Lesson 11

Reading: Larson, Section 6.1, Introduction to Linear Transformations; Section 6.2, The Kernel and Range of a Linear Transformation

Suggested exercises: Larson, Section 6.1: 1, 5, 9, 11, 27, 29, 37, 39; Section 6.2: 1, 9, 11, 15, 19, 27, 31, 39, 45, 47, 49.

Submit: Lesson 11: Kernel and range

Section 6.1: Introduction to Linear Transformations

A *linear transformation* is a function T from a vector space V to a vector space W that preserves the vector operations of V and W (vector addition and scalar multiplication). By "preserves," we mean that you can apply either of these operations in V and then apply T, and you will get the same result as if you applied T first and then applied the vector operation in W. In other words: for all vectors $\mathbf{u}, \mathbf{v}$ and scalar c, we have

$$\begin{aligned} T(\mathbf{u}+\mathbf{v}) &= T(\mathbf{u}) + T(\mathbf{v}) \\ T(c \cdot \mathbf{v}) &= c \cdot T(\mathbf{v}). \end{aligned}$$

On the left-hand side of the first equation, we added $\mathbf{u}$ and $\mathbf{v}$ first and then applied T, and on the right-hand side, we applied T first to each vector, and then added the results. These must be equal for a linear transformation. Similarly, on the left of the second equation, we multiplied $\mathbf{v}$ by a scalar c first and then applied T, and on the right, we applied T first, and then multiplied by c. Again, these must be the same.

Note that the sum rule applies to any finite sum, so for any vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$, we have

$$T(\mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_n) = T(\mathbf{v}_1) + T(\mathbf{v}_2) + \cdots + T(\mathbf{v}_n).$$

These requirements mean that linear transformations will respect linear relationships. For example, if $\mathbf{w}$ is a linear combination of vectors $\mathbf{v}_1, \dots, \mathbf{v}_n$, then $T(\mathbf{w})$ will be a linear combination of the vectors $T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)$. This is seen by applying the properties of a linear transformation:

$$\begin{aligned} T(\mathbf{w}) &= T(c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n) \\ &= T(c_1\mathbf{v}_1) + \cdots + T(c_n\mathbf{v}_n) \\ &= c_1T(\mathbf{v}_1) + \cdots + c_nT(\mathbf{v}_n). \end{aligned}$$

A linear transformation also has the important property of preserving subspaces, that is, if $T : V \to W$ is linear, and U is a subspace of the vector space V, then $T(U)$ will be a subspace of W. We will see this result in the next section of the text.

Note that a linear transformation $T : V \to W$ *must* send the zero vector in V to the zero vector in W: if $\mathbf{0}_V$ denotes the zero vector in V, then since $\mathbf{0}_V + \mathbf{0}_V = \mathbf{0}_V$, we have

$$T\left(\mathbf{0}_V\right) = T\left(\mathbf{0}_V + \mathbf{0}_V\right) = T\left(\mathbf{0}_V\right) + T\left(\mathbf{0}_V\right),$$

and subtracting $T\left(\mathbf{0}_V\right)$ from both sides gives

$$\mathbf{0}_W = T\left(\mathbf{0}_V\right).$$

Consequently, functions that you have probably heard referred to as linear functions, such as $f(x) = 2x + 1$, are actually not linear. This is a function from R to R, and its graph is a line, which is why it is frequently called linear, but $f(0) = 1$, so this is not a linear function. (Note: the words "function" and "transformation" are interchangeable.)

It's relatively harmless to misuse the terminology in this context, but if you want to enhance your reputation as a nitpicker, you can say things like: "Excuse me, but that function is not technically a linear function. It should be called an *affine* function, which is the correct term for a translation of a linear function." It helps if you can sound like Comic Book Guy from The Simpsons when you say this.

In fact, the only linear functions from R to R are the functions of the form $f(x) = ax$, where a is a real constant.

The fact that a linear transformation preserves linear combinations means that a linear transformation has a remarkable property: if you know its value on the basis vectors of your vector space, then you know it everywhere! This is vastly different behavior from a nonlinear function. Imagine you were told you that f is a function from R^2 to R^2, that $f(1,0) = (1,1)$, and that $f(0,1) = (-1,2)$. Could you then deduce the value of $f(3,5)$? Of course not! You know *nothing* about what a general function f might be doing there.

If you were also told that the function f was linear, then you would know *exactly* what f was doing at $(3,5)$! This is because $(3,5)$ is a linear combination of the basis vectors, so

$$\begin{aligned} f(3,5) &= f(3 \cdot (1,0) + 5 \cdot (0,1)) \\ &= 3 \cdot f(1,0) + 5 \cdot f(0,1) \\ &= 3 \cdot (1,1) + 5 \cdot (-1,2) \\ &= (-2,13). \end{aligned}$$

We are shown linear transformations $T : R^n \to R^m$ of the form

$$T(\mathbf{v}) = A\mathbf{v}$$

in this section. Note that the matrix A must be $m \times n$: $A\mathbf{v}$ is only defined if the vector $\mathbf{v}$ in R^n has as many components as there are columns in A, so A must have n columns. The result of computing $A\mathbf{v}$ must be a vector in R^m, so A must have m rows. Hence, A must be $m \times n$.

We will see later that these are actually all of the linear transformations from R^n to R^m: any linear transformation T can be associated with a matrix A and put into this form.

Section 6.2: The Kernel and Range of a Linear Transformation

The following ideas correspond:

$$\begin{array}{rcl} \text{linear transformation} & \leftrightarrow & \text{matrix} \\ \text{kernel} & \leftrightarrow & \text{nullspace} \\ \text{range} & \leftrightarrow & \text{column space} \end{array}$$

The nullspace of a matrix A is the set of vectors $\mathbf{v}$ satisfying $A\mathbf{v} = \mathbf{0}$; the kernel of a linear transformation T is the set of vectors $\mathbf{v}$ satisfying $T(\mathbf{v}) = \mathbf{0}$.

The column space of a matrix A is the set of all vectors $\mathbf{b}$ that are linear combinations of the columns of A, in other words, all vectors $\mathbf{b}$ such that

$A\mathbf{v} = \mathbf{b}$ for some $\mathbf{v}$. (Remember that $A\mathbf{v}$ can be considered as a linear combination of the columns of A, with the components of $\mathbf{v}$ as the scalars.) The range of T is the set of all vectors $\mathbf{b}$ such that $T(\mathbf{v}) = \mathbf{b}$ for some vector $\mathbf{v}$.

In fact, for a linear transformation T of the form $T(\mathbf{v}) = A\mathbf{v}$, the ideas correspond exactly: the kernel of T is the nullspace of A, and the range of T is the column space of A.

Since the ideas correspond, we use the terms rank and nullity for linear transformations to mean the corresponding things: $\text{rank}(T)$ is the dimension of the range of T, and $\text{nullity}(T)$ is the dimension of the kernel of T. If $T(\mathbf{v}) = A\mathbf{v}$ for some matrix A, then $\text{rank}(T) = \text{rank}(A)$ and $\text{nullity}(T) = \text{nullity}(A)$.

For example, consider the linear transformation $T : R^4 \to R^3$:

$$T(x_1, x_2, x_3, x_4) = (x_4, x_3 - x_2, x_1 + x_3 + x_4).$$

We find the kernel of T by setting $T(x_1, x_2, x_3, x_4) = (0, 0, 0)$:

$$\begin{aligned} x_4 &= 0 \\ x_3 - x_2 &= 0 \\ x_1 + x_3 + x_4 &= 0, \end{aligned}$$

or

$$\begin{aligned} x_1 \qquad + x_3 + x_4 &= 0 \\ -x_2 + x_3 \qquad &= 0 \\ x_4 &= 0, \end{aligned}$$

which is easily put into its reduced row-echelon form

$$\begin{aligned} x_1 \qquad + x_3 \quad &= 0 \\ x_2 - x_3 \quad &= 0 \\ x_4 &= 0. \end{aligned}$$

Setting the free variable x_3 equal to t, we get solutions

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -t \\ t \\ t \\ 0 \end{bmatrix} = t \begin{bmatrix} -1 \\ 1 \\ 1 \\ 0 \end{bmatrix}.$$

Therefore,

$$\ker(T) = \operatorname{span}\left(\begin{bmatrix} -1 \\ 1 \\ 1 \\ 0 \end{bmatrix}\right)$$

and $\operatorname{nullity}(T) = 1$.

Applying the rank-nullity theorem for linear transformations, we have

$$\operatorname{rank}(T) + \operatorname{nullity}(T) = 4,$$

so

$$\operatorname{rank}(T) + 1 = 4,$$

and therefore, $\operatorname{rank}(T) = 3$.

We will put off actually finding the range of T to the next section, when we will write this linear transformation in matrix form and apply methods developed for finding a basis for the column space of the corresponding matrix.

A transformation $T : V \to W$ is called *one-to-one* if every vector $\mathbf{b}$ in its range has exactly one preimage in V. The transformation above has many vectors whose image is the zero vector in R^3. The kernel of T is a one-dimensional subspace of R^4, and everything in this subspace is sent to $(0, 0, 0)$, so this transformation can't possibly be one-to-one.

In order for a transformation $T : V \to W$ to be one-to-one, it is, at the very least, necessary that only one vector in V is sent to the zero vector in W (and it has to be the zero vector in V). This is actually enough to guarantee that the linear transformation is one-to-one! In other words, if

$$\ker(T) = \{\mathbf{0}\},$$

then T is one-to-one. This is another property that linear transformations have that is not shared by functions in general. For example, the function $f(x) = x^2$ (which is *not* linear) has the property that $f(x) = 0$ only for $x = 0$, but f is certainly not one-to-one!

Why do linear transformations have this wonderful property, that behaving like a one-to-one function at $\mathbf{0}$ guarantees that they are one-to-one everywhere they are defined? Here's why: if $\ker(T) = \{\mathbf{0}\}$ and if for some vector $\mathbf{b}$ we had two vectors $\mathbf{u}$ and $\mathbf{v}$ with

$$T(\mathbf{u}) = T(\mathbf{v}) = \mathbf{b},$$

then

$$T(\mathbf{u}) = T(\mathbf{v}) \Rightarrow T(\mathbf{u} - \mathbf{v}) = \mathbf{0},$$

which means that $\mathbf{u} - \mathbf{v}$ is in $\ker(T)$ and is therefore $\mathbf{0}$, so

$$\mathbf{u} - \mathbf{v} = \mathbf{0} \Rightarrow \mathbf{u} = \mathbf{v}.$$

Therefore, the vector $\mathbf{b}$ actually has only one preimage, and T is therefore one-to-one.

The term *isomorphism* is used for a linear transformation $T : V \to W$ that is both one-to-one and onto, meaning every vector in W corresponds with exactly one vector in V. Since T is linear, the linear properties of V are identical to those of W. In other words, the vector spaces V and W can be considered to be the same, differing in appearance, perhaps, but not in any essential way.

For example, the vector space $M_{2,2}$ consisting of all 2×2 matrices of real numbers with standard addition of matrices and scalar multiplication is isomorphic to R^4 with its standard operations. An isomorphism is the linear transformation $T : M_{2,2} \to R^4$ given by

$$T\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}.$$

You can check that T is linear by showing that it satisfies the two properties of a linear transformation. We see that T is one-to-one by observing that

$$T\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

precisely when $a = b = c = d = 0$, that is,

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix},$$

so that

$$\ker(T) = \left\{\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}\right\},$$

which implies that T is one-to=one.

We find T to be onto by observing that, for any vector (w, x, y, z) in R^4,

$$T\left(\begin{bmatrix} w & x \\ y & z \end{bmatrix}\right) = \begin{bmatrix} w \\ x \\ y \\ z \end{bmatrix}.$$

Conclusion: T is an isomorphism, and there is no essential difference between $M_{2,2}$ and R^4 as vector spaces.

This idea of isomorphism is why, sometimes, we don't worry too much about, for example, the difference in representing vectors in R^4 as 4-tuples (x_1, x_2, x_3, x_4), as row vectors

$$\begin{bmatrix} x_1 & x_2 & x_3 & x_4 \end{bmatrix},$$

or as column vectors

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}.$$

They are different in appearance, but the vector space of 4-tuples, the vector space of 1×4 matrices, and the vector space of 4×1 matrices, each with their standard operations, are all isomorphic. (Of course, if we are planning on computing $A\mathbf{v}$ for a matrix A and a vector $\mathbf{v}$ in R^4, then we *do* have to worry about the representation.)

Isomorphism is also why we can focus on R^n in this class without feeling much guilt: it is a pleasant fact that every n-dimensional real vector space is isomorphic to R^n, so we are not really missing out on anything.