves only four coin flips free — a discrete variable with 16 values, which can be counted. At that choice one path needs Girard's theorem and an integral; the other needs a count to 2.

> *The uniform distribution on the sphere doesn't care whether I pick P or $-P$. So each point is really a random line through the centre plus a coin flip for which end. Fix the four lines. Now there are only sixteen tetrahedra left, and I can ask how many of them contain the centre.*

> *"The centre is inside" means 0 is a positive combination of the four vectors. Four vectors in three dimensions satisfy exactly one linear relation. Flipping P_i to $-P_i$ flips the sign of its coefficient in that relation and changes nothing else. So the question is: of the sixteen sign patterns, how many have all four coefficients of the same sign? Two.*

Layer 3 THE WHY

Start from definitions. A tetrahedron is the convex hull of its vertices; a point is in the convex hull when it is a combination of the vertices with nonnegative weights summing to 1. So "the centre is inside" is not a geometric statement about angles and areas at all. It is the statement that there is a linear relation $\sum r_i P_i = 0$ with every $r_i > 0$. The moment you write the event this way, the problem is about *signs of coefficients*, and signs are exactly what the symmetry $P \mapsto -P$ acts on.

Now read the parenthesis in the statement. The problem goes out of its way to say each point is *uniform* and *independent*. Ask what the uniform distribution on a sphere centred at O is symmetric under. Rotations, yes — but the natural attempt already used rotations and they bought an integral. The other symmetry, the one specific to a sphere *about its own centre*, is the antipodal map $x \mapsto -x$. And the centre is exactly the point the question is about. That is the

same logic as the Blueprint's circle: the problem is about the centre, the centre is the fixed point of $x \mapsto -x$, therefore the antipodal map is the object of interest. Applying it to one point at a time is not a trick; it is using the symmetry the question has put in front of you, point by point because the points are independent.

Why does counting to 2 work so cleanly? Because four vectors in $\mathbb{R}^3$ satisfy one linear relation, up to scale. Flipping one vector negates one coefficient of that relation. So the 16 sign-flipped tetrahedra correspond to the 16 sign patterns of $(w_1, \ldots, w_4)$, and "same sign" is two of them. This is the whole content; the "$1/8$" is $2/2^4$, and nothing about spheres beyond central symmetry and genericity was used.

That also explains why the natural attempt, if you push it through, gives the same $\frac{1}{8}$: Girard's theorem is secretly the same symmetry. Kedlaya, Poonen and Vakil record a second solution (in the project's MAA book, 1985–2000) which makes this visible: the three great circles through the sides of T cut the sphere into eight triangles, T and its images under flipping some of P_1, P_2, P_3; the flips preserve the joint distribution, so the eight have equal expected area, and together they cover the sphere; hence $\mathbb{E}[\text{area}(T)] = 4\pi/8$. The expected-area computation you were trying to do by integration is a symmetry count in disguise.

The general form (noted in the same source) is worth owning: if a distribution on $\mathbb{R}^n$ is symmetric under $x \mapsto -x$ and $n+1$ independent samples are almost surely in general position, the probability that their simplex contains 0 is $2/2^{n+1} = 1/2^n$. For $n = 2$ — three points on a circle — that is $\frac{1}{4}$, a good warm-up to test the argument on before trusting it in three dimensions.

Verification. This problem was solved here independently before consulting the source, and the argument above agrees with the source's Solution 1. Checks run: a Monte Carlo estimate over 3 independent runs of 2×10^6 samples each (barycentric coordinates by solving the 4×4 linear system) gave 0.12496, 0.12497, 0.12525, all within about one standard error (2.3×10^{-4}) of 0.125; an earlier run's 0.1249 is consistent. The key lemma was tested directly: for 20,000 random quadruples of lines, exactly 2 of the 16 sign choices contained the centre in every case. The planar analogue gave 0.2501 against the predicted $\frac{1}{4}$.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "the centre of the sphere" together with "uniform distribution on the sphere". Write down: *the distribution is invariant under $P \mapsto -P$, and the centre is the fixed point of that map.* The second cue is "inside the tetrahedron"; write down its definition, $0 = \sum r_i P_i$ with $r_i > 0$.

2. A symmetry-conditioning argument: split each random point into an unsigned part and a sign, condition on the unsigned parts, and count the favourable sign patterns.

3. Hesitation occurs after conditioning on three points: you have $\Pr = \text{area}(T)/4\pi$ and no way to average the area. That is the signal that you conditioned on the wrong thing. Reinforce: *if conditioning leaves a hard continuous average, try conditioning on more, so that what is left is discrete.*

4. The trap is the geometric picture itself. A tetrahedron inside a sphere invites angles, solid angles and areas, and the student spends the time on spherical trigonometry. The event is linear-algebraic (signs of coefficients in a linear relation) and should be translated before any geometry is attempted. A lesser trap is ignoring the measure-zero degenerate cases; one sentence dismissing them is required.

5. Slogan: *when the distribution has a symmetry of finite order, condition on the orbit and count within it.* Here the orbit of a quadruple under independent sign flips has 16 members, and the answer is (favourable members)/(orbit size). And: *"inside a convex hull" means "positive coefficients in a linear relation".*

6. Decision rule: in a probability problem where the answer is suspected to be a simple rational number, look for a finite group of measure-preserving transformations of the sample space, fix an orbit, and count the favourable points in it. If every orbit contains the same number of favourable points, that ratio is the answer and no integration is needed.

- ▷ **Cue:** random points uniform on a sphere, question about the centre $\longrightarrow$ **Move:** use $P \mapsto -P$ on each point separately; each point is a random line plus a fair sign
- ▷ **Cue:** "the origin lies inside the convex hull of $v_1, \ldots, v_k$" $\longrightarrow$ **Move:** rewrite as a linear relation $\sum r_i v_i = 0$ with all $r_i > 0$
- ▷ **Cue:** $n+1$ generic vectors in $\mathbb{R}^n$ $\longrightarrow$ **Move:** there is a unique linear relation up to scale; negating v_i negates its coefficient
- ▷ **Cue:** conditioning leaves an unmanageable continuous average $\longrightarrow$ **Move:** condition on more (an orbit of a symmetry group) so the remaining randomness is finite and can be counted
- ▷ **Cue:** a geometric-probability answer that looks like $1/2^k$ $\longrightarrow$ **Move:** look for k independent binary symmetries; count favourable members of each orbit of size 2^k

Putnam 1992 — Paper B

1992 B1

1992 B1. Let S be a set of n distinct real numbers. Let A_S be the set of numbers that occur as averages of two distinct elements of S. For a given $n \geq 2$, what is the smallest possible number of elements in A_S?

Fields: combinatorics (extremal), inequalities (ordering) **Position:** B1 – opening problem of the second session **Answer:** The minimum is $2n-3$, attained for example by $S = \{1, 2, \ldots, n\}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You start with small cases, which is right. For $n = 2$ there is one average. For $n = 3$, $S = \{a < b < c\}$, the three averages $\frac{a+b}{2} < \frac{a+c}{2} < \frac{b+c}{2}$ are forced to be distinct, so the answer is 3. For $n = 4$ you try $\{1, 2, 3, 4\}$: the pair sums are $3, 4, 5, 5, 6, 7$, so the averages are $\frac{3}{2}, 2, \frac{5}{2}, 3, \frac{7}{2}$ – five of them, because $1 + 4 = 2 + 3$ is the only coincidence. You guess the answer is $2n - 3$, and the construction $\{1, \ldots, n\}$ confirms it as an upper bound: every pair sum $i + j$ with $i < j$ lies between 3 and $2n - 1$, and every integer in that range is achieved, which gives $2n - 3$ averages.

Then comes the lower bound, and the natural attempt goes wrong here. You try to argue that coincidences are limited: "an average $\frac{a+b}{2} = \frac{c+d}{2}$ forces $a + b = c + d$, and the arithmetic progression has as many such coincidences as possible". That sentence is a hope, not a proof. To make it rigorous you would have to count the coincidences for an arbitrary set, and show that $\binom{n}{2} - (\text{number of distinct averages}) \leq \binom{n}{2} - (2n - 3)$ for every S – a statement about additive structure (sets with small sumset) that is a genuine theorem in additive combinatorics. You can spend the rest of the session there.

The model chain. Order the set: $S = \{s_1 < s_2 < \cdots < s_n\}$. Walk along the first row and then down the last column of the addition table:

$$\frac{s_1 + s_2}{2} < \frac{s_1 + s_3}{2} < \cdots < \frac{s_1 + s_n}{2} < \frac{s_2 + s_n}{2} < \frac{s_3 + s_n}{2} < \cdots < \frac{s_{n-1} + s_n}{2}.$$

Each inequality holds because consecutive terms share one summand and the other summand strictly increases (in the first stretch s_1 is shared, in the second s_n is shared, and at the join $s_1 < s_2$ with s_n shared). Every term is the average of two distinct elements of S. The chain has $(n - 1) + (n - 2) = 2n - 3$ terms, all distinct, so $|A_S| \geq 2n - 3$ for every S.

For $S = \{1, 2, \ldots, n\}$, each average $\frac{i+j}{2}$ with $1 \leq i < j \leq n$ has $3 \leq i + j \leq 2n - 1$, so $A_S \subseteq \{\frac{3}{2}, 2, \ldots, \frac{2n-1}{2}\}$, a set of $2n - 3$ numbers. Hence $|A_S| \leq 2n - 3$, and with the lower bound, $|A_S| = 2n - 3$. The minimum is $2n - 3$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths agree on the small cases and on the construction. They split at the lower bound. The natural attempt asks "how many coincidences can there be?" and tries to bound collisions from above. The model asks the opposite question: "which averages can I *prove* are different, without knowing anything about S?" – and answers it with one strictly increasing chain.

> *I cannot control coincidences in an arbitrary set; I have no information about S except that it is ordered. So I will not count collisions. I will find a list of averages that are forced to be distinct by the ordering alone, and make that list as long as I can.*

> *Two averages are comparable for free when they share a summand: $\frac{a+c}{2} < \frac{b+c}{2}$ whenever $a < b$. So I want a path through the pairs in which each step changes one element and increases it. Start at the smallest pair, end at the largest pair: along the bottom row, then up the last column.*

Layer 3 THE WHY

Why is the chain the natural move rather than a lucky one? Look at what a lower bound on $|A_S|$ is, logically. It says: *every* set produces at least $2n-3$ distinct averages. To prove "at least k distinct values" you must exhibit k values and prove them pairwise distinct. There is no other way to certify distinctness. And the only information available about an arbitrary S is the order of its elements. So the proof must consist of averages whose distinctness follows from order alone – which means a strictly increasing chain, because order is transitive: prove each consecutive step is strict, and all $2n-3$ are distinct at once.

Now, which steps are provable from order alone? Comparing $\frac{a+b}{2}$ with $\frac{c+d}{2}$ in general needs information about sizes (is $a-c$ bigger than $d-b$?). But if the two pairs share an element, the comparison collapses to comparing the other elements, which the ordering answers. This is the axiom-level reason: the only inequalities between sums you get for free are the monotonicity ones, $x < y \Rightarrow x+z < y+z$. A chain whose every step is a single monotone replacement is the complete list of what you can say without assumptions.

How long can such a chain be? Think of the $n \times n$ grid of pairs (i, j). A step that replaces one index by a larger one moves right or up. The smallest pair is $(1, 2)$ and the largest is $(n-1, n)$; the "index sum" $i+j$ goes from 3 to $2n-1$ and increases by at least 1 each step. So no such chain has more than $2n-3$ terms, and the L-shaped path along row 1 and column n achieves that. This is also why the answer is exactly $2n-3$: the arithmetic progression is the set in which *only* the index sum matters, so every chain of this kind hits every value and nothing else. The upper and lower bounds are the same object seen twice.

The general pattern – the additive combinatorics result behind the natural attempt – is that $|A+A| \geq 2|A|-1$ for a set of reals, with equality exactly for arithmetic progressions. The proof of that theorem is this same L-shaped chain. You do not need the theorem; you need to see that the chain *is* the theorem.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "smallest possible number of elements" of a set built from arbitrary reals. On seeing it, write $s_1 < s_2 < \cdots < s_n$: order is the only structure an arbitrary set of reals has.

2. It suggests a two-part extremal proof: a construction for the upper bound (the arithmetic progression) and, for the lower bound, a list of values whose distinctness is forced by the ordering.

3. Hesitation occurs after the construction, at the sentence "now show you cannot do better". The instinct is to count collisions. What reinforces it away is the rule that "at least k distinct" is proved by exhibiting k values in a strict chain, never by bounding the number of collisions.

4. The trap is trying to characterise all coincidences $a+b=c+d$ in an arbitrary set, which is a real theorem and a time sink. A smaller trap is forgetting that the elements averaged must be distinct, which excludes the diagonal $\frac{s_i+s_i}{2}$; with the diagonal the count would be $2n-1$.

5. Any time a problem asks for the minimum number of distinct values of a symmetric function of pairs (sums, averages, differences) over an arbitrary set, look for a monotone path through the index grid from the smallest pair to the largest.

6. The reusable rule: to prove a lower bound on the number of distinct values, exhibit a strict chain whose steps are each justified by monotonicity in one variable; the length of the longest such chain is usually the answer, and the extremal example is the configuration in which every such chain is tight.

▷ **Cue:** "smallest possible number of distinct values" over an arbitrary set of reals ⟶ **Move:** order the set $s_1 < \cdots < s_n$ and look for a strictly increasing chain of values

▷ **Cue:** two sums or averages share a summand ⟶ **Move:** compare the remaining summands; the ordering decides it with no further information

▷ **Cue:** a quantity indexed by pairs (i, j) with $i < j$ ⟶ **Move:** draw the index grid; a monotone path from the smallest pair to the largest gives a chain of length (last index-sum) − (first index-sum) +1

▷ **Cue:** a guess that the arithmetic progression is extremal for a sum-type count ⟶ **Move:** the upper bound is the progression; the lower bound is the L-shaped chain along the first row and last column

1992 B2

1992 B2. For nonnegative integers n and k, define $Q(n,k)$ to be the coefficient of x^k in the expansion of $(1+x+x^2+x^3)^n$. Prove that

$$Q(n,k) = \sum_{j=0}^{k} \binom{n}{j}\binom{n}{k-2j},$$

where $\binom{a}{b}$ is the standard binomial coefficient. (Reminder: For integers a and b with $a \geq 0$, $\binom{a}{b} = \frac{a!}{b!(a-b)!}$ for $0 \leq b \leq a$, with $\binom{a}{b} = 0$ otherwise.)

Fields: algebra (polynomials, generating functions), combinatorics (binomial identities) **Position:** B2 – early second-session problem **Answer:** Factor $1+x+x^2+x^3 = (1+x)(1+x^2)$; the identity is the coefficient of x^k in $(1+x^2)^n(1+x)^n$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You see a power of a four-term sum and reach for the multinomial theorem. Writing a_0, a_1, a_2, a_3 for the number of factors contributing $1, x, x^2, x^3$,

$$Q(n,k) = \sum_{\substack{a_0+a_1+a_2+a_3=n \\ a_1+2a_2+3a_3=k}} \frac{n!}{a_0!\,a_1!\,a_2!\,a_3!}.$$

That is correct and it is a wall. The right-hand side you are aiming for is a single sum over j of a product of two binomials, and there is no visible way to collapse a three-parameter multinomial sum into it. You might try induction on n instead: $(1+x+x^2+x^3)^{n+1} = (1+x+x^2+x^3)^n(1+x+x^2+x^3)$ gives $Q(n+1,k) = Q(n,k) + Q(n,k-1) + Q(n,k-2) + Q(n,k-3)$, and you would have to show the binomial sum satisfies the same four-term recurrence. Each term splits by Pascal's rule into two, the index $k-2j$ shifts differently from j, and after a page of reindexing the bookkeeping is still not closed. Both routes are "zombie maths": mechanical expansion begun before asking what the right-hand side is telling you.

The model chain. Factor the base:

$$1+x+x^2+x^3 = (1+x)+x^2(1+x) = (1+x)(1+x^2).$$

Hence

$$(1+x+x^2+x^3)^n = (1+x^2)^n(1+x)^n = \left(\sum_{j=0}^{n}\binom{n}{j}x^{2j}\right)\left(\sum_{i=0}^{n}\binom{n}{i}x^{i}\right).$$

The coefficient of x^k in a product of two polynomials is the sum, over all ways of writing $k = 2j + i$, of the product of the corresponding coefficients. Since $i = k - 2j$ is determined by j,

$$Q(n,k) = \sum_j \binom{n}{j}\binom{n}{k-2j},$$

where j runs over integers with $0 \le j \le n$ and $0 \le k - 2j \le n$. Every such j satisfies $0 \le j \le k$ (because $2j \le k$). Conversely, for integers $0 \le j \le k$ outside those constraints, one of the factors is 0 by the stated convention ($\binom{n}{j} = 0$ if $j > n$; $\binom{n}{k-2j} = 0$ if $k - 2j < 0$ or $k - 2j > n$). So extending the range to $0 \le j \le k$ adds only zero terms, and

$$Q(n,k) = \sum_{j=0}^{k} \binom{n}{j}\binom{n}{k-2j}. \qquad \blacksquare$$

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths agree that $Q(n,k)$ is a coefficient of a power. They split at the first manipulation. The natural attempt expands $(1 + x + x^2 + x^3)^n$ as it stands. The model first factors the base and only then expands, each factor by the binomial theorem.

> *The answer is a product of two binomial coefficients summed over one index. A product of two binomial coefficients is what you get from multiplying two binomial expansions. So the four-term polynomial must be a product of two two-term polynomials – can I factor* $1 + x + x^2 + x^3$*?*

> *The second binomial has bottom entry* $k - 2j$*, with a* 2 *in front of* j*. So one of the two factors contributes in steps of* x^2*: it is* $(1 + x^2)^n$*, and the other is* $(1 + x)^n$*.*

Layer 3 THE WHY

Why read the right-hand side first? Because in a "prove that" problem the target is data. The statement gives you the answer, and its shape tells you the mechanism. A sum $\sum_j c_j d_{k-2j}$ is, by definition of polynomial multiplication, a coefficient of a product: if $C(x) = \sum c_j x^{2j}$ and $D(x) = \sum d_i x^i$, then the coefficient of x^k in $C(x)D(x)$ is exactly $\sum_j c_j d_{k-2j}$. This is the Cauchy product rule, and it is the axiom-level fact here: multiplication of polynomials *is* convolution of coefficient sequences. Read in reverse, a convolution of binomial coefficients says "these two binomial expansions were multiplied". With $c_j = \binom{n}{j}$ and $d_i = \binom{n}{i}$, the two expansions are $(1 + x^2)^n$ and $(1 + x)^n$, and their product is $\big((1 + x)(1 + x^2)\big)^n$. At that point the only question left is whether $(1 + x)(1 + x^2) = 1 + x + x^2 + x^3$, which is a one-line expansion.

So the factorisation is not a trick you have to spot in the four-term polynomial. It is forced by the right-hand side. If you had looked only at $1 + x + x^2 + x^3$, the grouping $(1 + x) + x^2(1 + x)$ is still available (it is the standard "factor by grouping" move, and also the geometric-series fact $1 + x + x^2 + x^3 = \frac{x^4-1}{x-1} = \frac{(x^2-1)(x^2+1)}{x-1}$), but the right-hand side makes it unavoidable.

Kedlaya-style solutions to this problem are one line long: "$(1 + x + x^2 + x^3)^n = (1 + x)^n(1 + x^2)^n$; compare coefficients." Here is where that line comes from: the binomial product in the target, decoded as a Cauchy product. The only care it needs is the summation range, which the problem signals by spelling out the convention for $\binom{a}{b}$ outside $0 \le b \le a$. That reminder is itself a cue: the examiners are telling you that terms with $k - 2j < 0$ or $j > n$ appear in the stated sum and must be shown to vanish. A proof that silently uses a different range has not proved the identity as stated.

A combinatorial reading confirms the same structure. Choose, from each of n factors, an exponent in $\{0, 1, 2, 3\}$ summing to k. Each exponent $e \in \{0, 1, 2, 3\}$ has a unique binary

decomposition $e = 2\beta + \alpha$ with $\alpha, \beta \in \{0, 1\}$. Choosing the β's means choosing a set of j factors, $\binom{n}{j}$ ways; choosing the α's then means choosing a set of $k - 2j$ factors, $\binom{n}{k-2j}$ ways. The factorisation $(1 + x)(1 + x^2)$ is binary notation for $0, 1, 2, 3$.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the right-hand side: a sum over j of $\binom{n}{j}\binom{n}{k-2j}$. On seeing it, write "coefficient of x^k in $(1 + x^2)^n(1 + x)^n$".

2. It suggests generating functions: identify each binomial with a binomial expansion, multiply, and match the product with the given polynomial – here by factoring $1 + x + x^2 + x^3 = (1 + x)(1 + x^2)$.

3. Hesitation occurs at the four-term base, where the multinomial theorem feels like the right tool. What reinforces it away is the rule: before expanding a power of a polynomial, try to factor the polynomial.

4. The trap is the multinomial expansion (or induction on n), which is correct but does not simplify to the target. A second trap is the summation range: the stated sum runs to $j = k$, and you must say why the extra terms are zero.

5. When a target identity is a sum of products of coefficients with indices adding to a fixed total (weighted or not), it is a coefficient of a product of generating functions. Read the factors off the target.

6. The reusable rule: in a "prove this identity" problem, decode the target's shape first; a convolution $\sum_j a_j b_{k-mj}$ is the coefficient of x^k in $A(x^m)B(x)$, and the task reduces to recognising that product.

▷ **Cue:** a sum $\sum_j \binom{n}{j}\binom{n}{k-2j}$ (or any $\sum_j a_j b_{k-mj}$) $\longrightarrow$ **Move:** it is the coefficient of x^k in $A(x^m)B(x)$; here $(1 + x^2)^n(1 + x)^n$

▷ **Cue:** a power of a polynomial with several terms, $(1 + x + \cdots + x^r)^n$ $\longrightarrow$ **Move:** factor the base before expanding; $1 + x + x^2 + x^3 = (1 + x)(1 + x^2)$

▷ **Cue:** the statement spells out a convention such as $\binom{a}{b} = 0$ outside $0 \le b \le a$ $\longrightarrow$ **Move:** the proof must use it: show the extra terms in the stated range vanish

▷ **Cue:** exponents $0, 1, 2, 3$ chosen independently from n factors $\longrightarrow$ **Move:** split each exponent in binary, $e = 2\beta + \alpha$; one binomial counts the β's, another the α's

1992 B3

1992 B3. For any pair (x, y) of real numbers, a sequence $(a_n(x, y))_{n\ge 0}$ is defined as follows:

$$a_0(x, y) = x,$$
$$a_{n+1}(x, y) = \frac{(a_n(x, y))^2 + y^2}{2}, \qquad \text{for } n \ge 0.$$

Find the area of the region

$$\{(x, y) | (a_n(x, y))_{n\ge 0} \text{ converges}\}.$$

Fields: analysis (iteration of a real map, monotone sequences), geometry (area) **Position:** B3 – middle of the second session **Answer:** The region is $\{|y| \le 1,\ |x| \le 1 + \sqrt{1 - y^2}\}$ (the square $[-1, 1]^2$ together with the unit discs centred at $(\pm 1, 0)$), of area $4 + \pi$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You fix y and compute. With $y = 0$ the recursion is $a_{n+1} = a_n^2/2$; starting at $x = 1$ you get $1, \frac{1}{2}, \frac{1}{8}, \cdots \to 0$; starting at $x = 2$ you get $2, 2, 2, \ldots$; starting at $x = 3$ you get $3, 4.5, 10.1, \ldots$, diverging. So for $y = 0$ the threshold is $|x| \le 2$. You try $y = 1$: $a_{n+1} = (a_n^2 + 1)/2$,

and from $x = 0$ you get $0, 0.5, 0.625, 0.695, \ldots$, creeping up slowly – towards 1? From $x = 1$, constant 1. From $x = 1.1$, $1.105, 1.11, \ldots$, drifting upward, and you cannot tell from a few terms whether it escapes. The numerics are slow near $y = 1$ precisely because that is where the behaviour changes. You can guess the region is some lens or oval, but the boundary for general y does not come out of tables, and without a formula for the boundary there is no area.

The model chain. Fix y and write $f(t) = \frac{t^2+y^2}{2}$, so $a_{n+1} = f(a_n)$. Since the sequence $(a_n)_{n\geq 0}$ converges if and only if $(a_n)_{n\geq 1}$ does, and $a_n \geq y^2/2 \geq 0$ for $n \geq 1$, it suffices to study f on $[0, \infty)$, where f is increasing. Compute

$$f(t) - t = \frac{t^2 - 2t + y^2}{2} = \frac{(t-1)^2 - (1-y^2)}{2}.$$

Case $|y| > 1$. Then $f(t) - t \geq \frac{y^2-1}{2} > 0$ for all t, so $a_{n+1} - a_n \geq \frac{y^2-1}{2}$ and $a_n \to \infty$. No convergence.

Case $|y| \leq 1$. Put $s = \sqrt{1-y^2} \in [0,1]$, $r = 1 - s$, $R = 1 + s$. Then $f(t) - t = \frac{1}{2}(t-r)(t-R)$ and $0 \leq r \leq 1 \leq R$. Any limit L of the sequence satisfies $L = f(L)$ (take limits in $a_{n+1} = f(a_n)$, f continuous), so $L \in \{r, R\}$.

- If $0 \leq a_1 \leq r$: f is increasing with $f(0) = y^2/2 \geq 0$ and $f(r) = r$, so f maps $[0, r]$ into itself, and $f(t) \geq t$ there. The sequence is non-decreasing and bounded by r, hence converges.
- If $r \leq a_1 \leq R$: f maps $[r, R]$ into itself and $f(t) \leq t$ there. The sequence is non-increasing and bounded below by r, hence converges.
- If $a_1 > R$: $f(t) > t$ for $t > R$ and f maps (R, ∞) into itself, so the sequence increases. A limit L would satisfy $L \geq a_1 > R$ and $L \in \{r, R\}$, a contradiction. So it diverges.

Hence the sequence converges iff $|y| \leq 1$ and $a_1 = \frac{x^2+y^2}{2} \leq 1 + s$, i.e.

$$x^2 \leq 2 + 2s - y^2 = 1 + 2s + s^2 = (1+s)^2 \iff |x| \leq 1 + \sqrt{1-y^2}.$$

The region is $\{(x,y) : |y| \leq 1,\ |x| \leq 1 + \sqrt{1-y^2}\}$, with area

$$\int_{-1}^{1} 2\left(1 + \sqrt{1-y^2}\right) dy = 4 + 2 \cdot \frac{\pi}{2} = 4 + \pi,$$

using $\int_{-1}^{1} \sqrt{1-y^2}\, dy = \frac{\pi}{2}$ (half the unit disc). Geometrically, $|x| \leq 1 + \sqrt{1-y^2}$ with $|y| \leq 1$ is the square $[-1,1]^2$ together with the right half of the disc $(x-1)^2 + y^2 \leq 1$ and the left half of $(x+1)^2 + y^2 \leq 1$: area $4 + \frac{\pi}{2} + \frac{\pi}{2}$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths agree on fixing y. They split immediately after: the natural attempt iterates numbers; the model writes down $f(t) - t$ and factors it. From that line on, the model never computes a single term of the sequence.

> *This is a one-dimensional iteration $a_{n+1} = f(a_n)$ with f increasing. Such a sequence can only do one thing: move monotonically towards the nearest fixed point in its direction, or run off to infinity. So I do not need the terms; I need the fixed points of f and the sign of $f(t) - t$ between them.*

> *The quadratic $t^2 - 2t + y^2$ has a real root only if $y^2 \leq 1$. That is the y-boundary of the region, for free. For $|y| \leq 1$ the boundary in x is where a_1 equals the larger fixed point: beyond it, nothing stops the sequence.*

Layer 3 THE WHY

Why is "find the fixed points and the sign of $f(t) - t$" the natural first line? Go back to what convergence of $a_{n+1} = f(a_n)$ means when f is continuous: if $a_n \to L$ then $f(a_n) \to f(L)$, and $f(a_n) = a_{n+1} \to L$, so $L = f(L)$. The limit, if any, is a fixed point. That is not a technique, it is the definition of limit pushed through a continuous function. So the fixed points are the only candidates, and they are the objects of interest in the same way the centre is the object of interest for a circle.

The second axiom-level fact is about increasing maps. If f is increasing and $a_1 \le a_2$, then $a_2 = f(a_1) \le f(a_2) = a_3$, and by induction the whole sequence is non-decreasing; the same argument with the inequality reversed gives non-increasing. So for an increasing f the sequence is monotone, and whether it goes up or down is decided by the sign of $a_2 - a_1 = f(a_1) - a_1$. A monotone sequence converges iff it is bounded, and the fixed points are the barriers it cannot cross (an increasing f with $f(p) = p$ maps $[0, p]$ into $[f(0), p]$ and $[p, \infty)$ into $[p, \infty)$). The entire dynamics is therefore encoded in one sign chart of $f(t) - t$ on $[0, \infty)$. This is the "cobweb diagram" of A-level iteration, promoted from an illustration to a proof.

Why restrict to $t \ge 0$? Because f is not increasing on all of $\mathbb{R}$ – it is even. The first step squares x, so a_1 forgets the sign of x and lands in $[0, \infty)$, where f is increasing. Recognising that the problem's only non-monotone step is the very first one is what makes the monotone argument legitimate, and it explains the symmetry of the region in x.

Finally, the region's shape. The condition $x^2 + y^2 \le 2(1 + \sqrt{1 - y^2})$ looks awkward until you complete the square: $x^2 \le (1 + \sqrt{1 - y^2})^2$. The quantity $1 + \sqrt{1 - y^2}$ is the x-coordinate of the right half of the unit circle centred at $(1, 0)$, and that is why π appears in an answer to a question about a recursion. The official solution states the region and the area; the reasoning above – fixed points as the only candidates, monotonicity from an increasing map – is what makes that region the only possible answer rather than a guess from pictures.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is a recursion $a_{n+1} = f(a_n)$ with the question "for which starting values does it converge?" On seeing it, write $f(t) - t$ and factor it.

2. It suggests fixed-point analysis for a monotone map: limits must be fixed points, an increasing f gives a monotone sequence, and the sign of $f(t) - t$ decides the direction.

3. Hesitation occurs in the numerics near $y = \pm 1$ and x near $1 + \sqrt{1 - y^2}$, where convergence is very slow and you cannot tell if the sequence escapes. What reinforces it away is refusing to decide convergence numerically: decide it from the sign chart.

4. The trap is treating f as increasing on all of $\mathbb{R}$ (it is even), or forgetting the boundary case $a_1 = R$, where the sequence is constant and does converge. A lesser trap is the case $|y| > 1$, where there are no fixed points; the uniform lower bound on $f(t) - t$ disposes of it.

5. Whenever convergence of an iterated map is asked as a function of the starting point or a parameter, the answer set is bounded by preimages of fixed points; find the fixed points before computing a single iterate.

6. The reusable rule: for $a_{n+1} = f(a_n)$ with f continuous and increasing on an invariant interval, draw the sign chart of $f(t) - t$; the sequence moves monotonically towards the next fixed point in the direction of that sign, and diverges if there is none.

▷ **Cue:** a recursion $a_{n+1} = f(a_n)$ and a question about convergence $\longrightarrow$ **Move:** any limit is a fixed point: solve $f(L) = L$ first

▷ **Cue:** an iteration map f that is increasing on an invariant interval $\longrightarrow$ **Move:** the sequence is monotone; the sign of $f(a_1) - a_1$ gives the direction, and fixed points are barriers

▷ **Cue:** a quadratic fixed-point equation $t^2 - 2t + y^2 = 0$ depending on a parameter $\longrightarrow$ **Move:** the discriminant condition $|y| \le 1$ is a boundary of the parameter region

▷ **Cue:** an even map such as $t \mapsto (t^2 + c)/2$ $\longrightarrow$ **Move:** after one step the orbit is in $[0, \infty)$; study f there, where it is monotone

▷ **Cue:** a region defined by $x^2 \le (1+\sqrt{1-y^2})^2 \longrightarrow$ **Move:** it is bounded by arcs of unit circles centred at $(\pm 1, 0)$; area by square plus half-discs

1992 B4

1992 B4. Let $p(x)$ be a nonzero polynomial of degree less than 1992 having no nonconstant factor in common with $x^3 - x$. Let

$$\frac{d^{1992}}{dx^{1992}}\left(\frac{p(x)}{x^3-x}\right) = \frac{f(x)}{g(x)}$$

for polynomials $f(x)$ and $g(x)$. Find the smallest possible degree of $f(x)$.

Fields: algebra; calculus of rational functions; partial fractions **Position:** B4 — mid-paper; the difficulty is refusing to differentiate a quotient 1992 times **Answer:** 3984, attained by $p(x) = 1$. Partial fractions reduce the derivative to $1992!\left[ax^{-1993} + b(x-1)^{-1993} + c(x+1)^{-1993}\right]$ with a, b, c all nonzero; the numerator over $(x^3-x)^{1993}$ has degree 3986, and exactly two of its top three coefficients can be killed, by $b = c$, $a = -2b$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The student sees a quotient and a derivative and starts with the quotient rule. One derivative of $p/(x^3-x)$ gives $\big(p'(x^3-x) - p(3x^2-1)\big)/(x^3-x)^2$; a second gives a numerator built from p'', p', p over $(x^3-x)^3$. The pattern "denominator $(x^3-x)^{k+1}$, numerator of degree about $\deg p + 2k$" appears, so the student guesses that after 1992 steps the numerator has degree about $\deg p + 3984$ and tries to make $\deg p$ small.

This is where it jams. First, the degree bookkeeping is wrong in general: the top coefficients of the numerator can cancel, and whether they do depends on p in a way the quotient rule hides completely. Second, $\deg p$ can be as large as 1991, and nothing in the recursion tells you whether a large p might give *more* cancellation. Third, you must also worry whether f/g is in lowest terms — a common factor with $(x^3-x)^{1993}$ would lower $\deg f$ further. Iterating the quotient rule gives no handle on any of these three questions. The wall is real: the computation is 1992 steps long and the answer lives in cancellations the method cannot see.

The model chain. Write $N = 1993$ throughout.

Step 1: strip the polynomial part. Divide: $p(x) = q(x)(x^3-x) + r(x)$ with $\deg r \le 2$. Then $p/(x^3-x) = q + r/(x^3-x)$ and $\deg q \le 1991 - 3 = 1988 < 1992$, so $q^{(1992)} = 0$. Since $r \equiv p \pmod{x^3-x}$, the values of r and p at $0, \pm 1$ agree.

Step 2: partial fractions. As $x^3 - x = x(x-1)(x+1)$ with distinct roots,

$$\frac{r(x)}{x^3-x} = \frac{a}{x} + \frac{b}{x-1} + \frac{c}{x+1}, \qquad a = -p(0), \quad b = \frac{p(1)}{2}, \quad c = \frac{p(-1)}{2},$$

by the cover-up rule. The hypothesis that p shares no factor with $x^3 - x$ says exactly that $p(0), p(1), p(-1) \neq 0$, so a, b, c are all nonzero.

Step 3: differentiate. $\dfrac{d^k}{dx^k}(x-t)^{-1} = (-1)^k k!\,(x-t)^{-k-1}$, and 1992 is even, so

$$\frac{d^{1992}}{dx^{1992}}\frac{p}{x^3-x} = 1992!\left[\frac{a}{x^N} + \frac{b}{(x-1)^N} + \frac{c}{(x+1)^N}\right] = \frac{1992!\,F(x)}{(x^3-x)^N},$$

$$F(x) = a(x^2-1)^N + b\,x^N(x+1)^N + c\,x^N(x-1)^N.$$

Step 4: lowest terms. $F(0) = -a$, $F(1) = b\,2^N$, $F(-1) = c\,2^N$ (as N is odd, $(-1)^N(-2)^N = 2^N$), all nonzero. So F shares no root with $(x^3-x)^N$ and $1992!F/(x^3-x)^N$ is in lowest terms. Any

representation f/g then has $f = h \cdot 1992!F$ for some nonzero polynomial h (from $fg_0 = gf_0$ with $\gcd(f_0, g_0) = 1$), so $\deg f \geq \deg F$, with equality for $f = 1992!F$.
Step 5: the top three coefficients of F. Since N is odd, $(x^2-1)^N = x^{2N} - Nx^{2N-2} + \cdots$ has no x^{2N-1} term. Expanding,

$$[x^{2N}]F = a + b + c, \qquad [x^{2N-1}]F = N(b-c), \qquad [x^{2N-2}]F = -Na + \tbinom{N}{2}(b+c).$$

The first two vanish iff $c = b$ and $a = -2b$; then the third is $2Nb + N(N-1)b = N(N+1)b \neq 0$. So $\deg F \geq 2N - 2 = 3984$ always.
Step 6: attainment. $p(x) = 1$ is nonzero, of degree $0 < 1992$, coprime to $x^3 - x$, and $\frac{1}{x^3-x} = -\frac{1}{x} + \frac{1/2}{x-1} + \frac{1/2}{x+1}$: here $b = c = \frac{1}{2}$, $a = -1 = -2b$. So $\deg F = 3984$.
The smallest possible degree of f is $\boxed{3984}$. □

I checked the analogue with 1992 replaced by $m \in \{2, 3, 4, 6\}$ by computer algebra (answer $2m$ in each case, over random admissible p), and computed F for $N = 1993$, $b = c = 1$, $a = -2$ exactly: degree 3984, leading coefficient $N(N+1)$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start from the same object $p/(x^3 - x)$. They split at the very first line: (N) writes "$\frac{d}{dx}\frac{p}{x^3-x} = \ldots$" by the quotient rule; (M) writes "$\frac{p}{x^3-x} = q + \frac{a}{x} + \frac{b}{x-1} + \frac{c}{x+1}$" and only then differentiates. Everything after that is forced on each side — a swamp on one, three lines on the other.

> *I am being asked to differentiate 1992 times. The only rational functions I can differentiate 1992 times in my head are polynomials and $1/(x-t)$. So before I touch a derivative, I'll break $p/(x^3-x)$ into exactly those pieces.*

> *The polynomial part has degree under 1992, so it dies. Everything about p that survives is three numbers a, b, c — so the question is really about three numbers, and "degree less than 1992" is there only to make the polynomial part vanish.*

> *"Smallest degree of the numerator" means "how many leading coefficients can I cancel". Three free parameters, top coefficients are linear in them: write down the top three and solve.*

- ▷ **Cue:** a high-order derivative (or integral) of a rational function ⟶ **Move:** partial fractions first; each $1/(x-t)$ differentiates to $(-1)^k k!(x-t)^{-k-1}$
- ▷ **Cue:** "no nonconstant factor in common with" a product of distinct linear factors ⟶ **Move:** translate to: p is nonzero at each root, so every partial-fraction coefficient is nonzero
- ▷ **Cue:** minimise the degree of a polynomial depending linearly on a few parameters ⟶ **Move:** write the top coefficients as linear forms; count how many can vanish simultaneously
- ▷ **Cue:** "$\frac{f}{g}$ for polynomials f, g" with f to be minimised ⟶ **Move:** prove the fraction is in lowest terms by evaluating the numerator at the roots of the denominator

Layer 3 THE WHY

Why is partial fractions not a trick here but the only sensible first line? Go back to what differentiation does. The derivative is linear, and the k-th derivative of $(x-t)^{-1}$ is a single term. The quotient rule, by contrast, is a rule for a product, and a product does not stay a product under repeated differentiation: Leibniz's formula for the k-th derivative of $u \cdot v$ has $k+1$ terms. Asking for the 1992-th derivative of a quotient is asking you to choose a basis in which differentiation is diagonal. Partial fractions is exactly that basis for rational functions with simple poles: each basis element $1/(x-t)$ is sent to a multiple of $1/(x-t)^2$, and so on, never

mixing. That is the circle-and-centre level of "of course": the operation is linear, so decompose the input into pieces the operation treats one at a time.

The same reasoning explains the constraint $\deg p < 1992$, which the natural attempt treats as a vague size bound. In the decomposition it has one job: the polynomial quotient q has degree at most 1988, and 1992 derivatives annihilate it. The problem-setter made $\deg p$ large on purpose, to tempt you into thinking a big p matters. After decomposition, only the residues a, b, c survive, so a large p buys nothing a constant cannot.

And the coprimality hypothesis? The cover-up rule says the residue at a simple pole t is $p(t)$ divided by the derivative of the denominator at t. "No common factor with $x^3 - x$" is the statement $p(0)p(1)p(-1) \neq 0$, i.e. all three residues are nonzero. That hypothesis is doing two jobs: it prevents a pole from disappearing (which would change the denominator), and in Step 4 it is exactly what makes the reduced fraction have the full denominator $(x^3 - x)^{1993}$.

Finally, why do the degrees work out as they do? Over the common denominator, each piece contributes a polynomial of degree $2N$, and their leading behaviour at infinity is $a + b + c$, then the $1/x$ correction $N(b - c)$, and so on. At the level of the original function this is the expansion at infinity: $\dfrac{a}{x^N} + \dfrac{b}{(x-1)^N} + \dfrac{c}{(x+1)^N} = \dfrac{1}{x^N}\left[(a+b+c) + \dfrac{N(b-c)}{x} + \cdots\right]$ for large x, since $(1 \mp 1/x)^{-N} = 1 \pm N/x + \cdots$. Killing the top coefficient means the function decays faster at infinity. With three residues you get two linear conditions for free (after one normalisation), which is why two orders drop and not three. Parameter count is the reason; the computation only confirms it.

Layer $+\alpha$ BACKTESTING

1. The first cue is $\frac{d^{1992}}{dx^{1992}}$ applied to a rational function. Write $\frac{p}{x^3-x} = q + \frac{a}{x} + \frac{b}{x-1} + \frac{c}{x+1}$ on sight.
2. Partial fractions, then linear algebra on the top coefficients of the combined numerator.
3. Hesitation: "p has degree up to 1991; surely some large p does better?" It is reinforced away by noticing that the polynomial part is annihilated, so p enters only through $p(0), p(1), p(-1)$.
4. Two traps. The first is the quotient rule. The second is forgetting that f/g need not be reduced: you must show the reduced numerator is F, or the "minimum" might be lower via cancellation with the denominator. The evaluation $F(0), F(\pm 1) \neq 0$ closes that gap.
5. Any request to apply a linear operator many times (derivative, difference, shift) to a structured object should trigger the question: which decomposition makes this operator act on each piece separately?
6. Reusable decision rule: *before iterating a linear operation, decompose the input into eigen-pieces (or pieces the operation keeps simple); then a question about the output becomes a question about the finitely many coefficients of that decomposition.*

1992 B5

1992 B5. Let D_n denote the value of the $(n-1) \times (n-1)$ determinant

$$\begin{bmatrix} 3 & 1 & 1 & 1 & \cdots & 1 \\ 1 & 4 & 1 & 1 & \cdots & 1 \\ 1 & 1 & 5 & 1 & \cdots & 1 \\ 1 & 1 & 1 & 6 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & 1 & \cdots & n+1 \end{bmatrix}.$$

Is the set $\left\{\frac{D_n}{n!}\right\}_{n \geq 2}$ bounded?

Fields: linear algebra; determinants; series **Position:** B5 — late on the paper but one of the lighter B5s; the whole difficulty is reading the matrix as "diagonal plus all-ones" **Answer:** No. $D_n = n!\,H_n$ where $H_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n}$, and the harmonic numbers are unbounded.

Layer 1 THE WHAT — surface analysis

The natural attempt. Compute small cases. $D_2 = \det[3] = 3$. $D_3 = \det\begin{pmatrix} 3 & 1 \\ 1 & 4 \end{pmatrix} = 11$. D_4 by cofactor expansion along the first row of $\begin{pmatrix} 3 & 1 & 1 \\ 1 & 4 & 1 \\ 1 & 1 & 5 \end{pmatrix}$: $3(20-1) - 1(5-1) + 1(1-4) = 57 - 4 - 3 = 50$. So $D_n/n!$ runs $\frac{3}{2}$, $\frac{11}{6}$, $\frac{50}{24} = \frac{25}{12}$. The increments are $\frac{1}{3}$ and $\frac{1}{4}$, so a strong student guesses $D_n/n! = 1 + \frac{1}{2} + \cdots + \frac{1}{n}$ and the answer "unbounded".

Now the proof. The standard reflex is to subtract row 1 from every other row. Row $i \geq 2$ becomes $(-2, 0, \ldots, 0, i+1, 0, \ldots, 0)$, with $i+1$ in column i, and row 1 stays $(3, 1, \ldots, 1)$. This is an arrowhead matrix. To finish, clear the first row using each lower row: add $-\frac{1}{i+1} \times$(row i) to row 1, which kills the 1 in column i and changes the corner entry by $+\frac{2}{i+1}$. The result is upper triangular with corner $3 + \sum_{i=2}^{n-1} \frac{2}{i+1}$ and diagonal $3, 4, \ldots, n$ below it, so

$$D_n = \frac{n!}{2}\Big(3 + 2\sum_{j=3}^{n} \frac{1}{j}\Big) = n!\Big(\frac{3}{2} + \sum_{j=3}^{n} \frac{1}{j}\Big) = n!\,H_n.$$

This works. It costs a guess from three data points, an arrowhead elimination whose signs are easy to fumble, and a final index shift ($i+1$ runs over $3, \ldots, n$) that is exactly the kind of place an off-by-one error lands. It also treats row 1 as special when nothing about the matrix is special in row 1.

The model chain. Write A_n for the $(n-1) \times (n-1)$ matrix, J for the all-ones matrix, and $\Delta = \mathrm{diag}(2, 3, \ldots, n)$. The diagonal entries are $3, 4, \ldots, n+1$, i.e. $2+1, \ldots, n+1$, so

$$A_n = \Delta + J.$$

Step 1: split the last column. The last column of A_n is $\mathbf{1} + n\,\mathbf{e}_{n-1}$, where $\mathbf{1}$ is the all-ones vector. The determinant is linear in each column, so $D_n = \det B + n \det C$, where B is A_n with last column replaced by $\mathbf{1}$ and C is A_n with last column replaced by $\mathbf{e}_{n-1}$.
Step 2: $\det C = D_{n-1}$**.** Expand $\det C$ along its last column, which is $\mathbf{e}_{n-1}$: the cofactor is the top-left $(n-2) \times (n-2)$ block, which is A_{n-1}. (For $n = 3$ this block is $[3] = A_2$.)
Step 3: $\det B = (n-1)!$**.** In B every column $j < n-1$ equals $\mathbf{1} + (j+1)\mathbf{e}_j$ and the last column is $\mathbf{1}$. Subtract the last column from each other column (this does not change the determinant): column j becomes $(j+1)\mathbf{e}_j$. The matrix is now lower triangular with diagonal $2, 3, \ldots, n-1, 1$, so $\det B = (n-1)!$.
Step 4: the recursion. For $n \geq 3$, $D_n = nD_{n-1} + (n-1)!$. Divide by $n!$:

$$\frac{D_n}{n!} = \frac{D_{n-1}}{(n-1)!} + \frac{1}{n}, \qquad \frac{D_2}{2!} = \frac{3}{2} = 1 + \frac{1}{2}.$$

By induction $D_n/n! = H_n := \sum_{k=1}^{n} \frac{1}{k}$.
Step 5: H_n **is unbounded.** Group the terms between consecutive powers of 2: for $m \geq 1$, $\sum_{k=2^{m-1}+1}^{2^m} \frac{1}{k} \geq 2^{m-1} \cdot \frac{1}{2^m} = \frac{1}{2}$. Hence $H_{2^M} \geq 1 + \frac{M}{2}$, which exceeds any bound.
So the set $\{D_n/n!\}$ is **not** bounded. □

I checked $D_n/n! = H_n$ with exact arithmetic for $2 \leq n \leq 11$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains face the same matrix. (N) sees "a determinant" and reaches for the general-purpose tool, row reduction pivoting on row 1. (M) first describes the matrix:

"all ones, except the diagonal is bigger by $2, 3, \ldots, n$" — i.e. $\Delta + J$. The split is the sentence before any arithmetic. After it, (M) knows that one column operation turns J into something rank-one-free, and that each matrix contains the previous one as a corner block, so a recursion in n is available.

> *Before I row-reduce, what is this matrix? It's all ones plus a diagonal. The all-ones part is rank one — if I subtract one column from the others it disappears from them. So one column should be treated as special, and the natural one is the last, because that's where A_{n-1} sits inside A_n.*

> *The answer is asked about $D_n/n!$, and $n!$ is the product of the diagonal extras $2 \cdot 3 \cdots n$. So the problem is telling me $D_n = n! \times (\text{something})$. I want a recursion for $D_n/n!$, which will be a sum.*

- ▷ **Cue:** a matrix that is all ones (or all c) except on the diagonal $\longrightarrow$ **Move:** write it as $\Delta + J$; subtract a column (or row) of ones to kill J, or use $\det(\Delta + J) = \det\Delta\,(1 + \sum 1/\delta_i)$
- ▷ **Cue:** a family of matrices where each is the top-left block of the next $\longrightarrow$ **Move:** split the last column (or row) by linearity to get a recursion for D_n in terms of D_{n-1}
- ▷ **Cue:** a quantity normalised by $n!$ (or by a product) in the question $\longrightarrow$ **Move:** divide the recursion by that product; a multiplicative recursion becomes additive
- ▷ **Cue:** "is $\sum 1/k$-type sequence bounded?" $\longrightarrow$ **Move:** block terms between powers of 2: each block contributes at least $\frac{1}{2}$

Layer 3 THE WHY

Why is "diagonal plus all-ones" the right first sentence rather than "row-reduce"? Because row reduction is a procedure and this matrix has a structure, and the structure decides which procedure is short. A determinant is the unique alternating multilinear function of the columns with $\det I = 1$. "Multilinear" means you may split any column as a sum. "Alternating" means adding one column to another does nothing. Those two properties are the whole of Steps 1 and 3. The model chain uses the definition of the determinant; the natural attempt uses an algorithm for computing it. When the entries have a pattern, the definition sees it and the algorithm does not.

Why split the *last* column? Because A_{n-1} is the top-left corner of A_n. The problem defines a sequence of determinants, and a sequence invites a recursion; the recursion comes from peeling off whatever A_n has that A_{n-1} lacks, which is its last row and column. The extra diagonal entry $n+1$ is $1+n$, and splitting it as "the 1 that matches the pattern plus the n that is extra" is the only split that produces D_{n-1}.

Why does $n!$ appear in the question? Because the diagonal extras multiply to $2 \cdot 3 \cdots n = n!$. That is $\det\Delta$. The problem-setter divided by $\det\Delta$ to hand you the correction factor $1 + \sum_i \frac{1}{\delta_i} = 1 + \frac{1}{2} + \cdots + \frac{1}{n}$. This is the matrix determinant lemma, $\det(\Delta + \mathbf{u}\mathbf{u}^\mathsf{T}) = \det\Delta\,(1 + \mathbf{u}^\mathsf{T}\Delta^{-1}\mathbf{u})$, and the recursion in the model chain is its proof in this special case. So the harmonic number is not a coincidence you guess from $\frac{3}{2}, \frac{11}{6}, \frac{25}{12}$; it is $\mathbf{1}^\mathsf{T}\Delta^{-1}\mathbf{1}$ plus one.

Finally, the question is "bounded?" and not "compute". That wording is a signal that the value itself will be something whose growth is well known but slow. H_n grows like $\ln n$: unbounded, but so slowly that numerical evidence ($H_{11} \approx 3.02$) would never convince you. Proving divergence needs a lower bound, and the powers-of-2 grouping is the cheapest one available without calculus.

Layer $+\alpha$ BACKTESTING

1. The first cue is the matrix itself: every off-diagonal entry is 1. Write $A_n = \mathrm{diag}(2, \ldots, n) + J$ on sight, and note that $\det\mathrm{diag}(2, \ldots, n) = n!$ is the normaliser in the question.

2. Multilinearity in the last column to get a recursion, plus a column operation to evaluate the leftover determinant.

3. Hesitation occurs after computing D_2, D_3, D_4: the guess H_n from three values feels flimsy, and the student either stops or grinds D_5. It is reinforced away by deriving the recursion $D_n = nD_{n-1} + (n-1)!$, which turns a guess into an induction.

4. The trap is treating the question as "compute D_n" and producing a closed form without then proving H_n unbounded. The secondary trap is the index shift: the matrix is $(n-1) \times (n-1)$ with diagonal starting at 3, so off-by-one errors in the arrowhead sum are likely.

5. Whenever a matrix is "simple plus rank one", expect the determinant to be $\det(\text{simple}) \times (1 + \text{a sum})$. The normaliser in the question usually is $\det(\text{simple})$.

6. Reusable decision rule: *describe the object in words before choosing an algorithm; if a determinant is a structured matrix plus a rank-one correction, use multilinearity and column operations (or the matrix determinant lemma) rather than generic elimination, and read any normalising factor in the question as the determinant of the structured part.*

1992 B6

1992 B6. Let $\mathcal{M}$ be a set of real $n \times n$ matrices such that

(i) $I \in \mathcal{M}$, where I is the $n \times n$ identity matrix;

(ii) if $A \in \mathcal{M}$ and $B \in \mathcal{M}$, then either $AB \in \mathcal{M}$ or $-AB \in \mathcal{M}$, but not both;

(iii) if $A \in \mathcal{M}$ and $B \in \mathcal{M}$, then either $AB = BA$ or $AB = -BA$;

(iv) if $A \in \mathcal{M}$ and $A \neq I$, there is at least one $B \in \mathcal{M}$ such that $AB = -BA$.

Prove that $\mathcal{M}$ contains at most n^2 matrices.

Fields: linear algebra; algebraic structure (groups up to sign) **Position:** B6 — the last problem on the paper, and hard; but the difficulty is almost all in choosing the right target, not in the computation **Answer:** $\mathcal{M}$ is linearly independent in the n^2-dimensional space of real $n \times n$ matrices, so $|\mathcal{M}| \leq n^2$. The bound is attained, for example, by the four matrices I, $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$, $\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$, $\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ when $n = 2$.

Layer 1 THE WHAT — surface analysis

The natural attempt. A strong student reads four axioms and starts deriving consequences, which is right. Put $A = I$ in (ii): $IB = B \in \mathcal{M}$, so $-B \notin \mathcal{M}$. Hence $\mathcal{M} \cap (-\mathcal{M}) = \varnothing$, and in particular $0 \notin \mathcal{M}$. Next, (ii) says products of elements of $\mathcal{M}$ land in $\mathcal{M} \cup (-\mathcal{M})$, so $G = \mathcal{M} \cup (-\mathcal{M})$ is closed under multiplication. That looks like a group, with $|G| = 2|\mathcal{M}|$.

Now the student who has met groups sees "a group of matrices" and reaches for group theory: the elements pairwise commute or anticommute, so G is something like a Pauli group, and perhaps its order can be controlled by the representation theory of such groups ($\sum d_i^2 = |G|$ over irreducible representations). Pushed honestly, this route meets three walls in a row. First, nothing says $\mathcal{M}$ is finite, so "$|G|$" and the character-theoretic counting are not yet available; that is part of what is to be proved. Second, nothing yet says the elements are invertible, so G is not known to be a group. Third, even granting a finite group, turning $\sum d_i^2 = |G|$ into $|\mathcal{M}| \leq n^2$ needs an argument about which irreducible representations occur inside this particular n-dimensional real one, and that argument is not short. The student has the right raw material — (a) $\mathcal{M} \cap -\mathcal{M} = \varnothing$, (b) closure up to sign — and is building a cathedral on it.

The other common attempt is to try to classify $\mathcal{M}$: show it "must be" a set of tensor products of Pauli matrices. That is true in spirit and hopeless in forty-five minutes.

The model chain. Write $\mathrm{M}_n(\mathbb{R})$ for the real vector space of $n \times n$ matrices; it has dimension n^2. The plan is to show that $\mathcal{M}$ is linearly independent. Then every finite subset of $\mathcal{M}$ has at most n^2 elements, so $\mathcal{M}$ is finite with $|\mathcal{M}| \leq n^2$.

Step 1: $\mathcal{M} \cap (-\mathcal{M}) = \varnothing$, so $0 \notin \mathcal{M}$. For $B \in \mathcal{M}$, apply (ii) to the pair (I, B), legitimate by (i): exactly one of $B, -B$ lies in $\mathcal{M}$, and B does, so $-B \notin \mathcal{M}$. If $0 \in \mathcal{M}$ then $-0 = 0 \in \mathcal{M}$ too, a contradiction.
Step 2: every $A \in \mathcal{M}$ satisfies $A^2 = \pm I$. Let $A \in \mathcal{M}$ and let S be whichever of $A^2, -A^2$ lies in $\mathcal{M}$ (by (ii) with the pair (A, A)). For any $B \in \mathcal{M}$, (iii) gives $AB = \varepsilon BA$ with $\varepsilon = \pm 1$, so

$$A^2 B = A(AB) = \varepsilon ABA = \varepsilon(\varepsilon BA)A = \varepsilon^2 BA^2 = BA^2.$$

So S commutes with every element of $\mathcal{M}$. Suppose $S \neq I$. By (iv) some $B \in \mathcal{M}$ has $SB = -BS$; together with $SB = BS$ this gives $2SB = 0$, so $SB = 0$. But $S, B \in \mathcal{M}$, so by (ii) one of $\pm SB = 0$ lies in $\mathcal{M}$, contradicting Step 1. Hence $S = I$, i.e. $A^2 = \pm I$. In particular every $A \in \mathcal{M}$ is invertible, with $A^{-1} = \pm A$.
Step 3: a minimal dependence. Suppose $\mathcal{M}$ is linearly dependent. Among all relations

$$c_1 A_1 + c_2 A_2 + \cdots + c_r A_r = 0, \qquad A_i \in \mathcal{M} \text{ distinct}, \ c_i \in \mathbb{R} \setminus \{0\}, \tag{$*$}$$

choose one with r as small as possible. Since each A_i is invertible, $A_i \neq 0$, so $r \geq 2$.
Step 4: normalise so that I appears. Multiply $(*)$ on the left by $A_1^{-1} = \pm A_1$. For each i, $A_1 A_i = \delta_i B_i$ for a unique $B_i \in \mathcal{M}$ and sign δ_i, by (ii). So $(*)$ becomes $\sum_i c_i' B_i = 0$ with every $c_i' = \pm c_i \neq 0$. Here $B_1 = I$, since $A_1^{-1} A_1 = I \in \mathcal{M}$ and $-I \notin \mathcal{M}$. The B_i are distinct: if $B_i = B_j$ then $A_1^{-1} A_i = \pm A_1^{-1} A_j$, so $A_i = \pm A_j$, which for $i \neq j$ contradicts either distinctness or Step 1. So we have a relation of the same minimal length

$$c_1' I + c_2' B_2 + \cdots + c_r' B_r = 0, \qquad B_2, \ldots, B_r \neq I. \tag{$**$}$$

Step 5: conjugate to kill a term. Since $B_2 \neq I$, (iv) gives $C \in \mathcal{M}$ with $B_2 C = -C B_2$. By (iii), $B_i C = \varepsilon_i C B_i$ with $\varepsilon_i = \pm 1$, so $C^{-1} B_i C = \varepsilon_i B_i$; also $\varepsilon_1 = 1$ and $\varepsilon_2 = -1$. Conjugating $(**)$ by C (legitimate: C is invertible by Step 2) gives $c_1' I + \sum_{i \geq 2} \varepsilon_i c_i' B_i = 0$. Subtract this from $(**)$:

$$\sum_{i=2}^{r} (1 - \varepsilon_i)\, c_i'\, B_i = 0.$$

Each coefficient is 0 or $2c_i'$, and the B_2 coefficient is $2c_2' \neq 0$. Dropping the zero terms leaves a nontrivial relation among distinct elements of $\mathcal{M}$ with at most $r - 1$ terms, which contradicts the minimality of r.

Therefore $\mathcal{M}$ is linearly independent, and $|\mathcal{M}| \leq \dim \mathrm{M}_n(\mathbb{R}) = n^2$. $\square$

Sharpness. With $X = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$, $Z = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$, $Y = XZ$, the set $\{I, X, Z, Y\}$ satisfies (i)–(iv) with $n = 2$, and the 16 Kronecker products $P \otimes Q$ with $P, Q \in \{I, X, Z, Y\}$ do the same with $n = 4$. I checked both by direct computation, including that each set is linearly independent of full size n^2.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains derive the same first facts: $-B \notin \mathcal{M}$, and products stay in $\pm\mathcal{M}$. The split is at the sentence "prove that $\mathcal{M}$ contains at most n^2 matrices". (N) reads it as a counting problem about a group and looks for the group's order. (M) reads n^2 as a dimension, the dimension of $\mathrm{M}_n(\mathbb{R})$, and turns the problem into "show $\mathcal{M}$ is linearly independent". Everything after that is forced. A minimal relation is the standard way to attack independence. The anticommuting partner from (iv), used to conjugate, is the only tool the axioms supply that can change some coefficients and not others. The one extra lemma, $A^2 = \pm I$, is exactly what the conjugation and the normalisation need: inverses.

> *The bound is n^2, and n^2 is the dimension of the space of $n \times n$ matrices. A bound on the size of a set of vectors equal to the dimension means linear independence. So I'm not counting a group; I'm proving these matrices are independent.*

> *To prove independence, take a shortest relation and make a shorter one. What operation changes the signs of some terms but not others? Conjugating by a matrix that commutes with some and anticommutes with others. That's what (iv) is for: it hands me, for any non-identity term, a matrix that flips its sign.*

> *Conjugation leaves I fixed, so if I subtract the conjugated relation, I drops out and a flipped term survives with coefficient doubled. I need I in the relation, so I first multiply through by A_1^{-1}. That needs A_1 invertible, so I need a lemma: $A^2 = \pm I$.*

> *Why should A^2 be $\pm I$? Squaring cancels every sign in (iii): A^2 commutes with everything in $\mathcal{M}$. By (iv) the only element of $\mathcal{M}$ commuting with everything is I — unless the anticommuting partner kills it, and a zero product is forbidden by (ii).*

- ▷ **Cue:** a bound on the size of a set of $n \times n$ matrices equal to n^2 (or of vectors equal to the dimension) $\longrightarrow$ **Move:** prove the set is linearly independent
- ▷ **Cue:** need to prove linear independence of a structured set $\longrightarrow$ **Move:** take a relation of minimal length and manufacture a strictly shorter nontrivial one
- ▷ **Cue:** elements that pairwise commute or anticommute $\longrightarrow$ **Move:** conjugate a relation by one of them: each term is multiplied by ± 1; add or subtract to kill the fixed terms
- ▷ **Cue:** a relation where one term is special (the identity, a constant, the trivial character) $\longrightarrow$ **Move:** first multiply through so that the special term appears, then use an operation fixing it
- ▷ **Cue:** $AB = \pm BA$ for all pairs $\longrightarrow$ **Move:** square: A^2 commutes with everything, because the signs appear twice and cancel
- ▷ **Cue:** an axiom "either X or $-X$ but not both" $\longrightarrow$ **Move:** apply it with the identity: the set is disjoint from its negative, so 0 is excluded, so any product that vanishes is a contradiction

Layer 3 THE WHY

Why is "n^2 means dimension" not a trick? Because of what the number is. The problem gives you no information about $\mathcal{M}$ that could produce a combinatorial count — no generators, no finiteness, no order. The only object in the statement that has size n^2 is the ambient space $\mathrm{M}_n(\mathbb{R})$. And the only general reason a subset of an n^2-dimensional space has at most n^2 elements is linear independence. When a bound in a problem equals a dimension and nothing else in the problem is that size, the bound *is* the dimension. Reading it any other way means building the number n^2 from scratch out of axioms that never mention it. That is the cathedral the natural attempt starts building.

Why a minimal relation? Linear independence is a universal statement ("every relation is trivial"), and the axioms are local statements about pairs. The standard bridge is to take the worst counterexample in some measurable sense — the shortest relation — and show that pairwise information makes it shorter. This is the same engine that proves distinct characters are linearly independent (Dedekind/Artin) and that eigenvectors for distinct eigenvalues are independent. In the eigenvector proof you apply T and subtract λ_1 times the relation; here you conjugate by C and subtract. In both, you apply an operation that multiplies each term by a scalar depending on the term, then subtract to kill one term while keeping another alive. If you have seen the eigenvector proof, this problem is that proof with the eigenvalues ± 1 supplied by (iii).

Why conjugation specifically? Condition (iii) is a statement about AB versus BA, i.e. about $C^{-1}BC$ versus B: $C^{-1}BC = \pm B$. So each element of $\mathcal{M}$ is an eigenvector, with eigenvalue ± 1, of the linear map $X \mapsto C^{-1}XC$ on $\mathrm{M}_n(\mathbb{R})$. That sentence is the whole solution in miniature. The eigenvector-independence argument needs different eigenvalues, and (iv) guarantees that for any non-identity B there is a C separating B (eigenvalue -1) from I (eigenvalue $+1$). The four axioms are not a random list: (iii) makes the conjugation maps diagonal on $\mathcal{M}$, (iv) makes them separate points from I, and (i)–(ii) let you move any element into the I slot and keep the set disjoint from its negative.

Why is $A^2 = \pm I$ needed, and why is it true? It is needed because you cannot conjugate or normalise with a matrix you cannot invert. It is true for the reason the fourth critical thought gives: in $A^2B = \varepsilon^2 BA^2$ the sign appears twice. The step "$SB = BS$ and $SB = -BS$ give $SB = 0$" is where the "but not both" in (ii) earns its place. Without it, $\mathcal{M}$ could contain 0, and the zero matrix commutes and anticommutes with everything. The phrase "but not both" is a quiet way of saying $0 \notin \mathcal{M}$, and it is worth decoding explicitly the moment you read it.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the number n^2 in the conclusion. Write down: "$\dim \mathrm{M}_n(\mathbb{R}) = n^2$; aim to prove $\mathcal{M}$ linearly independent." The second is "but not both" in (ii): write down $-B \notin \mathcal{M}$, $0 \notin \mathcal{M}$.

2. A minimal-length linear relation, shortened by an operation that scales terms by ± 1: here, conjugation by an anticommuting element. This is the eigenvector-independence proof in another setting.

3. Hesitation occurs at the normalisation: the student has a minimal relation, sees that conjugation fixes commuting terms, but does not see how to get a term guaranteed to be fixed. It is reinforced away by asking "which element commutes with everything?" — the identity — and then "how do I force I into the relation?" — multiply by an inverse. That in turn forces the lemma $A^2 = \pm I$.

4. The trap is the group structure. $\mathcal{M} \cup -\mathcal{M}$ looks so much like a finite group that students reach for group orders or representation theory. That route assumes finiteness and invertibility, which are not given; it cannot even state the count before the problem is solved. A secondary trap is forgetting to check that the normalised B_i are still distinct, which is exactly where $\mathcal{M} \cap -\mathcal{M} = \varnothing$ is used a second time.

5. Whenever the claimed bound equals the dimension of the space the objects live in, and nothing else in the problem has that size, the problem is asking for linear independence. Whenever a family of operators acts diagonally on your objects with scalars that separate them, independence follows by the minimal-relation argument.

6. Reusable decision rule: *if a set must have at most d elements and d is the dimension of its ambient space, prove linear independence; do so by taking a shortest relation and applying a linear operation under which each element is an eigenvector, choosing the operation so that two terms of the relation get different eigenvalues, then subtracting.*