Notes on Change of Bases

Northwestern University, Summer 2014

Let V be a finite-dimensional vector space over a field $\mathbb{F}$, and let T be a linear operator on V. Given a basis $(v_1, \dots, v_n)$ of V, we've seen how we can define a matrix which encodes all the information about T as follows.

For each i, we can write

$$Tv_i = a_{1i}v_1 + \cdots + a_{ni}v_n$$

for a unique choice of scalars $a_{1i}, \dots, a_{ni} \in \mathbb{F}$. In total, we then have n^2 scalars a_{ij} which we put into an $n \times n$ matrix called the *matrix of* T relative to $(v_1, \dots, v_n)$:

$$\mathcal{M}(T)_v := \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \in M_{n,n}(\mathbb{F}).$$

In the notation $\mathcal{M}(T)_v$, the v showing up in the subscript emphasizes that we're taking this matrix relative to the specific bases consisting of v's. Given any vector $u \in V$, we can also write

$$u = b_1v_1 + \cdots + b_nv_n$$

for a unique choice of scalars $b_1, \dots, b_n \in \mathbb{F}$, and we define the *coordinate vector* of u relative to $(v_1, \dots, v_n)$ as

$$\mathcal{M}(u)_v := \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} \in \mathbb{F}^n.$$

In particular, the columns of $\mathcal{M}(T)_v$ are the coordinates vectors of the Tv_i. Then the point of the matrix $\mathcal{M}(T)_v$ is that the coordinate vector of Tu is given by

$$\mathcal{M}(Tu)_v = \mathcal{M}(T)_v\mathcal{M}(u)_v,$$

so that from the matrix of T and the coordinate vectors of elements of V, we can in fact reconstruct T itself. It is in this sense that $\mathcal{M}(T)_v$ encodes all information about T.

Here is a "higher-level" way of saying all this. The map

$$S : u \mapsto \mathcal{M}(u)_v$$

associating to a vector $u \in V$ its coordinate vector relative to $(v_1, \dots, v_n)$ defines an *isomorphism* $S : V \to \mathbb{R}^n$. The inverse map $S^{-1} : \mathbb{R}^n \to V$ sends a specific vector in $\mathbb{R}^n$ to the element of V determined by using the entries of that vector as the coefficients in

$$b_1v_1 + \cdots + b_nv_n.$$

Then the matrix $\mathcal{M}(T)_v$ of T fits into the following diagram:

$$\begin{tikzcd} V \arrow[r, "T"] & V \arrow[d, "S"] \\ \mathbb{R}^n \arrow[u, "S^{-1}"] \arrow[r, "\mathcal{M}(T)_v"] & \mathbb{R}^n \end{tikzcd}$$

This diagram says that the composition of the left, top, and right maps—which all together takes a vector in $\mathbb{R}^n$, applies S^{-1} to get a vector in V, then applies T to get something else in V, and finally applies S to get a vector in $\mathbb{R}^n$ again—gives the same result as matrix multiplication by $\mathcal{M}(T)_v$. That is,

$$\mathcal{M}(T)_v = STS^{-1}$$

as a composition of linear maps.

The point is that the matrix $\mathcal{M}(T)_v$ defines the map $\mathbb{R}^n \to \mathbb{R}^n$ which $T : V \to V$ becomes after "identifying" V with $\mathbb{R}^n$ using the isomorphism $S : V \to \mathbb{R}^n$. Because $S : V \to \mathbb{R}^n$ is an isomorphism, $\mathcal{M}(T)_v$ carries the same information as T.

Change of Bases

We would now like to know the extent to which the above depends on the choice of basis of V. In particular, given bases $(v_1, \ldots, v_n)$ and $(w_1, \ldots, v_n)$, we would like to know how the resulting matrices of T, $\mathcal{M}(T)_v$ and $\mathcal{M}(T)_u$, are related. To answer this question, we must first determine how coordinates vectors of elements of V depend on the choice of basis. This leads to the following definition:

Definition. The change of basis matrix from $(v_1, \ldots, v_n)$ to $(w_1, \ldots, w_n)$ is the $n \times n$ matrix defined by

$$S_{wv} := \begin{pmatrix} \mathcal{M}(v_1)_w & \cdots & \mathcal{M}(v_n)_w \end{pmatrix} \in M_{n,n}(\mathbb{F}).$$

In other words, the columns of S_{wv} are the coordinate vectors of the basis vectors $v_1, \ldots, v_n$ with respect to the basis $(w_1, \ldots, w_n)$.

The point of this matrix is that it allows us to convert coordinate vectors with respect to $(v_1, \ldots, v_n)$ into coordinate vectors with respect to $(w_1, \ldots, w_n)$ according to the following:

Proposition. *The coordinate vector of $u \in V$ with respect to $(w_1, \ldots, w_n)$ is given by*

$$\mathcal{M}(u)_w = S_{wv}\mathcal{M}(u)_v.$$

In other words, to get the coordinate vector of u with respect to $(w_1, \ldots, w_n)$ we can simply multiply the coordinate vector of u with respect to $(v_1, \ldots, v_n)$ by the change of basis matrix S_{wv}.

The subscripts in the notation S_{wv} should be read from right-to-left, so that coordinates vectors relative to v are transformed into coordinate vectors relative to w. Given the change of basis matrix S_{wv} from $(w_1, \ldots, w_n)$ to $(v_1, \ldots, v_n)$, we can ask what the change of basis matrix from $(v_1, \ldots, v_n)$ to $(w_1, \ldots, w_n)$ would be. The answer is given by:

Proposition. *The change of basis matrix S_{wv} is invertible, and the change of basis matrix S_{vw} from $(w_1, \ldots, w_n)$ to $(v_1, \ldots, v_n)$ is given by*

$$S_{vw} = S_{wv}^{-1}.$$

With these change of bases matrices at hand, we can now give an answer to our previous question, that of how the matrix of T with respect to a basis $(v_1, \ldots, v_n)$ relates to the matrix of T with respect to another basis $(w_1, \ldots, w_n)$. We have the following:

Theorem. *Let $\mathcal{M}(T)_v$ and $\mathcal{M}(T)_w$ denote the matrices of T with respect to $(v_1, \ldots, v_n)$ and $(w_1, \ldots, w_n)$ respectively. Then*

$$\mathcal{M}(T)_w = \mathcal{S}_{wv}\mathcal{M}(T)_v\mathcal{S}_{wv}^{-1}.$$

Recalling that $S_{wv}^{-1} = S_{vw}$, we should read this from right to left as follows: starting with the coordinate vector of u with respect to $(w_1, \ldots, w_n)$, first multiplying by S_{wv}^{-1} gives the coordinate vector of u with respect to $(v_1, \ldots, v_n)$, then multiplying by $\mathcal{M}(T)_v$ gives the coordinate vector of Tu with respect to $(v_1, \ldots, v_n)$, and finally multiplying by S_{wv} gives the coordinate vector of Tu with respect to $(w_1, \ldots, w_n)$, which is what we would get if we had multiplied the coordinate vector of u with respect to $(w_1, \ldots, w_n)$ by $\mathcal{M}(T)_w$.

Recall the following definition:

Definition. Two square $n \times n$ matrices A and B are similar if there exists an invertible $n \times n$ matrix S such that

$$A = SBS^{-1}.$$

The above theorem then says that the matrices of T with respect to different bases are similar; the converse is also true, and we summarize all this in the final theorem:

Theorem. *Two $n \times n$ matrices A and B represent the same linear map (with respect to possibly different bases) if and only if they are similar.*

This is how we should think about similar matrices, as representing the same linear map with respect to (possibly) different bases. Let us finish with a notion from a previous linear algebra course:

Definition. An $n \times n$ matrix A is diagonalizable if it is similar to a diagonal matrix.

From our new point of view, we see that a matrix A being diagonalizable means that there is a basis with respect to which the matrix of the linear map given by A is diagonal. The idea is that the orignal linear map may be difficult to work with (especially if we're dealing with some huge dimensional space), but finding the right basis to work with may allow us instead to deal with a diagonal matrix, which is much simpler.

Finally, you may recall the following fact (which we will review when needed): A square matrix $A \in M_{n,n}(\mathbb{R})$ is diagonalizable if and only if there exists a basis of $\mathbb{R}^n$ consisting of eigenvectors of A. This reflects what we've mentioned to be an ultimate goal for this course: studying how the structure of a vector space (in this case, the structure given by having a basis of eigenvectors of A) relates to properties of linear operators acting on it (in this case, the diagonalizability of A). We will return to this idea time and time again.