FUNCTIONAL EQUATIONS
for the
OLYMPIAD ENTHUSIAST

Graeme West

INTRODUCTION

The South African Mathematical Society has the responsibility for selecting and training teams to represent South Africa in the annual International Mathematical Olympiad (IMO).

The process of finding a team to go to the IMO is a long one. It begins with a nationwide Mathematical Talent Search, in which students are sent sets of problems to solve. Their submissions are marked and returned with comments, full solutions and a further set of problems. The principle behind the Talent Search is straightforward: the more problems you solve, the higher up the ladder you climb and the closer you get to selection.

The best students in the Talent Search are invited to attend Mathematical Camps in which specialised problem–solving skills are taught. The students also write a series of challenging Olympiad–level problem papers, leading to selection of a team of six to go to the IMO.

The bookets in this series cover topics of particular relevance to Mathematical Olympiads. Though their primary purpose is preparing students for the International Mathematical Olympiad, they can with profit be read by all interested high school students who would like to extend their mathematical horizons beyond the confines of the school syllabus. They can also be used by teachers and university mathematicians who are interested in setting up Olympiad training programmes and need ideas on topics to cover and sample Olympiad problems.

Titles in the series published to date are:

No. 1 *The Pigeon-hole Principle,* by Valentin Goranko
No. 2 *Topics in Number Theory,* by Valentin Goranko
No. 3 *Inequalities for the Olympiad Enthusiast,* by Graeme West
No. 4 *Graph Theory for the Olympiad Enthusiast,* by Graeme West
No. 5 *Functional Equations for the Olympiad Enthusiast,* by Graeme West
No. 6 *Mathematical Induction for the Olympiad Enthusiast,* by David Jacobs

Details of the South African Mathematical Society's Mathematical Talent Search may be obtained by writing to

Mathematical Talent Search
Department of Mathematics and Applied Mathematics
University of Cape Town
7700 RONDEBOSCH

The International Mathematical Olympiad Talent Search is sponsored by the Old Mutual.

J H Webb
March 1997

Functional Equations for the Olympiad Enthusiast

Graeme West

This particular booklet in the series of training manuals has been fermenting in my mind for some time, and put off a number of times because I felt that I didn't have very much to say about the topic. And in the end I didn't say very much, I merely confirmed a suspicion that I had been harbouring for quite some time : more than any other 'topic' in the IMO 'syllabus', functional equations are simply a matter of doing. We are obliged to tackle often difficult problems with our bare hands, without the hammers we have in other topics such as number theory or geometry.

The corollary is that this is mostly a booklet of exercises. Do them. If you're not going to do them, don't bother with any of it.

The order of events is quite important in places. Of course problems are never solved by means of one idea only. But every problem is positioned so that it can be solved by means of the techniques already developed in the booklet. It follows that you should work through things more or less in sequence.

To avoid misunderstanding, let me point out that numbering within exercises of the form (a), (b), (c) indicates distinct problems. (For example, Exercise 1.) On the other hand, numbering of the form (i), (ii), (iii) indicates a single problem : the conditions enumerated are to be considered simultaneously. (For example, Exercise 7.)

I hope you derive as much enjoyment from solving these problems as I did from trying to.

Graeme West
January 1995
Kent Ohio USA

Contents

1 The vocabulary of functions

First some notation, without which we can get nowhere.
$\mathbf{N}$ is the set of natural numbers, $\{1, 2, 3, \ldots\}$;
$\mathbf{Z}$ is the set of integers, $\{\ldots -3, -2, -1, 0, 1, 2, 3, \ldots\}$;
$\mathbf{R}$ is the set of real numbers;
$\mathbf{Q}$ is the set of rational numbers;
and $\mathbf{C}$ is the set of complex numbers.

Given some set A of real numbers, A^+ will denote the subset of A of members which are > 0. The word 'positive' means > 0; 'non-negative' means ≥ 0.

There is a special notation for intervals of the real line. (An interval is a subset of $\mathbf{R}$ with 'no gaps' in it.) (a, b) denotes the set $\{x : a < x < b\}$, while $[a, b)$ denotes the set $\{x : a \leq x < b\}$. Similarly $(a, b] = \{x : a < x \leq b\}$ and $[a, b] = \{x : a \leq x \leq b\}$. All of these intervals are termed bounded intervals.

Note that if we are talking about Cartesian planes and things then (a, b) would not denote an interval but would mean the point in the plane with x-coordinate a and y-coordinate b. We can never get confused though - the meaning is always clear from the context.

The following are also intervals : $(a, \infty) = \{x : a < x\}$, and $[a, \infty) = \{x : a \leq x\}$. Likewise $(-\infty, b) = \{x : x < b\}$ and $(-\infty, b] = \{x : x \leq b\}$. These are called unbounded intervals.

$\mathbf{R}$ could be denoted $(-\infty, \infty)$, but this would be ugly.

The empty set is denoted $\emptyset$.

Definition 1.1 *A function is an assignment of points in a given set (called the domain of the function) to another given set (called the codomain of the function) with the following property : each member of the domain set is assigned to exactly one member of the codomain set.*

Thus a function is specified by three things : the domain set, the codomain set, and the rule of assignment.

Typically the domain set of a function might be denoted by an X and the codomain set by a Y. The rule of assignment might be denoted by the letter f or g. This is all summarised in the mathematical sentence '$f : X \to Y$'. Given $x \in X$, the point x is assigned to by f will be denoted $f(x)$. We can now write $f : X \to Y : x \to f(x)$. Since members of the codomain set Y will typically be denoted with a y, the expression '$y = f(x)$' is a common way of describing the function.

Examples 1.2 *(a) Let $X = Y = \mathbf{R}$. The assignment $f(x) = x^2$ describes a function.*

(b) Let $X = Y = [-1, 1]$. The rule $x^2 + y^2 = 1$ does not describe a function. $x = 0$, for example, is assigned to to both $y = 1$ and $y = -1$.

(c) Let $X = Y = \mathbf{R}$. The rule $y = \sqrt{1 - x^2}$ does not describe a function. Why?

It is very important to be aware of what the specified domain of the function is when solving functional equations. Too many times I have seen attempted solutions of functional equations $f : \mathbf{N} \to \mathbf{N}$ where $f(0)$ is happily calculated. This leads to a zero of a different kind.

Now for something harder :-

Definition 1.3 *(a) The range of a function $f : X \to Y$ is all those points in the codomain Y which have members of the domain X mapped to them by the function, that is*

$$\text{range}(f) = \{y \in Y : \text{ there exists } x \in X \text{ with } f(x) = y\} \quad (1)$$

(b) a function is said to be onto (or surjective) if every point of the codomain is mapped to by some or other point in the domain, that is, the range is all of the codomain.

(c) In general, a point in the range could have more than one point in the domain which is mapped to it. A function is said to be one-to-one (or injective) if every point of the range has exactly one point in the domain mapped to it, in other words, $f(x_1) = y = f(x_2) \Rightarrow x_1 = x_2$.

(d) A function is said to be bijective if it is both surjective and injective.

Exercise 1 *Decide which of the following are injective, surjective and bijective functions.*

(a) $f : \mathbf{R} \to \mathbf{R} : x \to x^2$.

(b) $f : \mathbf{R} \to [0, \infty) : x \to x^2$.

(c) $f : [0, \infty) \to [0, \infty) : x \to x^2$.

(d) $f : \mathbf{R} \to \mathbf{R} : x \to 2x + 7$.

(e) $f : \mathbf{R} \to \mathbf{R} : x \to \sqrt{x}$.

Exercise 2 *In the following exercises, the domain and codomain of an unknown function is given, as well as a functional equation. Without solving the functional equation (you can if you are feeling ambitious, but some of them can't be solved with only the information given) determine if the function is injective or surjective.*

(a) $f : \mathbf{R} \to \mathbf{R};\ f(f(x)) = x$.

(b) $f : \mathbf{N} \to \mathbf{N};\ f(f(n) + f(m)) = m + n$.

(c) $f : \mathbf{Z} \to \mathbf{Z};\ f(f(n + 2) + 2) = n$.

(d) $f : \mathbf{R} \to \mathbf{R};\ f(x^2 + f(y)) = y + (f(x))^2$.

(e) $f : (0, \infty) \to (0, \infty);\ f(xf(y)) = yf(x)$.

(f) $f : \mathbf{Z} \to \mathbf{Z};\ f(m + f(f(n))) = -f(f(m + 1)) - n$.

Definition 1.4 *A function f whose domain and codomain are subsets of $\mathbf{R}$ is said to be increasing if for all x_1, x_2 belonging to the domain of f, $x_1 < x_2 \Rightarrow f(x_1) \leq f(x_2)$. It is strictly increasing if we have $f(x_1) < f(x_2)$.*
Decreasing and strictly decreasing are similarly defined. A function which is (strictly) increasing or decreasing is said to be (strictly) monotone.

Exercise 3 *Determine which of the functions in Exercise 1 are increasing, strictly increasing, decreasing, strictly decreasing.*

Exercise 4 *Show that a strictly monotone function is injective. Is a monotone function injective?*

Exercise 5 *Find all $f : \mathbf{R} \to \mathbf{R}$ that are increasing and which satisfy $f(f(x)) = x$.*

Definition 1.5 *Suppose $A \subset \mathbf{R}$ has the property that $A = -A$, that is, $x \in A \iff -x \in A$. A function f with domain A is said to be even if $f(x) = f(-x)$ for every $x \in A$; and odd if $f(x) = -f(-x)$ for every $x \in A$.*

We'll only consider this idea for functions defined on $\mathbf{R}$ and $\mathbf{Z}$. It is often useful to establish that a function is even or odd, because that literally cuts out half of the subsequent work.

Exercise 6 *Which functions are odd or even?*

(a) $f : \mathbf{R} \to \mathbf{R} : x \to x^2$.

(b) $f : \mathbf{R} \to \mathbf{R} : x \to 12x$.

(c) $f : \mathbf{R} \to \mathbf{R} : x \to 7x + 34$.

(d) $f : \mathbf{R} \to \mathbf{R} : x \to 0$.

You'll need to brush up on your trig for this one.

Exercise 7 *Find all functions $f : \mathbf{R} \to \mathbf{R}$ which satisfy*

(i) $f(2x) = f\left(\sin\left(\frac{\pi x}{2} + \frac{\pi y}{2}\right)\right) + f\left(\sin\left(\frac{\pi x}{2} - \frac{\pi y}{2}\right)\right)$;

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

Definition 1.6 (Composition of functions) *Suppose $f : X \to Y$ and $g : Y \to Z$. Then the composition of g and f is the function $g \circ f : X \to Z$ given by $g \circ f(x) = g(f(x))$.*

The clumsy notation seen in the solution of Exercise 2 makes us appreciate the following definitions and notations.

Definition 1.7 (Iterates of functions) *Suppose $f : X \to X$, that is, the codomain of f is the same as the domain. Then f composed with itself, $f \circ f$, is denoted f^2. Thus*

$$f \circ f(x) = f(f(x)) = f^2(x)$$

In general, f composed with itself n times will be denoted f^n, that is

$$f^n(x) = \underbrace{f(f(\ldots f(x)))}_{n}$$

It is to be understood that $f^1(x) = f(x)$ and $f^0(x) = x$.

Given $x \in X$, the (ordered) set

$$\{x,\ f(x),\ f^2(x),\ f^3(x),\ \ldots\}$$

is called the orbit of x under f.

Exercise 8 *Make sure you fully understand the differences and similarities between $f^2(x)$, $f(x)^2$, $f(x^2)$.*

There aren't two many similarities, so half of the problem is easy.

2 Pointers to solving functional equations

This section serves as a brainstorm of ideas for solving functional equations. There are no exercises here : the idea is that you will refer back here for advice when doing the exercises in the other sections of this booklet. (Your inspiration will have to come from elsewhere.)

Certain points are dealt with in great detail elsewhere, and so are only mentioned briefly here.

1. Find trivial solutions to the functional equation. $f = 0$ is often a solution. (That means the function for which $f(x) = 0$ for every x.) Very often finding trivial solutions takes out insurance for what you do later : for example, you might later want to divide by some particular $f(x)$, not very convincing if $f(x)$ is always 0!

   Make it clear to the reader of your solution that you have found these trivial solutions.

2. Try at the beginning to find at least one non-trivial solution, if you think this is appropriate. You may even feel that you have found all solutions. If so, this can help in forming your subsequent strategy, because you can test your ideas against the solutions found.

   There is always credit for correctly guessing the actual solutions, as long as they're not very obvious.

3. Make a mental note of the domain and make sure your calculations do too.

4. Put variables (singly, and in pairs, etc) equal to 0, 1, other suggested numbers, each other. This is the basic labour from which everything else needs to follow.

5. Test if the function is injective. This should develop into a reflex action.

6. Check to see if the function is monotone, could be monotone, or if monotonicity would enable further important conclusions to be drawn. Remember strict monotonicity implies injectivity.

7. Check if possible for surjectivity. Often this is not possible and often it's unimportant. But often it is useful to know that a certain value belongs to the range, because this allows for useful substitutions.

8. If 0 belongs to the domain then 'division by x' is illegal; if 0 belongs (or may belong) to the range then 'division by $f(x)$' is illegal. Treat division with care!

9. Consider orbits, iterates, and fixed points. This is dealt with in some detail later.

10. Think about transformations (also mentioned later). This doesn't come up often, but it's very powerful when it does.

11. Try to calculate the value of the function at a point in two different ways. (Especially common if the functional equation has two defining equations.) The results, even though the expressions are different, will be equal.

    The idea of symmetric substitutions is related to this, and is dealt with briefly later.

12. Having made some discovery, start again : (a) rewrite all the known equations in terms of the new information; (b) check if possible strategies which were stumped before now have a better chance of bearing fruit; (c) make new guesses at solutions.

13. At some point in your argument you might have eliminated 'parasitic' solutions e.g. $f = 0$. At the end of your solution, summarise, stating all the solutions, both trivial and non-trivial.

14. Check solutions that are found. Your arguments will find all possible solutions. That does not mean that what you have found is really a solution, so finding a 'solution' which isn't doesn't necessarily mean you've made an error in your calculations.

Checking the solution means verifying that the function you have found does indeed satisfy the conditions of the question.

Usually boring but always important, and usually worth the last piece of credit in an examination.

3 Inductive arguments

In this section we will consider functional equations on sets like **N** or **Z** or **Q** which are solved by use of inductive techniques.

Theorem 3.1 (Principle of mathematical Induction)
Suppose $T \subset \mathbf{N}$, $1 \in T$, and $n \in T \Rightarrow n+1 \in T$. Then $T = \mathbf{N}$. ∎

Some functional equations are solved by a no-strings-attached inductive argument. On other occasions, induction can be made more subtle : for example, the induction might be performed in 'blocks', for example, assuming the statement true for $\{1, 2, \ldots, 3n, 3n+1, 3n+2\}$, we establish it true for $\{3n+3, 3n+4, 3n+5\}$. In this case, induction is performed in blocks of three. By the way, in this example, it is necessary to start the induction by establishing 'by hand' that the statement is true for 1, 2, 3, 4, 5. (Why?)

By the way, in many places in the solutions I have waved my hands about inductive proofs if I believe that the induction is routine. I'm lazy, duplication costs of this booklet are high, the rainforests are dying, and this is not a competition. In a competition, you need to provide the details.

Exercise 9 *Determine all functions $f : \mathbf{Z} \to \mathbf{Z}$ which satisfy*

(i) $f(k+n) + f(k-n) = 2f(k)\,f(n)$ for all integers k and n;

(ii) there exists an integer N such that $-N < f(n) < N$ for all n.

Exercise 10 *The function $f(x, y)$ satisfies*

(i) $f(0, y) = y + 1$;

(ii) $f(x+1, 0) = f(x, 1)$;

(iii) $f(x+1, y+1) = f(x, f(x+1, y))$

for all non-negative integers x and y. Find $f(4, 1981)$.

Exercise 11 *The function $f(n)$ is defined for all positive integers n and takes on non-negative integer values. Also, for all m, n*

(i) $f(m+n) - f(m) - f(n) \in \{0, 1\}$;

(ii) $f(2) = 0$;

(iii) $f(3) > 0$;

(iv) $f(9999) = 3333$.

Determine $f(1982)$.

Exercise 12 *Define a function $f : \mathbf{N} \to \mathbf{N}$ iteratively by :-*

(i) $f(1) = 1$;

(ii) For $n \geq 2$, $f(n) = \begin{cases} f(n-1) - n & \text{if } f(n-1) > n \\ f(n-1) + n & \text{if } f(n-1) \leq n \end{cases}$.

Let $S = \{n \in \mathbf{N} : f(n) = 1993\}$.

(a) Prove that S is an infinite set.

(b) Find the least member of S.

(c) If all the members of S are written in ascending order as $n_1 < n_2 < \cdots$, show that $\lim_{i\to\infty} \frac{n_{i+1}}{n_i} = 3$.

Exercise 13 *A function f is defined on the positive integers by*

(i) $f(1) = 1$;

(ii) $f(3) = 3$;

(iii) $f(2n) = f(n)$;

(iv) $f(4n+1) = 2f(2n+1) - f(n)$;

(v) $f(4n+3) = 3f(2n+1) - 2f(n)$.

for all positive integers n. Determine the number of positive integers n, less than or equal to 1988, for which $f(n) = n$.

A very important result in mathematics is the following. It is logically equivalent to the principle of mathematical induction.

Theorem 3.2 (Well Ordering Principle)
Every non-empty subset of **N** *has a least element.* ■

We use this result quite often. For example, if we are trying to establish a certain formula for a function on **N**, we can suppose that it doesn't always hold, look at the least member of the set of points where it does not hold (extreme case principle), and then attempt to derive some sort of contradiction.

Exercise 14 *Let $f, g : \mathbf{N} \to \mathbf{N}$ be functions such that f is surjective, g is injective and $f(n) \geq g(n)$ for every $n \in \mathbf{N}$. Show that $f = g$.*

Exercise 15 *Suppose $f : \mathbf{N} \to \mathbf{N}$. Prove that if $f(n+1) > f(f(n))$ for all $n \in \mathbf{N}$ then $f(n) = n$ for all $n \in \mathbf{N}$.*

Note that the well ordering theorem and the induction principle fail in **Z**. But it's common practice to break **Z** up into its positive and negative parts and perform induction or use the well ordering principle on each part. Alternatively, induction on **Z** can be performed something like follows : given that the required statement is true for $\{-n, -n+1, \ldots, -1, 0, 1, \ldots, n\}$, show that it is true for $-n-1$ and for $n+1$. (To get things started in this example, one would show it true for 0.)

Exercise 16 *Find all $f : \mathbf{Z} \to \mathbf{Z}$ which satisfy*

(i) $f(f(n)) = n$;

(ii) $f(f(n+2)+2) = n$;

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

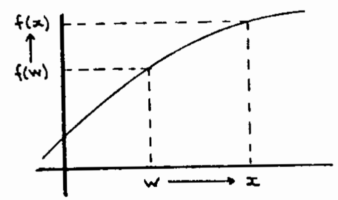
The well ordering theorem fails spectacularly on **Q** and on $\mathbf{Q}^+$. Nevertheless, there is an important inductive trick here that comes up quite often : see Example 4.1.

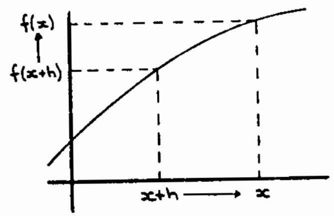
As discussed in §4, many of the techniques of this section are applicable there too.

4 Solving for continuous functions

In certain cases you will be asked to solve a given functional equation with the added information that the answer is known to be continuous. This allows some new strategies which we are going to discuss here.

For a function to be continuous at the point x means that if a variable point w approaches x then $f(w)$ approaches $f(x)$. This can be written as '$w \to x \Rightarrow f(w) \to f(x)$'.





An alternative, clearly equivalent formulation : a function is continuous at x if $f(x+h)$ approaches $f(x)$ as h approaches 0. This can be written as '$h \to 0 \Rightarrow f(x+h) \to f(x)$'.

Thus there is no 'jump' in the graph of f near to x. To say that a function is continuous is to say that it is continuous for every point x in its domain, in other words, there are no jumps anywhere in the graph of f.

Next we need the notion of 'dense' set in **R** : a set is dense if it is very thickly spread throughout the real line. Here is the precise definition : a set is dense if its complement does not contain any intervals.

A useful and very common test for denseness could be called the midpoint test : if a set contains arbitrarily large positive and negative members, and for any two points in the set the midpoint is also in the set, then that set is dense. (Warning : there are sets which are dense that fail the midpoint test. Okay?)

An obvious example of a dense set is the set of rationals **Q**. Check for yourself that **Q** is dense by definition, and also passes the midpoint test.

It follows that if we know that a function is continuous, and we can establish that the required function agrees with a certain continuous function on a dense set, then the required function *is* that function.

(This follows from the fact that a continuous function defined on a dense set admits at most one extension to the whole set.)

On the other hand, the following two points should be noted :-

- Continuity plays no role with functions defined on **N** or **Z**.
- Usually continuity plays no role with functional equations on **Q**. Let me be more precise : suppose you are asked to find a function defined on **Q** with certain properties. *A priori* we can bet that this function will be discontinuous, and cannot be extended to a continuous function on **R**. For if it could, the original question would surely have asked for a function defined on **R** with the given properties, and expected the solver to realise that the appropriate technique would be to work on **Q**, and then extend by continuity to **R**. Exercise 23 is a demonstration of all of this.

In the example and the exercises we follow the same strategy : we are told the function is continuous, so we establish a formula for the function on **Q** and then deduce that the same formula must hold on all of **R**. The methods for establishing the required formula on **Q** are an extension of those seen in §3.

Example 4.1 *Find all $f : \mathbf{R} \to \mathbf{R}$ which are continuous and satisfy*

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

for all $x,\ y \in \mathbf{R}$.

Solution : Putting $x = y = 0$ we get that $f(0) = f(0) + f(0)$, and so $f(0) = 0$. Now a little thought will convince you that we cannot determine any more values directly. However, putting $x = y = 1$ we get $f(2) = f(1) + f(1) = 2\,f(1)$, and putting $x = 2,\ y = 1$ we get $f(3) = f(2) + f(1) = 3\,f(1)$. In general we have $f(n) = n\,f(1)$ for $n \in \mathbf{N}$ which we establish by induction. Let us write $f(1) = a$.

So $f(n) = an$ for all $n \in \mathbf{N}$. But $f(0) = f(n) + f(-n)$, and so $f(-n) = -f(n)$. The function is odd. It now follows that $f(n) = an$ for all $n \in \mathbf{Z}$.

If $m \in \mathbf{N}$ and $x \in \mathbf{R}$ then

$$f(mx) = m\, f(x) \tag{2}$$

which is also established by induction. Thus if $\frac{n}{m} \in \mathbf{Q}$ then

$$an = f(n) = f\left(m\,\frac{n}{m}\right) = m\, f\left(\frac{n}{m}\right) \tag{3}$$

and so

$$f\left(\frac{n}{m}\right) = a\,\frac{n}{m}$$

In other words, $f(x) = ax$ for all $x \in \mathbf{Q}$. It now follows by the continuity of f that $f(x) = ax$ for all $x \in \mathbf{R}$. This checks for any $a \in \mathbf{R}$:

$$f(x+y) = a(x+y) = ax + ay = f(x) + f(y)$$

(It was an important problem in the early part of this century to know whether these were the only solutions to this functional equation, even without the continuity assumption. In fact it was eventually shown that there are discontinuous solutions to this functional equation. This had important ramifications in the development of the axiomatics for the study of vector spaces. The construction of discontinuous solutions goes a little beyond our interests.) ■

The key step in the above example - the argument around (2) and (3) - should be re-read, understood, remembered.

Exercise 17 *Find all continuous functions $g : \mathbf{R} \to \mathbf{R}$ which satisfy*

$$g(x+y) + g(x-y) = 2\,g(x) + 2\,g(y)$$

Exercise 18 *Suppose a function $f : \mathbf{R} \to \mathbf{R}$ satisfies*

(i) $f(1) = 1$;

(ii) $f(x+y) = f(x) + f(y)$ *for* $x,\ y \in \mathbf{R}$;

(iii) $f(x)\, f\left(\frac{1}{x}\right) = 1$ *for* $x \neq 0$.

Show that $f(x) = x$ for all $x \in \mathbf{R}$.

Exercise 19 *Suppose $a \in \mathbf{R}$, and f is a continuous function on $[0,1]$ satisfying*

(i) $f(0) = 0$;

(ii) $f(1) = 1$;

(iii) $f\left(\frac{x+y}{2}\right) = (1-a)f(x) + af(y)$ *for all* $x, y \in [0,1]$ *with* $x \leq y$.

Find the possible values of a.

Recall that a zero of a function f is a point x for which $f(x) = 0$. Sometimes such a number is also referred to as a root, although this terminology is usually reserved for polynomials. The most important result about continuous functions is possibly the following :-

Theorem 4.2 (Intermediate Value Theorem)
Suppose f is a continuous function defined on some interval I and $a,\ b \in I$. If $f(a) \neq f(b)$ then for every $y \in (f(a), f(b))$ there exists $x \in (a,b)$ such that $f(x) = y$.

If $f(a)$ and $f(b)$ are of different signs then there exists a zero of f between a and b. ■

Of course the second statement is (the most often used) special case of the first. Location of the zeros of a continuous function can be very important in solving functional equations because of their 'destructive nature' in the equation.

Exercise 20 *Suppose $f : \mathbf{R} \to \mathbf{R}$ is continuous and*

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

for all $x,\ y \in \mathbf{R}$. Show that either $f = 0$ or f has no zeros.

Exercise 21 *Suppose $f : \mathbf{R} \to \mathbf{R}$ is continuous and*

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

for all $x,\ y \in \mathbf{R}$. Show that either $f = 0$ or f has no zeros.

5 Analytic arguments

In this section we solve for functions that are not necessarily continuous. That doesn't mean that they are discontinuous. What it does mean, however, is that our arguments must be purely algebraic and analytic, and not use any continuity arguments. So, we can (and do) check for injectivity, surjectivity and so on; but we may not, for example, use the intermediate value theorem, which is true only for continuous functions.

Exercise 22 *Give an example of a function which changes sign infinitely often, but has no zeros.*

Any amount of experience in solving functional equations will tell you that 'most' solutions are continuous. But be warned : there are badly discontinuous functions which arise as the solutions to quite natural functional equations. The remarks immediately after Example 4.1 mentioned that there are discontinuous solutions to the very natural functional equation $f : \mathbf{R} \to \mathbf{R}$, $f(x+y) = f(x) + f(y)$, but that these are difficult to construct. Some loose ends there. Now try the following problem :-

Exercise 23 *Construct a function $f : \mathbf{Q}^+ \to \mathbf{Q}^+$ which satisfies*

$$f\left(xf(y)\right) = \frac{f(x)}{y}$$

for all $x,\ y \in \mathbf{Q}^+$.

Hopefully this example of a highly discontinuous solution to the functional equation $f(xy) = f(x)\,f(y)$ makes the possibly disturbing statement that there are discontinuous solutions to the functional equation $f(x+y) = f(x) + f(y)$ more plausible.

We are going to consider various things that you *can* do. Of course, all of these techniques also apply to functions which are known to be continuous.

5.1 Recognising straight line equations

In this section we will see what type of functional equations allow us to deduce that the required function is a straight line.

A straight line has equation $f(x) = mx + c$. Here m is the slope of the straight line and c is the value of the y-intercept, often just called the intercept. This form of the equation is often called the slope-intercept form.

A straight line can also be specified by means of the point-slope form. If a straight line is known to pass through the point (a, b) and is known to have slope m, then its equation is $y - b = m(x - a)$. Okay?

We will consider the problem of functional equations for straight line functions defined on any set. Usually though, for functions defined on a set like $\mathbf{R}$, the hypothesis of continuity is necessary to conclude from the typical linear functional equation that the function is a straight line.

For example, recall that in Example 4.1 we showed that the functional equation $f : \mathbf{R} \to \mathbf{R}$, $f(x+y) = f(x)+f(y)$ had $f(x) = mx$ as solutions for any $m \in \mathbf{R}$, but we pointed out that there are other solutions. This would not be an issue if the function was defined as a function on $\mathbf{Q}$, $\mathbf{Z}$, or $\mathbf{N}$. (This would however affect the admissible values of m, right?)

Okay. The functions $f(x) = mx$ all pass through the origin, but of course there are straight lines that do not. How do we recognise a straight line? Usually by the various ways of characterising the slope. For example, if $\frac{f(y)-f(x)}{y-x}$ is constant, then the function is a straight line, with that constant being its slope. (Such a function is automatically continuous, by the way - okay?) On the other hand, a straight line could be indicated by 'constant change' formulas : a formula that shows that the amount of change in the function between two points is the same for any two points whose distance apart is the same.

The equations in the following two exercises, after some manipulations, indicate a straight line solution because of a constant change formula. In Exercise 24, the additional hypothesis of continuity is necessary, because any solution the the functional equation $f(x+y) = f(x) + f(y)$ is a solution to this one, (check it!) and there are discontinuous solutions to $f(x+y) = f(x)+f(y)$. On the other hand, in Exercise 25 the function is defined on $\mathbf{N}$ and so continuity is not an issue.

Exercise 24 *Find all continuous functions $g : \mathbf{R} \to \mathbf{R}$ which satisfy*

$$g(x+y) + g(x-y) = 2\,g(x)$$

Exercise 25 *Find all $f : \mathbf{N} \to \mathbf{N}$ satisfying*

$$f(f(n) + f(m)) = m + n$$

for all $m, n \in \mathbf{N}$.

5.2 Substitutions and transformations

A substitution is a manipulation with the variables in the functional equation. Substitutions have already come up repeatedly; they are the bread and butter of functional equations, and I don't believe there is much that I can say which is not obvious. I mention them here only for the sake of some warnings of the serious shortcomings of sloppy substitutions :-

Suppose we are asked to solve an equation like

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

You might be tempted to say that

$$f(m+k) = -k - (m+1) = -(k+m) - 1$$

where $k = f^2(m+1)$. And you might be further tempted to put $n = m + k$, so $f(n) = -n - 1$. One temptation is human, two are unforgivable. The first conclusion is valid, but perhaps not useful. The second conclusion is invalid.

Exercise 26 *Why?*

If the equation had read

$$f(m + f^2(n+1)) = -f^2(n+1) - (m+1)$$

the argument would have been valid. (Of course *this* equation is rather feeble, but hopefully the point is made.)

Exercise 27 *Find all $f : \mathbf{R} \to \mathbf{R}$ such that*

(i) $f(x+y) + f(x-y) = 2f(x)\,f(y)$ for all $x, y \in \mathbf{R}$;

(ii) if $x \to \infty$ then $f(x) \to 0$.

Next we deal with transformations. A transformation is a change of the function, as opposed to a substitution, which is a change of the variable(s). The idea of a transformation is to convert the function into a more user-friendly form, achieve something there, and then transform back to the original function.

It follows that all transformations we perform need to be invertible. With this caveat, anything goes.

In the following two problems, we apply a logarithmic transformation (we take logarithms on both sides of the equation). This is natural to do because taking logarithms converts products into sums and powers into scalar multiples. Of course, logarithms are inverted by exponents.

There is one thing to worry about when performing transformations, and it is a real issue in these two problems. I hope you spot it; it is discussed in the solutions.

Exercise 28 *Suppose $f : \mathbf{R} \to \mathbf{R}$ is continuous and*

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

for all $x,\ y \in \mathbf{R}$. Find f.

Exercise 29 *Suppose $f : \mathbf{R} \to \mathbf{R}$ is continuous and*

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

for all $x,\ y \in \mathbf{R}$. Find f.

5.3 Symmetry in two-variable expressions

Suppose a functional equation is given where there are two (or possibly more) variables appearing in the equation. A very powerful strategy we should attempt is to introduce some symmetry into one side of the given functional equation. It follows that, since this side does not change when the variables are swopped, neither does the other. The two 'other' sides are equal.

All those words are best expressed in an example. Suppose we have a functional equation like

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

Now the left hand side is symmetric in the variables x and y, and so we see that

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

We now have the new information that $x + f(y) = y + f(x)$. The problem is practically finished.

Exercise 30 *Finish it.*

Now for the real thing.

Exercise 31 *Let $\alpha,\ \beta \in \mathbf{R}$, not necessarily distinct. Find all functions $f : \mathbf{R}^+ \to \mathbf{R}$ such that*

$$f(x)\,f(y) = y^{\alpha} f\left(\frac{x}{2}\right) + x^{\beta} f\left(\frac{y}{2}\right)$$

for all $x,\ y \in \mathbf{R}^+$.

Exercise 32 *Consider $f : \mathbf{Z} \to \mathbf{Z}$ satisfying*

$$f(m + f(f(n))) = -f(f(m+1)) - n$$

for all $m, n \in \mathbf{Z}$. Find f.

5.4 Fixed points, iterates and orbits

Definition 5.1 *A fixed point of a function is a point x for which $f(x) = x$.*

Necessarily, the point x must belong to both the domain and the codomain of the function. It follows that for all practical purposes, we would only seek out fixed points for functions whose domain and codomain coincide.

Note that the orbit of a fixed point is just that fixed point. Use of fixed points is a very powerful weapon in solving functional equations because of the simplifying effect it has in iterative equations.

It seems that the use of fixed points is usually in an abstract sense. We deduce certain facts about the set of fixed points (for example, given an arbitrary fixed point its inverse is a fixed point, or given two fixed points their product is a fixed point) and then attempt to derive some conclusions or contradictions. In both of the exercises that follow, we have information that tells us that there are 'few' fixed points, yet other information tell us that there are 'many'. All of this is abstract, in the sense that it is very late in the day that we find out what the fixed points actually are.

In both exercises we see a certain symmetry in the defining functional equation : this is exactly what should suggest a fixed point argument. In both cases, the second condition places a limitation on the possible number or location of the fixed points.

Exercise 33 *Find all functions $f : (-1,\infty) \to (-1,\infty)$ satisfying the two conditions*

(i) $f(x + f(y) + xf(y)) = y + f(x) + yf(x)$ *for all* $x,\ y \in (-1,\infty)$;

(ii) $\dfrac{f(x)}{x}$ *is strictly increasing on each of the intervals* $(-1,\ 0)$ *and* $(0,\ \infty)$.

Exercise 34 *Find all functions $f : (0,\infty) \to (0,\infty)$ which satisfy the conditions*

(i) $f(xf(y)) = yf(x)$ *for all positive* $x,\ y$;

(ii) $f(x) \to 0$ *as* $x \to \infty$.

Now we turn to the topic of iterates in functional equations, and start with characteristic equations. This quite specialised technique is relevant when the functional equation has only one variable, but includes iterates of the function.

Recall that a sequence of real numbers is defined by a recurrence relation if there is a known formula relating a term or terms to the succeeding term, and the first term or terms of the sequence (the initial values) are also known. (In general, if the recurrence formula involves the m previous terms, then the initial m values need to be given in order to properly define the recurrence relation.)

The following result is standard. For more details and examples, see the University of Otago Problem Solving Series Booklet #9.

Theorem 5.2
Let a recurrence relation be given by $a_{n+2} = Aa_{n+1} + Ba_n$ where A and B are known constants.

- *If $\lambda^2 = A\lambda + B$ has distinct roots α and β then the recurrence has the solution $a_n = K\alpha^n + L\beta^n$ where K and L are constants determined by the initial value conditions.*
- *If $\lambda^2 = A\lambda + B$ has the repeated root α then the recurrence has the solution $a_n = (K + nL)\,\alpha^n$ where K and L are constants determined by the initial value conditions.* ■

What is the relevance of this for us? Here is an example.

Example 5.3 *Find all $f : \mathbf{R} \to \mathbf{R}$ satisfying $f^2(x) = 3f(x) - 2x$*

Let us fix an $x \in \mathbf{R}$, and keep it fixed throughout. For $n \geq 0$ let us put $a_n = f^n(x)$. Then making the appropriate substitutions in the functional equation we get the recurrence relation $a_{n+2} = 3a_{n+1} - 2a_n$. The initial conditions (we need two, right?) are that $a_0 = x$ and $a_1 = f(x)$. We can't hope for any better than that since x is any real number. But now the above theorem tells us to examine the quadratic $\lambda^2 = 3\lambda - 2$. (We can call this the characteristic equation of the functional equation.) It has roots 1, 2 and so the theorem tells us that

$$f^n(x) = a_n = K + L\,2^n$$

Now substituting in $n = 0$ we get that $x = K + L$ and substituting in $n = 1$ we get that $f(x) = K + 2L$. Taking differences we get that $L = f(x) - x$ and hence $f(x) = 2x - K$. This is it.

Checking, we get $f^2(x) = 4x - 3K = 3f(x) - 2x$. Magic, or what? ■

It was a surprise to me that the following was the only IMO-submitted problem of this nature that I could find :-

Exercise 35 *Suppose a, b are positive real numbers. Find all functions $f : [0, \infty) \to [0, \infty)$ which satisfy*

$$f(f(x)) + af(x) = b(a + b)x$$

Finally we turn to the topic of orbits of points. Sometimes a certain member of the domain is clearly of great importance in a functional equation, and we would like to know what the orbit of the point looks like.

In the following example, we find the orbit of an unknown but clearly important point. Every member of the orbit is expressible as a function of the original point, and this enables us to determine what the point actually is.

Exercise 36 *Find all functions $f : \mathbf{R} \to \mathbf{R}$ such that*

$$f\left(x^2 + f(y)\right) = y + (f(x))^2$$

for all x, y in $\mathbf{R}$.

Here's a clue for the following rather difficult exercise : the idea is that if α is a root of $p(x)$ then so are all the members of the orbit of α under the equation $f(t) = t^2 - 1$ (okay?), and yet $p(x)$, being a polynomial, can have only finitely many roots.

Exercise 37 *Describe the family of polynomials whose roots are real and for which*

$$p(x^2 - 1) = p(x)p(-x)$$

for all $x \in \mathbf{R}$.

6 Just do it

I couldn't think of a better place for these guys.

Exercise 38 *Find all functions $f : \mathbf{R} \setminus \{0\} \to \mathbf{R}$ which satisfy*

(i) $f(x) = x\, f\left(\frac{1}{x}\right)$ for $x \neq 0$;

(ii) $f(x) + f(y) = 1 + f(x + y)$ for $x \neq -y$.

Exercise 39 *Find the functions $f : [0, \infty) \to [0, \infty)$ such that*

(i) $f(xf(y))\, f(y) = f(x + y)$;

(ii) $f(2) = 0$;

(iii) $0 \leq x < 2 \Rightarrow f(x) \neq 0$.

7 Solutions to the exercises

1. (a) The function is not injective, since $f(-1) = 1 = f(1)$. The function is not surjective, since $-1 \notin \text{range}(f)$.

   (b) Likewise this function is not injective, but it is surjective : given $y \in [0, \infty)$, put $x = \sqrt{y}$. Then $f(x) = y$.

   (c) Suppose $x_1^2 = x_2^2$. Since $x_1,\ x_2 \geq 0$, we can take square roots. Then $x_1 = x_2$. Thus f is injective. f is surjective as in (b).

   (d) $f(x_1) = f(x_2) \Rightarrow 2x_1 + 7 = 2x_2 + 7 \Rightarrow x_1 = x_2$, and f is injective. If $y \in \mathbf{R}$, then put $x = \frac{y-7}{2}$, and $f(x) = y$. Thus f is surjective.

   (e) It's not a function! $\mathbf{R}$ wants to be the domain, but the rule of assignment has not been specified, or if you prefer has been invalidly specified, for $x < 0$.

2. (a) f is surjective : given $y \in \mathbf{R}$, put $x = f(y)$, then $f(x) = f(f(y)) = y$.
   f is injective : suppose $f(x) = f(w)$, then $x = f(f(x)) = f(f(w)) = w$.

   (b) f is injective : $f(a) = f(b) \Rightarrow f(1) + f(a) = f(1) + f(b) \Rightarrow 1 + a = f(f(1) + f(a)) = f(f(1) + f(b)) = 1 + b \Rightarrow a = b$.

   It's not at all clear, without more involved work, whether or not 1 belongs to the range of f. So surjectivity we'll have to leave unanswered right now. (Later we'll solve this problem completely.)

   (c) Obviously f is surjective.

   $f(n) = f(m) \Rightarrow f(n)+2 = f(m)+2 \Rightarrow n-2 = f(f(n)+2) = f(f(m)+2) = m-2 \Rightarrow n = m$. Thus f is injective.

   (d) We have $f(f(y)) = y + (f(0))^2$. Very much like (a) we get that f is bijective.

   (e) We have $f(f(y)) = yf(1)$; again, f is bijective. (Here we use the fact that $f(1) \neq 0$.)

   (f) $f(a) = f(b) \Rightarrow f(f(f(a))) = f(f(f(b))) \Rightarrow -f(f(1)) - a = -f(f(1)) - b \Rightarrow a = b$. Thus f is injective.

   The function is surjective, but this is a bit more delicate. Put $m = 0$ in the equation to get $f(f(f(n))) = -f(f(1)) - n$. Now $-f(f(1))$ is a constant, let's denote it by α. So $f(f(f(n))) = \alpha - n$. By substitution we get $f(f(f(-\alpha - n))) = n$. (Okay?) So f is surjective.

3. (c) and (d) are strictly increasing.

4. Suppose f is strictly increasing and $f(x_1) = f(x_2)$. Then the assumption that $x_1 < x_2$ leads to $f(x_1) < f(x_2)$. Similarly $x_1 > x_2$ is impossible. Thus $x_1 = x_2$.

   A constant function, believe it or not, is increasing : and that's as far away from injective as you're going to get.

5. Clearly $f(x) = x$ is a solution. We claim it is the only solution. Suppose that $f(x) > x$ for some $x \in \mathbf{R}$. Then, since f is increasing, $f(f(x)) \geq f(x)$, that is, $x \geq f(x)$. This is a contradiction. Similarly $x > f(x)$ leads to $f(x) \geq x$.

6. (a) is even, (b) is odd, (d) is both even and odd.

7. From (i) we get that

$$\begin{aligned} & f(-2x) \\ = \ & f\left(\sin\left(\frac{-\pi x}{2} + \pi\right)\right) + f\left(\sin\left(\frac{-\pi x}{2} - \pi\right)\right) \quad (y = 2) \\ = \ & f\left(\sin\left(\frac{\pi x}{2}\right)\right) + f\left(\sin\left(\frac{\pi x}{2}\right)\right) \\ = \ & f(2x) \quad (y = 0) \end{aligned}$$

   and so f is an even function.

   Putting $x = 0$ in (ii) we get $f(-y^2) = y\,f(-y) - y\,f(y) = 0$, since f is even. Thus f is 0 on the negative numbers, and since f is even, it is zero everywhere. Thus $f = 0$.

   (Australian Mathematics Olympiad 1990)

8. $f^2(x)$ means $f(f(x))$;
$f(x)^2$ means $f(x)\,f(x)$;
$f(x^2)$ means exactly what it says.

They're all different.

9. Put $k = n = 0$ in (i). Then we get that $f(0) \in \{0, 1\}$.

Suppose $f(0) = 0$. Put $n = 0$ in (i). We get $2f(k) = 2f(k)f(0) = 0$ for all $k \in \mathbf{Z}$, and so $f = 0$. This is one of the solutions to the problem.

Suppose from now on that $f(0) = 1$. Putting $k = 0$ in (i) we get that $f(n) = f(-n)$. So it suffices to consider positive integers.

Put $n = 1$ in (i). Then

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

which shows (by induction) that the values of f are fully determined once we know $f(1)$.

If $|f(1)| \geq 2$ then it is clear from

$$f(2n) = 2f(n)^2 - 1$$

that f grows without bound (induction, again) which contradicts (ii). Thus $f(1) \in \{-1, 0, 1\}$.

- In the case that $f(1) = -1$, a simple induction shows that $f(n) = (-1)^n$;
- In the case $f(1) = 1$, a simple induction shows that $f(n) = 1$;
- In the case $f(1) = 0$, one finds that

$$f(4m) = 1,\ f(4m+1) = 0,\ f(4m+2) = -1,\ f(4m+3) = 0$$

for all m.

(1989 Australian Inter-State finals - senior division)

10. By numerical experimentation we easily find the values in the following table :-

	0	1	2	3	4	$\cdots$	x
0	1	2	3	5	13		
1	2	3	5	13			
2	3	4	7	29			
3	4	5	9	61			
4	5	$\vdots$	$\vdots$	$\vdots$			
$\vdots$							
y							

We make the following hypotheses :-

- $f(1, y) = y + 2$;
- $f(2, y) = 2y + 3$;
- $f(3, y) = 2^{y+3} - 3$.

Of course these are all proved by induction, and they have to be established in the order listed. We leave it to the reader. Now

$f(4,1) = f(3, f(4,0)) = f(3,13) = 2^{16} - 3$;
$f(4,2) = f(3, f(4,1)) = f\left(3, 2^{16} - 3\right) = 2^{2^{16}} - 3$;
$f(4,3) = f\left(3, 2^{2^{16}} - 3\right) = 2^{2^{2^{16}}} - 3$.

We hypothesise that $f(4,n) = 2^{2^{\cdot^{\cdot^{2}}}} - 3$ where there are $n+3$ 2's in the stack. Again, this follows by a trivial induction.

Thus $f(4, 1981) = 2^{2^{\cdot^{\cdot^{2}}}} - 3$ where there are 1984 2's in the stack.

(IMO 1981 Question 6)

11. It will be convenient to express (i) in the following manner : $f(m+n) = f(m) + f(n) + \epsilon$, where $\epsilon \in \{0,\ 1\}$ is dependent on m and n. In other words, f is nearly linear, except for the error quantity ϵ. (ϵ is Greek for 'e' which is short for 'error'.)

In fact, the function is *superadditive* : that is, $f(m+n) \geq f(m) + f(n)$. It is then easy to establish inductively that $f\left(\sum_{i=1}^{n} x_i\right) \geq \sum_{i=1}^{n} f(x_i)$.

Let's do some experimentation.

$0 = f(2) \geq f(1) + f(1)$, and so $f(1) = 0$;
$0 < f(3) = f(1) + f(2) + \epsilon = \epsilon$, and so $f(3) = 1$;
$f(4) = f(2) + f(2) + \epsilon = \epsilon$, and also $f(4) \geq f(1) + f(3) = 1$, and so $f(4) = 1$.

It seems impossible to decide now whether $f(5) = 1$ or $f(5) = 2$. Are we in trouble?

$f(6) = f(3)+f(3)+\epsilon = 2+\epsilon$, while $f(6) = f(2)+f(4)+\epsilon = 1+\epsilon$. (These ϵ's are different.) Thus $f(6) = 2$.

One might hypothesise, even with incomplete information, that f jumps by 1 at every multiple of 3. This is also supported by the information that $f(9999) = 3333$. This is (more or less) what we now show.

We claim that $f(3n) = n$ for all $n \leq 3333$. Since $f(3) = 1$ and f is superadditive, we certainly have $f(3n) \geq n$. Suppose that $f(3n) > n$ for some n. Then $f(3(n+1)) = f(3n+3) \geq f(3n) + f(3) > n + 1$, and so the required property fails for $3(n+1)$ too. By induction it fails for all successors of $3n$ which are multiples of 3. This contradicts the fact that the required property holds for 9999, which is a successor of $3n$ and a multiple of 3.

We claim also that $f(3n-1) = n-1$ for all $n \leq 1111$. Certainly we have $n = f(3n) \geq f(3n-1) \geq f(3n-3) = n-1$. So suppose for a contradiction that $f(3n-1) = n$. Then

$$\begin{aligned}
3n &= f(9n) \\
&\geq f(3n-1) + f(3n-1) + f(3n-1) + f(3) \\
&= n + n + n + 1 \\
&= 3n + 1
\end{aligned}$$

a contradiction as required. Thus $f(3n-1) = n-1$ for $n \leq 1111$.

Similarly we could show that $f(3n-2) = n-1$ for $n \leq 1111$, but no matter. We have all we need. Putting $n = 661$ we get $f(1982) = 660$.

(IMO 1982 Question 1)

12.

n	$f(n)$
1	1
2	3
3	6
4	2
5	7
6	1
7	8
8	16
9	7
10	17
11	6
12	18
13	5
14	19
15	4
16	20
17	3
18	21
19	2
20	22
21	1
22	23
23	46
24	22
25	47
26	21
27	48
	...

We notice the following two crucial things in the table :-

- The values of n for which $f(n) = 1$ are 1, 6, 21, ... which can be determined as follows : the next term is 3 times the previous plus 3.
- After this occurs, the values of $f(n)$ follow a nice alternating pattern until the next occurrence of a 1.

Lets prove some things.
Let the values of n for which $f(n) = 1$ be listed in order as b_1, b_2, b_3, ... Then we see that

$$\begin{aligned} f(b_n + 2j - 1) &= b_n - j + 3 \qquad (4) \\ f(b_n + 2j) &= 2b_n + j + 3 \qquad (5) \end{aligned}$$

for every n and for small j. In fact, this pattern holds until $b_n - j + 3$ reaches 1, at which point we have reached b_{n+1} and the pattern starts over again. Now, if $b_n - j + 3 = 1$ then $j = b_n + 2$ and so $b_n + 2j - 1 = 3b_n + 3$. Thus, $b_{n+1} = 3b_n + 3$, as expected.
Thus we have an inductive formula for b_n, but we are going to need a closed formula. We have

$$\begin{aligned} b_1 &= 1 \\ b_2 &= 3 \cdot 1 + 3 = 3 + 3 \\ b_3 &= 3(3+3) + 3 = 3^2 + 3^2 + 3 \\ b_4 &= 3(3^2 + 3^2 + 3) + 3 = 3^3 + 3^3 + 3^2 + 3 \\ &\dots \end{aligned}$$

and so we hypothesize that

$$b_n = 3^{n-1} + 3 \cdot \frac{3^{n-1} - 1}{3 - 1} = \frac{5 \cdot 3^{n-1} - 3}{2}$$

which is established by a trivial induction.
We then have

n	b_n
1	1
2	6
3	21
4	66
5	201
6	606
7	1821
8	5466
	...

It is clear from this and the formulae (4) and (5) that we derive 1993 via (4) only and never via (5), the first occurrence being for b_8. To derive 1993, we need $1993 = b_n - j + 3$. that is, $j = b_n - 1990$. Thus

$$b_n + 2j - 1 = b_n + 2(b_n - 1990) - 1 = 3b_n - 3981$$

This does it all. The set of numbers which map to 1993 is $\{3b_n - 3981 : n \geq 8\}$ which is an infinite set; its first member is $3 \cdot b_8 - 3981 = 12417$; and the ratio of the terms approaches 3.

(Proposed at the IMO 1993)

13. The reader should calculate the values of $f(n)$ for $1 \leq n \leq 30$. The formulas and the results obtained suggest some connection with binary representations of the numbers involved. If we do our calculations in binary, we get the following :-

n	$f(n)$
1	1
10	1
11	11
100	1
101	101
110	11
111	111
1000	1
1001	1001
1010	101
1011	1101
...	

which suggests that the general formula for f should be that $f(n)$ is the 'mirror' of n.
Of course we prove this by induction, in steps of 4. The result has already been seen to be true for $1 \leq n \leq 7$. Suppose that is is true for $1, 2, \ldots, 4n-1$; we need to establish it for $4n,\ 4n+1,\ 4n+2,\ 4n+3$.
Suppose n has the binary representation $1\alpha_2\ldots\alpha_k$ where $\alpha_i \in \{0,\ 1\}$. Then

$$\begin{aligned}
2n &= 01\alpha_2\ldots\alpha_k0 \\
2n+1 &= 01\alpha_2\ldots\alpha_k1 \\
4n &= 1\alpha_2\ldots\alpha_k00 \\
4n+1 &= 1\alpha_2\ldots\alpha_k01 \\
4n+2 &= 1\alpha_2\ldots\alpha_k10 \\
4n+3 &= 1\alpha_2\ldots\alpha_k11
\end{aligned}$$

Using the induction hypothesis and the given formulas, we get

$$\begin{aligned}
f(4n) &= f(2n) \\
&= f(1\alpha_2\ldots\alpha_k0) \\
&= 0\alpha_k\ldots\alpha_21 \\
&= 00\alpha_k\ldots\alpha_21 \\
f(4n+1) &= 2f(2n+1) - f(n) \\
&= 2\cdot f(1\alpha_2\ldots\alpha_k1) - f(1\alpha_2\ldots\alpha_k) \\
&= 2\cdot 1\alpha_k\ldots\alpha_21 - \alpha_k\ldots\alpha_21 \\
&= 1\alpha_k\ldots\alpha_21 + 1\alpha_k\ldots\alpha_21 - \alpha_k\ldots\alpha_21 \\
&= 1\alpha_k\ldots\alpha_21 + 10\ldots00 \\
&= 10\alpha_k\ldots\alpha_21 \\
f(4n+2) &= f(2n+1) \\
&= f(1\alpha_2\ldots\alpha_k1) \\
&= 1\alpha_k\ldots\alpha_21 \\
&= 01\alpha_k\ldots\alpha_21
\end{aligned}$$

$$\begin{aligned}
f(4n+3) &= 3f(2n+1) - 2f(n) \\
&= 3\cdot f(1\alpha_2\ldots\alpha_k1) - 2f(1\alpha_2\ldots\alpha_k) \\
&= 3\cdot 1\alpha_k\ldots\alpha_21 - 2\cdot\alpha_k\ldots\alpha_21 \\
&= 1\alpha_k\ldots\alpha_21 + 1\alpha_k\ldots\alpha_21 + 1\alpha_k\ldots\alpha_21 \\
&\quad +\alpha_k\ldots\alpha_21 + \alpha_k\ldots\alpha_21 \\
&= 1\alpha_k\ldots\alpha_21 + 10\ldots00 + 10\ldots00 \\
&= 11\alpha_k\cdots\alpha_21
\end{aligned}$$

thus completing all the induction steps.

Thus we will have $f(n) = n$ if and only if n is a palindrome when expressed in binary representation. Since $2^{10} = 1024 < 1988 < 2048 = 2^{11}$, we are interested in palindromes of length at most 11.

1	1
2	1
3	2
4	2
5	4
6	4
7	8
8	8
9	16
10	16
11	32

A simple counting argument shows that the number of palindromes is as indicated. Furthermore, $1988 = 11111000100$, and so there are two palindromes between 1988 and 2048, namely 11111011111 and 11111111111, which should be excluded from our count. Thus the required amount is

$$1+1+2+2+4+4+8+8+16+16+32-2 = 92$$

(IMO 1988 Question 3)

14. Suppose that $f(n) \neq g(n)$ for some $n \in \mathbf{N}$. Then $A = \{g(n) : f(n) \neq g(n)\}$ is non-empty. By the well ordering principle it has a least element, $g(n_0)$. Since f is surjective, there exists $n_1 \in \mathbf{N}$ such that $f(n_1) = g(n_0)$. Then

$$g(n_1) \leq f(n_1) = g(n_0) < f(n_0)$$

and so $n_0 \neq n_1$. Then by injectivity of g we have that $g(n_0) \neq g(n_1)$. Thus

$$g(n_1) < f(n_1) = g(n_0) < f(n_0)$$

and so $g(n_1) \in A$, and $g(n_1) < g(n_0)$, contradicting the choice of $g(n_0)$ as being the least element there.

(Romanian selection test for IMO team 1986)

15. Note that range(f), because it is a non-empty subset of $\mathbf{N}$, has a least member. Since

$$f(2) > f(f(1)) \ , \quad f(3) > f(f(2)) \ , \quad \ldots$$

it is not any of $f(2)$, $f(3)$, ... Thus $f(1)$ is the least member of range(f), and it is determined uniquely i.e. no other number is mapped to $f(1)$ by f. Since $f(1) \geq 1$ we have $f(n) > 1$ for $n > 1$. So we can, by restriction, consider the function

$$f : \mathbf{N} \setminus \{1\} \to \mathbf{N} \setminus \{1\}$$

Now the exact same argument as before yields the fact that $f(2)$ is the least member of the range of *this* function.

Thus $f(1) < f(2)$ and $f(n) > 2$ for $n > 2$ and we now consider the function

$$f : \mathbf{N} \setminus \{1,2\} \to \mathbf{N} \setminus \{1,2\}$$

Iterating (the inductive details are left to the reader) we see that

$$f(1) < f(2) < f(3) < \cdots$$

and in particular that $f(n) \geq n$ for all $n \in \mathbf{N}$.

Now assume that $f(n) > n$ for some $n \in \mathbf{N}$. Then $f(n) \geq n+1$ and so $f(f(n)) \geq f(n+1)$, since f has been shown to be increasing. But this last statement contradicts the hypothesis.

Thus $f(n) = n$ for all $n \in \mathbf{N}$.

(IMO 1977 Question 6)

16. We have from (iii) that $f(0) = 1$ and then (i) implies that $0 = f(f(0)) = f(1)$. We know what $f(0)$ and $f(1)$ are, so putting $n = -2$ and $n = -1$ in (ii) we get that $f(3) = -2$ and $f(2) = -1$. Then using (i) we get that $f(-2) = 3$ and $f(-1) = 2$. These results suggest that $f(n) = 1 - n$ which we establish by induction.

The statement $f(n) = 1 - n$ is true for $\{-2, -1, 0, 1, 2, 3\}$. Suppose the statement is true for $\{-2k, -2k+1, \ldots, 0, 1, \ldots, 2k, 2k+1\}$. To complete the induction step we need to establish that the statement is true for $\{-2k-2, -2k-1, 2k+2, 2k+3\}$. Now via the induction hypothesis and (ii) we have that

$$\begin{array}{rcccl} f(2k+2) & = & f(f(-2k+1)+2) & = & -2k-1 \\ f(2k+3) & = & f(f(-2k)+2) & = & -2k-2 \\ f(-2k-1) & = & f(f(2k+2)) & = & 2k+2 \\ f(-2k-2) & = & f(f(2k+3)) & = & 2k+3 \end{array}$$

and that completes the proof.

(Putnam Mathematics Competition 1992)

17. Putting $x = y = 0$, we get $2g(0) = 4g(0)$, and so $g(0) = 0$. Putting $x = 0$, we get $g(y) + g(-y) = 2g(y)$, and so $g(y) = g(-y)$ i.e. g is even.

Let us put $g(1) = a$. Then putting $x = y = 1$, we get $g(2) + g(0) = 2g(1) + 2g(1)$, and so g(2) = 4a. Putting $x = 2$, $y = 1$ we get $g(3) + g(1) = 2g(2) + 2g(1)$, which simplifies to g(3) = 9a.

We hypothesise that $g(n) = an^2$ for all $n \in \mathbf{N}$. This is of course proved by induction. Suppose it is true for 1, 2, ..., n. Then putting $x = n$, $y = 1$ we get

$$\begin{array}{rcl} g(n+1) + g(n-1) & = & 2g(n) + 2g(1) \\ g(n+1) & = & 2an^2 + 2a - a(n-1)^2 \\ & = & an^2 + 2an + a \\ & = & a(n+1)^2 \end{array}$$

In fact, we can improve this result : we can show

$$g(nx) = n^2 g(x)$$

for all $n \in \mathbf{N}$ and all $x \in \mathbf{R}$. The induction is identical, and so is omitted. Hence

$$an^2 = g(n) = g\left(m\,\frac{n}{m}\right) = m^2\, g\left(\frac{n}{m}\right)$$

and so

$$g\left(\frac{n}{m}\right) = a\,\frac{n^2}{m^2}$$

Thus $g(x) = ax^2$ for all $x \in \mathbf{Q}^+$. But since the function is even, it follows that $g(x) = ax^2$ for all $x \in \mathbf{Q}$.

It now follows by the continuity of g and the denseness of $\mathbf{Q}$ in $\mathbf{R}$ that $g(x) = ax^2$ for all $x \in \mathbf{R}$.

We check this solution :

$$\begin{aligned} g(x+y) + g(x-y) &= a(x+y)^2 + a(x-y)^2 \\ &= 2ax^2 + 2ay^2 \\ &= 2g(x) + 2g(y) \end{aligned}$$

and so this is a valid solution for any $a \in \mathbf{R}$.

18. From Example 4.1 we have that (i) and (ii) imply that $f(x) = x$ for all $x \in \mathbf{Q}$. To make the desired conclusion, it suffices to show that the function is continuous.

We need to show that if $h \to 0$, then $f(x+h) \to f(x)$. But $f(x+h) = f(x) + f(h)$, so it suffices to prove that $f(h) \to 0$.

First note that if a, b are of the same sign, then $|a+b| = |a| + |b|$. Now (iii) implies that for any $x \neq 0$, $f(x)$ and $f\left(\frac{1}{x}\right)$ are of the same sign. Hence

$$\begin{aligned} \left|f\left(x+\frac{1}{x}\right)\right| &= \left|f(x) + f\left(\frac{1}{x}\right)\right| \\ &= |f(x)| + \left|f\left(\frac{1}{x}\right)\right| \\ &\geq 2\sqrt{|f(x)|\left|f\left(\frac{1}{x}\right)\right|} \\ &= 2 \end{aligned}$$

by the arithmetic-geometric mean inequality. Now the range of $x + \frac{1}{x}$ as x varies through $\mathbf{R}$ is $(-\infty, -2] \cup [2, \infty)$. Thus, we have shown that $|f(y)| \geq 2$ if $|y| \geq 2$.

Thus if $|y| \leq \frac{1}{2}$ then $\left|\frac{1}{y}\right| \geq 2$ and so

$$1 = \left|f(y)\, f\left(\frac{1}{y}\right)\right| \geq |f(y)| \cdot 2$$

that is, $\frac{1}{2} \geq |f(y)|$.

Now if $|y| \leq \frac{1}{4}$ then $\frac{1}{2} \geq |f(2y)| = 2|f(y)|$ and so $\frac{1}{4} \geq |f(y)|$.

By induction, $|f(y)| \leq \frac{1}{2^n}$ for $|y| \leq \frac{1}{2^n}$. Hence $f(h) \to 0$ as $h \to 0$.

(Proposed at the IMO 1989)

19. Using 0 and 1 in (iii) we find $f\left(\frac{1}{2}\right) = a$;
using 0 and $\frac{1}{2}$ in (iii) we find $f\left(\frac{1}{4}\right) = a^2$;
using $\frac{1}{2}$ and 1 in (iii) we find $f\left(\frac{3}{4}\right) = (1-a)a + a = 2a - a^2$;
using $\frac{1}{4}$ and $\frac{3}{4}$ in (iii) we find $f\left(\frac{1}{2}\right) = (1-a)a^2 + a(2a - a^2) = 3a^2 - 2a^3$.

Thus $a = 3a^2 - 2a^3$, or $-2a\left(a - \frac{1}{2}\right)(a-1) = 0$, or $a \in \{0, \frac{1}{2}, 1\}$.

Suppose $a = 0$. Then $f\left(\frac{x+y}{2}\right) = f(x)$ whenever $0 \leq x \leq y \leq 1$. Put $x = 0$; we get $f(y) = 0$ for $0 \leq y \leq \frac{1}{2}$. Now put $x = \frac{1}{2}$; we get $f(y) = 0$ for $\frac{1}{2} \leq y \leq \frac{3}{4}$. Iterate this procedure; inductively we find that $f(y) = 0$ for $0 \leq y < 1$. This is a contradiction to the continuity of f and the fact that $f(1) = 1$.

A similar argument eliminates the possibility that $a = 1$.

Hence $a = \frac{1}{2}$. It is then clear that $f(x) = x$ provides a solution to the equation, and so $a = \frac{1}{2}$ is admissible.

(Modified from a problem proposed at the IMO 1989)

20. Putting $x = y$, we get $f(2x)\,f(0) = f(x)^2$. It follows that if $f(0) = 0$, then $f(x) = 0$. We get the trivial solution $f = 0$. So let us now suppose that $f(0) \neq 0$.

Now if $f(2x) = 0$ then $f(x) = 0$. Repeating, we would get that $f\left(\frac{x}{2}\right) = 0$, and so on. Thus we have a sequence of numbers

x, $\frac{x}{2}$, $\frac{x}{4}$, ... which get close to 0 and which map to 0. By the continuity of f, $f(0) = 0$. This is a contradiction, and so $f(x) \neq 0$ for all $x \in \mathbf{R}$.

21. Putting $x = y = 0$, we get $f(0)^2 = f(0)^4$, and so $f(0) \in \{-1, 0, 1\}$.

Putting $x = y$, we get $f(2y)\, f(0) = f(y)^4$. It follows that if $f(0) = 0$, then $f(y) = 0$. We get the trivial solution $f = 0$. So let us now suppose that $f(0) \neq 0$.

Now if $f(2y) = 0$ then $f(y) = 0$. Repeating, we would get that $f\left(\frac{y}{2}\right) = 0$, and so on. Thus we have a sequence of numbers y, $\frac{y}{2}$, $\frac{y}{4}$, ... which get close to 0 and which map to 0. By the continuity of f, $f(0) = 0$. This is a contradiction, and so $f(y) \neq 0$ for all $y \in \mathbf{R}$.

22. Consider the function $f(x) = \begin{cases} -1 & \text{if } x \in \mathbf{Q} \\ 1 & \text{if } x \notin \mathbf{Q} \end{cases}$

23. Putting $x = 1$, we get $f^2(y) = \frac{f(1)}{y}$. This shows that the required function is bijective. Putting x, $y = 1$ we get $f^2(1) = f(1)$. By the injectivity we have $f(1) = 1$. Thus $f^2(y) = \frac{1}{y}$.

Suppose $w \in \mathbf{Q}^+$, and $w = f(y)$. Then $f(w) = f^2(y) = \frac{1}{y}$. Thus $f(xw) = f(xf(y)) = \frac{f(x)}{y} = f(x)f(w)$. (A function with this property is called a multiplicative homomorphism.)

We have shown that the required function satisfies the following two 'natural' properties :-

$$\begin{aligned} f^2(y) &= \frac{1}{y} \\ f(xy) &= f(x)f(y) \end{aligned}$$

Conversely, one can easily check that a function with these two properties satisfies the given functional equation. (In actual fact, one can start the solution to the problem with this observation; all the stuff above not being needed for the solution.) Therefore we focus on these two equations rather than the given one.

Every member of $\mathbf{Q}^+$ can be expressed in the form $\prod p_i^{\alpha_i}$ where the p_i are primes and $\alpha_i \in \mathbf{Z}$. (Fractions are derived via negative α_i.) By the homomorphism property of the required function f,

$$f\left(\prod p_i^{\alpha_i}\right) = \prod f(p_i)^{\alpha_i}$$

It therefore suffices to determine the value of the function on the primes. There are many possibilities. The most elementary is as follows :-

$$f(2) = 3,\ f(3) = \frac{1}{2},\ f(5) = 7,\ f(7) = \frac{1}{5},\ \ldots$$

and then, as indicated, extend to all of $\mathbf{Q}^+$ by means of the homomorphism property.

We need to verify that this function does satisfy the equations. This is more an exercise in notation than one with serious mathematical content. It is easy to verify that the function is well defined (that is, if a member of $\mathbf{Q}^+$ is represented in two different ways then the function can recognise this and assigns the same value under either representation). By definition the function satisfies the homomorphic property. Finally,

$$f^2\left(\prod p_i^{\alpha_i}\right) = \prod f^2(p_i)^{\alpha_i} = \prod \left(\frac{1}{p_i}\right)^{\alpha_i} = \frac{1}{\prod p_i^{\alpha_i}}$$

i.e. $f^2(y) = \frac{1}{y}$.

(IMO 1990 Question 4)

24. We are going to show that $g(x) = mx + c$ where m and c are arbitrary. So suppose that $g(0) = c$ and $g(1) = m + c$. We are going to analyse the set $\mathcal{G} = \{x : g(x) = mx + c\}$.

Certainly $0,\ 1 \in \mathcal{G}$ by the way we've set things up. Putting $x = y = 1$ we get that $g(2) + c = 2(m + c)$, or $g(2) = m2 + c$, and so $2 \in \mathcal{G}$. Proceeding by induction, we show that $\mathbf{N} \subset \mathcal{G}$. Then by putting $x = 0$, $y = n$ we find that $\mathbf{Z} \subset \mathcal{G}$.

By substitution we have that $g(x)+g(y) = 2g\left(\frac{x+y}{2}\right)$ from which it easily follows that if $x,\ y \in \mathcal{G}$ then $\frac{x+y}{2} \in \mathcal{G}$. Thus $\mathcal{G}$ is a dense set. It now follows by the continuity of g that $g(x) = mx + c$ for all $x \in \mathbf{R}$. This does check :-

$$g(x+y)+g(x-y) = m(x+y)+c+m(x-y)+c = 2(mx+c) = 2g(x)$$

25. Suppose $f(n) = f(m)$. Then

$$n+m = f(f(n)+f(m)) = f(f(n)+f(n)) = n+n$$

and so $n = m$. Thus f is injective. Therefore

$$f(f(n)+f(n)) = n+n = n-1+n+1 = f(f(n-1)+f(n+1))$$

and so by the injectivity we get that

$$f(n)+f(n) = f(n-1)+f(n+1)$$

and so

$$f(n)-f(n-1) = f(n+1)-f(n)$$

Thus the slope of the function is constant, so

$$f(n) = bn + c$$

for some $b,\ c$. Substituting into the original expression,

$$m+n = f(f(n)+f(m)) = f(bn+c+bm+c) = b^2(n+m)+2bc+c$$

from which it follows that $b = 1,\ c = 0$. Thus $f(n) = n$.

(There are numerous solutions to this problem, most of them miserable. This very elegant solution is due to Richard Schneider.)

(Proposed at the 1988 IMO)

26. For the second conclusion, n needs to be any number. In other words, $m+k$ needs to be arbitrary. This would be fine if both m and k were arbitrary, but they are not. k is a function of m, and hence $m+k$ is not arbitrary. Even though m was arbitrary,

k might act in such a way as to cause $m+k$ to not be arbitrary. In this example, if we close our eyes and pretend that our final answer is correct, then we can see that $k = f^2(m+1) = m+1$, and so $m+k$ is never even, so $m+k$ is not arbitrary.

27. We have after making the substitution $a = x - y$ that

$$f(a+2y)+f(a) = 2f(a+y)\,f(y)$$

for any $y \in \mathbf{R}$. (This is not a random substitution - it is done to make $f(a)$ the subject of the equation. In the original equation, $f(x)$ cannot immediately be made the subject - we cannot divided by $2f(y)$ there, since this may be division by 0.)

Allowing $y \to \infty$ (and keeping a fixed), we get from (ii) that

$$0 + f(a) = 2 \cdot 0 \cdot 0$$

This happens because as $y \to \infty$ so too $a+2y \to \infty$ and $a+y \to \infty$. Thus $f(a) = 0$. Since a was any fixed real number, we have that $f(x) = 0$ for all $x \in \mathbf{R}$.

(Proposed at the 1985 IMO)

28. Clearly $f = 0$ is a solution. For non-trivial solutions, we have from Exercise 20 that f has no zeros. It now follows that f is always positive or always negative (for otherwise by the intermediate value theorem f would have a zero). Actually, if f is a solution to the problem then so too is $-f$, so we may suppose that $f(x) > 0$ for all $x \in \mathbf{R}$.

With this information, it is now legal to transform by taking logarithms. (Recall that logarithms are only defined for positive numbers.) Let $g(x) = \ln(f(x))$. Then the original function equation is transformed to

$$g(x+y)+g(x-y) = 2g(x)$$

which has previously been considered in Exercise 24. The solution to this function equation is $g(x) = mx+c$ for any $m,\ c \in \mathbf{R}$. Thus

$$\ln(f(x)) = mx + c$$

and so by taking exponents on both sides

$$f(x) = e^{mx+c} = K\, e^{mx}$$

where $K = e^c$ is now some positive constant. But recalling that if f is a solution then so too is $-f$, and recalling that $f = 0$ is a solution, we get

$$f(x) = Ke^{mx}$$

for any $K,\ m \in \mathbf{R}$. This does check :-

$$\begin{aligned} f(x+y)\, f(x-y) &= K\, e^{m(x+y)}\, K\, e^{m(x-y)} \\ &= K^2 e^{2mx} \\ &= (K\, e^{mx})^2 \\ &= f(x)^2 \end{aligned}$$

29. Clearly $f = 0$ is a solution. For non-trivial solutions, we have from Exercise 21 that f has no zeros. It now follows that f is always positive or always negative. If f is a solution to the problem then so too is $-f$, so we may suppose that $f(x) > 0$ for all $x \in \mathbf{R}$. Let $g(x) = \ln(f(x))$. Then the original functional equation is transformed to

$$g(x+y) + g(x-y) = 2g(x) + 2g(y)$$

which has previously been considered in Exercise 17. The solution to this function equation is $g(x) = ax^2$ for any $a \in \mathbf{R}$. Thus

$$\ln(f(x)) = ax^2$$

and so by taking exponents on both sides

$$f(x) = e^{ax^2} = K\, e^{x^2}$$

where $K = e^a$ is now some positive constant. But recalling that if f is a solution then so too is $-f$, and recalling that $f = 0$ is a solution, we get

$$f(x) = Ke^{x^2}$$

for any $K \in \mathbf{R}$. Lets check :-

$$\begin{aligned} f(x+y)\, f(x-y) &= Ke^{(x+y)^2}\, Ke^{(x-y)^2} \\ &= K^2 e^{x^4+y^4} \\ &= \left(Ke^{x^2}\right)^2 \left(Ke^{y^2}\right)^2 \\ &= f(x)^2\, f(y)^2 \end{aligned}$$

only if $K^2 = K^4$, that is, $K \in \{-1,\ 0,\ 1\}$. So the solutions are

$$f(x) = e^{x^2}, \quad f(x) = 0, \quad f(x) = -e^{x^2}$$

(Proposed at an IMO)

30. Since $x + f(y) = y + f(x)$, we get that $f(y) - f(x) = y - x$, in other words, f is a straight line function with gradient 1. Thus the general solution is $f(x) = x + c$, which checks.

31. Let us first consider the case where $\alpha = \beta$. Then $f(x)\, f(y) = y^\alpha f\left(\frac{x}{2}\right) + x^\alpha f\left(\frac{y}{2}\right)$, and so $f(x)^2 = 2x^\alpha f\left(\frac{x}{2}\right)$. Thus

$$\begin{aligned} f(x)\, f(y) &= y^\alpha \frac{f(x)^2}{2x^\alpha} + x^\alpha \frac{f(y)^2}{2y^\alpha} \\ \Rightarrow 2x^\alpha y^\alpha f(x) f(y) &= (y^\alpha f(x))^2 + (x^\alpha f(y))^2 \end{aligned}$$

and so

$$\begin{aligned} 0 &= (y^\alpha f(x))^2 - 2y^\alpha f(x) x^\alpha f(y) + (x^\alpha f(y))^2 \\ &= (y^\alpha f(x) - x^\alpha f(y))^2 \\ \Rightarrow 0 &= y^\alpha f(x) - x^\alpha f(y) \end{aligned}$$

Thus $y^\alpha f(x) = f(y) x^\alpha$, and so $f = 0$ or $f(x) = K\, x^\alpha$ for some constant K. (K is determined by choosing any y for which $f(y) \neq 0$.) In the latter case we need to determine the value of K.

By substituting into the equation $f(x)^2 = 2x^\alpha f\left(\frac{x}{2}\right)$, we get that $K^2 x^{2\alpha} = 2x^\alpha K \left(\frac{x}{2}\right)^\alpha$ and so $K = 2^{1-\alpha}$.

We need to check this solution :

$$\begin{aligned} y^\alpha f\left(\frac{x}{2}\right) + x^\alpha f\left(\frac{y}{2}\right) &= y^\alpha\, 2^{1-\alpha}\left(\frac{x}{2}\right)^\alpha + x^\alpha\, 2^{1-\alpha}\left(\frac{y}{2}\right)^\alpha \\ &= 2\cdot 2^{1-\alpha} x^\alpha y^\alpha 2^{-\alpha} \\ &= 2^{1-\alpha}x^\alpha \cdot 2^{1-\alpha} y^\alpha \\ &= f(x)f(y) \end{aligned}$$

We now consider the case where $\alpha \neq \beta$. Using symmetry, we get that

$$y^\alpha f\left(\frac{x}{2}\right) + x^\beta f\left(\frac{y}{2}\right) = x^\alpha f\left(\frac{y}{2}\right) + y^\beta f\left(\frac{x}{2}\right)$$

Hence

$$(x^\beta - x^\alpha) f\left(\frac{y}{2}\right) = (y^\beta - y^\alpha) f\left(\frac{x}{2}\right)$$

Therefore $f = 0$ or $f\left(\frac{x}{2}\right) = K\,(x^\beta - x^\alpha)$ for every $x \in \mathbf{R}^+$ and some constant K. (K is determined by choosing any y for which $f\left(\frac{y}{2}\right) \neq 0$.) Then

$$\begin{aligned} f(x)f(y) &= y^\alpha f\left(\frac{x}{2}\right) + x^\beta f\left(\frac{y}{2}\right) \\ &= K\, y^\alpha (x^\beta - x^\alpha) + K\, x^\beta (y^\beta - y^\alpha) \\ &= K(x^\beta y^\beta - x^\alpha y^\alpha) \end{aligned}$$

and so $f(x)f\left(\frac{1}{x}\right) = 0$; thus f has many zeros. However, the function $y^\beta - y^\alpha$ has only one zero. Thus $f = 0$ is the only solution in this case.

(Proposed at the IMO 1994)

32. We want to introduce symmetry into the given expression so we replace m with $f^2(m)$ and we get

$$f(f^2(m) + f^2(n)) = -f^2(f^2(m)+1) - n$$

and now by the symmetry this must also be equal to

$$-f^2(f^2(n)+1) - m$$

Hence

$$m - n = f^2(f^2(m)+1) - f^2(f^2(n)+1)$$

But, from the original equation,

$$f^2(f^2(n)+1) = f(-f^2(1+1) - n) = f(-k-n)$$

where $k = f^2(2)$. Thus

$$\begin{aligned} m - n &= f(-k-m) - f(-k-n) \\ \Rightarrow\ m + k &= f(-k-m) - f(0) \\ \Rightarrow\ -n &= f(n) - f(0) \\ \Rightarrow\ f(n) &= -n + f(0) \end{aligned}$$

for all $n \in \mathbf{N}$. It now follows that

$$f^2(n) = f(-n + f(0)) = -(-n + f(0)) + f(0) = n$$

and hence

$$f(m+n) = f(m + f^2(n)) = -f^2(m+1) - n = -(m+n) - 1$$

Thus

$$f(n) = -n - 1$$

33. The second condition implies that the fixed point equation $f(x) = x$ has at most three solutions : one in $(-1,0)$, 0 itself, and one in $(0,\infty)$. Suppose $u \in (-1,0)$ is a fixed point of f. Putting $x = y = u$ in the functional equation, we have $f(2u+u^2) = 2u + u^2$. Moreover, $2u + u^2 \in (-1,0)$. (Inspect the graph of the quadratic $2u+u^2$.) Hence $2u + u^2 = u$, but then $u \notin (-1,0)$, a contradiction.

Likewise there can be no fixed point in $(0,\infty)$. Thus the only *possible* fixed point is 0.

Are there *any* fixed points? Yes! Putting $x = y$ in the functional equation we get that $x + (1+x)f(x)$ is a fixed point for any x, and hence $x + (1+x)f(x) = 0$ for any x. Thus

$$f(x) = \frac{-x}{1+x}$$

for any x. It remains to check that all is well i.e. that this function does indeed satisfy the given properties. Certainly $\frac{f(x)}{x} = \frac{-1}{1+x}$ is strictly decreasing and (ii) is satisfied. Also by substituting into (i) we get $\frac{y-x}{1+x}$ on both the left and right hand sides - you should check this. Thus all is well.

34. We have $f^2(y) = y\,f(1)$, and since $f(1) \neq 0$, it follows that f is bijective. Hence there is a value y such that $f(y) = 1$. This together with $x = 1$ in (i) gives

$$f(1 \cdot 1) = f(1) = yf(1)$$

and since $f(1) > 0$ by hypothesis, it follows that $y = 1$, and so $f(1) = 1$. When we set $y = x$ in (i) we get

$$f(xf(x)) = xf(x)$$

for all $x > 0$. Hence $xf(x)$ is a fixed point of f.

Now if x and y are fixed points of f then (i) implies that

$$f(xy) = yx$$

so xy is also a fixed point of f. Thus the set of fixed points is closed under multiplication. Furthermore, if x is fixed point then

$$1 = f(1) = f\left(\frac{1}{x}f(x)\right) = x\,f\left(\frac{1}{x}\right)$$

and so $f\left(\frac{1}{x}\right) = \frac{1}{x}$, that is, $\frac{1}{x}$ is a fixed point. The set of fixed points is closed under inversion.

Thus if there are any fixed points besides 1, then either it or its inverse is bigger than 1 (and is a fixed point), and then the powers of this number become arbitrarily big and are all fixed points. This contradicts (ii).

Thus 1 is the only fixed point, and since $xf(x)$ is a fixed point for every x, we get that $1 = xf(x)$, or $f(x) = \frac{1}{x}$.

This checks.

(IMO 1983 Q1)

35. Fix x and keep it fixed throughout.

The characteristic polynomial is

$$\begin{aligned} 0 &= \lambda^2 + a\lambda - b(a+b) \\ &= (\lambda + a + b)(\lambda - b) \end{aligned}$$

so $\lambda = -a-b$ or $\lambda = b$, and so

$$f^n(x) = \alpha b^n + \beta(-a-b)^n$$

for some quantities α, β. By putting $n = 0$ and $n = 1$ respectively, we get

$$\begin{aligned} x &= \alpha + \beta \\ f(x) &= \alpha b - \beta(a+b) \end{aligned}$$

Now $f^n : [0,\infty) \to [0,\infty)$ and so we necessarily have that

$$0 \leq \frac{f^n(x)}{(a+b)^n} = \alpha\left(\frac{b}{a+b}\right)^n + (-1)^n\beta$$

Letting $n \to \infty$, we have that $\left(\frac{b}{a+b}\right)^n \to 0$ and so it must be that $\beta = 0$.

Thus $\alpha = x$ and $f(x) = bx$, which checks.

(Proposed at the IMO 1992)

36. Putting $y = 0$, we get

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

Putting $x = 0$, we get

$$f^2(y) = y + f(0)^2$$

The quantity $f(0)$ appears to be important. Let us set it equal to q. Thus

$$\begin{aligned} f(0) &= q & (6) \\ f(x^2+q) &= f(x)^2 & (7) \\ f^2(y) &= y + q^2 & (8) \end{aligned}$$

We are going to examine the orbit of 0. This is suggested by the previous three equations, which indicate that all elements in this orbit will be expressible in terms of q alone. This might enable us to determine q.

- $f(0) = q$
- $f(q) = f^2(0) = q^2$, from (8).
- $f(q^2) = f^3(0) = f^2(f(0)) = f^2(q) = q + q^2 = q^2 + q$, from (8). But the quantity $q^2 + q$ suggests making use of (7) in the next step.
- $f^4(0) = f^2(f^2(0)) = f^2(q^2) = 2q^2$ from (8);
  $f^4(0) = f(f^3(0)) = f(q^2 + q) = f(q)^2 = q^4$ from (7).
  Thus $2q^2 = q^4$, and $q \in \{0,\ \sqrt{2},\ -\sqrt{2}\}$.
- $f^5(0) = f^2(f^3(0)) = f^2(q^2 + q) = q^2 + q + q^2 = 2q^2 + q = q^4 + q$, again suggesting use of (7) in the next step.
- $f^6(0) = f^2(f^4(0)) = f^2(2q^2) = 3q^2$ from (8);
  $f^6(0) = f(f^5(0)) = f(q^4 + q) = f(q^2)^2 = (q^2 + q)^2$ from (7).
  Thus $3q^2 = (q^2 + q)^2 = q^4 + 2q^3 + q^2$, which is not satisfied for either $q = \sqrt{2}$ or $q = -\sqrt{2}$.

We conclude that $q = 0$. Let's rewrite the equations :-

$$f(0) = 0 \tag{9}$$
$$f(x^2) = f(x)^2 \tag{10}$$
$$f^2(y) = y \tag{11}$$

It is clear from (11) that f is surjective : given $w \in \mathbf{R}$, put $y = f(w)$, and then $f(y) = w$. Then for any $x \in \mathbf{R}$,

$$\begin{aligned} f(x^2 + w) &= f(x^2 + f(y)) \\ &= y + f(x)^2 \\ &= f(w) + f(x)^2 \\ &\geq f(w) \end{aligned}$$

which shows that f is increasing.

Referring to Exercise 5 we see that $f(x) = x$ is the only possible solution. This does check.

(IMO 1992 Question 2)

37. The desired polynomials are 0 and those of the form

$$p_{j,k,l}(x) = (x^2 + x)^j (\phi - x)^k (\overline{\phi} - x)^l$$

where $\phi = \frac{1+\sqrt{5}}{2}$, $\overline{\phi} = \frac{1-\sqrt{5}}{2}$ are the roots of $x^2 - x - 1 = 0$ and $j,\ k,\ l \geq 0$ are integers.

It is straightforward to check that the $p_{j,k,l}(x)$ do satisfy the functional equation $p(x^2 - 1) = p(x)p(-x)$. To see how they were found, observe that if α is a root of $p(x)$, then $\alpha^2 - 1$ is also. But then $(\alpha^2 - 1)^2 - 1$ is, in turn, a root; more generally, each term of the iterative sequence

$$\alpha,\ \alpha^2 - 1,\ (\alpha^2 - 1)^2 - 1 \tag{12}$$

is a root of $p(x)$. Since $p(x)$ has only finitely many roots, this sequence must eventually become periodic. The easiest way for this to happen is to have $\alpha^2 - 1 = \alpha$, which yields $\alpha = \phi$ or $\alpha = \overline{\phi}$. In this case $p(x)$ is divisible by $\phi - x$ or $\overline{\phi} - x$, respectively.

The next case to consider is $(\alpha^2 - 1)^2 - 1 = \alpha$. Besides ϕ and $\overline{\phi}$, this fourth degree equation for α has roots 0 and -1. Note that if either 0 or -1 is a root of $p(x)$, then so is the other, because $0^2 - 1 = -1$ and $(-1)^2 - 1 = 0$. So in this case $p(x)$ is divisible by $x^2 + x$.

We now show that it is not necessary to study any further cases, that is, that every non-zero polynomial $p(x)$ satisfying the given functional equation is one of the $p_{j,k,l}(x)$. Note that if $p(x)$ satisfies the given functional equation and is divisible by $p_{j,k,l}(x)$, then the polynomial $\frac{p(x)}{p_{j,k,l}(x)}$ also satisfies the functional equation. So we can divide out by any factors $\phi - x$, $\overline{\phi} - x$ and $x^2 + x$ that $p(x)$ may have, and assume that $p(x)$ does not have any of $-1,\ 0,\ \phi,\ \overline{\phi}$ as roots.

If $p(x)$ is nonconstant, let α_0 be the smallest root of $p(x)$. (Extreme case principle!) If $\alpha_0 > \phi$, then ${\alpha_0}^2 - 1 > \alpha_0$, and the sequence (12) will be strictly increasing, yielding infinitely many roots, a contradiction. If $\overline{\phi} < \alpha_0 < \phi$, then ${\alpha_0}^2 - 1 < \alpha_0$, contradicting our choice of α_0. Thus, we must have $\alpha_0 < \overline{\phi}$. However, since α_0 is a root of $p(x)$, $p(x)$ has a factor $x - \alpha_0$, and so $p(x^2 - 1)$ has a factor $x^2 - 1 - \alpha_0$. Since $p(x)$ has real roots, $p(x^2 - 1) = p(x)p(-x)$ factors into linear factors, hence $1 + \alpha_0 \geq 0$, and we have $-1 < \alpha_0 < \overline{\phi}$.

The third term in the sequence (12) is

$$\begin{aligned}({\alpha_0}^2 - 1)^2 - 1 &= \alpha_0 + (({\alpha_0}^2 - 1)^2 - 1 - \alpha_0) \\ &= \alpha_0 + \alpha_0(\alpha_0 + 1)(\alpha_0 - \phi)(\alpha_0 - \overline{\phi})\end{aligned}$$

Because $-1 < \alpha_0 < \overline{\phi}$, the product on the right is negative, which contradicts our choice of α_0. We conclude that $p(x)$ is constant. The functional equation now shows that $p(x) = 1 = p_{0,0,0}(x)$, and we are done.

(The Wohascum County Problem Book #112)

38.

$$\begin{aligned}2\,f(x) &= 2\,x\,f\left(\frac{1}{x}\right) \\ &= x\left[f\left(\frac{1}{x}\right) + f\left(\frac{1}{x}\right)\right] \\ &= x\left[1 + f\left(\frac{2}{x}\right)\right] \\ &= x\left[1 + \frac{2}{x}\,f\left(\frac{x}{2}\right)\right] \\ &= x + 2\,f\left(\frac{x}{2}\right) \\ &= x + 1 + f(x)\end{aligned}$$

and so $f(x) = x + 1$. Lets check :-

(i) $f(x) = x + 1 = 1 + x = x\left(\frac{1}{x} + 1\right) = x\,f\left(\frac{1}{x}\right)$;

(ii) $f(x) + f(y) = x + 1 + y + 1 = 1 + f(x + y)$.

(Australian Mathematics Olympiad 1991)

39. It is easy to see that $x \geq 2 \iff f(x) = 0$. Furthermore, with $0 \leq y < 2$ we have that

$$\begin{aligned}x \geq 2 - y &\iff x + y \geq 2 \\ &\iff f(x + y) = 0 \\ &\iff f(xf(y))\,f(y) = 0 \\ &\iff f(xf(y)) = 0 \\ &\iff x\,f(y) \geq 2 \\ &\iff x \geq \frac{2}{f(y)}\end{aligned}$$

and so it must be that $2 - y = \frac{2}{f(y)}$ for $0 \leq y < 2$.

Thus the only possible solution for the equation is

$$f(y) = \begin{cases} \frac{2}{2-y} & \text{if } \; 0 \leq y < 2 \\ 0 & \text{if } \; y \geq 2 \end{cases}$$

It is necessary to check that this formula is in fact a solution to the given equation. Clearly (ii) and (iii) are satisfied, so it remains only to check (i).

We will need the following key observation : if $0 \leq y < 2$ then $x + y \diamond 2 \iff xf(y) \diamond 2$, where $\diamond$ denotes any of the the five order relationships. To see this : if $x + y \diamond 2$, then $x \diamond 2 - y$, and so $xf(y) \diamond (2 - y)f(y) = 2$. (Here we use the fact that $y < 2$.)

There are now three possibilities :-

- $y \geq 2$. Then both sides of the functional equation are 0.
- $y < 2$ and $x + y \geq 2$. Then we have seen that $xf(y) \geq 2$ and so again both sides are 0.
- $x + y < 2$. Then $xf(y) < 2$ and so

$$f(xf(y))\,f(y) = \frac{2}{2 - xf(y)}\,\frac{2}{2 - y}$$

$$
\begin{aligned}
&= \frac{2}{2 - x\frac{2}{2-y}} \, \frac{2}{2-y} \\
&= \frac{4}{2(2-y) - 2x} \\
&= \frac{2}{2-(x+y)} \\
&= f(x+y)
\end{aligned}
$$

(IMO 1986 Question 5)

THE SOUTH AFRICAN COMMITTEE FOR THE IMO

Professor J H Webb (University of Cape Town) - Chairman
Professor N J H Heideman (Rhodes University)
Dr L R le Riche (University of Stellenbosch)
Professor P Persens (University of Western Cape)
Dr G P West (University of the Witwatersrand)

Regional Organizers

Mr R Anguelov (Vista University–Pretoria)
Dr W Bartoszek (University of South Africa)
Professor G L Booth (University of Transkei)
Professor I Broere (Rand Afrikaans University)
Dr F Bullock (University of South Africa)
Dr M H Burton (Rhodes University)
Dr J de Klerk (University of Potchefstroom)
Professor J C Engelbrecht (University of Pretoria)
Mr D Fish (University of Zululand)
Dr W Greybe (University of Pretoria)
Professor M J Glencross (University of Transkei)
Mr H Glover (University of Port Elizabeth)
Mr D Jacobs (University of Cape Town)
Mr P Jeganathan (University of Transkei)
Professor P J Laridon (University of the Witwatersrand)
Professor D P Laurie (University of Potchefstroom)
Dr J H Meyer (University of the Orange Free State)
Professor J Moori (University of Natal, Pietermaritzburg)
Professor P Pillay (University of Durban-Westville)
Professor J N Ridley (University of the Witwatersrand)
Professor T Schultz (University of the North-West)
Dr L C Slammert (University of the Western Cape)
Dr A J M Snyders (University of Port Elizabeth)

Tutors

Mr M Berman, Mr D Fraser, Mr D Hatton, Mr T Lawrance,
Mr J van Zyl Smit, Mr I A Webb (University of Cape Town),
Ms E Mitova, Mr A Skeen (University of the Witwatersrand)

Administration

Mrs C Sher (University of Cape Town)