Review of basic concepts about functions.

This is a short review of some general stuff that you will need to know about functions. Mainly, we are going to apply these ideas in a very specific context: linear functions (or as the book prefers, linear *transformations.* The basic concepts are

1. Definition of a function.
2. Pictorial notation $f : X \to Y$ means a function that is applied to elements of the set X and produces elements of the set Y.
3. Associative law; picture of associative law.
4. Domain, co-domain (or, as I prefer, target), range of a function.
5. Image (of a subset of the domain), preimage (of a subset of the target).
6. One-to-one functions.
7. Onto functions.
8. Invertible functions.

I will assume that the idea of a function is known to you as a rule f that associates to every element of one set X (of numbers, letters, vectors, or whatever) an element of another set Y. The set X is known as the **domain** of f, and the book calls Y the **codomain** of f. I prefer the name **target** for Y because it works with the following pictorial notation:

$$f : \ X \to Y$$

This is to be read as a phrase of English: the function f *from* X *to* Y.

Example 1. Let X be the set of numbers $\{-1, 0, 1, 2, 3\}$ and let Y be the set of numbers $\{0, 1, 2, 4, 6, 9\}$. Then a function from X to Y is defined by $f(x) = x^2$. The domain is X and the target is Y.

Example 2. Let A be the set of all letters $\{a, b, \ldots, z\}$ and let B be the set of numbers $\{1, 2, \ldots, 26\}$. Then the 'secret code' $c : A \to B$ defined by $c(a) = 26, c(b) = 25, \ldots, c(z) = 1$ defines a function from A to B.

The composition of two functions is a function: if $f : A \to B$ is a function and $g : B \to C$ is a function, then the composition $h = g \circ f : A \to C$ is defined by $h(x) = g(f(x))$. Note how well this fits with the pictorial notation; the arrow that indicates h is gotten by following (consecutively) the arrows for g and f:

$$A \xrightarrow{f} B \xrightarrow{g} C$$

Composition of functions satisfies the associative law: if we have functions as indicated in the diagram

$$A \xrightarrow{f} B \xrightarrow{g} C \xrightarrow{k} D$$

then $k \circ (g \circ f) = (k \circ g) \circ f$. The reason for this is that both of these, when applied to an element $x \in A$ give $k(g(f(x)))$.

The **range** of a function $f : X \to Y$ is a subset of Y. It consists of those elements of Y that can be written as $f(x)$ for some $x \in X$. In example 1 above, the values of f are the numbers $0, 1, 4$ and 9, so the range of f is the set $\{0, 1, 4, 9\}$. In the second example, every number from 1 to 26 is g of some letter, so the range is all of B.

Example 3. This is the kind of example we'll mainly discuss. Suppose that A is an $m \times n$ matrix, then we get a function $T : \mathbb{R}^n \to \mathbb{R}^m$ defined by $T(\vec{x}) = A\vec{x}$ (matrix multiplication). We've met the range of T already: it is the set of vectors in $\mathbb{R}^m$ that can be written as $A\vec{x}$. In other terms, it is the set of vectors $\vec{b}$ for which the equation $A\vec{x} = \vec{b}$ has a solution (is consistent). If we remember that $A\vec{x}$ is the linear combination of the columns of A with coefficients given by the entries in $\vec{x}$, then we see that the range of T is the set of vectors that can be written as linear combinations of the columns of A. That is, the range of T is the same as the column space of the matrix A.

If an equation of the form $A\vec{x} = \vec{b}$ is consistent (or in the language we just learned, $\vec{b}$ is in the range of the function T) then we usually ask for all of the vectors $\vec{x}$ that satisfy the equation. There is a general idea about functions that expresses this question. Suppose that $f : X \to Y$ is a function, and that C is a subset of Y, denoted by $C \subset Y$. (That means that every element of C is an element of Y.) The **preimage** of C is the set of those elements of X that get sent to an element of C by our function. We write $f^{-1}(C)$ for the preimage. The notation is a little confusing, because there may not be any inverse function f^{-1} around. In symbols $f^{-1}(C) = \{x \in X \mid f(x) \in C\}$.

In example 1, let $C = \{0, 1, 2, 4\}$. Then $-1 \in f^{-1}(C)$ because $f(-1) = 1 \in C$, and similarly $0, 1, 2$ are in $f^{-1}(C)$. But 3 is not in $f^{-1}(C)$, because $f(3) = 9$ is not in C. Using the same function, let $D = \{2, 6\}$. Then there are no elements in X with $f(x) \in D$, so we say that $f^{-1}(D)$ is the set with no elements, otherwise known as the **empty set**. This is denoted $\emptyset$, so we would write $f^{-1}(D) = \emptyset$. Make sure not to confuse the symbol for the empty set with 0!

An important special case is when C is a set consisting of a single element c, or in symbols $C = \{c\}$. Then we would write the awkward looking $f^{-1}(\{c\})$. This comes up in example 3, when we take C to be the set containing just the vector $\vec{0}$. Then $T^{-1}(\{\vec{0}\})$ consists of all vectors $\vec{x} \in \mathbb{R}^n$ for which $A\vec{x} = \vec{0}$. Again, this is something we've studied already: the null space of A.

If we have a function $f : X \to Y$, then there are two general questions that we want to answer. The first is whether, for every choice of y, there is an x for which $f(x) = y$. The second question is how many elements of X get sent to any given y. More formally, we describe the first one by the following definition: A function is **onto** if for every $y \in Y$, there is an $x \in X$ with $f(x) = y$. Another way to say this is to say that the range of the function is all of the target. So in example 1, the function f is not onto, because there is no $x \in X$ with $f(x) = 2$. On the other hand, in example 2, every number between 1 and 26 encodes some letter in the alphabet, so the function g is onto. For our matrix example, the function T is onto when for every $\vec{b}\ in \mathbb{R}^m$, the equation $A\vec{x} = \vec{b}$ has a solution, or in other words when the column space of A is all of $\mathbb{R}^m$.

The second question (about uniqueness) leads to this definition: A function $f : X \to Y$ is **one-to-one** if $f(x_1) = f(x_2)$ implies that $x_1 = x_2$. So for example the first example f is not one-to-one, because $f(-1) = f(1)$ (but of course $-1 \neq 1$). Here's an interesting example of a one-to-one function that is not onto: consider the set N of natural numbers ($N = \{0, 1, 2, 3, 4, \ldots\}$), and define the function $d : N \to N$ to be $d(n) = 2n$. To see that d is one-to-one, suppose that $2n = 2m$. Then by dividing, we have $n = m$. On the other hand, d is not onto, because its image is just the even numbers. By the way, this example illustrates why we should specify the domain of our function; the same formula would define a function from $\mathbb{R} \to \mathbb{R}$ (ie the real numbers) which is both one-to-one and onto.

In example 3, the function $T(\vec{x}) = A\vec{x}$ is one-to-one means that if $A\vec{x} = \vec{b}$ is consistent, then there is a unique solution to the equation. We will prove a very special property of this function: T is one-to-one if and only if the equation $A\vec{x} = \vec{0}$ has a unique solution (or in other words if the null space of A consists of just the 0 vector). This follows from theorem 1.18 in section 1.6.

Finally, a function that is both one-to-one and onto is called invertible. The reason for this is that if $f : X \to Y$ has both of these properties, then we can define a function $g : Y \to X$ by saying $g(y)$ is the *unique* $x \in X$ with $f(x) = y$. We need the 'onto' property to know that for every y there is such an x, and we need the 'one-to-one' property to specify an x to choose for the value $g(y)$. To put it another way, if f were not one-to-one, then there would be x_1, x_2 with $f(x_1) = y = f(x_2)$. If this happens, then we don't know whether to take $g(y) = x_1$ or $g(y) = x_2$. If f is invertible, then we write f^{-1} for the function g and call it the inverse function of f.

The main properties of inverse functions are the two composition rules: $f(f^{-1}(y)) = y$ and $f^{-1}(f(x)) = x$. The most important example for us will be example 3: we will show that the function T is invertible if and only if the matrix A is invertible. Moreover, the inverse of the function T turns out to be given by matrix multiplication by the matrix A^{-1}.