MAT 2550, Introduction to Linear Algebra: Properties of Functions

March 27, 2020

1 Functions and relations: fundamental concepts

Learn these definitions by Wednesday, April 1!

As you have undoubtedly noticed, functions and relations are key concepts in mathematics, whether we are talking about linear maps, continuous and differentiable functions in calculus, order in inequalities, or any other subject. Functions and relations are concepts that can apply to any set of things. They are not specific to any subject, and it is invaluable to understand their general properties. Let's first review what functions and relations are.

In what follows, let A and B be sets, that is, collections of objects. These objects could be numbers, vectors, points in the plane, or whatever. We are not now concerned with anything specific about them (except in examples). A *relation* is a more general concept than a *function*; a function is a special type of relation. As an example, consider the equation $x^2 + y^2 = 1$, which relates pairs of real numbers x and y. So A is, in this specific example, the set of real numbers, and so is B, and we can picture the relation by thinking of the pair (x, y) as *belonging* to the relation or *satisfying* the relation (which are just other ways of saying that x is related to y in this specific way) if the point (x, y) lies on the unit circle in the plane. This picture follows from the fact that $x^2 + y^2$ is the square of the distance of the point (x, y) from the origin in the plane. As another example, in linear algebra we have seen linear relations given by linear equations, for example $y - x = 0$. The pairs (x, y) satisfying this relation may be viewed as the points on the the diagonal line through the origin.

So in order to talk about relations, we need to first talk about pairs of things. Unlike those in the examples above, not all relations allow us to exchange x and y. (The ones that do are called *symmetric*.) For example, $(1, 2)$ satisfies $y = 2x$, but $(2, 1)$ does not. We need to know which number is x and which is y; we do this by putting x first and y second. The order of the coordinates in the pair matters: we are talking about *ordered* pairs. The set of all possible ordered pairs, with the first coordinate taken from set A and the second from set B, is the *Cartesian product* $A \times B$, as defined below:

Definition. Given sets A and B, the *Cartesian product* $A \times B = \{(a, b) : a \in A \text{ and } b \in B\}$.

A relation from A to B, then, is simply a subset of the Cartesian product: the ordered pairs that satisfy the relation are in this subset, whereas the others are not.

Definition. Given sets A and B, a *relation R from A to B* is a subset $R \subseteq A \times B$.

If $A = B$, we refer to a relation *on A*. For example, order in the real numbers is given by the relation $\{(x, y) \in \mathbb{R} \times \mathbb{R} : x < y\} \subseteq \mathbb{R} \times \mathbb{R}$.

Of course there are also relations among more than two things, such as the one given by $x+2y-3z = 1$, but this can be reduced to a relation between the ordered pair (x, y) and z. This idea should be intuitively clear, and the formal details need not concern us. In general, we usually specify relations in words or mathematics expressions, such as equations and inequalities, rather than using the formality of the Cartesian product, but understanding the formal definition clarifies the concept.

Here are some examples of relations from geometry:

- *Incidence*: in geometry, a points P is incident with a line l if P lies on l. This is an undefined relation – we know it when we see it! – that is assumed to satisfy certain axioms, such as that any two distinct points are both incident with exactly one line.

- *Congruence* is another undefined relation in geometry. Actually there are two undefined congruence relations, the one between segments and the one between angles. The axioms assume that every segment or angle is congruent to itself, and that congruence is symmetric and transitive. We define two triangles to be congruent if their corresponding sides and angles are congruent There are various theorems telling us how much we need to know to conclude that two triangles are congruent; for example, if all the corresponding sides are congruent, so are the angles.

Now let's discuss functions. Here is the definition:

Definition. A relation f from set A to set B is a *function* if, for every element $a \in A$, there is a unique element $b \in B$ such that $(a, b) \in f$.

In less formal terms: a function relates every element of the first set to a unique element of the second set. It is a sort of assignment, like assigning names to people. Every person has a name, and (leaving aside nicknames) that name is unique. (However, two people can have the same name, and many potential names – after all, any word could be used as a name – go unused.) You are probably more familiar with functions in the usual notation that results from this definition: since there is only one element $b \in B$ associated to each element $a \in A$, we can write $b = f(a)$ to express this relation, instead of saying $(a, b) \in f$. We also write $f : A \to B$ to denote a function from A to B.

Set A is always called the *domain* of the function. Some additional terminology is customary in linear algebra. A linear function $\lambda : V \to W$, where V and W are vector spaces, is often called a *map*. (This is also true for continuous functions in calculus and topology.) The space W is called the *target space*, and the set of outputs $\{\lambda(v) : v \in V\}$ is, as you know, called the *range* of λ. (We can emphasize that the range is a subspace of W by writing it in the following more complicated way: $\text{Ran}\lambda = \{w \in W : \text{ there exists some } v \in V \text{ such that } w = \lambda(v)\}$. Note that there is no reason this element v has to be unique.) Unfortunately, the word "range" is used differently in other contexts, so you always have to pay attention to context and the terminology used by individual authors. I will use the words "domain," "target," and "range" consistently in talking about all functions, whether they are linear maps between vector spaces or not.

If $f : A \to B$ and $g : B \to C$, the *composition* $g \circ f : A \to C$ is defined by $g \circ f(a) = g(f(a))$. Composition is a binary operation on functions with suitable domains and targets. In general it is not commutative, and in fact the composition in the reverse order may well not even be defined. However, it is associative; if $h : C \to D$, we have

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

(To show to functions are equal, just show they have the same output for each input.)

You probably know all or most of what has been discussed so far. I hope this has been a useful, relaxed review. Now let's discuss some notions that may be less familiar.

There is nothing in the definition of a function that says two domain elements cannot be associated with the same element of the range. (It only says that two elements of the range cannot be associated with the same element of the domain.) For example, if $f : \mathbb{R} \to \mathbb{R}$ is the function defined by $f(x) = x^2$, $f(2) = 4 = f(-2)$. If distinct domain elements are always associated to distinct range elements, the function is called *injective*. (The word "one-to-one" is also used, but it's meaning is ambiguous, so I will not use it.) Here is a simple, logical definition:

Definition. A function $f : A \to B$ is *injective* if, for any two domain elements a and a' of the domain A, $f(a) = f(a') \Rightarrow a = a'$.

In other words, if the outputs are equal, so are the inputs (which means the same thing as, "if the inputs are unequal, so are the outputs"). We can see from the example given above that the function defined by $f(x) = x^2$ is *not* injective. By contrast, the linear function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = 2x$ is injective. Here is the proof: $f(x) = f(x') \Leftrightarrow 2x = 2x' \Leftrightarrow x = x'$. For the second equivalence, we just divide or multiply both sides by 2. The function $\lambda : \mathbb{R}^2 \to \mathbb{R}$ defined by $\lambda(x, y) = x + y$ is not injective, since, for example, $\lambda(1, 2) = \lambda(2, 1) = \lambda(3, 0) = \cdots$.

There is also nothing in the definition of a function $f : A \to B$ that says every element of B has to be associated to some domain element a. In other words, the range is a subset of the target but need not encompass all of the target. If $f : \mathbb{R} \to \mathbb{R}$ is the function defined by $f(x) = x^2$, there is no x such that $f(x) = -1$, for example. If every element of B is associated by the function with some element of A (in other words, the range is the whole target), we say the function is *surjective*. (The word "onto" is also used, and that is a good description, but I will use "surjective" since it contrasts naturally with "injective." Keep in mind, however, that these words are not opposites; a function can be both injective and surjective.) Here is a simple, logical defintion:

Definition. A function $f : A \to B$ is *surjective* if, for every $b \in B$, there is some $a \in A$ such that $f(a) = b$.

Note that there is no reason this element $a \in A$ need be unique, but if the function is also injective it will be. As an example, the linear map $\lambda : \mathbb{R}^2 \to \mathbb{R}$ defined by $\lambda(x, y) = x + y$ is surjective: for any $z \in \mathbb{R}$, we can take $x = z$ and $y = 0$, for example. We could also take $x = z + 1$ and $y = z - 1$, etc.

Definition. A function that is both injective and surjective is called *bijective*.

Definition. A bijective linear map between vector spaces is called an *isomorphism*, and the two vector spaces are said to be *isomorphic*.

Isomorphic vector spaces are essentially the same, just described in different ways. A vector space can also be isomorphic to itself in various ways, by having different vectors play the role of a basis. An example is the isomorphism $\lambda : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $\lambda(x, y) = (2x, y)$. You can visualise this isomorphism as a rescaling of the x-axis. Furthermore, an injective linear map provides an isomorphism between the domain and the range, even if it is not surjective. A surjective linear map that is not injective shows that the domain space has some of the structure of the target, but additional structure as well. For example, consider the map $\lambda : \mathbb{R}^2 \to \mathbb{R}$ defined by $\lambda(x, y) = x$. The *pre-image* of each particular value x_0, meaning the set of points $\{(x, y) : \lambda(x, y) = x_0\}$, is the vertical line passing through all the possible y values; we can visualize this map as arranging these vertical lines in a linear spectrum along the x-axis; instead of points arranged in a line, as in $\mathbb{R}$, we have lines arranged in a line. These are some of the reasons that motivate us to know if a map is injective, surjective, or bijective.

2 Application to linear maps

The following theorems make it relatively easy to tell if a linear map is injective, surjective, or bijective.

Proposition. *A linear map $\lambda : V \to W$ is injective if and only if $\lambda v = 0 \Rightarrow v = 0$.*

Proof. First we prove that λ is injective *only if* $\lambda v = 0 \Rightarrow v = 0$. (This is a useful thing to know about injective linear maps.) To prove this implication, we assume λ is injective. Assume in addition that $\lambda v = 0$. We have already proven that, for any linear map, the image of 0 is 0. So $\lambda v = \lambda 0 = 0$, and since λ is injective, $v = 0$. (Only one vector maps to 0.)

Next we prove that λ is injective *if* $\lambda v = 0 \Rightarrow v = 0$. (This is the part of the result we can use to show a given map is injective.) To prove this implication, we assume as our hypothesis that $\lambda v = 0 \Rightarrow v = 0$. (Note that this hypothesis is itself a conditional statement.) Using the defintion of injectivity, suppose in addition that $\lambda v = \lambda v'$. Then $\lambda(v - v') = \lambda v - \lambda v' = 0$, by linearity. Applying our hypothesis yields $v - v' = 0$, from which it immediately follows that $v = v'$. So λ is injective by definition. □

With this proposition we can prove another way to decide if a linear map is injective:

Theorem. *Let $\{v_1, v_2, v_3, \ldots, v_n\}$ be a basis for V, and let $\lambda : V \to W$ be a linear map. The map λ is injective if and only if the images $\{\lambda v_1, \lambda v_2, \lambda v_3, \ldots, \lambda v_n\}$ of these basis vectors are linearly independent.*

Proof. First we prove λ is injective only if $\{\lambda v_1, \lambda v_2, \lambda v_3, \ldots, \lambda v_n\}$ is a linearly independent set. Assume λ is injective and, using the definition of linear independence, assume $\sum_{i=1}^n \alpha_i \lambda v_i = 0$. By linearity, $\sum_{i=1}^n \alpha_i \lambda v_i = \lambda\left(\sum_{i=1}^n \alpha_i v_i\right)$, and the previous proposition then implies that $\sum_{i=1}^n \alpha_i v_i = 0$. But then, since $\{v_1, v_2, v_3, \ldots, v_n\}$ is a linearly independent set, $\alpha_i = 0$ for $i = 1, 2, 3, \ldots, n$. (Note that we only used that $\{v_1, v_2, v_3, \ldots, v_n\}$ was a linearly independent set of vectors, so we've actually proved something more general: a linear map is injective only if the image of any linearly independent set is linearly independent.)

Next we prove λ is injective if $\{\lambda v_1, \lambda v_2, \lambda v_3, \ldots, \lambda v_n\}$ is a linearly independent set. Assume as hypothesis that $\{\lambda v_1, \lambda v_2, \lambda v_3, \ldots, \lambda v_n\}$ is a linearly independent set. By the previous proposition, it is sufficient to check that, if $\lambda v = 0$, for some $v \in V$, then $v = 0$, so suppose $\lambda v = 0$. Since $\{\lambda v_1, \lambda v_2, \lambda v_3, \ldots, \lambda v_n\}$ is a basis, $v = \sum_{i=1}^n \alpha_i v_i$ for some choice of scalars α_i, $i = 1, 2, 3, \ldots, n$. Hence, $0 = \lambda v = \lambda\left(\sum_{i=1}^n \alpha_i v_i\right) = \sum_{i=1}^n \alpha_i \lambda v_i$. Since by hypothesis $\{\lambda v_1, \lambda v_2, \lambda v_3, \ldots, \lambda v_n\}$ is a linearly independent set, it follows that $\alpha_i = 0$ for $i = 1, 2, 3, \ldots, n$. But that means $v = \sum_{i=1}^n \alpha_i v_i = 0$. (If you carefully examine the assumptions used in the proof in this direction, you will see that we only used the fact that $\{v_1, v_2, v_3, \ldots, v_n\}$ is a spanning set for V and that their images $\{\lambda v_1, \lambda v_2, \lambda v_3, \ldots, \lambda v_n\}$ are linearly independent. So we have actually proven the more general result that a linear map is injective if the image of some spanning set is linearly independent.) $\square$

Exercise for you: Prove that if the image of a spanning set under a linear map is a linearly independent set, then that spanning set is in fact a basis. (This is a bit challenging, but it uses arguments similar to the proof above. Try it, and we'll discuss the proof in class.)

Here's how to prove a linear map is surjective:

Theorem. *A linear map $\lambda : V \to W$ is surjective if and only if the image of any basis for V spans W.*

Proof. ($\Rightarrow$) Assume λ is surjective, and let $\{v_1, v_2, v_3, \ldots, v_n\}$ be a basis for V. Let $w \in W$; we want to show that w is a linear combination of $\lambda v_1, \lambda v_2, \lambda v_3, \ldots, \lambda v_n$. Since λ is surjective, $w = \lambda v$ for some $v \in V$, and since $\{v_1, v_2, v_3, \ldots, v_n\}$ is a basis, and therefore spans V, $v = \sum_{i=1}^n \alpha_i v_i$ for some choice of scalars α_i, $i = 1, 2, 3, \ldots, n$. So $w = \lambda v = \lambda\left(\sum_{i=1}^n \alpha_i v_i\right) = \sum_{i=1}^n \alpha_i \lambda v_i$.

($\Leftarrow$) *Left to the reader.* $\square$

Finally, putting the previous two theorems together, we obtain:

Theorem. *A linear map is bijective, and hence a linear isomorphism, if and only if the image of a basis is a basis.*

Proof. Left to the reader. $\square$

Example: The linear map $\lambda : \mathbb{R}^2 \to \mathbb{R}^2$ defined by the matrix $M_\lambda = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ is a linear isomorphism. The two column vectors $\lambda \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\lambda \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ are linearly independent, since neither is a multiple of the other. Thus λ takes a basis for the domain to a basis for the target.

3 Exercises

Try these on your own, and we'll discuss them in detail in class on Friday, April 3.

1. Prove that if $\lambda : V \to W$ is injective, then $\dim V \leq \dim W$.
2. Prove that if $\lambda : V \to W$ is surjective, then $\dim V \geq \dim W$.

(Hint for the two problems above: remember that a linearly independent set can be extended to a basis, and a spanning set can be cut down to a basis.)

Determine if each of the linear maps is injective, surjective, or bijective (that is, both).

3. The map $\lambda : P_2 \to P_3$ defined by $\lambda(1) = x$, $\lambda(x) = x^2$, and $\lambda(x^2) = x^3$.

4. The map $\lambda : P_3 \to P_2$ defined by $\lambda(p) = p'$.
5. The map $\lambda : P_2 \to P_2$ defined by $\lambda(p) = p + p'$.
6. The map $\lambda : P_2 \to \mathbb{R}_3$ defined by $\lambda(1) = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $\lambda(x) = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, and $\lambda(x^2) = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$.

4 Inverses

Consider the map $\lambda : \mathbb{R}^2 \to \mathbb{R}$ defined by $\lambda(x, y) = x + y$, which is surjective (but not injective). If this map is preceded by the map $\mu : \mathbb{R} \to \mathbb{R}^2$ defined by $\mu(x) = (x, 0)$, the composition is the identity map on $\mathbb{R}$: $\lambda \circ \mu(x) = \lambda(\mu(x)) = \lambda(x, 0) = x + 0 = x$. We say that μ is a *right inverse*, or *inverse applied before*, for λ. This one-sided inverse is by no means unique; for example, $\nu(x) = (x - 1, 1)$ also defines a right inverse for λ. We also say that λ is a *left inverse* or *inverse applied after* for μ and for ν.

In terms of the standard bases for $\mathbb{R}$ and $\mathbb{R}^2$, the maps λ and μ above are given by the matrices

$$M_\lambda = \begin{bmatrix} 1 & 1 \end{bmatrix} \text{ and } M_\mu = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

You can easily check that

$$M_\lambda M_\mu = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \end{bmatrix},$$

the 1×1 identity matrix.

More generally, let $\lambda : V \to W$ be a surjective map, and let $\{w_1, w_2, \ldots, w_m\}$ be a basis for W. We can provide a linear right inverse for λ by simply choosing, for each basis element w_i, one of its *pre-images* v_i, that is, an element $v_i \in V$ such that $\lambda v_i = w_i$. The map $\mu : W \to V$ defined by $\mu w_i = v_i$ is a right inverse for λ, as you can easily see. Conversely, in order to have a right inverse, a map must be surjective, since every element of the target must have some pre-image for the inverse to map it to. This is actually a general result about functions, not just linear functions, and is a consequence of the following propostion.

Proposition. *Suppose $f : A \to B$, $g : B \to A$, and $g \circ f = id_A$, the identity map on A (that is, the map defined by $id_A(a) = a$). Then f is injective, and g is surjective.*

Proof. To prove f is injective, suppose $f(a) = f(a')$. Then $a = id_A(a) = g \circ f(a) = g(f(a)) = g(f(a')) = g \circ f(a') = id_A(a') = a'$. To prove g is surjective, let $a \in A$, and let $b = f(a)$. Then $g(b) = g(f(a)) = g \circ f(a) = id_A(a) = a$. $\square$

You can check that the right inverse maps defined in the previous examples are indeed injective, as this proposition tells us they must be. Now consider an arbitrary injective linear map $\lambda : V \to W$. Let $\{v_1, v_2, \ldots, v_n\}$ be a basis for V. Let $w_i = \lambda v_i$, for $i = 1, 2, \ldots, n$. (This will make the notation nicer.) Since λ is injective, $\{w_1, w_2, \ldots, w_n\}$ is a linearly independent set, so we can extend it to a basis $\{w_1, w_2, \ldots, w_n, w_{n+1}, w_{n+2}, \ldots w_m\}$, where m is the dimension of W. Choose any elements $v_{n+1}, v_{n+2}, \ldots, v_m \in V$. (For example, you can choose them all to be the 0 vector.) Then you can easily check that the map $\mu : W \to V$ defined by $\mu w_i = v_i$, for $i = 1, 2, \ldots, m$ is a linear left inverse for λ: $\mu \circ \lambda = id_V$. Obviously, this left inverse is not unique, since we had many choices for $v_{n+1}, v_{n+2}, \ldots, v_m$. It is surjective, as it must be by the preceding proposition, since the image of the basis $\{w_1, w_2, \ldots, w_n, w_{n+1}, w_{n+2}, \ldots w_m\}$ contains a basis for V, so it is clearly a spanning set.

Example: Consider the map $\lambda : \mathbb{R}^2 \to \mathbb{R}^3$ defined by $\lambda \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ and $\lambda \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$. You can easily check (do it!) that $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$ are linearly independent, so λ is injective. I can choose

$\left\{\begin{bmatrix}1\\1\\0\end{bmatrix}, \begin{bmatrix}0\\1\\1\end{bmatrix}, \begin{bmatrix}1\\0\\1\end{bmatrix}\right\}$ as my basis for $\mathbb{R}^3$ and define a right inverse $\mu : \mathbb{R}^3 \to \mathbb{R}^2$ for λ by $\mu \begin{bmatrix}1\\1\\0\end{bmatrix} = \begin{bmatrix}1\\0\end{bmatrix}$ (as I must), $\mu \begin{bmatrix}0\\1\\1\end{bmatrix} = \begin{bmatrix}0\\1\end{bmatrix}$ (as I must), and $\mu \begin{bmatrix}1\\0\\1\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix}$ (my choice).

To summarize, a linear map has a linear left inverse if and only if it is injective, and any left inverse is surjective. A linear map has a linear right inverse if and only if it is surjective, and any right inverse is injective. Thus, a linear map has both a left and a right inverse if and only if it is bijective. Furthermore, the following proposition shows that these one-sided inverses are the same and are unique. Thus, for any linear isomorphism there is a single unique *inverse map*, which is also an isomorphism. The argument in fact applies to one-sided inverses for any associative operation (composition in this case).

Proposition. *Let $g : A \to B$ be a bijective function, let $f : B \to A$ be a left inverse for g, and let $h : B \to A$ be a right inverse for g. Then $f = h$. It follows, furthermore, that the function f is unique, since any left inverse will be equal to the same right inverse h. (Equivalently, one could argue that any right inverse would be equal to the same left inverse f.)*

Proof. Note that $f = f \circ id_B$, since $f \circ id_B(b) = f(id_B(b)) = f(b)$. Similarly, $id_A \circ h = h$. Identity functions act as identity elements for composition. Thus we have

$$f = f \circ id_B = f \circ (g \circ h) = (f \circ g) \circ h = id_A \circ h = h.$$

□

Notation. If f is a bijective function, its unique inverse function is denoted by f^{-1}.

5 Exercises

Construct a linear right inverse for each surjective map in the previous set of exercises. Construct a linear left inverse for each injective map in the previous set of exercises. For each bijective function, you will obtain a unique inverse map.