CMM 2026 Team Round Solutions

CALTECH MATH MEET

January 24, 2026

1 Problem 1

Problem 1. Call a binary string *encompassing* if by erasing some digits from the string, one can obtain any binary string of length 2026.

(a) Find, with proof, the minimum possible length N of an encompassing string.

(b) Find, with proof, the number of length N encompassing strings (where N is the answer to part (a) of the question).

Proposed by Justin Lee

1.1 Solution

(a) First, observe that any encompassing string must must at least 2026 0's and at least 2026 1's. This forces $N \geq 4052$. Now we shall show that $N = 4052$ is possible by showing that the string $\boldsymbol{x} = 1010\cdots10$ is an encompassing string. Indeed, suppose we wish to obtain a string $a_1a_2\cdots a_{2026}$ from $\boldsymbol{x}$. Let x_n denote the n^{th} digit of the string $\boldsymbol{x}$. Then, for each $1 \leq i \leq 2026$, exactly one of x_{2i-1} or x_{2i} equals a_i. If we erase the digit among $\{x_{2i-1}, x_{2i}\}$ which is *not* equal to a_i, then we obtain a string that is exactly $a_1a_2\cdots a_{2026}$.

(b) We claim that there are 2^{2026} encompassing strings, namely the ones $\boldsymbol{x}$ that satisfy $\{x_{2i-1}, x_{2i}\} = \{0,1\}$ for every $1 \leq i \leq 2026$. Notice that these strings are encompassing, for exactly the same reason as in part (a): to obtain any string $a_1a_2\cdots a_{2026}$, one can erase, for each $1 \leq i \leq 2026$, the number out of $\{x_{2i-1}, x_{2i}\}$ that does not equal a_i. It remains to show that every length 4052 encompassing string must satisfy $\{x_{2i-1}, x_{2i}\} = \{0,1\}$. Note that in order to be able to obtain the string $0\cdots01\cdots1$ where there are i consecutive 0's followed by $2026-i$ consecutive 1's, one must have that the i^{th} occurrence of a 0 in $\boldsymbol{x}$ must be before the $i+1^{\text{st}}$ occurrence of a 1 in $\boldsymbol{x}$. Similarly, in order to be able to obtain the string $1\cdots10\cdots0$, the i^{th} occurrence of a 1 must be before the $i+1^{\text{st}}$ occurrence of a 0. This means that in the first $2i$ digits, we cannot have more than i 0s, else the $i+1^{\text{th}}$ 0 would be among the first $2i$ digits of the string whereas the i^{th} 1 would not; similarly, we cannot have more than i 1s among the first $2i$ digits. Hence, for every $1 \leq i \leq 2026$, the first $2i$ digits contain exactly i 0s and exactly i 1s, which implies $\{x_{2i-1}, x_{2i}\} = \{0,1\}$, as desired.

2 Problem 2

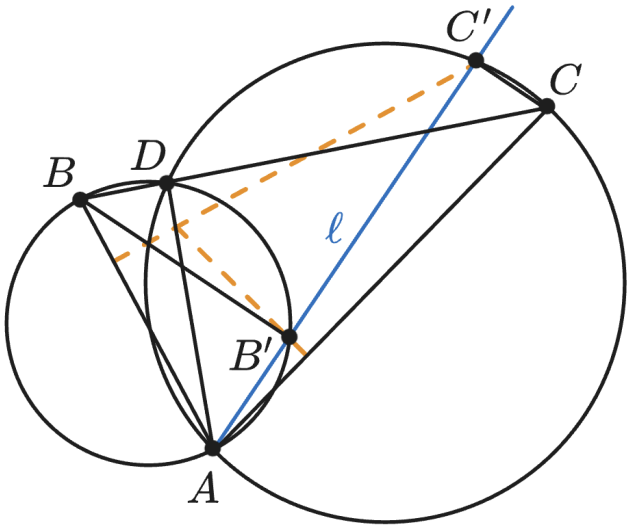
Problem 2. In acute scalene triangle ABC, the foot of the altitude from A is D. ℓ is a line passing through A, and the feet of the altitudes from B and C to ℓ are B' and C', respectively. Show that the line through B' perpendicular to AC and the line through C' perpendicular to AB intersect on AD.

Proposed by Vivian Loh

2.1 Solution

We will show that both lines intersect AD on the circumcircle of $\triangle DB'C'$. First of all, the circles with diameters AB and AC intersect at D, and their intersections with ℓ other than A are points B' and C' respectively. Consider the intersection point of the line through B' that is perpendicular to AC with line AD; let this point be P. Then $\angle C'B'P = 90^\circ - \angle C'AC$. Since AC is a diameter, this implies that $\angle C'B'P = 180^\circ - \angle ADC'$, and that P is on the circumcircle of $\triangle DB'C'$.

We can make a similar argument for the intersection Q of the line through C' that is perpendicular to AB with AD also being on the circumcircle of $\triangle DB'C'$: $\angle QC'B' = 90^\circ - \angle BAC' = \angle ABB' = \angle ADB' = \angle QDB'$. Thus, both the perpendicular line passing through B' and the perpendicular line passing through C' intersect AD on the circumcircle of $\triangle DB'C'$, and so they must intersect AD at the same point. $\square$



3 Problem 3

Problem 3. Justin and Eduardo play a game. There is a calculator with a display and two buttons: the first button replaces the number x on the display with $\log_2 x$ (unless $x \leq 0$ in which case it displays ERROR), and the second button replaces x with N^x, where N is a fixed positive integer greater than 2. Justin chooses the value of N, and a positive integer to be initially displayed on the calculator screen. What is the smallest integer k such that Eduardo can press a sequence of k buttons that result in the displayed number not being a positive integer, regardless of Justin's choices?

Proposed by Justin Lee

3.1 Solution

We claim that the answer is 5.

Lemma 3.1 (Upper Bound)
Eduardo is always able to obtain a number that is not a positive integer within 5 calculator presses.

Proof. Let the two calculator buttons be denoted E and L, where E for "exponent" is the button that replaces x with N^x, and L for "logarithm" is the button that replaces x with $\log_2 x$. Then we will show that the sequence of buttons $EELLL$ always obtains a number that is not a positive integer.

Let Justin's choices be $\{N, x_0\}$ where the positive integer x_0 is the starting number. Then pressing the E buttons at the beginning always results in positive integers, since N to the power of a positive integer is a positive integer. However, if N is not a power of 2 then pressing the L button after that will result in a number that is not a positive integer, since the value of (non-power of 2) raised to an integer power will never be a power of 2. So if N is not a power of 2, then we already know $EELLL$ will eventually work. Now

we just have to consider the case in which N is a power of 2, say $N = 2^{n_0}$ for some **positive** integer n_0. Then after pressing the buttons EE we have

$$(2^{n_0})^{(2^{n_0})^{x_0}} = (2^{n_0})^{(2^{n_0 x_0})} = 2^{n_0 \cdot 2^{n_0 x_0}}.$$

After pressing L we obtain $n_0 \cdot 2^{n_0 x_0}$, and after pressing L again we obtain $\log_2 n_0 + n_0 x_0$. If $\log_2 n_0$ is not an integer then this is not an integer, and if $\log_2 n_0$ *is* an integer then the only way for this sum to be a power of 2 is for each piece ($\log_2 n_0$ and $n_0 x_0$) to contain the same number of factors of 2, i.e. the largest power of 2 that divides $\log_2 n_0$ is equal to the largest power of 2 that divides $n_0 x_0$. But this is impossible, since $2^{\log_2 n_0} \mid n_0 x_0$ but $2^{\log_2 n_0} \nmid \log_2 n_0$. So there is no way for $\log_2 n_0 + n_0 x_0$ to be a power of 2, and thus pressing the L button for the third time obtains a number that is not a positive integer.

Thus we have proven that regardless of the value of N and the initial positive integer displayed on the screen, during the process of pressing $EELLL$ Eduardo will obtain a number that is not a positive integer. □

Lemma 3.2 (Lower Bound)
There exists a choice of $\{N,$ initial number to be displayed on the calculator$\}$ such that every sequence of 4 calculator presses made by Eduardo yields a positive integer.

Proof. We will show that $\{N, x_0\} = \{2^{65536}, 65536\}$ works, where the first number is the value of N that Justin chooses and the second number is the initial positive integer to be displayed that Justin chooses. (Note that $65536 = 2^{16}$.)

We just have to verify that any string of 4 or fewer button presses made by Eduardo will keep the displayed number a positive integer. Note that if there is a sequence of button presses which obtains a not-(positive integer), then we can remove E's from the end and it will still result in a non-(positive integer). This is because N to the power of a positive integer is clearly a positive integer. Thus, we only have to ensure that all strings of length ≤ 4 **ending in** L result in a positive integer for $\{N, x_0\} = \{2^{65536}, 65536\}$.

We can just manually check all of the following strings, and **expedite the process of checking by noting that the sequence** NL **is the same as multiplication by** 65536**, while the sequence** LN **is the same as taking the number to the** 65536**'th power**:

- L: works because $65536 = 2^{16}$
- NL: $65536 \cdot 65536$
- LL: works because $65536 = 2^{16}$
- NNL: (some power of N) $\cdot 65536$
- NLL: $\log_2(65536 \cdot 65536) = \log_2(2^{32}) = 32$
- LNL: $\log_2(65536^{65536})$ = integer.
- LLL: works because $65536 = 2^{16}$
- $NNNL$: NN clearly gives an integer, and then we are just multiplying that by 65536.
- $NNLL$: Pressing N gives $2^{65536 \cdot 65536}$, and pressing NL after that gives $65536 \cdot 2^{65536 \cdot 65536} = 2^{16+65536^2}$, which is clearly a power of 2 so pressing L afterwards results in a positive integer.
- $NLNL$: Multiplying by 65536 and then multiplying again by 65536 clearly returns an integer.
- $NLLL$: Pressing NL results in $65536 \cdot 65536 = 2^{32}$, pressing L after that results in 32, and pressing L one more time results in 5.
- $LNNL$: A integer power of 65536 multiplied to 65536 is clearly an integer.
- $LNLL$: Pressing L results in 16, and then pressing NL results in $16 \cdot 65536 = 2^{4+16}$, which is a power of 2 so pressing L after that clearly results in a positive integer.

- $LLNL$: Pressing LL results in 4, and then pressing NL after that results in $4 \cdot 65536$, which is clearly an integer.
- $LLLL$: works because $65536 = 2^{16}$.

Thus, if Justin chooses $\{N, x_0\} = \{2^{65536}, 65536\}$ then any sequence of ≤ 4 button presses made by Eduardo results in the printed number being a positive integer. □

The first lemma (Upper Bound) gives that the smallest integer k is ≤ 5, and the second lemma (Lower Bound) gives that it is ≥ 5, so the answer is 5.

4 Problem 4

Problem 4. Let S be a set of 2026 points in 3D space placed by Ana, such that no three points in S lie on a line. Ana chooses a starting point P_0 in S. Then Ana and Bob play a game on these points, with Bob moving first.

(a) Bob chooses a point P_1 in S distinct from P_0 and draws the segment $L_1 = \overline{P_0P_1}$.

(b) Ana then chooses a point P_2 in S distinct from P_1 and draws the segment $L_2 = \overline{P_1P_2}$, such that the lengths satisfy $|L_2| > |L_1|$.

(c) The players continue to alternate: if the previous segment drawn was $L_k = \overline{P_{k-1}P_k}$, the next player chooses a point P_{k+1} in S distinct from P_k such that $|L_{k+1}| = |\overline{P_kP_{k+1}}| > |L_k|$.

(d) A player who cannot make a valid move loses.

Over all sets S of 2026 points that Ana can choose, what is the maximum number of starting points $P_0 \in S$ that she can pick so that she will have a winning strategy?

Proposed by Luke Jin

4.1 Solution

Lets prove that for n points, there exists at most one point can be the point with winning strategy for Ana. Specifically, lets induct on n. $n = 0, 1$: It is an automatic winning of Bob in the case $n = 0$, and Ana in the case $n = 1$.
Lets assume that the hypotheses hold for $0 \leq n < k$.
Now, given an appropriate configuration S of $n = k$, lets define the sets $\partial S, S^\circ$. Define

$$\operatorname{diam}(S) = \max_{s_1, s_2 \in S} \{d(s_1, s_2)\} > 0$$

$$\partial S := \{s \in S | \exists t \in S : d(s, t) = \operatorname{diam}(S)\}, S^\circ := S \setminus \partial S$$

In other words, ∂S is the set of points which are included in any of the longest line segments.
Also, note that $|\partial S| \geq 2$.
Case 1. Ana starts on a point $s \in \partial S$:
There exists $t \neq s \in \partial S : d(s, t) = \operatorname{diam}(S) > 0$.
If Bob chooses point t, by the maximality of diameter, there are no more points that Ana can move on, so Ana loses.
Case 2. Ana starts on a point $s \in S^\circ$:
From case 1, we can observe that in the firstmost moment when some player reaches to some $s \in \partial S$, the opponent can choose $t \neq s \in \partial S : d(s, t) = \operatorname{diam}(S) > 0$ and win. The maximality of diameter invalidates further moves and even guarantees the validness of the current move, so the opponent wins. So, both players should work on S° until they don't have anymore valid points inside it. Then, this game is equivalent to the game on S°, and $|S^\circ| \leq |S| - 2$. Thus, the inductive hypothesis implies that Ana can win on at most one point.

Summing up case 1 and 2, there is at most one possible starting point that Ana can win.
If 2025 points form a regular polygon and the left point is the center of that polygon, starting from the center obviously makes Ana win, so the maximum number of winning starting points is $\boxed{1}$.

5 Problem 5

Problem 5. Find all functions $f : \mathbb{N} \to \mathbb{N}$ satisfying $f(n+1)^2 - f(n)^2 = 2f(f(n)) + 1$ for all $n \in \mathbb{N}$. Here, $\mathbb{N}$ denotes the set of positive integers (greater than or equal to 1).

Proposed by Justin Lee

5.1 Solution

Written with significant contributions from Yoll (Gurt) Feng.

We claim that the answer is $f(n) = cn + \frac{c^2-1}{2}$ where c is some odd positive integer. It is easy to verify that these functions satisfy the equation.

Note that $f(n+1) > f(n)$ for all $n \in \mathbb{N}$. If we let $c(n) = f(n+1) - f(n)$, then the equation can be rewritten as

$$f(f(n)) = c(n)f(n) + \frac{c(n)^2 - 1}{2}$$

or equivalently

$$\frac{f(f(n)) - f(n)}{f(n) - n} = (c(n) - 1) \cdot \frac{f(n) + \frac{c(n)+1}{2}}{f(n) - n}$$

Observe that if $c(n) > 1$ and $f(n) > n + 1$ then

$$\begin{aligned} & f(f(n)) - f(n) > (c(n) - 1)(f(n) - n) + 1 \\ \implies & f(f(n)) - f(n+1) = f(f(n)) - f(n) - c(n) > (f(n) - n - 1)(c(n) - 1) \end{aligned}$$

or in other words, $\sum_{i=n+1}^{f(n)-1} c(i) > (c(n) - 1)(f(n) - n - 1)$. Hence, there exists some $n < m < f(n)$ such that $c(m) \geq c(n)$. Hence we have shown

Proposition. If n is such that $f(n) > n + 1$, then there exists m with $n < m < f(n)$ such that $c(m) \geq c(n)$.

Note that unless $f(n) \equiv n$, the condition $f(n) > n + 1$ must hold for all sufficiently large n. In particular, we obtain an infinite (nonconstant) sequence of n such that the values of $c(n)$ are nondecreasing.

Case 1. If $c(n)$ is bounded, then $c(n)$ must attain some value, say c, an infinite number of times. Let us suppose we have chosen the largest such c with this property. Then, by the above, c must be the maximum value of $c(n)$. This yields an infinite sequence a_i such that $f(f(a_i)) = cf(a_i) + \frac{c^2-1}{2}$. For any $i < j$, note that

$$\sum_{b=f(a_i)}^{f(a_j)-1} c(b) = f(f(a_j)) - f(f(a_i)) = c \cdot (f(a_j) - f(a_i))$$

If $c = 1$ then we immediately deduce $c(b) = 1$ for all $f(a_i) < b < f(a_j) - 1$, and hence $c(b) = 1$ for all sufficiently large b. The only such function satisfying the original equation is $f(n) = n$. On the other hand, if $c > 1$, then we see that $f(b) > b + 1$ is satisfied for any $b > f(a_2)$, and by the proposition, we must have $c(b) \leq c$, otherwise we would have an infinite number of n such that $c(n) > c$, which will contradict the maximality of c. Combining this with the above equation yields $c(b) = c$ for all b with $f(a_2) \leq f(a_i) \leq b < f(a_j)$, and hence $c(b) = c$ for all sufficiently large b. This implies $f(b) = cb + \frac{c^2-1}{2}$ for sufficiently large b (since it holds for $b = f(a_i)$). Now we can show that this in fact holds for all $b \in \mathbb{N}$. Indeed, if not, consider the maximal b for which this does not hold. Then, it holds for $b + 1$ and for $f(b)$,

and the original equation would imply that it must hold for b, a contradiction. Hence, the possible functions arising from this case are $f(n) = cn + \frac{c^2-1}{2}$ for some odd c.

Case 2. If $c(n)$ is unbounded, then we shall first show the following:

Lemma. If n and k are positive integers such that $f(n) > kn + \frac{k^2-1}{2}$, then $c(n) > k$.

Proof. Suppose otherwise, so that $c(n) = k$ and $f(n) > kn + \frac{k^2-1}{2}$. Then we know that $f(f(n)) = kf(n) + \frac{k^2-1}{2}$, and hence $\frac{f(f(n))-f(n)}{f(n)-n} < k$. Consequently, there exists some n_1 with $n \le n_1 < f(n)$ such that $c(n_1) < k$. Choosing the smallest such n_1, we note that $f(n_1) > kn_1 + \frac{k^2+1}{>} c(n_1)n_1 + \frac{c(n_1)^2+1}{2}$ and hence iterating this yields an infinite sequence n_i with $c(n_i)$ decreasing, which is impossible. This proves the lemma.

A consequence of the lemma is that if $f(n) \ge kn + \frac{k^2-1}{2}$, then $f(m) \ge km + \frac{k^2-1}{2}$ for all $m \ge n$. Indeed, this follows from $c(m) > k-1$ and hence $c(m) \ge k$ for all $m \ge n$.

Now for any $n > 1$, the corollary to the lemma applied to $f(n-1)$ and $k = c(n-1)$ tells us that since $f(f(n-1)) = kf(n-1) + \frac{k^2-1}{2}$, then we must have $c(n)f(n) + \frac{c(n)^2-1}{2} = f(f(n)) \ge kf(n) + \frac{k^2-1}{2}$ and hence $c(n) \ge c(n-1)$.

We claim that in the $c(n)$ is unbounded case, we can strengthen this to $c(n) > c(n-1)$. Indeed, suppose $c(n) = c(n-1) = k$. Then, $f(f(n-1)) = kf(n-1) + \frac{k^2-1}{2}$ and $f(f(n)) = kf(n) + \frac{k^2-1}{2}$, yet for every $f(n-1) \le m < f(n)$ we have $c(m) \ge c(n-1) = k$, so equality must hold: that is, $c(m) = k$ and $f(m) = km + \frac{k^2-1}{2}$. Applying this argument again to $f(n) - 1$ and iterating will yield an infinite sequence of numbers satisfying $f(n) = kn + \frac{k^2-1}{2}$, and now the monotonicity of $c(n)$ forces $c(n)$ to be constant, i.e., $f(n)$ is linear, as encompassed by Case 1. Hence, in Case 2, we must have $c(n-1) < c(n)$ for all $n > 1$.

Now it follows that for sufficiently large n, we have $f(n) > 100n$, and furthermore

$$f(f(n)) \ge f(0) + 1 + 2 + \cdots + f(n) - 1 > \frac{f(n)^2 - f(n)}{2}$$

which implies $f(n+1)^2 - f(n)^2 = 2f(f(n)) + 1 > f(n)^2 - f(n)$ and hence $f(n+1) > 1.4f(n)$ for sufficiently large n. It follows in particular that

$$f(n+1)^2 > f(f(n)) \ge f(n+1) \times 1.4^{f(n)-n-1} \implies f(n+1) > 1.4^{0.99f(n)}$$

but this means $f(f(n)) > f(n+2) > 1.4^{0.99f(n+1)} > f(n+1)^2$, which contradicts the original equation. This concludes the proof, and thus the solutions we found in Case 1 are all solutions.

6 Problem 6

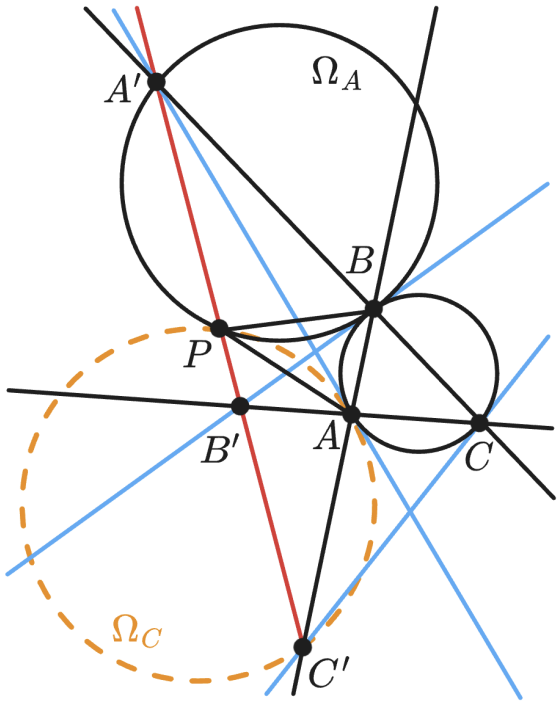
Problem 6. Given acute scalene triangle ABC with circumcircle Ω, points A', B', and C' are chosen on lines BC, AC, and AB respectively such that AA', BB', and CC' are tangent to Ω. Ω_A is the circle passing through A' which is tangent to Ω at B, Ω_B is the circle passing through B' that is tangent to Ω at C, and Ω_C is the circle passing through C' that is tangent to Ω at A. Prove that Ω_A, Ω_B, and Ω_C share a point.

Proposed by Vivian Loh

6.1 Solution

Lemma 6.1
A', B', and C' lie on a line.

Proof. By Menelaus's Theorem, this is true if and only if $\frac{BA'}{A'C} \cdot \frac{CB'}{B'A} \cdot \frac{AC'}{C'B} = 1$. Because of the tangent lines AA', BB', and CC', we have $\triangle A'AB \sim \triangle A'CA$ (and the analogous similarities $\triangle B'BA \sim \triangle B'CB$ and $\triangle C'CA \sim \triangle C'BC$). So, $\frac{BA'}{A'C} = \left(\frac{AB}{AC}\right)^2$. Thus, the cyclic product is $\left(\frac{AB}{AC}\right)^2 \cdot \left(\frac{AC}{BC}\right)^2 \cdot \left(\frac{BC}{AB}\right)^2$, which is clearly equal to 1. □



Let the line through A', B', and C' be ℓ. We will show that Ω_A, Ω_B, and Ω_C intersect at a point on ℓ.

First, we will focus only on Ω_A and Ω_C, and show that they intersect on ℓ (see the diagram below). Let Ω_A intersect ℓ at point P; we will show that P is also on Ω_C. Note that $P \in \Omega_C \iff \angle APC' = \frac{\widehat{AC'}}{2}$, and because of the tangency between Ω_C and (ABC) at A, $\widehat{AC'}$ of Ω_C has the same measure as $\widehat{AB}$ of (ABC). So $P \in \Omega_C \iff \angle APC' = \angle ACB$, which is equivalent to $PA'CA$ being cyclic. Since B' is on the radical axis BB' of Ω_A and (ABC), Power of a Point at B' gives that $PA'CA$ is cyclic. So indeed, $P \in \Omega_C$, and Ω_A and Ω_C intersect on ℓ.

By the same logic, Ω_C and Ω_B intersect on ℓ. Thus, all three circles intersect at a point on ℓ. □

7 Problem 7

Problem 7. Let p be a prime, and let $\mathbb{Z}/p\mathbb{Z} = \{0, 1, \ldots, p-1\}$ be the set of residues modulo p. For any two 2026-tuples $a = (a_1, \cdots, a_{2026}), b = (b_1, \cdots, b_{2026}) \in (\mathbb{Z}/p\mathbb{Z})^{2026}$, and an integer $1 \le i \le 2026$, set $g_i(a,b) \equiv a_i + b_i + \sum_{j=1}^{i-1} a_j b_{i-j} \mod p$ and $g(a,b) = (g_1(a,b), \cdots, g_{2026}(a,b)) \in (\mathbb{Z}/p\mathbb{Z})^{2026}$. Show that for any $p \neq 2027$, there exists a function

$$f : (\mathbb{Z}/p\mathbb{Z})^{2026} \to \mathbb{Z}/p\mathbb{Z}$$

such that for any $a, b \in (\mathbb{Z}/p\mathbb{Z})^{2026}$, we have that

$$f(g(a,b)) \equiv f(a) + f(b) + \sum_{i=1}^{2026} a_i b_{2027-i} \mod p.$$

Proposed by Samuel Goodman

7.1 Solution

Let $a = (a_1, \cdots, a_{n-1}), b = (b_1, \cdots, b_{n-1}) \in (\mathbb{Z}/p\mathbb{Z})^{n-1}$ and define $g(a,b) = (g_1(a,b), \cdots, g_{n-1}(a,b))$. We prove the much stronger claim that $f(g(a,b)) \equiv f(a) + f(b) + \sum_{i=1}^{n-1} a_i b_{n-i} \mod p$ has a solution if $p \nmid n$ (in fact this is an iff). We start with two lemmas:

Lemma 7.1
For a partition p_j of j, let $\ell(p_j)$ be the length of the partition, let p_{ji} be the number of occurrences of i in the partition, and let $m(p_j)$ be the number of permutations of the partition. Let P_j be the set of partitions of a positive integer j ($P_{j,k}$ for those with length k) and for p_j a partition of j. Then we have that $\log(1 + \sum_{i=1}^{\infty} a_i T^i) = \sum_{n=1}^{\infty} (\sum_{p_n \in P_n} (-1)^{\ell(p_n)} \frac{m(p_n)}{\ell(p_n)} \prod_{i=1}^{n} a_i^{p_{ni}}) T^n$.

Proof. To start, recall the calculus fact that $\log(1+x) = \sum_{m=1}^{\infty} (-1)^m \frac{x^m}{m}$. From this, note that

$$\log(1 + \sum_{i=1}^{\infty} a_i T^i) = \sum_{m=1}^{\infty} \frac{(-1)^m}{m} (\sum_{e_1, \cdots, e_m} (\prod_{i=1}^{m} a_{e_i}) T^{e_1 + \cdots + e_m})$$

Fixing an exponent $e_1 + \cdots + e_m = n$, the formal sum rearranges as

$$\sum_{n=1}^{\infty} (\sum_{k=1}^{n} \frac{(-1)^k}{k} \sum_{e_1 + \cdots + e_k = n} \prod_{m=1}^{k} a_{e_m}) T^n$$

Now letting f_i be the number of occurrences of i among the e_ms, the product can be written $\prod_{m=1}^{j} a_i^{f_i}$ for each tuple $e_1, \cdots, e_k$. The exponents f_i satisfy $\sum_{i=1}^{n} f_i = k$ and $\sum_{i=1}^{k} i f_i = n$, giving a corresponding partition p of j with f_i occurrences of i so that $\ell(p) = k$ and $m(p)$ the number of ways in which the product can occur. Thus breaking down $\sum_{e_1 + \cdots + e_k = n} \prod_{m=1}^{k} a_{e_m}$ by partitions gives

$$\sum_{p_n \in P_{n,k}} m(p_n) a_i^{p_{ni}}$$

Finally, summing over k, corresponding to partition lengths, gives the desired coefficient for T^j. □

Let $r_n(x_1, \cdots, x_n) = \sum_{p_n \in P_n} (-1)^{\ell(p_n)} \frac{m(p_n)}{\ell(p_n)} \prod_{i=1}^{n} x_i^{p_{ni}}$.

Lemma 7.2
For each partition $p_n \in P_n$, the rational number $\frac{m(p_n)}{\ell(p_n)}$ cannot have reduced denominator divisible by any $q \nmid n$.

Proof. Choose some $q | \ell(p_n)$. For simplicity, let $\ell = \ell(p_n)$. We prove that $q^{v_q(\ell(p_n))} | m(p_n)$. Indeed, consider the set T_{p_n} of all distinct permutations of the partition p_n. Let σ be the operation which takes a list $(s_1, \cdots, s_\ell)$ and reorders it as $(s_2, s_2, \cdots, s_\ell, s_1)$. We break T_{p_n} into subsets where two elements are in the same subset if one is obtained from the other by iterating σ.

We claim that the size of each such subset S is divisible by $q^{v_q(n)}$. Fix a given element $p' = (p'_1, \cdots, p'_\ell)$ of S. Let m be the least positive integer such that $\sigma^m(p') = p'$. We claim that S has exactly m elements. Indeed it has at most m elements because $\sigma^k(p') = \sigma^{k \mod m}(p')$ and has at least m since if $\sigma^j(p') = \sigma^k(p')$ for some $1 \le j < k \le m$, then $p'_{i+j} = p'_{i+k}$ for all i, where indices are taken $\mod \ell$, but this means that $p'_{i+k-j} = p'_i$ for all i and thus $k - j \ge m$, contradiction.

Next let m' be the number of $1 \le j \le \ell$ such that $\sigma^j(p') = p'$. We claim that $mm' = \ell$. Indeed, note that if $\sigma^j(p') = p'$, then $p'_{i+j} = p'_i$ for all i. We claim that this happens precisely when $m | j$. Note that if $m | j$, then $p'_{i+j} = p'_i$, meaning j works. If $m \nmid j$, then writing $j = mk + r$ for some $1 \le r < m$, we get $p'_{i+r} = p'_i$ or $\sigma^r(p') = p'$, contradicting the minimality of m. Thus there are $\frac{\ell}{m}$ working values up to ℓ and so $m' = \frac{\ell}{m}$.

Finally, we show that $q \nmid m'$. Suppose that $q | m'$. Write $m' = qq'$ and so $\ell = qq'm$ and thus $\frac{\ell}{q} = q'm$.

Thus $m|\frac{\ell}{q}$ and so $\sigma^{\frac{\ell}{q}}(s') = s'$, which implies that s' breaks into q equal blocks of size $\frac{\ell}{q}$. But then the sum of the elements in s', which by definition is n, would be a multiple of q, contradicting $q \nmid n$.

Thus q cannot divide m' and so $q^{v_q(\ell(p_n))}|m$. But the set T_{p_j} is a disjoint union of such sets S, and thus $q^{v_q(\ell(p_n))}||T_{p_n}| = m(p_n)$, as desired. □

Let $z_i = x_i + y_i + \sum_{j=1}^{i-1} x_j y_{i-j}$. Then we claim that in the polynomial ring $\mathbb{Q}[x_1, \cdots, x_n, y_1, \cdots, y_n]$, we have the identity $r_n(x_1, \cdots, x_n) + r_n(y_1, \cdots, y_n) = r_n(z_1, \cdots, z_n)$. Indeed, first note by additivity of the formal logarithm that

$$\log(1 + \sum_{i=1}^{\infty} x_i T^i) + \log(1 + \sum_{i=1}^{\infty} y_i T^i) = \log(1 + \sum_{i=1}^{\infty} (x_i + y_i + \sum_{j=1}^{i-1} x_j y_{i-j}) T_i) = \log(1 + \sum_{i=1}^{\infty} z_i T^i)$$

The claim then follows from Lemma 7.1.

Since $p \nmid n$, Lemma 7.2 implies that the polynomial $f(x_1, \cdots, x_n) \in \mathbb{Q}[x_1, \cdots, x_n]$ has all coefficients with reduced denominators coprime to p, and so we can view it as an element in $(\mathbb{Z}/p\mathbb{Z})[x_1, \cdots, x_n]$. Then considering $r_n(x_1, \cdots, x_n) - x_n$ and invoking Lemma 7.1, removing the term corresponding to the partition $\{n\}$ of n, corresponding to the unique partition of n containing n, the resulting polynomial is now a function of $x_1, \cdots, x_{n-1}$, and so set $h(x_1, \cdots, x_{n-1}) = r_n(x_1, \cdots, x_n) - x_n$. Then we have that

$$r_n(x_1 + \cdots, x_n) + r_n(y_1 + \cdots, y_n) = r_n(z_1, \cdots, z_n) \in \mathbb{Q}[x]$$

It follows that

$$h(x_1, \cdots, x_{n-1}) + x_n + h(y_1, \cdots, y_{n-1}) + y_n \equiv h(z_1, \cdots, z_{n-1}) + z_n \mod p$$

Since $z_n = x_n + y_n + \sum_{i=1}^{n-1} x_i y_{n-i}$, we see that

$$h(x_1, \cdots, x_{n-1}) + h(y_1, \cdots, y_{n-1}) \equiv h(z_1, \cdots, z_{n-1}) + \sum_{i=1}^{n-1} x_i y_{n-i} \mod p$$

and so substituting $x_i = a_i, y_i = b_i$, we see that $z_i = g_i(a, b)$, and thus $-h$ provides a solution, as desired.