

International Baccalaureate

November 2022 Examination Session

Extended Essay

Mathematics

An Investigation on IMO'1988 Problem 6 and Its Extended Generalized Form

Personal code: jqz749

Word count: 3285

Research question: What are the properties of the generalized extension of IMO'1988 problem 6 which includes more than 2 variables?

Table of Contents

1. Introduction	3
2. Solution for IMO 1988 Problem 6.....	5
3. Solutions to the Extended Problem for 3 Variables	10
4. Solutions to the Extended Problem for n Variables.....	19
5. Conclusion	24
6. Bibliography.....	25
7. Appendices	26

1. Introduction

As a student who has been preparing for mathematical olympiads since middle school, I have known about a popular olympiad problem for a long time: the 6th problem of International Mathematical Olympiads (IMO) in 1988. A mathematical paper that I read recently, pushed me to investigate this problem and its extensions for my Extended Essay.

The original statement that was given the students which needed to be proven in the IMO is:

“If a, b and $q = \frac{a^2 + b^2}{ab + 1}$ are integers, then q is a perfect square.”

This famous problem that was submitted by the Federal Republic of Germany to the problem committee, was marked as “a very difficult problem” when given to the jury, as A. Engel states in his book *Problem Solving Strategies* (Engel 127). Since then, this problem interested many mathematicians and was subjected to many mathematical papers.

Apart from many different solutions using many different ideas of this problem, there are also different extensions to this problem that have been investigated, such as

$\frac{a^2 + b^2 + 1}{ab + 1}$, $\frac{a^2 + b^2}{ab - 1}$, $\frac{a^2 + b^2 + c^2}{abc}$ (Vandervelde). While reading about this problem, I came across

another extension was suggested at the end of the paper *Another Step Further...On a Problem of the 1988 IMO* (Laukó et al.). In that paper, the authors push the readers to investigate the properties of the equality where $q, a, b, c \in \mathbb{Z}$:

$$q = \frac{a^2 + b^2 + c^2}{ab + bc + ca + 1}$$

This problem intrigued me very much and I decided to seek the (a, b, c) solution triplets that made q an integer. So, in this essay I mainly seek solutions to 3 different problems: the first one is the original IMO problem, the second one is the extension problem right above and the third problem is the ultimate generalized problem that also covers the previous two problems:

$$I_n(x_1, x_2, \dots, x_n) = \frac{\sum_{i=1}^n x_i^2}{\sum_{1 \leq i < j \leq n} x_i x_j + 1}$$

I solve these problems in a logical order, denoting my observations and various corollaries. In all of the problems; a formed chain will be observed in the process of finding solutions, and the properties of this chain will be investigated afterwards using similar techniques for the problems.

2. Solutions to the IMO 1988 Problem 6

Approaching this problem, my initial aim will mainly be finding all solutions to the problem (not only proving the statement in the IMO) since my extension problems include finding solutions. Later on, I will simultaneously observe that finding all solutions to the IMO'1988 equation proves what the problem actually wants: the fact that q is a perfect square if it is an integer.

For simplicity, let's define a function $f: [\mathbb{Z}_0^+]^2 \rightarrow \mathbb{Z}$ (\mathbb{Z}_0^+ denotes non-negative integers throughout this essay):

$$f(x, y) = \frac{x^2 + y^2}{xy + 1} = q$$

I aim to find the integer pairs which satisfy $q \in \mathbb{Z}_0^+$.

Initial Observations

- $f(x, y) = f(y, x)$ so, it is possible to only observe the functions where $x \leq y$, and ignore the other cases for clearer calculations.
- $f(0, m) = m^2$ where $m \in \mathbb{Z}_0^+$.
- If $f(x, y) = 0 \Rightarrow x^2 + y^2 = 0 \Rightarrow x = y = 0$
- If $f(x, y) = 1 \Rightarrow x^2 + y^2 = xy + 1 \geq 2xy \Rightarrow 1 \geq xy \Rightarrow x = 0, y = 1$ or $x = 1, y = 1$

This means it is enough to observe $f(x, y) \geq 2$ to find integer solutions. From now on, $f(x, y) \geq 2$ will be assumed.

Let's try to find more pairs where $f(x, y) \in \mathbb{Z}^+$.

As a first step, I will call (x, y) pairs that satisfy $f(x, y) = n$ for a constant $n \geq 2$ integer “the solution pairs for n ”. Since it is already known that $f(0, m) = m^2$, I start by observing the solution pairs for m^2 .

$f(x, m) = m^2$ can be written as a second-degree polynomial where x is the variable: It can have 2 roots at most (x_1 and x_2). Since $x_1 = 0$, finding x_2 should lead me to another solution pair.

$$m^2 = f(x, m) = \frac{x^2 + m^2}{1 + xm} \Rightarrow m^2 + xm^3 = x^2 + m^2 \Rightarrow x^2 - m^3x = 0 \Rightarrow x(x - m^3) = 0$$

The roots are $x_1 = 0$ and $x_2 = m^3$.

$$f(0, m) = f(m, m^3) = m^2$$

Let’s solve a similar equation for $f(x, m^3) = m^2$ where x is the variable in order to find another solution.

$$\begin{aligned} m^2 = f(x, m^3) &= \frac{x^2 + m^6}{1 + xm^3} \Rightarrow x^2 - xm^5 + m^6 - m^2 = 0 \\ &\Rightarrow x(x - m^5 + m) - m(x - m^5 + m) = (x - m)(x - m^5 + m) \end{aligned}$$

The roots are $x_1 = m$ and $x_2 = m^5 - m$.

$$f(0, m) = f(m, m^3) = f(m^3, m^5 - m)$$

Using the same procedure, it is obtained:

$$f(0, m) = f(m, m^3) = f(m^3, m^5 - m) = f(m^5 - m, m^7 - 2m^3) = m^2$$

Looking at the elements of the pairs, a forming chain can be seen: the first entry of a solution pair is the second entry of the preceding pair. This chain of numbers can be denoted as: $0, m, m^3, m^5 - m, m^7 - 2m^3$. Now, it should be observed if these numbers follow any rules.

It is possible to guess the following pattern:

$$m^3 = m^2 \cdot m - 0$$

$$m^5 - m = m^2 \cdot m^3 - m$$

$$m^7 - 2m^3 = m^2 \cdot (m^5 - m) - m^3$$

In other words, if $a_0 = 0, a_1 = m$ and $a_i = m^2 \cdot a_{i-1} - a_{i-2}$ for $i \in \{2,3,4\}$

$$f(a_0, a_1) = f(a_1, a_2) = f(a_2, a_3) = f(a_3, a_4) = m^2$$

This is a hint about a rule for generating other solution pairs for a constant integer from a solution pair, can be found.

Rule 1.1: Let us define an infinite sequence of integers: $a_0, a_1, a_2 \dots$ where a_0, a_1 are constant non-negative integers and $q = f(a_0, a_1)$ is an integer. The rest of the terms could be found by the recursive formula:

$$a_k = q \cdot a_{k-1} - a_{k-2} \text{ for } k \geq 3$$

$$\Rightarrow f(a_i, a_{i+1}) = q \text{ for } i \in \{0,1,2, \dots\}.$$

Proof: The steps of induction can be seen:

i) $f(a_0, a_1) = q$ holds.

ii) Assume $f(a_{i-1}, a_i) = q$ holds for $i \in \{0,1,2, \dots\}$;

iii) $q = \frac{qa_i}{a_i} = \frac{a_{i+1}+a_{i-1}}{a_i} = \frac{a_{i+1}^2 - a_{i-1}^2}{a_i(a_{i+1} - a_{i-1})} = \frac{(a_i^2 + a_{i+1}^2) - (a_{i-1}^2 + a_i^2)}{(1+a_i a_{i+1}) - (1+a_{i-1} a_i)} =$

$$f(a_{i-1}, a_i) = \frac{a_i^2 + a_{i+1}^2}{1 + a_i a_{i+1}}$$

$$\left(\frac{X - Z}{Y - T} = \frac{Z}{T} \Leftrightarrow \frac{X}{Y} = \frac{Z}{T} \right)$$

$$\Rightarrow \frac{a_{i-1}^2 + a_i^2}{1 + a_{i-1} a_i} = \frac{a_i^2 + a_{i+1}^2}{1 + a_i a_{i+1}} \Rightarrow f(a_{i-1}, a_i) = f(a_i, a_{i+1}) = q \blacksquare.$$

The motivation for this proof can clearly be seen if the proof is done backwards.

However, this ordering seems more elegant.

Corollaries:

1.1 A chain of integers is obtained “ $a_0, a_1, a_2 \dots$ ” where $f(a_i, a_{i+1}) = q$ for $i \in \{0, 1, 2, \dots\}$. The “initial term” of this chain is evidently a_0 and the chain has infinitely many terms.

1.2 If integers a_0 and a_1 , that satisfy $f(a_0, a_1) = q$, are taken, a chain can be constructed using those constants: a_0, a_1, a_2, \dots . This means every solution pair for an integer q is within a chain.

1.3 The roots of $f(a_i, x) = q$ for variable x are a_{i-1} and a_{i+1} for any integer $i \geq 1$. Since the mentioned equation is of second degree, there exists at most two roots so a_{i-1} and a_{i+1} are the only ones.

1.4 There are infinitely many solution pairs. When $a_0 = 0$ there is a chain for every $a_i \in \mathbb{Z}$ that is chosen.

In summary, all the solution pairs for an integer q is in a chain, and some of them has 0 as the first term of the chain (e.g., 0, 2, 8, 30 ...) Now it will be proven that all the chains have 0 as the first term. So far, greater terms in the chain were being searched, approaching infinity. Now smaller terms in the chain will be searched in order to figure if 0 can always be reached.

The motivation for the following proof comes from Corollary 1.3, it is known that

$x_1 = a_{i-1} \leq a_i$ and $x_2 = a_{i+1} > a_i$. In other words, solving the equation $f(a_i, x) = q$ with respect to x we find one root smaller than a_i for sure. This may mean that smaller terms that can be added to the chain can be found similarly:

To prove that, let's take any solution pair (a, b) and $f(a, b) = q \in \mathbb{Z}^+$ ($a, b \in \mathbb{Z}^+$). Solve the equation $f(a, x) = q$ with respect to x :

$$\frac{a^2 + x^2}{ax + 1} = q \Rightarrow a^2 + x^2 = aqx + q = 0 \Rightarrow x^2 - aqx + a^2 - q = 0$$

$$\Rightarrow \Delta = a^2q^2 - 4(a^2 - q) \Rightarrow x_1 = \frac{aq + \sqrt{a^2q^2 - 4(a^2 - q)}}{2}$$

(It is evident that $a^2q^2 - 4(a^2 - q) \geq 0$ because $q \geq 2$)

It is also clear $x_1 \geq \frac{aq}{2} \geq a$. It is found that one root that is greater than or equal to a .

Now, let us observe if x_2 is non-negative and $x_2 < a$.

$$x_2 = \frac{aq - \sqrt{a^2q^2 - 4(a^2 - q)}}{2}$$

$$(q \geq 2) \Rightarrow 8a^2 \leq 4a^2q \Rightarrow 8a^2 < 4a^2q + 4q$$

$$\Rightarrow 8a^2 + (a^2q^2 - 4a^2q - 4a^2) < 4a^2q + 4q + (a^2q^2 - 4a^2q - 4a^2)$$

$$\Rightarrow a^2q^2 - 4a^2q + 4a^2 = a^2(q - 2)^2 < a^2q^2 - 4(a^2 - q)$$

$$\Rightarrow a(q - 2) < \sqrt{a^2q^2 - 4(a^2 - q)}$$

$$\Rightarrow aq - \sqrt{a^2q^2 - 4(a^2 - q)} < 2a$$

$$\Rightarrow x_2 = \frac{aq - \sqrt{a^2q^2 - 4(a^2 - q)}}{2} < a$$

It is also seen that $\frac{a^2 + x_2^2}{ax_2 + 1} = q \in \mathbb{Z}^+$ so $ax_2 + 1 \geq 0 \Rightarrow x_2 \geq 0$. This means that another solution pair (x_2, a) is found. Since $a > x_2 \geq 0$, a smaller term can be added to this aforementioned chain. The chain can now be: $x_2, a, b \dots$. It is clear that the same

procedure can be implemented to every 2 consecutive terms in the chain to find a smaller term to eventually reach 0.

∴ Every chain must begin with a zero.

This also means the elements of solution sets (x, y) for a constant $q \in \mathbb{Z}^+$ exist in one distinct chain. The first element of the chain must be 0.

As planned before, the original IMO'88 Problem 6 is solved in the meantime: Assuming the second element of the chain is $k \in \mathbb{Z}_0^+$ (the first element is always 0), every pair in the chain will give k^2 as the output ($f(0, k) = k^2$) because every consecutive integer pair in a chain gives the same value when plugged into the function f . This means every solution pair for an integer q gives a perfect square(Laukó et al.). For example, if the chain is 0, 3, 27, 240 ... this means that $3^2 = f(0,3) = f(3,27) = f(27,240) = \dots$

3. Solutions to the Extended Problem for 3 Variables

As it can be recalled, the newly generated expression is $q = \frac{x^2 + y^2 + z^2}{xy + yz + zx + 1}$. In this section, I will be looking into the properties of the (x, y, z) triplets which make q an integer. This expression is merely the original problem with three variables, the similarity between them can be seen if $z = 0$ is plugged in:

$$\frac{x^2 + y^2 + 0^2}{x \cdot y + y \cdot 0 + 0 \cdot x + 1} = \frac{x^2 + y^2}{xy + 1}$$

Therefore, the IMO'1988 problem is the special case of this extended problem and I can use similar steps to investigate this expression too.

Similarly, let's define a function g so that $g: [\mathbb{Z}_0^+]^3 \rightarrow \mathbb{Z}$ and

$$g(x, y, z) = \frac{x^2 + y^2 + z^2}{xy + yz + zx + 1}$$

Initial observations:

- $g(x, y, z) = g(y, z, x) = g(x, z, y) = \dots$ Since the expression is symmetric, I can only look into the solution triplets which satisfy $x \leq y \leq z$ and ignore the others without loss of generalization.
- For $q = 0$: $\frac{x^2+y^2+z^2}{xy+yz+zx+1} = 0 \Rightarrow x^2 + y^2 + z^2 = 0 \Rightarrow x = y = z = 0$
- The only case that satisfies $x = y = z$ is described above. If there were any other solution triplet, let it be $x = y = z = k$ where $k \neq 0$ so $g(x, y, z) = \frac{3k^2}{3k^2+1}$. Since $1 < 3k^2 < 3k^2 + 1$, $g(x, y, z)$ can never be an integer. $\rightarrow\leftarrow$
- For $q = 1$: $\frac{x^2+y^2+z^2}{xy+yz+zx+1} = 1 \Rightarrow x^2 + y^2 + z^2 = xy + yz + zx + 1$

$$\Rightarrow 2x^2 + 2y^2 + 2z^2 - 2xy - 2yz - 2zx = 2$$

$$\Rightarrow (x - y)^2 + (x - z)^2 + (z - y)^2 = 2$$

Since $z \geq y \geq x$,

$$\checkmark \text{ If } x \neq y \neq z \Rightarrow (x - y)^2 + (x - z)^2 + (z - y)^2 \geq 3 \rightarrow\leftarrow$$

$$\checkmark \text{ If } x = y \Rightarrow z = y + 1$$

$$\checkmark \text{ If } y = z \Rightarrow x = y - 1$$

\therefore Every $(k, k, k + 1)$ and $(k, k + 1, k + 1)$ triplets satisfy the condition ($k \in \mathbb{Z}_0^+$).

- The case of $x = 0$ was explained before, the problem becomes the IMO'88 problem. So, for every (a, b) that satisfy $f(a, b) \in \mathbb{Z}_0^+$, $g(0, a, b)$ is an integer too.

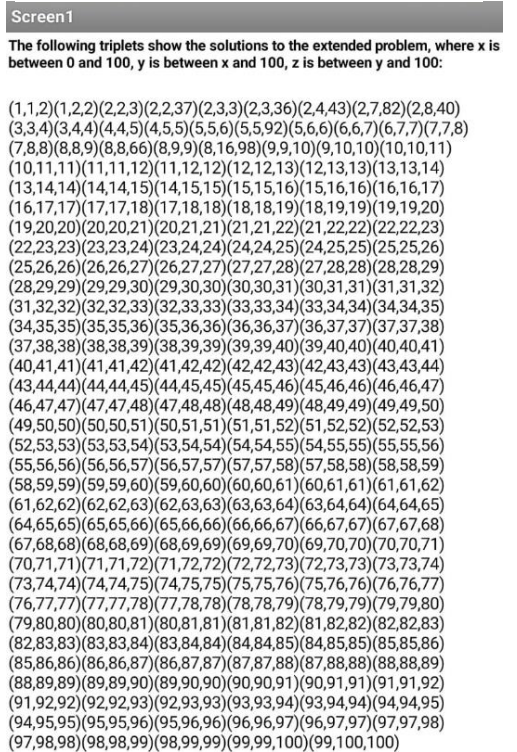
Trying to find how to approach the problem, I figured that finding examples that satisfy the condition is harder than the original IMO problem. This led me into seeking help from technology. From the MIT's coding app, MIT App Inventor, I wrote some codes (see Appendix A) in order to see the solution triplets that make the function an integer. In the code, I defined (x, y, z) integers where $0 < x \leq y \leq z \leq 100$. Choosing 100 as the bound seemed enough to observe primitive patterns. The $x = 0$ case is already observed so I didn't include it in the code.

The most remarkable pattern in the screen (Figure 1) is the existence of such solutions:

$(1,1,2), (1,2,2), (2,2,3), (2,3,3), \dots, (99,100,100)$.

$(k, k, k + 1)$ and $(k, k + 1, k + 1)$ solution triplets ($k \in \mathbb{Z}^+$) are already shown to satisfy the condition, as a part of "initial observations".

Figure 1:



However, I can see a couple more triplets that are not in this pattern: (2,3,36), (2,4,43), (5,5,92), (8,16,98) ...

Changing the computer code I wrote in the MIT App Inventor a little bit to be able to observe more solutions; I could see some more patterns. Let's observe solutions for $0 < x \leq 100, x < y \leq z \leq 300$ (Figure 2).

The existence of the triplets like (2,3,36) and (3,36,271); (2,8,40), (8,40,190) indicates that there can be a chain here too. In order to have a clearer view, other triplets of this potential chain will be found:

$$\frac{40^2 + 190^2 + x^2}{40x + 190x + 40 \cdot 190 + 1} = 4$$

$$40^2 + 190^2 + x^2 = 4 \cdot (40x + 190x + 40 \cdot 190 + 1)$$

$$x^2 - 920x + 7296 = 0$$

$$\Delta = \frac{920^2 - 4 \cdot 7296}{4} \Rightarrow x_1 = 8 \quad x_2 = 912$$

This means now the "chain" is:

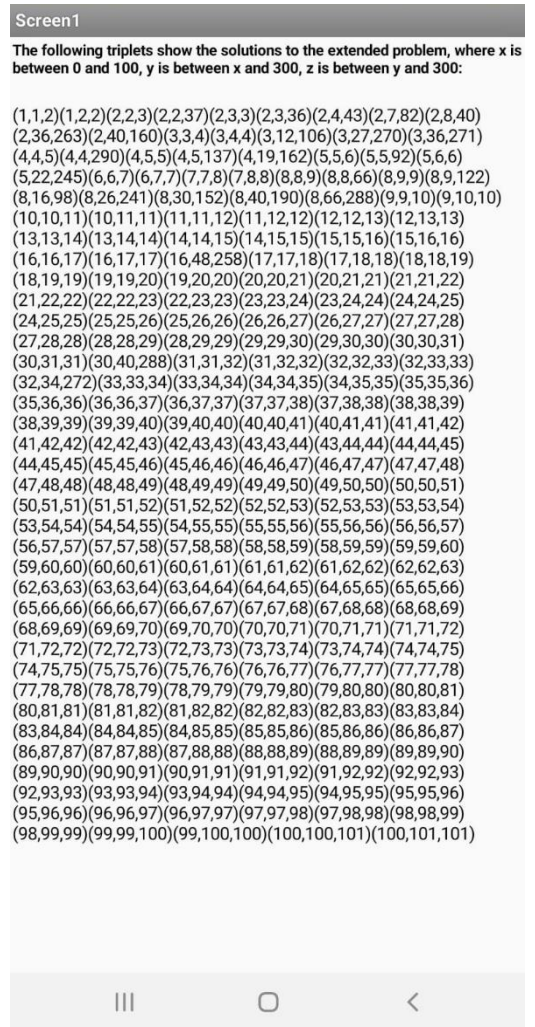
$$2,8,40,190,912.$$

If the similar step is done for $g(x, 2,8) = 4$ another solution triplet can be found: (0,2,8)

The "chain" is now:

$$0,2,8,40,190,912.$$

Figure 2



I could guess the pattern here now, also considering the pattern for the original IMO problem:

$$40 = 4(2 + 8) - 0$$

$$190 = 4(40 + 8) - 2$$

$$912 = 4(190 + 40) - 8$$

Now the following calculations will prove the existence of a chain and also prove the pattern above.

Rule 2.1: Let's define an infinite sequence of integers a_0, a_1, a_2, \dots where a_0, a_1, a_2 are constant non-negative integers that satisfy:

$$g(a_0, a_1, a_2) = q \text{ and } a_k = q(a_{k-1} + a_{k-2}) - a_{k-3} \text{ for } k \geq 3$$

$$\Rightarrow g(a_i, a_{i+1}, a_{i+2}) = q \text{ for every } i \in \{0, 1, 2, \dots\}.$$

PROOF:

The steps of induction:

- i) $g(a_0, a_1, a_2) = q$ holds.
- ii) Assume $g(a_i, a_{i+1}, a_{i+2}) = q$ holds for $i \in \{0, 1, 2, \dots\}$.
- iii) Let's prove $g(a_{i+1}, a_{i+2}, a_{i+3}) = q$

$$\begin{aligned} q &= \frac{q(a_{i+2} + a_{i+1})}{a_{i+2} + a_{i+1}} = \frac{a_i + a_{i+3}}{a_{i+2} + a_{i+1}} = \frac{a_{i+3}^2 - a_i^2}{a_{i+2}a_{i+3} + a_{i+1}a_{i+3} - a_i a_{i+2} - a_i a_{i+1}} \\ &= \frac{(a_{i+3}^2 + a_{i+2}^2 + a_{i+1}^2) - (a_{i+2}^2 + a_{i+1}^2 + a_i^2)}{(a_{i+2}a_{i+3} + a_{i+1}a_{i+3} + a_{i+1}a_{i+2} + 1) - (a_i a_{i+2} + a_i a_{i+1} + a_{i+1}a_{i+2} + 1)} \\ &= \frac{a_{i+2}^2 + a_{i+1}^2 + a_i^2}{a_i a_{i+2} + a_i a_{i+1} + a_{i+1}a_{i+2} + 1} \\ &\Rightarrow \frac{a_{i+3}^2 + a_{i+2}^2 + a_{i+1}^2}{a_{i+2}a_{i+3} + a_{i+1}a_{i+3} + a_{i+1}a_{i+2} + 1} = \frac{a_{i+2}^2 + a_{i+1}^2 + a_i^2}{a_i a_{i+2} + a_i a_{i+1} + a_{i+1}a_{i+2} + 1} \end{aligned}$$

$$\Rightarrow g(a_i, a_{i+1}, a_{i+2}) = g(a_{i+1}, a_{i+2}, a_{i+3})$$

Corollaries:

2.1 There appears to be a chain “ $a_0, a_1, a_2 \dots$ ” where $g(a_i, a_{i+1}, a_{i+2}) = q$ for

$i \in \{0, 1, 2, \dots\}$. The “initial term” of this chain is evidently a_0 and the chain has infinitely many terms.

2.2 Taking any solution triplet for an integer q is within a chain.

2.3 The roots of $g(a_{i+1}, a_{i+2}, x) = g(a_{i+1}, a_{i+2}, a_{i+3})$ with respect to x are a_i and a_{i+3} for $i \geq 0$. These are the only roots since the simplified expression is a second-degree polynomial.

2.4 There can be infinitely many solution triplets for a constant q in a chain. There are also infinitely many distinct chains: for every integer k , taking $(a_0, a_1, a_2) = (0, 0, k)$ will produce a new chain.

Just like the solution to the IMO problem 6, chains are found and greater integers can be added to the chain infinitely many times. However; I, then, realized that not all the chains in this extended problem can have 0 as their first term, when smaller terms are sought.

An example can be the chain which (2,43,311) belongs to. When the equation $g(m, 2, 43) = g(2, 43, 311)$ is solved using discriminants, it is found that $m_1 = 4, m_2 = 311$. Since neither of these are smaller than or equal to 2, (2,43,311) triplet must be at the beginning of the chain, it doesn't start with 0.

The following calculations state the condition for an (a, b, c) triplet to be the beginning of the chain:

Assume that (a, b, c) holds $g(a, b, c) = q$ where $a, b, c \in \mathbb{Z}^+$ and $a \leq b \leq c$

Let the solutions to $g(x, a, b) = q$ be x_1 and x_2 . So $x_1 = c > a$.

- i. If $x_2 \leq a$, (a, b, c) can't be the beginning of the chain and (x_2, a, b) would be the beginning of the chain. So, if (a, b, c) is the beginning of the chain, it is crucial that $x_2 > a$.

$$\begin{aligned}
 x_1 &= c = q(a + b) - x_2 \\
 x_2 &= q(a + b) - c = \frac{a^2 + b^2 + c^2}{ab + bc + ca + 1} \cdot (a + b) - c \\
 &= \frac{a^3 + ab^2 + ac^2 + a^2b + b^3 + bc^2 - c - abc - bc^2 - c^2a}{ab + bc + ca + 1} \\
 &= \frac{a^3 + ab^2 + a^2b + b^3 - c - abc}{ab + bc + ca + 1} > a \\
 &\Rightarrow a^3 + ab^2 + a^2b + b^3 - c - abc > a + a^2b + abc + a^2c \\
 &\Rightarrow a^3 + ab^2 + b^3 - a > c(2ab + a^2 + 1) \\
 &\Rightarrow \frac{a^3 + ab^2 + b^3 - a}{2ab + a^2 + 1} > c
 \end{aligned}$$

However, I later recognized that this condition does not cover all the cases. It is claimed here that (x_2, a, b) can't be the beginning of the chain only if $x_2 > a$, ignoring the fact that $x_2 = q(a + b) - c$ might be a negative number and negative numbers are not defined in the domain of the function. It is also possible that even if $x_2 = q(a + b) - c < a$; if $q(a + b) - c < 0$ satisfies, (a, b, c) has to be the beginning of this chain anyways and the proof above does not include this.

An example for the described situation can be $g(2,7,82) = 9$. If this is plugged in the "condition" that is found:

$$\frac{2^3 + 2 \cdot 7^2 + 7^3 - 2}{2 \cdot 2 \cdot 7 + 2^2 + 1} \approx 14 < 82$$

This means that (2,7,82) is not the beginning of the chain according to our condition.

However, I figured that it is not true as shown below.

When another solution triplet is tried to be found such that $(m, 2, 7)$ and $0 \leq m \leq 2$, it can be seen that is not possible because the roots are $m_1 = 82$ and $m_2 = -1$. Since the definition of the function requires the elements of the triplets to be non-negative integers, this solution triplet $(-1, 2, 7)$ is not acceptable. $(2, 7, 82)$ is bound to be the beginning of the chain it belongs to.

ii. This means that additional condition should be found. If

$x_2 = q(a + b) - c < 0$, (a, b, c) is still the beginning of the chain:

$$q(a + b) - c = \frac{a^2 + b^2 + c^2}{ab + bc + ca + 1} \cdot (a + b) - c = \frac{a^3 + ab^2 + a^2b + b^3 - c - abc}{ab + bc + ca + 1} < 0$$

$$\Rightarrow a^3 + ab^2 + a^2b + b^3 - c - abc < 0 \Rightarrow a^3 + ab^2 + a^2b + b^3 < c(1 + ab)$$

$$\Rightarrow \frac{a^3 + ab^2 + a^2b + b^3}{1 + ab} < c$$

Now the overall condition for (a, b, c) to be the beginning of the chain becomes:

$$c < \frac{a^3 + ab^2 + b^3 - a}{2ab + a^2 + 1} \text{ or } \frac{a^3 + ab^2 + a^2b + b^3}{1 + ab} < c$$

As the (2,7,82) triplet satisfies the second condition, it can be said that it is the beginning of the chain.

Solutions being a perfect square

The original IMO problem claimed that if there is an (a, b) non-negative integer pair that made $f(a, b)$ an integer, it also had to be a perfect square. This proof is done with the

fact that every chain starts with 0. In this extended problem it is found that not all the chains can start with 0, this may also mean that not all triplets that make the function g an integer makes it a perfect square.

It is evident that the adjacent triplets from the chains starting with 0, should give a perfect square when plugged in the function. This is obvious because when the first adjacent integer triplet from the chain is plugged into the function g , this problem is the same as the original IMO problem. Since this function gives out an integer, it must be also a perfect square.

However, this is not the case for every triplet that makes the function an integer. In order to find the cases where the output of the function is an integer but not a perfect square, coding is used once again (see Appendix B for the codes). As a result of the computer calculations, I found multiple examples of triplets that yielded an integer but not a perfect square when plugged into the function. In order to observe a potential pattern, the yields were also calculated using a calculator:

$$g(2,3,36) = g(2,4,43) = g(2,36,263) = g(2,43,311) = g(3,12,106) = g(3,36,271) \\ = g(4,19,162) = g(4,43,327) = g(8,9,122) = 7$$

$$g(3,7,312) = 31$$

$$g(4,5,137) = 15$$

$$g(8,10,364) = 20$$

The prevalence of 7 as the outcome is noted; however, no other pattern is observed.

These numbers did not seem to be related and it was deduced that no rule for this criterion can be found.

4. Solutions to the Extended Problem for n Variables

In this section, I decided to extend the problem even further. Considering the two previous functions, I aimed to generate a more general function that will also cover those previously-studied functions. It is seen that the outcomes of the both of the previous functions happen to be a fraction: The numerator stands for a sum of squares and the denominator is the sum of product of 2-combinations. Using this information let's define a function:

$$I_n(x_1, x_2, \dots, x_n) = \frac{\sum_{i=1}^n x_i^2}{\sum_{1 \leq i < j \leq n} x_i x_j + 1} = q$$

With this notation it is clear that there is a connection between this function and the previously investigated functions, f and g : $f = I_2$ and $g = I_3$. Since these functions are the special cases of I_n the aim is reached.

It can be realized that I chose the expression $\sum_{1 \leq i < j \leq n} x_i x_j$ (the sum of products of 2-combinations) in the denominator. However, it should be noted that another distinct function could also be investigated, if I chose to create the function with a cyclic sum.

Let's look into these two possible definitions. Cyclic sum of a function refers to the sum of the functions in which the variables are "cycled through" ("Cyclic Sum"). For example, for a function l , $\sum_{cyc} l(x, y, z, t) = l(x, y, z, t) + l(y, z, t, x) + l(z, t, x, y) + l(t, x, y, z)$. In our case, a cyclic sum would be $\sum_{cyc} x_1 x_2 = x_1 x_2 + x_2 x_3 + x_3 x_4 + \dots + x_{n-1} x_n + x_n x_1$. On the other hand, the sum of product of 2-combinations is $\sum_{1 \leq i < j \leq n} x_i x_j$ which contains all the possible 2-combinations of the terms from the set $\{x_1, x_2 \dots x_n\}$.

This decision step in generating the function I_n was important. I want to show that there were two possible valid pathways that I could take at this step. Both of the distinct functions that I could generate would be a generalized version of the previously-

investigated functions from different perspectives, because the cyclic sums and the sums of products of 2-combination are equal:

$$\begin{aligned} \sum_{cyc} x_1 x_2 &= \sum_{1 \leq i < j \leq 2} x_i x_j = x_1 x_2 \Rightarrow f(x_1, x_2) = \frac{x_1^2 + x_2^2}{x_1 x_2 + 1} = \frac{x_1^2 + x_2^2}{\sum_{1 \leq i < j \leq 2} x_i x_j + 1} \\ &= \frac{x_1^2 + x_2^2}{\sum_{cyc} x_1 x_2 + 1} \end{aligned}$$

$$\begin{aligned} \sum_{cyc} x_1 x_2 &= \sum_{1 \leq i < j \leq 3} x_i x_j = x_1 x_2 + x_1 x_3 + x_2 x_3 \Rightarrow g(x_1, x_2, x_3) = \frac{x_1^2 + x_2^2 + x_3^2}{x_1 x_2 + x_2 x_3 + x_1 x_3 + 1} \\ &= \frac{x_1^2 + x_2^2 + x_3^2}{\sum_{cyc} x_1 x_2 + 1} = \frac{x_1^2 + x_2^2 + x_3^2}{\sum_{1 \leq i < j \leq 3} x_i x_j + 1} \end{aligned}$$

In other words, there were two ways to generalize these functions; however, I decided to move on with the 2-combination version.

Initial Observations

- In this symmetric expression, I may only look into the solution sets which satisfy $x_1 \leq x_2 \leq \dots \leq x_n$ and ignore the others without loss of generalization, just like I did for the functions g and f .
- For $q = 0$: $\Rightarrow x_1^2 + x_2^2 + \dots + x_n^2 = 0 \Rightarrow x_1 = x_2 = \dots = x_n = 0$
- The only case that provides $x_1 = x_2 = \dots = x_n$ for $n \geq 3$ is described above. If there were any other solution set, let it be $x_1 = x_2 = \dots = x_n = k$ where $k \neq 0$ so $I_n(k, k, \dots, k) = \frac{nk^2}{\binom{n}{2}k^2 + 1}$. Since $1 \leq nk^2 < \binom{n}{2}k^2 + 1$, $I_n(k, k, \dots, k)$ can never be an integer. $\rightarrow \leftarrow$
- It can be seen how the functions I_n for different n values are similar to each other. For example, plugging 0 into the function as a variable: $I_n(x_1, x_2, \dots, x_{n-1}, 0) =$

$I_{n-1}(x_1, x_2, \dots, x_{n-1})$. Plugging in multiple zeros in the function, other functions can be obtained: $I_n(x_1, x_2, \dots, x_{n-3}, 0, 0, 0) = I_{n-3}(x_1, x_2, \dots, x_{n-3})$ etc.

Just like the two previous chapters, I will first prove the existence of chains. At first, let us recall the previous chains from the previous functions f and g :

- The chain that was found for the original problem was $a_0, a_1, a_2 \dots$ where a_0 and a_1 are constant non-negative integers that satisfy $f(a_0, a_1) = q \in \mathbb{Z}_0^+$ and $a_k = qa_{k-1} - a_{k-2}$ for $k \geq 2$.
- The chain for the function g was a_0, a_1, a_2, \dots where a_0, a_1, a_2 are constant non-negative integers that satisfy

$$g(a_0, a_1, a_2) = q \in \mathbb{Z}_0^+ \text{ and } a_k = q(a_{k-1} + a_{k-2}) - a_{k-3} \text{ for } k \geq 3$$

Considering these rules, I guessed that the chain for function I_n might be:

a_0, a_1, a_2, \dots (where a_0, a_1, \dots, a_{n-1} are constant non-negative integers and $I_n(a_0, a_1, a_2, \dots, a_{n-1}) = q$) and for $k \geq n$

$$a_k = q(a_{k-1} + a_{k-2} + \dots + a_{k-n+1}) - a_{k-n}$$

Proof: An induction is followed:

- $I_n(a_0, a_1, \dots, a_{n-1}) = q$
- Assume $I_n(a_i, a_{i+1}, \dots, a_{i+n-1}) = q$ holds.
- Let's prove $I_n(a_{i+1}, a_{i+2}, \dots, a_{i+n}) = q$

$$\begin{aligned} q &= \frac{q(a_{i+1} + a_{i+2} + \dots + a_{i+n-1})}{a_{i+1} + a_{i+2} + \dots + a_{i+n-1}} = \frac{a_i + a_{i+n}}{a_{i+1} + a_{i+2} + \dots + a_{i+n-1}} \\ &= \frac{a_{i+n}^2 - a_i^2}{a_{i+n}(a_{i+1} + \dots + a_{i+n-1}) - a_i(a_{i+1} + \dots + a_{i+n-1})} \\ &= \frac{\sum_{l \in \{1, 2, \dots, n\}} a_{i+l}^2 - \sum_{l \in \{0, 1, \dots, n-1\}} a_{i+l}^2}{\sum_{1 \leq k < l \leq n} a_{i+k}a_{i+l} + 1 - \sum_{0 \leq k < l \leq n-1} a_{i+k}a_{i+l} - 1} \end{aligned}$$

$$(= q) = \frac{\sum_{l \in \{0,1,\dots,n-1\}} a_{i+l}^2}{\sum_{0 \leq k < l \leq n-1} a_{i+k} a_{i+l} + 1}$$

$$I_n(a_{i+1}, a_{i+2}, \dots, a_{i+n}) = I_n(a_i, a_{i+1}, \dots, a_{i+n-1}) = q \blacksquare$$

Corollaries:

3.1 I once again could form a chain “ $a_0, a_1, a_2 \dots$ ” where $I_n(a_i, a_{i+1}, \dots, a_n) = q$ for

$i \in \{0,1,2, \dots\}$. The chain has infinitely many terms.

3.2 Taking any solution set for an integer q is within a chain. (Using the same procedure as Corollary 2.2)

3.3 The roots of $I_n(a_{i+1}, \dots, a_{i+n-1}, x) = q$ with respect to x are a_i and a_{i+n} for $i \geq 0$.

These are the only roots since this expression can be simplified to a second-degree polynomial.

3.4 There can be infinitely many solution sets in a chain for a constant q . There are also infinitely many distinct chains: for every integer k , taking $(a_i, a_{i+1}, \dots, a_{i+n-1}) = (0,0, \dots, 0, k)$ will produce a new chain.

Now the beginning of such chains can be observed using a similar procedure used for the function g . It is known that the roots of $I_n(y_1, y_2, \dots, y_{n-1}, x) = I_n(y_1, y_2, \dots, y_n) = q$ with respect to x are $x_1 = y_n$ and $x_2 = q(y_1 + \dots y_{n-1}) - y_n$. It was felt necessary to first simplify this expression before moving on with the inequalities:

$$\begin{aligned} x_2 &= \frac{\sum_{i=1}^n y_i^2}{\sum_{1 \leq i < j \leq n} y_i y_j + 1} \cdot \sum_{i=1}^{n-1} y_i - y_n \\ &= \frac{(\sum_{i=1}^{n-1} y_i^2) \cdot (\sum_{i=1}^{n-1} y_i) + y_n^2 \sum_{i=1}^{n-1} y_i - y_n \sum_{1 \leq i < j \leq n} y_i y_j - y_n}{\sum_{1 \leq i < j \leq n} y_i y_j + 1} \\ &= \frac{(\sum_{i=1}^{n-1} y_i^2) \cdot (\sum_{i=1}^{n-1} y_i) + y_n^2 \sum_{i=1}^{n-1} y_i - (y_n^2 \sum_{i=1}^{n-1} y_i + y_n \sum_{1 \leq i < j \leq n-1} y_i y_j) - y_n}{\sum_{1 \leq i < j \leq n} y_i y_j + 1} \end{aligned}$$

$$\begin{aligned}
& \text{(Since } y_n \sum_{1 \leq i < j \leq n} y_i y_j = y_n^2 \sum_{i=1}^{n-1} y_i + y_n \sum_{1 \leq i < j \leq n-1} y_i y_j) \\
& = \frac{(\sum_{i=1}^{n-1} y_i^2) \cdot (\sum_{i=1}^{n-1} y_i) - y_n (\sum_{1 \leq i < j \leq n-1} y_i y_j) - y_n}{\sum_{1 \leq i < j \leq n} y_i y_j + 1}
\end{aligned}$$

The similar deductions that were used for the previous extension problem are also valid in this generalized function. If $0 \leq x_2 \leq y_1$, $I_n(x_2, y_1, y_2, \dots, y_{n-1}) = q$ too so x_2 would be the beginning of the chain. If y_1 is the beginning of the chain, $x_2 > y_1$ or $x_2 < 0$.

i) $x_2 > y_1$

$$\begin{aligned}
& \Rightarrow \frac{(\sum_{i=1}^{n-1} y_i^2) \cdot (\sum_{i=1}^{n-1} y_i) - y_n (\sum_{1 \leq i < j \leq n-1} y_i y_j) - y_n}{\sum_{1 \leq i < j \leq n} y_i y_j + 1} > y_1 \\
& \Rightarrow \left(\sum_{i=1}^{n-1} y_i^2 \right) \cdot \left(\sum_{i=1}^{n-1} y_i \right) - y_n \left(\sum_{1 \leq i < j \leq n-1} y_i y_j \right) - y_n > y_1 \left(\sum_{1 \leq i < j \leq n} y_i y_j \right) + y_1 \\
& = y_1 \left(\sum_{1 \leq i < j \leq n-1} y_i y_j \right) + y_1 y_n \left(\sum_{i=1}^{n-1} y_i \right) + y_1 \\
& \Rightarrow \frac{(\sum_{i=1}^{n-1} y_i^2) \cdot (\sum_{i=1}^{n-1} y_i) - y_1 (\sum_{1 \leq i < j \leq n-1} y_i y_j) - y_1}{(\sum_{1 \leq i < j \leq n-1} y_i y_j) + y_1 (\sum_{i=1}^{n-1} y_i) + 1} > y_n
\end{aligned}$$

ii) $x_2 < 0$

$$\begin{aligned}
& \Rightarrow \left(\sum_{i=1}^{n-1} y_i^2 \right) \cdot \left(\sum_{i=1}^{n-1} y_i \right) - y_n \left(\sum_{1 \leq i < j \leq n-1} y_i y_j \right) - y_n < 0 \\
& \Rightarrow \frac{(\sum_{i=1}^{n-1} y_i^2) \cdot (\sum_{i=1}^{n-1} y_i)}{(\sum_{1 \leq i < j \leq n-1} y_i y_j) + 1} < y_n
\end{aligned}$$

Ultimately, the condition for $y_1, y_2, \dots, y_{n-1}, y_n$ to be the terms at the beginning of the chain is found:

$$\frac{(\sum_{i=1}^{n-1} y_i^2) \cdot (\sum_{i=1}^{n-1} y_i) - y_1(\sum_{1 \leq i < j \leq n-1} y_i y_j) - y_1}{(\sum_{1 \leq i < j \leq n-1} y_i y_j) + y_1(\sum_{i=1}^{n-1} y_i) + 1} > y_n \text{ or } \frac{(\sum_{i=1}^{n-1} y_i^2) \cdot (\sum_{i=1}^{n-1} y_i)}{(\sum_{1 \leq i < j \leq n-1} y_i y_j) + 1} < y_n$$

The outcome being a perfect square

Since the previous part of this paper couldn't find a rule for the equation with 3 variables (function g), I figured that no rule could be found for the general function I_n either. Even though the solutions for the equation are observed, the condition in the original IMO problem which was "the outcome of the function being a perfect square" couldn't be adapted to a general function, and identified as a special property for

$$f(x, y) = \frac{x^2 + y^2}{xy + 1}.$$

5. Conclusion

In this essay, a famous problem that was originally included in the International Mathematical Olympiad 1988 was investigated and extended. The first extension problem turned the original equation to an equation with 3 variables and the second extension examined an equation with n variables for a more general perspective. The ultimate aim for all 3 problems was to investigate the solution sets that made the given fractional expression an integer. For this aim, "chains" of solutions are found and the beginnings/ ends of these chains were investigated with the help of different inequalities.

It should be noted that many other extensions can be made out of these investigated functions. I decided to work on non-negative integers in the domain; however, it is possible that similar/ more rules or patterns can be found if the negative numbers are considered too. Moreover, as stated while forming the generalized function, another generalized function using "cyclic sum" instead of the sum of product of 2-combinations in the denominator can be investigated.

6. Bibliography

“Cyclic Sum.” *Art of Problem Solving*,

https://artofproblemsolving.com/wiki/index.php/Cyclic_sum. Accessed 11 Sept. 2022.

Engel, Arthur. *Problem-Solving Strategies*. Springer, 1998.

Laukó, István G., et al. “Another Step Further... On a Problem of the 1988 IMO.”

Mathematics Magazine, vol. 79, no. 1, 2006, pp. 45–53. *JSTOR*,

<https://doi.org/10.2307/27642901>.

Vandervelde, Sam. *A Rational Function Whose Integral Values Are Sums of Two Squares*.

2013, p. 8.

7. Appendices

Appendix A

```
initialize global bound_x to 100
initialize global bound_y to 100
initialize global bound_z to 100

when Screen1.Initialize
do
  set Label2.Text to join
    " The following triplets show the solutions to the..."
    get global bound_x
    ", y is between x and "
    get global bound_y
    ", z is between y and "
    get global bound_z
    " : "

  for each x from 1
    to get global bound_x
    by 1
  do
    for each y from get x
      to get global bound_y
      by 1
    do
      for each z from get y
        to get global bound_z
        by 1
      do
        if round get x ^ 2 + get y ^ 2...
          then set Label1.Text to join
            " ("
            Label1.Text
            join
              " ("
              get x
              " "
              get y
              " "
              get z
              ")"
```

Appendix B

initialize global bound_x to 100

initialize global bound_y to 100

initialize global bound_z to 100

The code is a Scratch script for finding Pythagorean triplets. It starts with an initialization block for Screen1, followed by a 'do' block that sets Label2's text to a title and a list of instructions. The main logic is a nested loop structure: a 'for each x' loop from 0 to 100, a 'for each y' loop from x to 100, and a 'for each z' loop from y to 100. Inside the z loop, an 'if' block checks if $x^2 + y^2 = z^2$ using 'round' and 'square root' blocks. If true, it sets Label1's text to a formatted string of the triplet (x, y, z).

```
when Screen1 Initialize
do
  set Label2 . Text to
  join
  " The following triplets show the solutions to the..."
  get global bound_x
  ". y is between x and "
  get global bound_y
  ". z is between y and "
  get global bound_z
  ". "

  for each x from 0
  to get global bound_x
  by 1
  do
    for each y from get x
    to get global bound_y
    by 1
    do
      for each z from get y
      to get global bound_z
      by 1
      do
        if
        round get x ^ 2 + get y ^ 2 + get... = round square root get x ^ 2 + get...
        and
        false = round square root get x ^ 2 + get... = square root get x ^ 2 + get...
        then
          set Label1 . Text to
          join
          " ("
          join
          get x
          " "
          get y
          " "
          get z
          " ) "
```