CNCM Basic Functional Equations Handout

Minseok Eli Park

October 26, 2020

§1 Introduction

Functional equations are problems will ask you to find all functions that satisfy a set of conditions. In most classic functional equations, there are 2 conditions.

1. The domain and range of the function. For example, $f : \mathbb{R} \to \mathbb{Q}$ means the function f has domain of real numbers, and a range of rational numbers. If the problem statement says, for example, "solve over positive integers", this is equivalent to $f : \mathbb{N} \to \mathbb{N}$.
2. An equation, which is the heart of the problem. For example, the equation from USEMO 2020/4 is

$$f(x + f(y) + xy) = xf(y) + f(x + y)$$

This means that for every function f that is part of the solution must satisfy this equation for every valid x and y (be careful on what x and y can be!)

All functional equations are two part problems. This means that not only must you find all working functions, you must then confirm that they actually work. Thankfully, the second part is very easy. All you have to do is plug it in, and verify the equation is always true. For example, in the above example, after proving that $f(x) = x$ is the only solution, it is easy to see that it works.

$$f(x + f(y) + xy) = xf(y) + f(x + y)$$

$$x + f(y) + xy = xf(y) + f(x + y)$$

$$x + y + xy = xy + x + y$$

§2 Definitions

The basic sets of numbers are as follows.

$$\mathbb{N}, \mathbb{Z}_{>0} = \text{Positive Integers}$$

$$\mathbb{W}, \mathbb{Z}_{\geq 0} = \text{Non-negative Integers}$$

$$\mathbb{Z} = \text{Integers}$$

$$\mathbb{Q} = \text{Rational Numbers}$$

$$\mathbb{R} = \text{Real Numbers}$$

We also want to define what *injective, surjective*, and *bijective* functions are.

A function is injective if $f(x) = f(y)$ if and only if $x = y$ for all x and y. In other words, it is a one-to-one function.

A function is surjective if for each a in the range, there is some x such that $f(x) = a$. In other words, the function reaches the entire domain.

A function is bijective if it is both injective and surjective.

§3 Basic Tips

1. Find the answer. For most standard functional equations, the solutions will all be linear functions. It will also normally be obviously when they aren't, for example $f(x)f(y) = f(x+y)$. If necessary, plug in some higher degree polynomials.

2. Think about the domain and range. If your functional equation is over rationals or integers, there is likely a reason why the problem wasn't over reals. Often, this may mean you would want to induct.

3. Plug stuff in. Usually, plugging in stuff such as $x = 0, 1, y, f(x)$, etc. can give you some clues going forward.

4. Keep in mind that proving that f is injective and/or surjective is very useful, and can help solve most problems. Do note that this is not always feasible.

§4 Example 1

Enough talking, let's go through an example.

Problem 4.1. Find all functions $f\colon \mathbb{N} \to \mathbb{N}$ such that

$$f(f(n)) + (f(n))^2 = n^2 + 3n + 3$$

Our first instinct is to guess all the functions. We could just plug in $f(n) = mn + b$ and see what I get for the values of integer constants m and b.

$$m(mn+b) + b + (mn+b)^2 = n^2 + 3n + 3$$

$$m^2n + mb + b + m^2n^2 + 2mnb + b^2 = n^2 + 3n + 3$$

Matching up the n^2 terms, we have $m = -1$ or 1. If $m = -1$,

$$n - b + b + n^2 - 2nb + b^2 = n^2 + 3n + 3$$

$$-2nb + b^2 = 2n + 3$$

An alternative way to do this part is realize that if m were negative, then f must output a negative number for arbitrarily large n, which is not allowed under our conditions, meaning $m \neq -1$. Matching up the n terms, we get $b = -1$, giving $1 = 3$, a contradiction. So, we must check $m = 1$,

$$n + b + b + n^2 + 2nb + b^2 = n^2 + 3n + 3$$

$$2b + 2nb + b^2 = 2n + 3$$

Matching up the n terms, we get $b = 1$, giving

$$2 + 2n + 1 = 2n + 3$$

which is true. So, our only linear solution is $f(n) = n + 1$.

It doesn't seem very obvious that there will be any non-linear solutions, so we will guess that the only solution is $f(n) = n + 1$. If this guess is correct, then we could try to use the general motivation behind the unproven fact that every value must map to only one value over all possible solutions. For example, $f(2)$ must equal 3, because that is the only output over all possible solutions for f.

Since we are over positive integers, we have a lower bound. Letting $P(n)$ be the given assertion, let's see what $P(1)$ gives us. If you are unfamiliar notation, this just means that if $P(n)$ is the given assertion, or the given equation, then $P(1)$ is the equation when we plug in $n = 1$.

$$f(f(1)) + f(1)^2 = 7$$

Well, our guess is that $f(1) = 2$. How could we prove this? Well, since f only outputs positive integers, we cannot have $f(1) \geq 3$ because if it were, then

$$f(f(1)) + f(1)^2 \geq f(f(1)) + 9 > 7$$

a contradiction. So, $f(1) \leq 2$. Why can't we have $f(1) = 1$? If we plug in $f(1) = 1$, we get

$$f(f(1)) + f(1)^2 = f(1) + 1^2 = 1 + 1^2 = 2$$

another contradiction. So, we have now proven $f(1) = 2$.
It is natural to now ask, can we find $f(2)$ explicitly like we did $f(1)$? There is a messier way to do this with $P(2)$, but there is actually a much easier way right in front of us when we found $f(1) = 2$. $P(1)$ gives

$$f(f(1)) + f(1)^2 = 7$$
$$f(2) + 2^2 = 7$$
$$f(2) = 3$$

That was quick! We realize this became very easy, because there are multiple instances of $f(1)$ that simplify immediately and give our result of $f(2) = 3$.
We can now use our knowledge of $f(2)$ to find $f(3)$ from $P(2)$. We could keep repeating this process: finding what $f(a)$ is to find $f(a + 1)$... This sounds like induction!

Claim: $f(n) = n + 1$ for all n.
Base case: We have already shown that $f(1) = 2$.
Inductive step: Say our claim works for $n = k$. We want to prove that our claim works for $n = k + 1$. From $P(k)$, we get

$$f(f(k)) + f(k)^2 = k^2 + 3k + 3$$

From our inductive step, we know that $f(k) = k + 1$.

$$f(k + 1) + (k + 1)^2 = k^2 + 3k + 3$$
$$f(k + 1) + k^2 + 2k + 1 = k^2 + 3k + 3$$
$$f(k + 1) = k + 2$$

So our inductive step is complete. We are now done, because we have proven through this induction that our guess that $f(n) = n + 1$ is actually true!
I have written so far the general motivation and steps taken to reach the solution for this problem. An actual contest proof would be much shorter, and look something like this.

Solution. I claim the answer is $f(n) = n + 1$, which clearly works. I prove this by induction.

Base Case:
When $n = 1$, we get

$$f(f(1)) + f(1)^2 = 3$$

Clearly, $f(1) \neq 1$ or else the LHS would be 2. Also, $f(1) < 3$ because if it weren't, then the RHS would be at least 9, due to the $f(1)^2$ term.
Inductive step:
Say our claim works for $n = k$. We want to prove that our claim works for $n = k + 1$. From n=k,

$$f(f(k)) + f(k)^2 = k^2 + 3k + 3$$
$$f(k + 1) + (k + 1)^2 = k^2 + 3k + 3$$
$$f(k + 1) = k + 2$$

so our induction is complete, and our claim is proven. $\square$

Note what this solution did not contain: It did not contain the process for which we guessed the solutions, we just said them. We also didn't explicitly go through $f(2)$ or $f(3)$, we only did them earlier because they helped us motivate the induction. The actual induction only required the base case to be done manually.

§5 The $f(f(x)) = x$ Lemma

Lemma 5.1

If we have $f(f(x)) = x$, then f is a bijection.

Proof. We want to prove that f is both surjective and injective. First we prove that it is surjective. Clearly, we can let x be anything, so $f(f(x))$ must be able to "reach all values". But this also means that $f(x)$ must be able to "reach all values", meaning f is surjective.

Now we prove that is it injective. For the sake of contradiction, let a and b be numbers such that $f(a) = f(b)$ but $a \neq b$. Because $f(a) = f(b)$, we get that $f(f(a)) = f(f(b))$. So,

$$a = f(f(a)) = f(f(b)) = b$$

which is a contradiction since $a \neq b$. So, there do not exist numbers a and b such that $f(a) = f(b)$ and $a \neq b$, implying f is injective. □

Now, we're going to investigate a very instructive problem. The problem can be divided into two parts, and each are quite important.

Problem 5.2 (Iran TST 1996)**.** Find all functions $f : \mathbb{R} \to \mathbb{R}$ such that

$$f(x^2 + y) = f(f(x) - y) + 4f(x)y \qquad \forall x, y \in \mathbb{R}.$$

§6 The Cancellation Trick

This trick is summarized by a nice quote from Evan Chen: "*DURR WE WANT STUFF TO CANCEL*". How could we do that here? We attempt a very rude cancellation, by letting stuff be equal as barbaric as possible. Specifically, we will try setting $f(x^2 + y) = f(f(x) - y)$.

$$x^2 + y = f(x) - y$$

$$y = \frac{1}{2}(f(x) - x^2)$$

Letting y be this, we have cancelled the LHS and the first terms on the RHS. $P(x, \frac{1}{2}(x^2 - f(x)))$ gives us

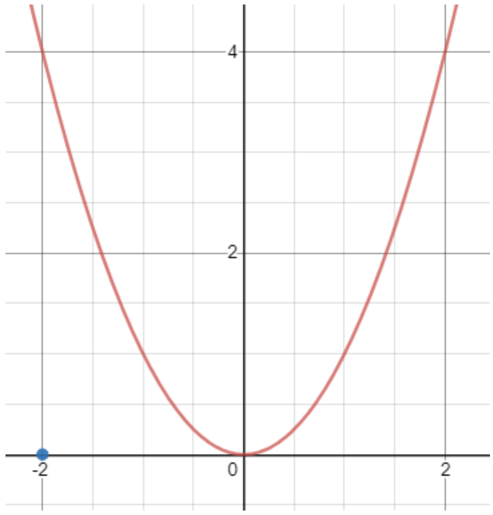
$$4(\frac{1}{2}(f(x) - x^2)f(x) = 0$$

$$f(x)(f(x) - x^2) = 0$$

Wow, that simplified a lot! And if you had guessed the solutions $f(x) = 0$ and $f(x) = x^2$, you are right? Wait, not yet...

§7 The Pointwise Trap

Now, we have fallen into the deep... dark... pointwise trap. Yes, $f(x) = 0$ and $f(x) = x^2$ both satisfy the equation $f(x)(f(x) - x^2) = 0$, but they aren't the only functions that satisfy it. Any function, that is equal to either 0 or x^2 at every point satisfies this condition. For example, the graph below is the graph of $f(x) = x^2$ except $f(-2) = 0$.



There's a small issue. The function with the graph above, and all of these, don't actually work when you plug them into the original functional equation, so we have to do extra work to prove they don't work.

Let $a \neq 0$ and $b \neq 0$ be numbers such that $f(a) = 0$ and $f(b) = b^2$ for the same function f. We want to run into a contradiction here, to prove that the entire function has to be $f(x) = 0$ or $f(x) = x^2$, and there aren't any naughty exceptions. We have the $a, b \neq 0$ conditions because the point $f(0) = 0$ is on any of these pathological solutions, and we want to prove that these two points a and b must not be in the same function f.

$P(a, b)$ gives $f(a^2 + b) = b^2$. Since $b^2 \neq 0$, we have $f(a^2 + b) = (a^2 + b)^2$, so setting this equal to b^2 we have

$$a^4 + 2a^2b + b^2 = b^2$$
$$a^2(a^2 + 2b) = 0$$
$$a^2 = -2b$$

where the last step comes from $a \neq 0$. Now, $P(a, -b)$ gives us $f(a^2 - b) = b^2$. Plugging in $a^2 = -2b$, we have $f(-3b) = b^2$. Since $b \neq 0$, $(-3b)^2 = b^2$, which is a contradiction since $b \neq 0$.

So, such numbers a, b don't exist in the same function, and we have proven that only the two functions $f(x) = 0$ and $f(x) = x^2$ work (which we should verify by plugging back in, a straightforward process).