Lecture 2c

Change of Coordinates with Orthonormal Bases

(pages 325-329)

Because it is so easy to find the coordinates with respect to an orthonormal basis $\mathcal{B} = \{\vec{v}_1, \ldots, \vec{v}_n\}$, we will easily be able to find the change of coordinates matrix from the standard basis $\mathcal{S}$to $\mathcal{B}$. In fact, it will be easier than we could have hoped! After all, the change of coordinates matrix Q from $\mathcal{S}$ to $\mathcal{B}$ satisfies the following:

$$Q = \left[\begin{array}{ccc} [\vec{e}_1]_{\mathcal{B}} & \cdots & [\vec{e}_n]_{\mathcal{B}} \end{array} \right]$$

But what is $[\vec{e}_j]_{\mathcal{B}}$? If we set $[\vec{e}_j]_{\mathcal{B}} = \begin{bmatrix} b_{1j} \\ \vdots \\ b_{nj} \end{bmatrix}$, then we know that $b_{ij} = \vec{e}_j \cdot \vec{v}_i$. (Note that I have chosen this numbering system so that b_{ij} is q_{ij}, the ij-th entry of Q.) But $\vec{e}_j \cdot \vec{v}_i$ is simply the j-th component of $\vec{v}_i$. So $\vec{e}_1$ will select all the first components of our vectors in $\mathcal{B}$, $\vec{e}_2$ will select all the second components of our vectors in $\mathcal{B}$, and so on. Thus we see that Q looks like this:

$$\begin{bmatrix} \text{1st entry of } \vec{v}_1 & \text{2nd entry of } v_1 & \cdots & n\text{th entry of } v_1 \\ \text{1st entry of } \vec{v}_2 & \text{2nd entry of } v_2 & \cdots & n\text{th entry of } v_2 \\ \vdots & \vdots & & \vdots \\ \text{1st entry of } \vec{v}_n & \text{2nd entry of } v_n & \cdots & n\text{th entry of } v_n \end{bmatrix}$$

While we created Q by looking at its columns, if we now turn our attention to its rows, we see that the first row of Q lists the components of $\vec{v}_1$, the second row of Q lists the components of $\vec{v}_2$, and so on. Using our notation, we see that the i-th ROW of Q is $\begin{bmatrix} b_{i1} \\ \vdots \\ b_{in} \end{bmatrix}^T$, where the j-th entry of the i-th row b_{ij} is the j-th component of $\vec{v}_i$. That is, the i-th row of Q is $\vec{v}_i^T$.

But this means, if $P = \left[\begin{array}{ccc} \vec{v}_1 & \ldots & \vec{v}_n \end{array} \right]$ (which, amongst other things, is the $\mathcal{B}$ to $\mathcal{S}$ change of basis matrix), then we see that $Q = P^T$. But we also know that $Q = P^{-1}$. And so we have that $P^{-1} = P^T$.

Course Author's Note: The textbook proves that $P^T = P^{-1}$ by showing that $P^T P = I$, while I chose instead to go through the creation of our change of basis matrix Q and then get the result that $Q = P^T$ and so $P^T = Q = P^{-1}$.

With this result in mind, we make the following definition:

Definition: An $n \times n$ matrix P such that $P^TP = I$ is called an **orthogonal matrix**. It follows that $P^{-1} = P^T$ and that $PP^T = I = P^TP$.

That's probably not what you thought the definition would be! But the theorem below shows that an orthogonal matrix really is what we think it should be.

Theorem 7.1.3 The following are equivalent for an $n \times n$ matrix P:

(1) P is orthogonal
(2) The columns of P form an orthonormal set
(2) The rows of P form an orthonormal set

Proof of Theorem 7.1.3: Usually, if I were proving a TFAE ("the following are equivalent") statement such as this one, I would prove it in a circular fashion, such as (1) $\Rightarrow$ (2) $\Rightarrow$ (3) $\Rightarrow$ (1). But in this case, I believe it is easier to prove that (1) $\Leftrightarrow$ (2) and (1) $\Leftrightarrow$ (3) separately.

But first, we want to recall our definition of matrix multiplication:

Definition: Let B be an $m \times n$ matrix with rows $\vec{b}_1^T, \ldots, \vec{b}_m^T$ and A be an $n \times p$ matrix with columns $\vec{a}_1, \ldots, \vec{a}_p$. Then we define BA to be the matrix whose ij-th entry is $(BA)_{ij} = \vec{b}_i \cdot \vec{a}_j$.

That is, when we multiply two matrices, we take the dot product of the i-th row of the matrix on the left with the j-th column of the matrix on the right.

(1) $\Leftrightarrow$ (2) Let's let $\vec{v}_1, \ldots, \vec{v}_n$ be the columns of P. Then, substituting P^T for B and P for A in the definition of matrix multiplication, we see that the ij-th entry of P^TP is the dot product of the i-th row of P^T with the j-th column of P. But, by the definition of the transpose, we have that the i-th row of P^T is the same as the i-th column of P. And so we see that the ij-th entry of P^TP is $\vec{v}_i \cdot \vec{v}_j$.

So P is an orthogonal matrix if and only if $P^TP = I$. But P^TP is the identity matrix if and only if its ii-entries equal 1 and its ij-entries equal 0 for $i \neq j$. This happens if and only if $\vec{v}_i \cdot \vec{v}_i = 1$ and $\vec{v}_i \cdot \vec{v}_j = 0$ when $i \neq j$. By definition, this happens if and only if the set $\{\vec{v}_1, \ldots, \vec{v}_n\}$ is orthonormal.

(1) $\Leftrightarrow$ (3) This proof proceeds as above, except that we want to look at the product PP^T. Let $\vec{w}_1^T, \ldots, \vec{w}_n^T$ be the rows of P. Then the ij-th entry of PP^T is the dot product of the i-th row of P with the j-th column of P^T. But, by the definition of the transpose, we know that the j-th column of P^T is the same as the j-th row of P. As such, the ij-th entry of PP^T is $\vec{w}_i \cdot \vec{w}_j$.

So P is an orthogonal matrix if and only if $P^TP = I$, which happens if and only if $PP^T = I$. But PP^T is the identity matrix if and only if its ii-entries equal 1 and its ij-entries equal 0 for $i \neq j$. This happens if and only if $\vec{w}_i \cdot \vec{w}_i = 1$ and $\vec{w}_i \cdot \vec{w}_j = 0$ when $i \neq j$. By definition, this happens if and only if the set $\{\vec{w}_1, \ldots, \vec{w}_n\}$ is orthonormal.

Based on this theorem, you would think that we would call these matrices "orthonormal matrices", but unfortunately that is not standard practice. So, please remember that an *orthogonal* matrix must have *orthonormal* columns (and rows).

Example: To see that the set $\mathcal{B} = \left\{ \begin{bmatrix} 1/\sqrt{14} \\ 2/\sqrt{14} \\ 3/\sqrt{14} \end{bmatrix}, \begin{bmatrix} -1/\sqrt{3} \\ -1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}, \begin{bmatrix} 5/\sqrt{42} \\ -4/\sqrt{42} \\ 1/\sqrt{42} \end{bmatrix} \right\}$ is an orthonormal set, we look at the matrix P^TP, where P is the matrix whose columns are the vectors in $\mathcal{B}$:

$$\begin{bmatrix} 1/\sqrt{14} & 2/\sqrt{14} & 3/\sqrt{14} \\ -1/\sqrt{3} & -1/\sqrt{3} & 1/\sqrt{3} \\ 5/\sqrt{42} & -4/\sqrt{42} & 1/\sqrt{42} \end{bmatrix} \begin{bmatrix} 1/\sqrt{14} & -1/\sqrt{3} & 5/\sqrt{42} \\ 2/\sqrt{14} & -1/\sqrt{3} & -4/\sqrt{42} \\ 3/\sqrt{14} & 1/\sqrt{3} & 1/\sqrt{42} \end{bmatrix} =$$
$$\begin{bmatrix} (1+11+9)/14 & (1-2+3)/\sqrt{42} & (5-8+3)/\sqrt{588} \\ (-1-2+3)/\sqrt{42} & (1+1+1)/3 & (-5+4+1)/\sqrt{126} \\ (5-8+3)/\sqrt{588} & (-5+4+1)/\sqrt{126} & (25+16+1)/42 \end{bmatrix} =$$
$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Since $P^TP = I$, by definition we see that P is an orthogonal matrix, and by Theorem 7.1.3 this means that the columns of P are an orthonormal set.

Example: To see that the set $\mathcal{B} = \left\{ \begin{bmatrix} 2/\sqrt{22} \\ -3/\sqrt{22} \\ -3/\sqrt{22} \end{bmatrix}, \begin{bmatrix} 3/\sqrt{10} \\ 1/\sqrt{10} \\ 1/\sqrt{10} \end{bmatrix}, \begin{bmatrix} -1/\sqrt{6} \\ 2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix} \right\}$ is NOT orthonormal, we still look at the matrix P^TP, where P is the matrix whose columns are the vectors in $\mathcal{B}$

$$\begin{bmatrix} 2/\sqrt{22} & -3/\sqrt{22} & -3/\sqrt{22} \\ 3/\sqrt{10} & 1/\sqrt{10} & 1/\sqrt{10} \\ -1/\sqrt{6} & 2/\sqrt{6} & 1/\sqrt{6} \end{bmatrix} \begin{bmatrix} 2/\sqrt{22} & 3/\sqrt{10} & -1/\sqrt{6} \\ -3/\sqrt{22} & 1/\sqrt{10} & 2/\sqrt{6} \\ -3/\sqrt{22} & 1/\sqrt{10} & 1/\sqrt{6} \end{bmatrix}$$
$$= \begin{bmatrix} (4+9+9)/22 & (6-3-3)/\sqrt{220} & (-2-6-3)/\sqrt{132} \\ (6-3-3)/\sqrt{220} & (9+1+1)/10 & (-3+2+1)/\sqrt{60} \\ (-2-6-3)/\sqrt{132} & (-3+2+1)/\sqrt{60} & (1+4+1)/6 \end{bmatrix}$$
$$= \begin{bmatrix} 1 & 0 & -11/\sqrt{132} \\ 0 & 11/10 & 0 \\ -11/\sqrt{132} & 0 & 1 \end{bmatrix}$$

Since $P^TP \neq I$, the columns of P do not form an orthonormal set. Moreover, if we recall that the ij-th entry of P^TP is $\vec{v}_i \cdot \vec{v}_j$, we can use the product P^TP to tell us *why* $\mathcal{B}$ is not an orthonormal set. The first thing we notice is that $p_{13} \neq 0$. This means that $\vec{v}_1 \cdot \vec{v}_3 \neq 0$, i.e. that the first and third vectors in $\mathcal{B}$ are not orthogonal. We also see that $p_{31} \neq 0$, but this simply also tells us that $\vec{v}_3 \cdot \vec{v}_1 \neq 0$, which again tells us that the first and third vectors in $\mathcal{B}$ are not orthogonal. There is another reason that $\mathcal{B}$ is not an orthonormal set, which we discover when we look at p_{22}. Since $p_{22} = ||\vec{v}_2||^2$, and since $p_{22} \neq 1$, we see that $||\vec{v}_2|| \neq 1$.