

The Competition on the Mathematics of Information 2025

The Chinese University of Hong Kong

1 March, 2025

SOLUTIONS

Question	Points
1	70
2	70
3	70
4	70
Total	280

Question 1: Sums of powers

Let x, y, z be real numbers that satisfy

$$\begin{aligned}x + y + z &= a, \\x^2 + y^2 + z^2 &= b^2, \\x^3 + y^3 + z^3 &= c^3,\end{aligned}$$

for some real numbers a, b, c .

- (a) Find the value of $x^4 + y^4 + z^4$ in terms of a, b, c .

Define the following notation:

$$\begin{aligned}T_1 &= x^4 + y^4 + z^4 \\T_2 &= x^3y + x^3z + xy^3 + zy^3 + xz^3 + yz^3 \\T_3 &= x^2y^2 + x^2z^2 + y^2z^2 \\T_4 &= x^2yz + xy^2z + xyz^2.\end{aligned}$$

Then we have

$$\begin{aligned}(x + y + z)^4 &= T_1 + 4T_2 + 6T_3 + 12T_4 \\(x^2 + y^2 + z^2)^2 &= T_1 + 2T_3 \\(x^2 + y^2 + z^2)(x + y + z)^2 &= T_1 + 2T_2 + 2T_3 + 2T_4 \\(x + y + z)(x^3 + y^3 + z^3) &= T_1 + T_2.\end{aligned}$$

Then we have the following linear equations:

$$\begin{aligned}T_1 + 4T_2 + 6T_3 + 12T_4 &= a^4 \\T_1 + 2T_3 &= b^4 \\T_1 + 2T_2 + 2T_3 + 2T_4 &= a^2b^2 \\T_1 + T_2 &= ac^3\end{aligned}$$

Solving for T_1 we obtain

$$T_1 = \frac{8ac^3 + 3b^4 + a^4 - 6a^2b^2}{6}.$$

(b) Let $S_n = x^n + y^n + z^n$, for integers $n \geq 4$. Show that

$$S_n = p_1(a, b, c)S_{n-1} + p_2(a, b, c)S_{n-2} + p_3(a, b, c)S_{n-3},$$

for some (multivariate) polynomials $p_1(a, b, c), p_2(a, b, c), p_3(a, b, c)$. Furthermore, determine the three polynomials.

An alternate way to do the first part is first to observe that x, y, z satisfy: $xy + yz + xz = \frac{a^2 - b^2}{2}$, and

$$xyz = (x^3 + y^3 + z^3 - (x + y + z)(x^2 + y^2 + z^2 - xy - yz - xz))/3 = \frac{2c^3 - 3ab^2 + a^3}{6}.$$

Therefore x, y, z are solutions of the cubic polynomial

$$t^3 - at^2 + \frac{a^2 - b^2}{2}t - \frac{2c^3 - 3ab^2 + a^3}{6} = 0.$$

Therefore each of x, y, z satisfy

$$t^n - at^{n-1} + \frac{a^2 - b^2}{2}t^{n-2} - \frac{2c^3 - 3ab^2 + a^3}{6}t^{n-3} = 0.$$

Summing them up, we obtain that

$$S_n = aS_{n-1} - \frac{a^2 - b^2}{2}S_{n-2} + \frac{2c^3 - 3ab^2 + a^3}{6}S_{n-3}.$$

Therefore

$$p_1(a, b, c) = a, \quad p_2(a, b, c) = \frac{b^2 - a^2}{2}, \quad p_3(a, b, c) = \frac{2c^3 - 3ab^2 + a^3}{6}.$$

Question 2: Guess the number

In a number guessing game, your task is to correctly guess an integer x in the range $3, 4, \dots, n$, where $n \geq 4$ is a given integer (you initially know n , but not x). You are allowed to ask two kinds of questions:

- A **threshold question** is in the form “is $x \geq a$?”. You are allowed to choose any integer a when you ask such a question. You will receive an answer “yes” if $x \geq a$, or “no” if $x < a$.
- The **smallest prime factor question** is the question “what is the smallest prime factor of x ?”. Let $f(x)$ be the smallest prime factor of x . You will receive the value of $f(x)$ as the answer.

Your goal is to guess x correctly after asking the fewest questions. An **m -question strategy** is a strategy that guarantees a correct guess after asking at most m questions. You are allowed to choose which question to ask next based on the answers to your previous questions.

For example, when $n = 6$, one strategy is to first ask the smallest prime factor question. If the answer is 3, you guess $x = 3$. If the answer is 5, you guess $x = 5$. If the answer is 2, you further ask whether $x \geq 6$. If yes, you guess $x = 6$. If no, guess $x = 4$. This is a 2-question strategy since you can guarantee a correct guess after at most $m = 2$ questions.

- a) For $n = 11$, what is the minimum number of questions needed (i.e., what is the smallest m such that there exists an m -question strategy)? Justify your assertion. (To justify your assertion, you should explain your strategy, and argue that there is no strategy that requires fewer questions.)

3. First ask the smallest prime factor:

- 2: The possibilities are 4,6,8,10. You need 2 more questions to find out the number.
- 3: The possibilities are 3, 9. You need 1 more question.
- 5: Guess 5.
- 7: Guess 7.
- 11: Guess 11.

We now argue that 3 is optimal. If you do not use the smallest prime factor question, then asking m questions gives you at most 2^m combinations of answers, and hence you can only tell 2^m numbers apart from each other. Now there are 9 possibilities, which means you need at least $\lceil \log_2 9 \rceil = 4$ questions if you do not use the smallest prime factor question. If you use the smallest prime factor question, then you still need to tell 4,6,8,10 apart from each other since they all have smallest prime factor 2. This requires $\lceil \log_2 4 \rceil = 2$ more questions. Hence the optimal number of questions is 3.

- b) Now consider general n , and you must ask the smallest prime factor question first and exactly once (i.e., your strategy must begin with the smallest prime factor question). Find the minimum number of questions needed, in terms of n . Justify your assertion.

Hint: You may express your answer in terms of the floor function $\lfloor t \rfloor$ (which is the greatest integer that is not greater than t), the ceiling function $\lceil t \rceil$ (which is the smallest integer that is not smaller than t), and logarithms. Note that $\lceil \log_2 t \rceil$ is the smallest integer a such that $t \leq 2^a$.

$$\lceil \log_2 \lfloor (n-2)/2 \rfloor \rceil + 1 = \lceil \log_2 (n-3) \rceil \text{ if } n \geq 5, 1 \text{ if } n = 4.$$

- c) Now consider general n , and you must ask the smallest prime factor question **last** and exactly once (i.e., you must ask the smallest prime factor question, and you must guess x right after the smallest prime factor question). Find the minimum number of questions needed, in terms of n . Justify your assertion.

Same answer $\lceil \log_2 \lfloor (n-2)/2 \rfloor \rceil + 1 = \lceil \log_2(n-3) \rceil$ if $n \geq 5$, 1 if $n = 4$.

- d) Now consider general n , and suppose you cannot ask the smallest prime factor question. Your goal now is not to guess x , but to guess $f(x)$ (the smallest prime factor of x) within a distance 1, that is, after the questions, you say an integer q , and you are considered correct if $|f(x) - q| \leq 1$ (i.e., the absolute difference between $f(x)$ and q is at most 1). Find the minimum number of questions needed, in terms of n . Justify your assertion.

Let $n = 6y + z$, $z \in \{0, \dots, 5\}$. The answer is:

$\lceil \log_2(4y - 1) \rceil$ if $z = 0$,

$\lceil \log_2(4y) \rceil$ if $z = 1$,

$\lceil \log_2(4y + 1) \rceil$ if $z = 2, 3, 4$,

$\lceil \log_2(4y + 2) \rceil$ if $z = 5$.

Question 3: Sequences with maximum pairwise distance

We say that (a_1, a_2, \dots, a_n) is a binary sequence of length n if each entry a_i is either 0 or 1. Given two binary sequences $\vec{a} = (a_1, a_2, \dots, a_n)$ and $\vec{b} = (b_1, b_2, \dots, b_n)$, their distance is equal to the number of positions in which they differ, i.e., the number of $1 \leq i \leq n$ where $a_i \neq b_i$. In other words,

$$d(\vec{a}, \vec{b}) = \sum_i |a_i - b_i|.$$

For instance, the distance between $(0, 0, 0, \dots, 0)$ and $(1, 1, 1, \dots, 1)$ is n because the two sequences differ in every index. The distance between binary sequences $\vec{a} = (a_1, a_2, \dots, a_n)$ and $\vec{b} = (b_1, b_2, \dots, b_n)$ is denoted by $d(\vec{a}, \vec{b})$. Note that $d(\vec{a}, \vec{b})$ is an integer between 0 and n .

- a) Let n be an integer divisible by 3. Find three sequences \vec{a}_1 , \vec{a}_2 and \vec{a}_3 of length n such that the minimum pairwise distance between them is maximized, i.e., find $\vec{a}_j = (a_{j1}, a_{j2}, \dots, a_{jn})$ for $j = 1, 2, 3$ such that

$$\min_{t \neq k} d(\vec{a}_t, \vec{a}_k) = \min(d(\vec{a}_1, \vec{a}_2), d(\vec{a}_1, \vec{a}_3), d(\vec{a}_2, \vec{a}_3))$$

is maximized.

Note that

$$\begin{aligned} d(\vec{a}_1, \vec{a}_2) + d(\vec{a}_1, \vec{a}_3) + d(\vec{a}_2, \vec{a}_3) &= \sum_i (|a_{1i} - a_{2i}| + |a_{1i} - a_{3i}| + |a_{2i} - a_{3i}|) \\ &\leq 2n. \end{aligned}$$

Thus,

$$\min(d(\vec{a}_1, \vec{a}_2), d(\vec{a}_1, \vec{a}_3), d(\vec{a}_2, \vec{a}_3)) \leq \frac{2n}{3}.$$

On the other hand, for the three sequences

$$\begin{aligned} \vec{a}_1 &= \underbrace{0, 0, \dots, 0}_{\frac{n}{3}}, \underbrace{1, 1, \dots, 1}_{\frac{n}{3}}, \underbrace{0, 0, \dots, 0}_{\frac{n}{3}} \\ \vec{a}_2 &= \underbrace{1, 1, \dots, 1}_{\frac{n}{3}}, \underbrace{0, 0, \dots, 0}_{\frac{n}{3}}, \underbrace{0, 0, \dots, 0}_{\frac{n}{3}} \\ \vec{a}_3 &= \underbrace{0, 0, \dots, 0}_{\frac{n}{3}}, \underbrace{0, 0, \dots, 0}_{\frac{n}{3}}, \underbrace{1, 1, \dots, 1}_{\frac{n}{3}} \end{aligned}$$

the minimum distance is $2n/3$. Therefore, the answer is $2n/3$.

- b) Assume that we have r sequences \vec{a}_i for $i = 1, 2, 3, \dots, r$ such that the minimum pairwise distance between them is at least c , i.e.,

$$\min_{u \neq w} d(\vec{a}_u, \vec{a}_w) \geq c.$$

Show that

$$r \left(\sum_{j=0}^t \binom{n}{j} \right) \leq 2^n$$

where $t = \lfloor \frac{c-1}{2} \rfloor$ and $\binom{n}{j} = n!/(j!(n-j)!)$ is the binomial coefficient.

Hint: For any sequence \vec{a}_u , $\sum_{j=0}^t \binom{n}{j}$ is the number of binary sequences whose distance from \vec{a}_u is at most t . This counts all possible ways of selecting up to t positions out of n positions to flip (change from 0 to 1, or 1 to 0).

For every \vec{a}_u , let \mathcal{B}_u be the set of sequences of distance at most t from \vec{a}_u . The set \mathcal{B}_u can be

viewed as a “ball” of radius t centered at \vec{a}_u . We have

$$|\mathcal{B}_u| = \sum_{j=0}^t \binom{n}{j}$$

If we have r sequences, we can create r such balls. If any two balls overlap, then the sequences would not satisfy the minimum distance d . Therefore, the total number of sequences occupied by these balls must not exceed the number of available codewords 2^n , leading to the desired inequality.

c) Let n be an integer divisible by 5. Find three sequences \vec{a}_1 , \vec{a}_2 and \vec{a}_3 of length n such that

$$\min(2d(\vec{a}_1, \vec{a}_2), d(\vec{a}_1, \vec{a}_3), d(\vec{a}_2, \vec{a}_3))$$

is maximized. Note the factor 2 in front of $d(\vec{a}_1, \vec{a}_2)$.

Note that given \vec{a}_1 , \vec{a}_2 and \vec{a}_3 , if we XOR the three sequences (bitwise) with some sequence \vec{b} , their pairwise distance will not change. If we take $\vec{b} = \vec{a}_3$, this shows that without loss of generality, we can assume that $\vec{a}_3 = (0, 0, 0, \dots, 0)$. For $j, k \in \{0, 1\}$, let $n_{j,k}$ be the number of indices $i \in \{1, 2, \dots, n\}$ where $\vec{a}_{1i} = j$ and $\vec{a}_{2i} = k$. Then, we have the following:

$$n_{0,0} + n_{0,1} + n_{1,0} + n_{1,1} = n$$

$$d(\vec{a}_1, \vec{a}_3) = n_{1,0} + n_{1,1} \tag{1}$$

$$d(\vec{a}_2, \vec{a}_3) = n_{0,1} + n_{1,1} \tag{2}$$

$$d(\vec{a}_1, \vec{a}_2) = n_{1,0} + n_{0,1} \tag{3}$$

Thus, we wish to maximize

$$\min(2(n_{1,0} + n_{0,1}), n_{0,1} + n_{1,1}, n_{1,0} + n_{1,1})$$

subject to

$$n_{0,0} + n_{0,1} + n_{1,0} + n_{1,1} = n$$

It is clear that setting $n_{0,0} = 0$ is optimal. Because of symmetry, let us assume that $n_{0,1} \geq n_{1,0}$. Note that if $n_{0,1} - n_{1,0} > 2$, we can reduce $n_{0,1}$ by one and increase $n_{1,0}$ by one. This will increase the minimum. Thus, two cases may occur for the optimizer: $n_{0,1} = n_{1,0}$ or $n_{0,1} = n_{1,0} + 1$.

If $n_{0,1} = n_{1,0}$: in this case, we wish to maximize

$$\min(4n_{1,0}, n_{1,0} + n_{1,1})$$

subject to

$$2n_{1,0} + n_{1,1} = n.$$

Equivalently, we wish to maximize

$$\min(4n_{1,0}, n - n_{1,0})$$

The solution here is to set $n_{1,0} = n/5$. The value becomes $4n/5$.

If $n_{0,1} = n_{1,0} + 1$: in this case, we wish to maximize

$$\min(4n_{1,0} + 2, n_{1,0} + n_{1,1})$$

subject to

$$2n_{1,0} + n_{1,1} = n - 1.$$

Equivalently, we wish to maximize

$$\min(4n_{1,0} + 2, n - n_{1,0} - 1)$$

One can see that the value of the above maximization does not exceed $4n/5$ because one cannot have

$$4n_{1,0} + 2 > n/5$$

and

$$n - n_{1,0} - 1 > n/5$$

meaning that the maximum does not occur in this case.

- d) Assume that we have r sequences \vec{a}_i for $i = 1, 2, 3, \dots, r$ such that the pairwise distance between any two of them is exactly k . Moreover, each sequence \vec{a}_i has exactly h ones and $n - h$ zeros. Assuming that $2h^2 - 2nh + nk \geq 0$, show that

$$r \leq \frac{nk}{2h^2 - 2nh + nk}.$$

Let m_i be the number of sequences that have one in their i -th coordinate. Note that $\sum_{i=1}^n m_i = rh$ is the total number of ones in all sequences.

The sum of all the pairwise distances can be computed in two ways: on the one hand, it equals

$$k \binom{r}{2}.$$

On the other hand, it equals

$$\sum_{i=1}^n m_i(r - m_i) = r \sum_{i=1}^n m_i - \sum_{i=1}^n m_i^2 = r^2h - \sum_{i=1}^n m_i^2.$$

Thus, equating the two values, we get

$$r^2h - \frac{kr(r-1)}{2} = \sum_{i=1}^n m_i^2.$$

By the Cauchy-Schwartz inequality we have

$$n \sum_{i=1}^n m_i^2 \geq \left(\sum_{i=1}^n m_i \right)^2 = r^2h^2$$

. Thus, we obtain

$$nr^2h - \frac{nkr(r-1)}{2} \geq r^2h^2.$$

This simplifies to

$$2nrh - nk(r-1) \geq 2rh^2$$

or

$$nk \geq 2rh^2 - 2nrh + nkr.$$

It implies the following inequality:

$$r \leq \frac{nk}{2h^2 - 2nh + nk}.$$

Question 4: Functional Equation

Find all functions $g : [0, 1] \mapsto \mathbb{R}$ that satisfy all the following conditions:

- (a) g is continuous.
- (b) $g(0) = g(1) = 0$.
- (c) $g(x) = g\left(\frac{x}{2}\right) + \left(1 - \frac{x}{2}\right)g\left(\frac{x}{2-x}\right)$, for all $x \in [0, 1]$.
- (d) $g(x) = g(1-x)$ for all $x \in [0, 1]$.

Remark: This is a functional equation characterizing the binary entropy function using the grouping axiom.

$g(x) = 0$ is the only solution.

Putting $x = 1$ in the third condition, we obtain that $g\left(\frac{1}{2}\right) = 0$

we have

$$g(x) = g\left(\frac{x}{2}\right) + \left(1 - \frac{x}{2}\right)g\left(\frac{x}{2-x}\right). \quad (4)$$

Note that if g is a solution of the functional equation, then $-g$ is also a solution.

Define $x_0 = \sup\{x : g(y) = 0 \ \forall y : 0 \leq y \leq x\}$. If $x_0 \geq \frac{1}{2}$, we are done (then $g(x) = 0$ on $[0, 1]$). If $g(x)$ is not the zero-function, we must have (due to $g(x) = g(1-x)$), $0 \leq x_0 < \frac{1}{2}$. By continuity of $g(x)$, note that $g(x_0) = 0$. Define $x_1 = \inf\{x : g(x) = 0, x > x_0\}$. Note that $g(x_1) = 0$ and $x_0 < x_1 \leq \frac{1}{2}$. By construction, either $g(x)$ or $-g(x)$ must be positive in the interval (x_0, x_1) . W.l.o.g., let us assume that it is $g(x)$. Therefore $g(x) \geq 0$ for $x \in [0, x_1]$.

When $x = x_1$ in (4), we have

$$0 = g(x_1) = g\left(\frac{x_1}{2}\right) + \left(1 - \frac{x_1}{2}\right)g\left(\frac{x_1}{2-x_1}\right).$$

As $\frac{x_1}{2-x_1} < x_1$, and $g(x) \geq 0$ for $x \in [0, x_1]$, we must have $g\left(\frac{x_1}{2}\right) = g\left(\frac{x_1}{2-x_1}\right) = 0$. However, from construction of x_0, x_1 , this also implies that $\frac{x_1}{2}, \frac{x_1}{2-x_1} \leq x_0$. Observe that for any $x < x_1$, we have $\frac{x}{2} < \frac{x_1}{2} \leq x_0$ and $\frac{x}{2-x} < \frac{x_1}{2-x_1} \leq x_0$. From (4), and that $g(x) = 0$ for $x \in [0, x_0]$, we obtain $g(x) = 0$ for any $x_0 < x < x_1$. This contradicts our assumption that $g(x)$ is not the zero function.