MEETING 17 - FUNCTIONS

In this lecture we will study the concept of *functions* more precisely than before. Earlier we said that a function is a rule that for each element in a set A produces an element in (possibly) another set B. It was written $f : A \to B$ and we accepted this as a definition. But this definition is not precise. We will present another treatment based on relations that will be precise.

Basic terminology

Even though the view of a *function* as a machine or automaton that takes an input and produces an output is imprecise it is very valuable because we can introduce a name for the elements that are involved. The input-elements are then called the functions *arguments* of the function and the output-elements (those that are produced) are called the *images* of the function. The reason for these names will become apparent later on.

A function can be specified by laying down any rule of how the output is to be produced, there is only one requirement and that is that the output must be uniquely determined by the arguments. Of course this was clear before, but we emphasize it to be able to be clear about what we mean when we make the precise definition of a function.

A mathematical function always has two associated sets, the *domain* and the *target* (sometimes called the *codomain*). If we denote the domain of a function by A and the target by B, then we sometimes write $f : A \to B$ to denote the function and it's associated sets.

Now we will study the precise definition of what a function is:

Definition: A function from a set A (called the *domain*) to a set B (called the *target* or *codomain*) is a binary relation f from A to B with the property that for every $a \in A$ there is exactly one $b \in B$ such that $(a, b) \in f$. If $(a, b) \in f$ then we say that f *assumes* the value b in a and this is also written $b = f(a)$. (Here b is also called the *image* of a.)

* The domain of f is sometimes denoted $domf$.
* The set of elements $b \in B$ that has $(a, b) \in f$ for some $a \in A$ is called the *range* of f and it is written $rngf$. Thus
$$rngf = \{b \in B; \exists a \in A : (a, b) \in f\}.$$
* If $rngf = B$ then the function f is called *onto*, this means that f assumes all values in its target. If we want to use a french word for this we would say that the function is *surjective*, the french word *sur* means "on".
* If different elements in $domf$ have different images in B, then the function is called *one-to-one*. If we want to use french we say that the function is *injective*.
* A function that is both one-to-one and onto (that is both injective and surjective) is called bijective.

We will now start to study lots of examples to illustrate all these concepts:

Example: Suppose $A = \{1, 2, 3, 4, 5\}$ and $B = \{x, y, z, w\}$ and
$$f = \{(1, x), (2, y), (3, z), (4, y), (5, x)\}$$
This is the same thing as saying $f(1) = x, f(2) = y, f(3) = z, f(4) = y, f(5) = x$. This function has $rngf = \{x, y, z\}$, it is not one-to-one (injective), since x (and y) is an image of two different element in A. No function between the sets A and B can be one-to-one... why?

Example: Suppose $A = \{1, 2, 3, 4\}$ and $B = \{x, y, z, w, u\}$ and
$$f = \{(1, x), (2, y), (3, z), (4, w)\}$$
Since $rngf = \{x, y, z, w\} \neq \{x, y, z, w, u\} = B$, then this function is not onto (surjective). No function from A to B can be onto. Why? The function is however one-to-one (injective) since each element in B is the image of at most one element in A. This is exactlythe same thing as saying that all the images $f(1), f(2), f(3), f(4)$

are different.

Example: Suppose $A = \{1, 2, 3, 4\}$ and $B = \{x, y, z, u\}$ and

$$f = \{(1, u), (2, y), (3, y), (4, x)\}, \text{ and } g = \{(1, u), (2, z), (3, y), (4, x)\}.$$

We are defining two functions with the same domain and target. However they are different. The function f is not one-to-one (since $y = f(2) = f(3)$ and 2 and 3 are two distinct elements of A). Further since f does not assume the value z, f is not onto. If we study g however, we see that it is both onto and one-to-one, this means that g i bijective.

We can define functions by means of a mathematical expression too (and this is perhaps more familiar).

Example: Define a function $f : \mathbb{Z} \to \mathbb{Z}$ by setting

$$f(x) = 2x + 1, \text{ for every } x \in \mathbb{Z}$$

Is this function one-to-one? One way to prove this is to prove that $x \neq y \Rightarrow f(x) \neq f(y)$ but this may be a bit unwieldy. It is often better to show the contraposition: $f(x) = f(y) \Rightarrow x = y$. (Remember: $\not p \to \not q \Leftrightarrow q \to p$.) We do this now by assument that x and y are arbitrary numbers, then

$$f(x) = f(y) \Leftrightarrow 2x + 1 = 2y + 1 \Rightarrow 2x = 2y \Rightarrow x = y$$

and since we see that $x = y$ follows from $f(x) = f(y)$ the implication $f(x) = f(y) \Rightarrow x = y$ is established showing that f is one-to-one. What is the range of this function? If we look at the expression defining f we find that this expression, which is $2x + 1$, is the general expression for an odd number. As x can vary through all integers we must conclude that the range of f is the set of all odd numbers. We could also go through the hassle of saying, "Let y be an arbitrary odd number, then $y = 2k + 1$ for some $k \in \mathbb{Z}$, but then clearly $y = f(k)$ and since y was an arbitrary odd number it is show that the odd numbers are a subset of the range of f. Conversely since every image of f is odd, the range is a subset of the set of all odd numbers and hence the two sets coincide."

Example: Define $f : \mathbb{Z} \to \mathbb{Z}$ by $f(x) = 2x^4 - x$. Decide if f is onto and/or one-to-one.

Solution: We can write $f(x) = x(2x^3 - 1)$ and if $x < 0$, then this expression is positive. Similarly, if $x \geq 0$ then, as $2x^4 \geq x$ for all positive integers, the values of f becomes positive also for all positive x. This means that the whole function only assumes positive values and therefore it cannot be onto. (Because an onto function assumes all the values in it's target and the target of this function is $\mathbb{Z}$ itself which contains negative values.) Can the function be one-to-one? If we can prove $f(x) = f(y) \Rightarrow x = y$ then we are done. Assume therefore that we have $f(x) = f(y)$ but that $x \neq y$. Then

$$f(x) = f(y) \Leftrightarrow 2x^4 - x = 2y^4 - y \Leftrightarrow 2(x^4 - y^4) = x - y$$

and as $x^4 - y^4 = (x - y)(x^3 + x^2y + xy^2 + y^3)$ we can cancel $(x - y)$ (since it was assumed to be none-zero) on both sides giving the equation

$$2(x^3 + x^2y + xy^2 + y^3) = 1$$

but this is impossible since x, y are integers, here we claim that 1 is equal to an integer times 2, which is false. We have reached a contradiction which means that we must have $f(x) = f(y) \Leftrightarrow x = y$ which proves that the function is one-to-one.

We will now introduce a very simple but necessary function:

Defintion: For any set A the *identity function* $\iota_A : A \to A$ is the function defined by $\iota_A(a) = a$ for all $a \in A$. In terms of ordered pairs,

$$\iota_A = \{(a, a); a \in A\}.$$

This function can have any target set, but one prerequisite is of course that the target set contains the set A.

Functions of a real variable. From studying calculus (mathematical analysis), we are pretty used to functions of a real variable. The ones that are easiest to handle are the continuous ones (which can be integrated) and those that even are differentiable. We will make very little use of such methods when studying discrete mathematics, but of course all the concepts that are layed down about functions are valid when we are dealing with functions of a real variable. We are often used to drawing graphs of functions of a real variable and the concepts of injectivity and surjectivity translate to this:

Injectivity: Each line through the y-axis parallel to the x-axis intersects the graph of the function in at most one point.

Surjectivity: Each line through the y-axis parallel to the x-axis intersects the graph of the function in at least one point.

Bijectivity: Each line through the y-axis parallel to the x-axis intersects the graph of the function in exactly one point.

If we draw the functions of $f(x) = e^x$, $f(x) = ln(x)$, and $f(x) = 2x+1$ we can see that these functions have these three properties, injectivity, surjectivity and bijectivity. Draw them please!

The absolute value function and the floor and ceiling functions. Study them independently.

Inverses and composition. If we view a function as a machine that tranforms an object (the argument) to another object (the image) the process of inverting a function is then to construct a new function which, regarded as a machine, tranforms an image of an element back to the element to which is was an image of. This is not possible with all functions, indeed, for this to work, the function must be one-to-one. We will take an example.

Example: The function of a real variable defined by

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

can be inverted. This means that we can define a function $g(y)$ such that g undoes which f does. We find the expression for g by solving the equation $y = 2x+1 \Leftrightarrow x = (y-1)/2$, and indeed, if we compute $g(f(x))$ we get

$$g(f(x)) = \frac{f(x)-1}{2} = \frac{2x+1-1}{2} = \frac{2x}{2} = x$$

so that we see that g does the opposite of what f does. We cannot always do this, sometimes there are not function such as g. We need a formal definition to be able to talk about this properly:

Definition: A function $f : A \to B$ is said to have an *inverse* if and only if the set obtained by reversing the ordered pairs of f is a function from B to A. If $f : A \to B$ has an inverse, then the function

$$f^{-1} = \{(b, a); (a, b) \in f\}$$

is called the *inverse* of f.

Forming in an inverse entails changing the roles of a, b in the pairs that determine the function. If this new relation is a function, then the function is invertible. This is possible if and only if the function is bijective (that is both injective and surjective, both onto and one-to-one), that is we have the following:

Proposition: A function $f : A \to B$ has an inverse $f^{-1}B \to A$ if and only f is one-to-one and onto.

Proof: Assume first that there is an inverse and denote the inverse with f^{-1}. Then $f^{-1} = \{(b, a); (a, b) \in f\}$. We need to prove that f is both onto and one-to-one. Assume that $f(x) = f(y)$. This means that $(x, f(x)), (y, f(y)) \in f$ which means that $(f(x), x), (f(y), y) \in f^{-1}$ (this is the existence of an inverse function). But as $z = f(x) = f(y)$ and f^{-1} is a function $(z, x), (z, y) \in f^{-1}$ means that $x = y$ because it is this element that is the image of $z = f(x) = f(y)$. Hence $x = y$ and we have shown that $f(x) = f(y) \Rightarrow x = y$ so that f is one-to-one. Next we must prove that f is also onto. Select an arbitrary element $b \in B$. As f^{-1} is a function from B to A, we must have $(b, a) \in f^{-1}$ for some $a \in A$, again this is because of the existence of an inverse function. But by the property of the inverse we must also have $(a, b) \in f$ which mjust means that $b = f(a)$. This means that we have shown that any element $b \in B$ is an image of an element in A. This is precisely the statement that f is onto. This concludes the first part of the proof, we have now shown that the existence of an inverse implies that the function must be bijective, that is both onto and one-to-one. Assume conversely that the function f is bijective. Our job now is to find the inverse. Since f is onto, every $b \in B$ has $b = f(a)$ for at least one a. But since f is injective there can only be one such a, this means that the relation $\{(b, a); (a, b) \in f\}$ is a function, but this exactly the statement that the inverse exists, and this relation is then of course the inverse. The proof is complete.

So for the function and its inverse we have the following chain of equivalences

$$a = f^{-1}(b) \Leftrightarrow (b, a) \in f^{-1} \Leftrightarrow (a, b) \in f \Leftrightarrow b = f(a)$$

and we can use these to calculate much as we are used to from earlier studies in mathematics.

Example: Solve the equation $e^{sin(x)} = 2$. Since the inverse of the function $f(x) = e^x$ is the natural logarithm function, ln, we have, by the above equivalence,

$$e^{\sin(x)} = 2 \Leftrightarrow \sin(x) = \ln(2).$$

Now since $\sin(x)$ is a 2π-periodic function, it assumes the same values in the whole of $\mathbb{R}$ as it does in the interval $[-\pi, \pi]$ with a period of 2π, we can therefore write the set of solutions of this equation like this

$$\{x + 2\pi k; \sin(x) = \ln(2) \wedge x \in [-\pi, \pi], k \in \mathbb{Z}\}$$

and as $\ln(2) > 0$ we only have solutions when $x \in (0, \pi]$, and the solutions here are given by $x = \arcsin(\ln(2))$ and $x = \pi - arcsin(\ln(2))$. In summary all solutions of $e^{sin(x)} = 2$ can be written

$$\{\arcsin(\ln(2)) + k \cdot 2\pi; k \in \mathbb{Z}\} \cup \{\pi - \arcsin(\ln(2)) + k \cdot 2\pi; k \in \mathbb{Z}\}$$

Composition of functions. We wish now to study what happens when we apply functions in combination. We have seen this, above we composed the exponential function with the sine function to form $e^{\sin(x)}$. In general we make the following definition:

Definition: If $f : A \to B$ and $g : B \to C$ are functions the the *composition* of g and f is the function which we denote by $g \circ f : A \to C$ which is defined by $g \circ f(a) = g(f(a))$ for all $a \in A$.

We will take an example which illustrates this concept when f and g are functions defined on finite sets to really look into the details of the definition.

Example: Let $f = \{(1, x), (2, y), (3, z), (4, w)\}$ and $g = \{(x, a), (y, b), (z, c), (w, d)\}$. Then, by the definition above, $g \circ f$ is defined by

$$g \circ f(1) = g(f(1)) = g(x) = a \quad g \circ f(2) = g(f(2)) = g(y) = b$$

$$g \circ f(3) = g(f(3)) = g(z) = c \quad g \circ f(4) = g(f(4)) = g(w) = d$$

so that the function $g \circ f$ is the relation $\{(1, a), (2, b), (3, c), (4, d)\}$.

Very rarely we have $g \circ f = f \circ g$, in the example above $f \circ g$ does not even exist so it cannot be $g \circ f$. But even if both the functions $f \circ g$ and $g \circ f$ exist, it is unlikely that they are the same function.

Example: The function ι defined earlier that always map an element onto itself behaves in an interesting way when we compose it with other functions. Take for instance the function f, above, given by $f = \{(1, x), (2, y), (3, z), (4, w)\}$. What is $\iota \circ f$? Well, it is the function that is defined by

$$\{(1, \iota(f(1))), (2, \iota(f(2))), (3, \iota(f(3))), (4, \iota(f(4)))\} = \{(1, f(1)), (2, f(2)), (3, f(3)), (4, f(4)\}$$

since we always have $\iota(y) = y$ for all y. But then the function is

$$\{(1, f(1)), (2, f(2)), (3, f(3)), (4, f(4)\} = \{(1, x), (2, y), (3, z), (4, w)\}$$

which of course is just f itself. So we have $\iota \circ f = f$. In a similar way we see that $f \circ \iota = f$ so that we always have $\iota \circ f = f = f \circ \iota$. This argument applies to any function, not just the f we have here so we always have

$$\iota \circ f = f = f \circ \iota$$

for every function f. This means that the function ι composed with other functions behaves just like the number 1 when it is multiplied with other numbers, $1 \cdot x = x = x \cdot 1$ for all x and this is why we give ι the name the *identity* function.

When are two functions equal? Easy! Since functions are relations, functions are really sets and set are only equal if they contain exactly the same elements, this means that two functions are equal if and only if they have the same domain and codomain and assume exactly the same image for the same element. Tahat is the functions f and g are equal if they have the same domains and codomains and $f(a) = g(a)$ for every a in their common domain.

Composition seen as an operation on functions. Notice that with the operation of composition, we have introduced an *operation.* Addition and multiplication are operations on numbers, that is we can combine two numbers by adding them or multiplying them, and then we get a new number. Union and intersection are operations on sets, and we can combine two sets by forming their union or intersection and these are new sets. Similarly we have the operation of composition which combines two functions to produce a new function. Operations have certain properties as we have seen before, for example the operations of addition of numbers and union of sets are commutative, that is for each pair of numbers x, y and each pair of sets A, B we have

$$x + y = y + x \quad \text{and} \quad A \cup B = B \cup A$$

that is we can change the order of the operands (operand = what we operate on) and the value of the result remains the same. This was *not* the case with functions, here we had, generally

$$f \circ g \neq g \circ f$$

we then say that the composition of functions is not a commutative operation.

Commutativity is a nice property. This means, in some sense that composition of functions is not such a nice operation. But there is one property that an operation almost always has and that is called *associativity.* For numbers and sets this property is expressed as

$$x + (y + z) = (x + y) + z \quad \text{and} A \cup (B \cup C) = (A \cup B) \cup C$$

for all numbers x, y, z and all sets A, B, C and this is a property that also composition of functions has, and we formulate this as a proposition:

Proposition: Composition of functions is an associative operation.

Proof: What we need to prove can be stated like this:

$$(f \circ g) \circ h = f \circ (g \circ h)$$

for any three functions that we can form compositions of in the studied way. This just means that the functions f, g, h are compatible, that is if $f : a \to B$, then g is defined everwhere in B and if $g : B \to C$, then h is defined in C, that is $h : C \to D$ (for some set D.) If we use these names on the sets forming the domains and codomains belinging to the functions f, g, h and denote the elements in A by a, the elements in B by b, the elements in C by c, and the elements in D by d, then the proof is really very simple if we observe that for every $a \in A$, if we form the elements $b = h(a)$, $c = g(b)$, and $d = f(c)$, then the three elements b, c, d are always uniquely determined for each a. This is due to that we are dealing with functions. But first we need some terminology, denote

$$(f \circ g) \circ h \text{ by } w_1 \text{ and } f \circ (g \circ h) \text{ by } w_2,$$

then $w_1 : A \to D$ and $w_2 : A \to D$ are two functions which we want to prove are one and the same. That two functions are the same is precisely the statement

$$\forall a \in A : w_1(a) = w_2(a)$$

so to prove this we assume the opposite, that is, we assume that there is an $a \in A$ with $d' = w_1(a) \neq d'' = w_2(a)$. This written out becomes $d' = w_1(a) = ((f \circ g) \circ h)(a) \neq (f \circ (g \circ h))(a) = w_2(a) = d''$. But what are d' and d''? If we go straight to the definition of composition we get

$$d' = ((f \circ g) \circ h)(a) = (f \circ g)(h(a)) = f(g(h(a))) = f(g(b)) = f(c) = d$$

$$d'' = (f \circ (g \circ h))(a) f((g \circ h)(a)) = f(g(h(a))) = f(g(b)) = f(c) = d$$

so that they both must be equal to the d which was uniquely determined by the successive application of the functions that we work with (this was seen in the beginning of the proof). Hence we have that $d' = d = d''$ but that is a contradiction. The conclusion must be that there cannot be any $a \in A$ with the property $w_1(a) \neq w_2(a)$, hence we must have $\forall a \in A : w_1(a) = w_2(a)$ which is what we wanted to prove.

Finally we just state, without proof a proposition:

Proposition: The functions $f : A \to B$ and $g : B \to A$ are each others inverses if and only if $g \circ f = \iota_A$ and $f \circ g = \iota_B$.

Example: Let A be the set of all positive real numbers and let B be the interval $(0,1)$. Show that the functions $f : A \to B$ and $g : B \to A$ defined by

$$f(x) = \frac{1}{x^2+1} \quad \text{and} \quad g(y) = \sqrt{\frac{1}{y} - 1}$$

are each others inverses.

Solution: It is enough to show that $f \circ g = \iota_B$ and $g \circ f = \iota_A$.

1. For each $y \in B$ (=the interval $(0,1)$) we have

$$(f \circ g)(y) = f(g(y)) = \frac{1}{(g(y)^2+1)} = \frac{1}{(\sqrt{\frac{1}{y}-1})^2+1} = \frac{1}{\frac{1}{y}-1+1} = 1/(\frac{1}{y}) = y.$$

   Since this holds for all $y \in B$ we have shown $f \circ g = \iota_B$.
2. For each $x \in A$ (the set of all positive real numbers) we have

$$(g \circ f)(x) = g(f(x)) = \sqrt{\frac{1}{f(x)} - 1} = \sqrt{1/(\frac{1}{x^2}+1) - 1} = \sqrt{x^2+1-1} = \sqrt{x^2} = x.$$

   Again, since x was arbitrary we have shown $g \circ f = \iota_A$.

Since both $f \circ g = \iota_B$ and $g \circ f = \iota_A$ conclude that are each others inverses.