Linear transformations and matrices
Math 130 Linear Algebra

D Joyce, Fall 2015

One of the principles of modern mathematics is that functions between objects are as important as the objects themselves. The objects we're looking at are vector spaces, and the functions that preserve the structure of vector spaces are called *linear transformations*.

The structure of vector space $V$ over a field $F$ is its addition and scalar multiplication, or, if you prefer, its linear combinations.

**Definition 1.** A *linear transformation* from one vector space $V$ to another $W$ is a function $T$ that preserves vector addition and scalar multiplication. In more detail, a *linear transformation* $T : V \to W$ is a function such that

$$T(\mathbf{v} + \mathbf{w}) = T(\mathbf{v}) + T(\mathbf{w}) \quad \text{and} \quad T(c\mathbf{v}) = cL(\mathbf{v})$$

for every vector $\mathbf{v}$ and $\mathbf{w}$ in $V$ and every scalar $c$.

The vector space $V$ is called the *domain* of $T$ while the vector space $W$ is called the codomain of $T$.

The term *linear operator* is also used when $W = V$, especially when the elements of the vector spaces are functions.

We've already discussed isomorphisms. An isomorphism $T : V \overset{\cong}{\to} W$ is a linear transformation which is also a bijection; its inverse function $T^{-1} : W \overset{\cong}{\to} V$ is also an isomorphism.

**Properties of linear transformations.** A few important properties follow directly from the definition. For instance, every linear transformation sends $\mathbf{0}$ to $\mathbf{0}$.

Also, linear transformations preserve subtraction since subtraction can we written in terms of vector addition and scalar multiplication.

A more general property is that linear transformations preserve linear combinations. For example, if $\mathbf{v}$ is a certain linear combination of other vectors $\mathbf{s}$, $\mathbf{t}$, and $\mathbf{u}$, say $\mathbf{v} = 3\mathbf{s} + 5\mathbf{t} - 2\mathbf{u}$, then $T(\mathbf{v})$ is the same linear combination of the images of those vectors, that is $T(\mathbf{v}) = 3T(\mathbf{s}) + 5T(\mathbf{t}) - 2T(\mathbf{u})$.

This property can be stated as the identity

$$T(c_1\mathbf{v_1} + c_2\mathbf{v_2} + \cdots + c_k\mathbf{v_k}) = c_1T(\mathbf{v_1}) + c_2T(\mathbf{v_2}) + \cdots + c_kT(\mathbf{v_k})$$

The property of preserving linear combinations is so important, that in some texts, a linear transformation is defined as a function that preserves linear combinations.

**Example 2.** Let $V$ and $W$ both be $\mathbf{R}[x]$, the vector space of polynomials with real coefficients. You're familiar with some linear operators on it. Differentiation $\frac{d}{dx} : \mathbf{R}[x] \to \mathbf{R}[x]$ is a linear operator since (1) the derivative of a polynomial is another polynomial, (2) the derivative of a sum is the sum of derivatives, and (3) the derivative of a constant times a polynomial is the constant times the derivative of the polynomial.

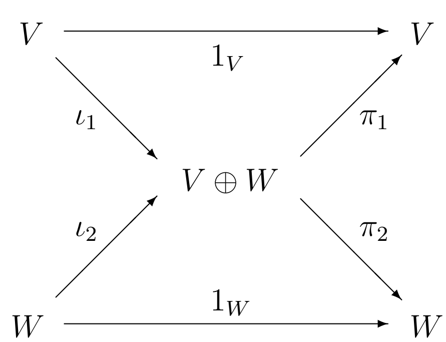
Integrals are also linear operators. The definite integral $\int_a^x$ over the interval $[a, x]$ where $a$ is any constant is also a linear operator.

The Fundamental Theorem of Calculus relates these two operators by composition. Let $D$ stand for $\frac{d}{dx}$, and $I$ stand for $\int_a^x$. Then the FTC says $IDf = f - f(a)$, while the inverse FTC says $DIf = f$. These don't quite say that $D$ and $I$ are inverse linear operators since $IDf$ is $f - f(a)$ instead of $f$ itself, but they're almost inverse operators.

**Example 3** (Projection and inclusion functions)**.** The projection functions from the product of two vector spaces to those vector spaces are linear transformations. They pick out the coordinates. The first projection $\pi_1 : V \times W \to V$ is defined by $\pi_1(\mathbf{v}, \mathbf{w}) = \mathbf{v}$, and the second projection $\pi_2 : V \times W \to W$ is defined by $\pi_1(\mathbf{v}, \mathbf{w}) = \mathbf{w}$. Since the operations on the product $V \times W$ were defined coordinatewise in the first place, these projections will be linear transformations.

There are also inclusion functions from the vector spaces to their product. $\iota_1 : V \to V \times W$ is defined by $\iota_1(\mathbf{v}) = (\mathbf{v}, \mathbf{0})$, and $\iota_2 : V \to V \times W$ is defined by $\iota_2(\mathbf{w}) = (\mathbf{0}, \mathbf{w})$.

There are some interesting connections among the projection and inclusion functions. For one thing, the composition $\pi_1 \circ \iota_1$ is the identity transformation $1_V : V \to V$ on $V$, and the composition $\pi_2 \circ \iota_2$ is the identity transformation $1_W : W \to W$ on $W$. The other compositions are 0, that is to say, $\pi_2 \circ \iota_1 : V \to W$ and $\pi_1 \circ \iota_2 : W \to V$ are both 0 transformations. Furthermore, summing the reverse compositions $(\iota_1 \circ \pi_1) + (\iota_2 \circ \pi_2) : V \times W \to V \times W$ gives the identity function $1_{V \times W} : V \times W \to V \times W$ on $V \times W$.




The symmetry of the equations and diagrams suggest that inclusion and projection functions are two halves of a larger structure. They are, and later we'll see how they're 'dual' to each other. Because of that, we'll use a different term and different notation for products of vector spaces. We'll call the product $V \times W$ the *direct sum* of the vector spaces $V$ and $W$ and denote it $V \oplus W$ from now on.

**The standard matrix for a linear transformations $T : \mathbf{R}^n \to \mathbf{R}^m$.** Our vector spaces usually have coordinates. We can treat any vector space of dimension $n$ as if it were $\mathbf{R}^n$ (or

$F^n$ if our scalar field is something other than $\mathbf{R}$). What do our linear transformations look like then?

Each vector is a linear combination of the basis vectors

$$\mathbf{v} = (v_1, \ldots, v_n) = v_1\mathbf{e}_1 + \cdots + v_n\mathbf{e}_n.$$

Therefore,

$$T(\mathbf{v}) = v_1 T(\mathbf{e}_1) + \cdots + v_n T(\mathbf{e}_n).$$

Thus, when you know where $T$ sends the basis vectors, you know where $T$ sends all the rest of the vectors. Suppose that

$$\begin{aligned} T(\mathbf{e}_1) &= (a_{11}, a_{21}, \ldots, a_{m1}) \\ T(\mathbf{e}_2) &= (a_{12}, a_{22}, \ldots, a_{m2}) \\ \ldots &= \ldots \\ T(\mathbf{e}_n) &= (a_{1n}, a_{2n}, \ldots, a_{mn}) \end{aligned}$$

Then knowing all the scalars $a_{ij}$ tells you not only where $T$ sends the basis vectors, but where it sends all the vectors.

We'll put all these scalars $a_{ij}$ in a matrix, but we'll do it column by column. We already have a notation for columns, so let's use it.

$$[T(\mathbf{e}_1)]_\epsilon = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix}, [T(\mathbf{e}_2)]_\epsilon = \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix}, \ldots, [T(\mathbf{e}_n)]_\epsilon = \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix}$$

where $\epsilon$ denotes the standard basis of $\mathbf{R}^m$. Now place them in a rectangular $m \times n$ matrix.

$$A = \left[\ [T(\mathbf{e}_1)]_\epsilon \;\middle|\; [T(\mathbf{e}_2)]_\epsilon \;\middle|\; \cdots \;\middle|\; [T(\mathbf{e}_n)]_\epsilon\ \right] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

Thus, the matrix $A$ determines the transformation $T$.

This matrix $A$, whose $j^{\text{th}}$ column is the $m$-vector $[T(\mathbf{e}_j)]_\epsilon$, is called the *standard matrix* representing the linear transformation $T$. We'll also denote it as $[T]_\epsilon^\epsilon$, where two $\epsilon$'s are used since we used one standard basis $\epsilon$ for $\mathbf{R}^n$ and another standard bases $\epsilon$ for $\mathbf{R}^m$.

Thus, a linear transformation $T : \mathbf{R}^n \to \mathbf{R}^m$ determines an $m \times n$ matrix $A$, and conversely, an $m \times n$ matrix $A$ determines a linear transformation $T : \mathbf{R}^n \to \mathbf{R}^m$.

**Evaluation.** Since the matrix $A$ holds all the information to determine the transformation $T$, we'll want to see how to use it to evaluate $T$ at a vector $\mathbf{v}$. Given $A$ and the $\mathbf{v}$, how do you compute $T(\mathbf{v})$?

Well, what are $\mathbf{v}$ and $T(\mathbf{v})$ in terms of coordinates? Let $\mathbf{v}$ have coordinates $(v_1, \ldots, v_n)$. Then

$$\begin{aligned} & T(\mathbf{v}) \\ = \; & v_1 T(\mathbf{e}_1) + v_2 T(\mathbf{e}_2) + \cdots + v_n T(\mathbf{e}_n) \\ = \; & v_1(a_{11}, a_{21}, \ldots, a_{m1}) + v_2(a_{12}, a_{22}, \ldots, a_{m2}) + \cdots + v_n(a_{1n}, a_{2n}, \ldots, a_{mn}) \\ = \; & (v_1 a_{11} + v_2 a_{12} + \cdots + v_n a_{1n}, v_1 a_{21} + v_2 a_{22} + \cdots + v_n a_{2n}, \ldots, v_1 a_{m1} + v_2 a_{m2} + \cdots + v_n a_{mn}) \end{aligned}$$

For various reasons, both historical and for simplicity, we're writing our vectors $\mathbf{v}$ and $T(\mathbf{v})$ as column vectors. Then

$$[\mathbf{v}]_\epsilon = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \quad \text{and} \quad [T(\mathbf{v})]_\epsilon = \begin{bmatrix} v_1 a_{11} + v_2 a_{12} + \cdots + v_n a_{1n} \\ v_1 a_{21} + v_2 a_{22} + \cdots + v_n a_{2n} \\ \vdots \\ v_1 a_{m1} + v_2 a_{m2} + \cdots + v_n a_{mn} \end{bmatrix}$$

Our job is to take the matrix $A$ and the column vector for $\mathbf{v}$ and produce the column vector for $T(\mathbf{v})$. Let's see what that the equation $A\mathbf{v} = T(\mathbf{v})$, or more properly, $A[\mathbf{v}]_\epsilon = [T(\mathbf{v})]_\epsilon$ looks like in terms of matrices.

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} v_1 a_{11} + v_2 a_{12} + \cdots + v_n a_{1n} \\ v_1 a_{21} + v_2 a_{22} + \cdots + v_n a_{2n} \\ \vdots \\ v_1 a_{m1} + v_2 a_{m2} + \cdots + v_n a_{mn} \end{bmatrix}$$

Aha! You see it! If you take the elements in the $i^{\text{th}}$ row from $A$ and multiply them in order with elements in the column of $\mathbf{v}$, then add those $n$ products together, you get the $i^{\text{th}}$ element of $T(\mathbf{v})$.

With this as our definition of multiplication of an $m \times n$ matrix by a $n \times 1$ column vector, we have $A\mathbf{v} = T(\mathbf{v})$.

So far we've only looked at the case when the second matrix of a product is a column vector. Later on we'll look at the general case.

**Linear operators on $\mathbf{R}^n$, eigenvectors, and eigenvalues.** Very often we are interested in the case when $m = n$. A linear transformation $T : \mathbf{R}^n \to \mathbf{R}^n$ is also called a linear transformation on $\mathbf{R}^n$ or a *linear operator* on $\mathbf{R}^n$. The standard matrix for a linear operator on $\mathbf{R}^n$ is a square $n \times n$ matrix.

One particularly important square matrix is the identity matrix $I$ whose $ij^{\text{th}}$ entry is $\delta_{ij}$, where $\delta_{ii} = 1$ but if $i \neq j$ then $\delta_{ij} = 0$. In other words, the identity matrix $I$ has 1's down the main diagonal and 0's elsewhere. It's important because it's the matrix that represents the identity transformation $\mathbf{R}^n \to \mathbf{R}^n$.

When the domain and codomain of a linear operator are the same, there are more questions you can ask of it. In particular, there may be some fixed points, that is, vectors $\mathbf{v}$ such that $T(\mathbf{v}) = \mathbf{v}$. Or there may be points sent to their negations, $T(\mathbf{v}) = -\mathbf{v}$.

When a linear operator $T$ sends a vector to a scalar multiple of itself, then that vector is called an *eigenvector* of $T$, and the scalar is called the corresponding *eigenvalue*. All the eigenvectors associated to a specific eigenvalue form an *eigenspace* for that eigenvalue.

For instance, fixed points are eigenvectors with eigenvalue 1, and the eigenspace for 1 consists of the subpace of fixed points.

We'll study eigenvectors and eigenvalues in detail later on, but we'll mention them as we come across them. They tell you something about the geometry of the linear operator.

Math 130 Home Page at `http://math.clarku.edu/~ma130/`