CHANGE OF BASIS AND ALL OF THAT

LANCE D. DRAGER

1. Introduction

The goal of these notes is to provide an apparatus for dealing with change of basis in vector spaces, matrices of linear transformations, and how the matrix changes when the basis changes. We hope this apparatus will make these computations easier to remember and work with.

To introduce the basic idea, suppose that V is vector space and $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ is an *ordered list* of vectors from V. If a vector $\mathbf{x}$ is a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$, we have

$$\mathbf{x} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n \tag{1.1}$$

for some scalars $c_1, c_2, \dots, c_n$. The formula (1.1) can be written in matrix notation as

$$\mathbf{x} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \dots & \mathbf{v}_n \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}. \tag{1.2}$$

The object $\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \dots & \mathbf{v}_n \end{bmatrix}$ might seem a bit strange, since it is a row vector whose entries are vectors from V, rather than scalars. Nonetheless the matrix product in (1.2) makes sense. If you think about it, it makes sense to multiply a matrix of vectors with a (compatibly sized) matrix of scalars, since the entries in the product are linear combinations of vectors. It would not make sense to multiply two matrices of vectors together, unless you have some way to multiply two vectors and get another vector. You don't have that in a general vector space.

More generally than (1.1), suppose that we have vectors $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$ that are all linear combinations of $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$. The we can write

$$\begin{aligned} \mathbf{x}_1 &= a_{11}\mathbf{v}_1 + a_{21}\mathbf{v}_2 + \cdots + a_{n1}\mathbf{v}_n \\ \mathbf{x}_2 &= a_{12}\mathbf{v}_1 + a_{22}\mathbf{v}_2 + \cdots + a_{n2}\mathbf{v}_n \\ &\vdots \\ \mathbf{x}_m &= a_{1m}\mathbf{v}_1 + a_{2m}\mathbf{v}_2 + \cdots + a_{nm}\mathbf{v}_n \end{aligned} \tag{1.3}$$

for scalars a_{ij}. The indices tell you that a_{ij} occurs as the coefficient of $\mathbf{v}_i$ in the expression of $\mathbf{x}_j$. This choice of the order of the indices makes things work out nicely later. These equations can be summarized as

$$\mathbf{x}_j = \sum_{i=1}^{n} \mathbf{v}_i a_{ij}, \qquad j = 1, 2, \dots, m, \tag{1.4}$$

where I'm writing the scalars a_{ij} to the right of the vector to make the equation easier to remember; it means the same thing as writing the scalars on the left.

Using our previous idea, we can write (1.3) in matrix form as

$$\begin{bmatrix}\mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_m\end{bmatrix} = \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \dots & \mathbf{v}_n\end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix} \tag{1.5}$$

To write (1.5) in briefer form, let

$$\mathcal{X} = \begin{bmatrix}\mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_m\end{bmatrix}, \qquad \mathcal{V} = \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \dots & \mathbf{v}_n\end{bmatrix}$$

and let A be the $n \times m$ matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix}.$$

Then we can write (1.5) as

$$\mathcal{X} = \mathcal{V}A. \tag{1.6}$$

The object $\mathcal{V}$ is a row vector whose entries are vectors. For brevity, we'll refer to this as "a row of vectors." We'll usually use upper case script letters for a row of vectors, with the corresponding lower case bold letters for the entries. Thus, if we say that $\mathcal{U}$ is a row of k vectors, you know that $\mathcal{U} = \begin{bmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_k\end{bmatrix}$.

You want to remember that the short form (1.6) is equivalent to (1.3) and and (1.4). It is also good to recall, from the usual definition of matrix multiplication, that (1.6) means

$$\mathbf{x}_j = \mathcal{V}\operatorname{col}_j(A), \tag{1.7}$$

where $\operatorname{col}_j(A)$ is the column vector formed by the j-th column of A.

The multiplication between rows of vectors and matrices of scalars has the following properties, similar to matrix algebra with matrices of scalars.

Theorem 1.1. *Let V be a vector space and let $\mathcal{U}$ be a row of vectors from V. Let A and B denotes scalar matrices, with sizes appropriate to the operation being considered. Then we have the following.*

1. *$\mathcal{U}I = \mathcal{U}$, where I is the identity matrix of the right size.*
2. *$(\mathcal{U}A)B = \mathcal{U}(AB)$, the "associative" law.*
3. *$(s\mathcal{U})A = \mathcal{U}(sA) = s(\mathcal{U}A)$ for any scalar s.*
4. *$\mathcal{U}(A + B) = \mathcal{U}A + \mathcal{U}B$, the distributive law.*

The proofs of these properties are the same as the proof of the corresponding statements for the algebra of scalar matrices. They're not illuminating enough to write out here. The associative law will be particularly important in what follows.

2. Ordered Bases

Let V be a vector space. An *ordered basis* of V is an ordered list of vectors, say $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$, that span V and are linearly independent. We can think of this ordered list as a row of vectors, say $\mathcal{V} = \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \dots & \mathbf{v}_n\end{bmatrix}$. Since $\mathcal{V}$ is an ordered basis, we know that any vector $\mathbf{v} \in V$ can be written *uniquely* as

$$\begin{aligned}\mathbf{v} &= c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n \\ &= \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \dots & \mathbf{v}_n\end{bmatrix}\begin{bmatrix}c_1\\ c_2\\ \vdots\\ c_n\end{bmatrix} \\ &= \mathcal{V}\mathbf{c},\end{aligned}$$

where $\mathbf{c}$ is the column vector

$$\mathbf{c} = \begin{bmatrix}c_1\\ c_2\\ \vdots\\ c_n\end{bmatrix}.$$

Thus, another way to state the properties of an ordered basis $\mathcal{V}$ is that for every vector $\mathbf{v} \in V$ there is a *unique* column vector $\mathbf{c} \in \mathbb{R}^n$ such that

$$\mathbf{v} = \mathcal{V}\mathbf{c}.$$

This unique column vector is called the *coordinate vector* of $\mathbf{v}$ with respect to the basis $\mathcal{V}$. We will use the notation $[\mathbf{v}]_{\mathcal{V}}$ for this vector. Thus, the coordinate vector $[\mathbf{v}]_{\mathcal{V}}$ is defined to be the unique column vector such that

$$\boxed{\mathbf{v} = \mathcal{V}[\mathbf{v}]_{\mathcal{V}}}. \tag{2.1}$$

The following proposition records some facts (obvious, we hope) about coordinate vectors. Proving this proposition for yourself would be an instructive exercise.

Proposition 2.1. *Let V be a vector space of dimension n and let $\mathcal{V}$ be an ordered basis of V. Then, the following statements hold.*

(1) $[s\mathbf{v}]_{\mathcal{V}} = s[\mathbf{v}]_{\mathcal{V}}$, *for all $\mathbf{v} \in V$ and scalars s.*

(2) $[\mathbf{v} + \mathbf{w}]_{\mathcal{V}} = [\mathbf{v}]_{\mathcal{V}} + [\mathbf{w}]_{\mathcal{V}}$ *for all vectors $\mathbf{v}, \mathbf{w} \in V$.*

(3) $\mathbf{v} \leftrightarrow [\mathbf{v}]_{\mathcal{V}}$ *is a one-to-one correspondence between V and $\mathbb{R}^n$.*

More generally, if $\mathcal{U}$ is a row of m vectors in V, our discussion of (1.7) shows

$$\mathcal{U} = \mathcal{V}A \iff \operatorname{col}_j(A) = [u_j]_{\mathcal{V}}, \quad j = 1, 2, \dots, m \tag{2.2}$$

Since the coordinates of a vector with respect to the basis $\mathcal{V}$ are unique, we get the following fact, which is stated as a proposition for later reference.

Proposition 2.2. *Let V be a vector space of dimension n, and let $\mathcal{V}$ be an ordered basis of V. Then, if $\mathcal{U}$ is a row of m vectors from V, there is a* unique *$n \times m$ matrix A so that*

$$\mathcal{U} = \mathcal{V}A.$$

We can apply the machine we've built up to prove the following theorem.

Theorem 2.3. *Let V be a vector space of dimension n and let $\mathcal{V}$ be an ordered basis of V. Let $\mathcal{U}$ be a row of n vectors from V and let A be the $n \times n$ matrix such that*

$$\mathcal{U} = \mathcal{V}A.$$

Then $\mathcal{U}$ is an ordered basis of V if and only if A is invertible.

Proof. Since $\mathcal{U}$ contains n vectors, $\mathcal{U}$ is a basis if and only if its entries are linearly independent. To check this, we want to consider the equation

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_n\mathbf{u}_n = \mathbf{0} \tag{2.3}$$

If we let $\mathbf{c}$ be the column vector with entries c_j, we have

$$\begin{aligned}\mathbf{0} &= c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_n\mathbf{u}_n \\ &= \mathcal{U}\mathbf{c} \\ &= (\mathcal{V}A)\mathbf{c} \\ &= \mathcal{V}(A\mathbf{c})\end{aligned}$$

By Proposition 2.2, $\mathcal{V}(A\mathbf{c})$ can be zero if and only if $A\mathbf{c} = \mathbf{0}$.

Suppose that A is invertible. Then $A\mathbf{c} = \mathbf{0}$ implies that $\mathbf{c} = \mathbf{0}$. Thus, all the c_i's in (2.3) must be zero, so $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$ are linearly independent.

Conversely, suppose that $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$ are linearly independent. Let $\mathbf{c}$ be a vector such that $A\mathbf{c} = \mathbf{0}$. Then (2.3) holds. By linear independence, $\mathbf{c}$ must be $\mathbf{0}$. Thus, the equation $A\mathbf{c} = \mathbf{0}$ has only the trivial solution, and so A is invertible. □

3. Change of Basis

Suppose that we have two ordered bases $\mathcal{U}$ and $\mathcal{V}$ of an n-dimensional vector space V. From the last section, there is a *unique* matrix A so that $\mathcal{U} = \mathcal{V}A$. We will denote this matrix by $S_{\mathcal{V}\mathcal{U}}$. Thus, the defining equation of $S_{\mathcal{V}\mathcal{U}}$ is

$$\boxed{\mathcal{U} = \mathcal{V}S_{\mathcal{V}\mathcal{U}}}. \tag{3.1}$$

Proposition 3.1. *Let V be an n-dimensional vector space. Let $\mathcal{U}$, $\mathcal{V}$, and $\mathcal{W}$ be ordered bases of V. Then, we have the following.*

(1) $S_{\mathcal{U}\mathcal{U}} = I$.
(2) $S_{\mathcal{U}\mathcal{W}} = S_{\mathcal{U}\mathcal{V}}S_{\mathcal{V}\mathcal{W}}$.
(3) $S_{\mathcal{U}\mathcal{V}} = [S_{\mathcal{V}\mathcal{U}}]^{-1}$.

Proof. For the first statement, note that $\mathcal{U} = \mathcal{U}I$ so, by uniqueness, $S_{\mathcal{U}\mathcal{U}} = I$.

For the second statement, recall the defining equations

$$\mathcal{W} = \mathcal{U}S_{\mathcal{U}\mathcal{W}}, \qquad \mathcal{V} = \mathcal{U}S_{\mathcal{U}\mathcal{V}}, \qquad \mathcal{W} = \mathcal{V}S_{\mathcal{V}\mathcal{W}}.$$

Compute as follows

$$\begin{aligned}\mathcal{W} &= \mathcal{V}S_{\mathcal{V}\mathcal{W}} \\ &= (\mathcal{U}S_{\mathcal{U}\mathcal{V}})S_{\mathcal{V}\mathcal{W}} \\ &= \mathcal{U}(S_{\mathcal{U}\mathcal{V}}S_{\mathcal{V}\mathcal{W}})\end{aligned}$$

and so, by uniqueness, $S_{\mathcal{U}\mathcal{W}} = S_{\mathcal{U}\mathcal{V}}S_{\mathcal{V}\mathcal{W}}$.

For the last statement, set $\mathcal{W} = \mathcal{U}$ in the second statement. This yields

$$S_{\mathcal{U}\mathcal{V}}S_{\mathcal{V}\mathcal{U}} = S_{\mathcal{U}\mathcal{U}} = I,$$

so $S_{\mathcal{U}\mathcal{V}}$ and $S_{\mathcal{V}\mathcal{U}}$ are inverses of each other. □

The matrices $S_{\mathcal{U}\mathcal{V}}$ tell you how to change coordinates from one basis to another, as detailed in the following Proposition.

Proposition 3.2. *Let V be a vector space of dimension n and let $\mathcal{U}$ and $\mathcal{V}$ be ordered bases of V. Then we have*

$$\boxed{[\mathbf{v}]_{\mathcal{U}} = S_{\mathcal{U}\mathcal{V}}[\mathbf{v}]_{\mathcal{V}}},$$

for all vectors $\mathbf{v} \in V$.

Proof. The defining equations are

$$\mathbf{v} = \mathcal{U}[\mathbf{v}]_{\mathcal{U}}, \qquad \mathbf{v} = \mathcal{V}[\mathbf{v}]_{\mathcal{V}}.$$

Thus, we have

$$\begin{aligned} \mathbf{v} &= \mathcal{V}[\mathbf{v}]_{\mathcal{V}} \\ &= (\mathcal{U}S_{\mathcal{U}\mathcal{V}})[\mathbf{v}]_{\mathcal{V}} \\ &= \mathcal{U}(S_{\mathcal{U}\mathcal{V}}[\mathbf{v}]_{\mathcal{V}}), \end{aligned}$$

and so, by uniqueness, $[\mathbf{v}]_{\mathcal{U}} = S_{\mathcal{U}\mathcal{V}}[\mathbf{v}]_{\mathcal{V}}$. $\square$

Thus, $S_{\mathcal{U}\mathcal{V}}$ tells you how to compute the $\mathcal{U}$-coordinates of a vector from the $\mathcal{V}$-coordinates.

Another way to look at this is the following diagram.

$$\begin{tikzcd} & V \arrow[dl, "{[\cdot]_{\mathcal{V}}}"'] \arrow[dr, "{[\cdot]_{\mathcal{U}}}"] & \\ \mathbb{R}^n \arrow[rr, "S_{\mathcal{U}\mathcal{V}}"'] & & \mathbb{R}^n \end{tikzcd} \tag{3.2}$$

In this diagram, $[\cdot]_{\mathcal{V}}$ represents the operation of taking coordinates with respect to $\mathcal{V}$, and $[\cdot]_{\mathcal{U}}$ represents taking coordinates with respect to $\mathcal{U}$. The arrow labeled $S_{\mathcal{U}\mathcal{V}}$ represents the operation of left multiplication by the matrix $S_{\mathcal{U}\mathcal{V}}$. We say that the diagram (3.2) *commutes*, meaning that the two ways of taking a vector from V to the right-hand $\mathbb{R}^n$ yield the same results.

Example 3.3. Consider the space P_3 of polynomials of degree less than 3. Of course, $\mathcal{P} = \begin{bmatrix} 1 & x & x^2 \end{bmatrix}$ is an ordered basis of P_3. Consider

$$\mathcal{V} = \begin{bmatrix} 2 + 3x + 2x^2 & 2x + x^2 & 2 + 2x + 2x^2 \end{bmatrix}.$$

Show that $\mathcal{V}$ is a basis of P_3 and find the transition matrices $S_{\mathcal{P}\mathcal{V}}$ and $S_{\mathcal{V}\mathcal{P}}$.

Let $p(x) = 1 + x + 5x^2$. Use the transition matrix to calculate the coordinates of $p(x)$ with respect to $\mathcal{V}$ and verify the computation.

Solution. By reading off the coefficients, we have

$$\begin{bmatrix} 2 + 3x + 2x^2 & 2x + x^2 & 2 + 2x + 2x^2 \end{bmatrix} = \begin{bmatrix} 1 & x & x^2 \end{bmatrix} \begin{bmatrix} 2 & 0 & 2 \\ 3 & 2 & 2 \\ 2 & 1 & 2 \end{bmatrix}. \tag{3.3}$$

Thus, $\mathcal{V} = \mathcal{P}A$, where A is the 3×3 matrix in the previous equation. A quick check with a calculator shows that A is invertible, so $\mathcal{V}$ is an ordered basis by

Theorem 2.3. Equation (3.3) shows that $S_{\mathcal{PV}} = A$. But then we have $S_{\mathcal{VP}} = S_{\mathcal{PV}}^{-1}$. A calculation shows

$$S_{\mathcal{VP}} = \begin{bmatrix} 1 & 1 & -2 \\ -1 & 0 & 1 \\ -1/2 & -1 & 2 \end{bmatrix}.$$

Since $p(x) = 1 + x + 5x^2$, we have

$$p(x) = \begin{bmatrix} 1 & x & x^2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 5 \end{bmatrix},$$

so we have

$$[p(x)]_{\mathcal{P}} = \begin{bmatrix} 1 \\ 1 \\ 5 \end{bmatrix}.$$

What we want is $[p(x)]_{\mathcal{V}}$. From Proposition 3.2 we have

$$[p(x)]_{\mathcal{V}} = S_{\mathcal{VP}}[p(x)]_{\mathcal{P}} = \begin{bmatrix} 1 & 1 & -2 \\ -1 & 0 & 1 \\ -1/2 & -1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 5 \end{bmatrix} = \begin{bmatrix} -8 \\ 4 \\ 17/2 \end{bmatrix}.$$

The significance of this calculation is that we are claiming that

$$\begin{aligned} 1 + x + 5x^2 &= p(x) \\ &= \mathcal{V}[p(x)]_{\mathcal{V}} \\ &= \begin{bmatrix} 2 + 3x + 2x^2 & 2x + x^2 & 2 + 2x + 2x^2 \end{bmatrix} \begin{bmatrix} -8 \\ 4 \\ 17/2 \end{bmatrix} \\ &= -8(2 + 3x + 2x^2) + 4(2x + x^2) + \frac{17}{2}(2 + 2x + 2x^2). \end{aligned}$$

The reader is invited to carry out the algebra to show this is correct. □

A few remarks are in order before we attempt computations in $\mathbb{R}^n$. If $\mathcal{U} = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_m \end{bmatrix}$ is a row of vectors in $\mathbb{R}^n$, each vector $\mathbf{u}_i$ is a column vector. We can construct an $n \times m$ matrix U by letting the first column of U have the entries in $\mathbf{u}_1$, the second column of U be $\mathbf{u}_2$, and so on. The difference between the row of vectors $\mathcal{U}$ and the matrix U is just one of point of view, depending on what we want to emphasize.

If $\mathcal{U}$ is a row of vectors in $\mathbb{R}^n$, we'll use the notation $\operatorname{mat}(\mathcal{U})$ for the corresponding matrix. Fortunately, the two possible notions of matrix multiplication are compatible, i.e.,

$$\mathcal{V} = \mathcal{U}A \iff \operatorname{mat}(\mathcal{V}) = \operatorname{mat}(\mathcal{U})A,$$

since one description of the matrix multiplication $\operatorname{mat}(\mathcal{U})A$ is that the jth column of $\operatorname{mat}(\mathcal{U})A$ is a linear combination of the columns of $\operatorname{mat}(\mathcal{U})$, using coefficients from the jth column of A; and that is exactly what we're thinking when we look at $\mathcal{U}A$.

In $\mathbb{R}^n$ we have the standard basis $\mathcal{E} = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \dots & \mathbf{e}_n \end{bmatrix}$, where $\mathbf{e}_j$ is the column vector with 1 in the jth row and all other entries zero. If $\mathbf{x} \in \mathbb{R}^n$, we have

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} + \cdots + x_n \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n = \mathcal{E}\mathbf{x}$$

This leads to the observation that the standard basis has the property that

$$\boxed{[\mathbf{x}]_{\mathcal{E}} = \mathbf{x}, \qquad \text{for all } \mathbf{x} \in \mathbb{R}^n.}$$

In terms of matrix multiplication, this is not surprising, since $\mathbf{x} = \mathcal{E}\mathbf{x}$ is equivalent to $\mathbf{x} = \operatorname{mat}(\mathcal{E})\mathbf{x}$ and $\operatorname{mat}(\mathcal{E})$ is the identity matrix.

Suppose that $\mathcal{V}$ is an ordered basis of $\mathbb{R}^n$. The transition matrix $S_{\mathcal{E}\mathcal{V}}$ is defined by

$$\mathcal{V} = \mathcal{E}S_{\mathcal{E}\mathcal{V}}.$$

This is equivalent to the matrix equation

$$\operatorname{mat}(\mathcal{V}) = \operatorname{mat}(\mathcal{E})S_{\mathcal{E}\mathcal{V}}$$

but

$$\operatorname{mat}(\mathcal{E})S_{\mathcal{E}\mathcal{V}} = IS_{\mathcal{E}\mathcal{V}} = S_{\mathcal{E}\mathcal{V}}$$

Thus, we have the important fact that

$$\boxed{S_{\mathcal{E}\mathcal{V}} = \operatorname{mat}(\mathcal{V})}.$$

Example 3.4. Let $\mathcal{U}$ be the ordered basis of $\mathbb{R}^2$ given by

$$\mathbf{u}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \qquad \mathbf{u}_2 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

(1) Show that $\mathcal{U}$ is a basis and find $S_{\mathcal{E}\mathcal{U}}$ and $S_{\mathcal{U}\mathcal{E}}$.
(2) Let $\mathbf{y} = \begin{bmatrix} 3 & 5 \end{bmatrix}^T$. Find $[\mathbf{y}]_{\mathcal{U}}$ and express $\mathbf{y}$ as a linear combination of $\mathcal{U}$.

Solution. Let A be the matrix such that $\mathcal{U} = \mathcal{E}A$. Then, as above,

$$A = \operatorname{mat}(\mathcal{U}) = \begin{bmatrix} 2 & 2 \\ 2 & 3 \end{bmatrix}.$$

The determinant of A is 2, so A is invertible. Then, since $\mathcal{E}$ is a basis, $\mathcal{U}$ is a basis. We have $S_{\mathcal{E}\mathcal{U}} = A$, and so

$$S_{\mathcal{U}\mathcal{E}} = [S_{\mathcal{E}\mathcal{U}}]^{-1} = \begin{bmatrix} 2 & 2 \\ 2 & 3 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{3}{2} & -1 \\ -1 & 1 \end{bmatrix}.$$

For the last part, we note that $[\mathbf{y}]_{\mathcal{E}} = \mathbf{y} = \begin{bmatrix} 3 & 5 \end{bmatrix}^T$ and we have

$$[\mathbf{y}]_{\mathcal{U}} = S_{\mathcal{U}\mathcal{E}}[\mathbf{y}]_{\mathcal{E}} = \begin{bmatrix} \frac{3}{2} & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 5 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} \\ 2 \end{bmatrix}.$$

By saying that

$$[\mathbf{y}]_{\mathcal{U}} = \begin{bmatrix} -\frac{1}{2} \\ 2 \end{bmatrix}.$$

we're saying that

$$\begin{aligned}\begin{bmatrix}3\\5\end{bmatrix} &= \mathbf{y}\\ &= \mathcal{U}[\mathbf{y}]_{\mathcal{U}}\\ &= \begin{bmatrix}\mathbf{u}_1 & \mathbf{u}_2\end{bmatrix}\begin{bmatrix}-\frac{1}{2}\\ 2\end{bmatrix}\\ &= -\frac{1}{2}\mathbf{u}_1 + 2\mathbf{u}_2\\ &= -\frac{1}{2}\begin{bmatrix}2\\2\end{bmatrix} + 2\begin{bmatrix}2\\3\end{bmatrix},\end{aligned}$$

the reader is invited to check the vector algebra to see that this claim is correct □

Example 3.5. Consider the ordered basis $\mathcal{U}$ of $\mathbb{R}^3$ given by

$$\mathbf{u}_1 = \begin{bmatrix}2\\1\\1\end{bmatrix}, \qquad \mathbf{u}_2 = \begin{bmatrix}3\\2\\1\end{bmatrix}, \qquad \mathbf{u}_3 = \begin{bmatrix}1\\1\\1\end{bmatrix}.$$

Let $\mathcal{V}$ be the ordered basis of $\mathbb{R}^3$ given by

$$\mathbf{v}_1 = \begin{bmatrix}-1\\2\\2\end{bmatrix}, \qquad \mathbf{v}_2 = \begin{bmatrix}-6\\2\\4\end{bmatrix}, \qquad \mathbf{v}_3 = \begin{bmatrix}-2\\0\\1\end{bmatrix}.$$

(1) Find $S_{\mathcal{U}\mathcal{V}}$ and $S_{\mathcal{V}\mathcal{U}}$.
(2) Suppose that $[\mathbf{x}]_{\mathcal{U}} = \begin{bmatrix}1 & 2 & -1\end{bmatrix}^T$. Find $[\mathbf{x}]_{\mathcal{V}}$ and express $\mathbf{x}$ as a linear combination of the basis $\mathcal{V}$. Verify.

Solution. Since it's easier to find the transition matrix between a given basis and $\mathcal{E}$, we go through $\mathcal{E}$. Thus, we have

$$S_{\mathcal{E}\mathcal{U}} = \operatorname{mat}(\mathcal{U}) = \begin{bmatrix}2 & 3 & 1\\1 & 2 & 1\\1 & 1 & 1\end{bmatrix}, \qquad S_{\mathcal{E}\mathcal{V}} = \operatorname{mat}(\mathcal{V}) = \begin{bmatrix}-1 & -6 & -2\\2 & 2 & 0\\2 & 4 & 1\end{bmatrix}.$$

But then we can compute

$$S_{\mathcal{U}\mathcal{V}} = S_{\mathcal{U}\mathcal{E}}S_{\mathcal{E}\mathcal{V}} = [S_{\mathcal{E}\mathcal{U}}]^{-1}S_{\mathcal{E}\mathcal{V}} = \begin{bmatrix}2 & 3 & 1\\1 & 2 & 1\\1 & 1 & 1\end{bmatrix}^{-1}\begin{bmatrix}-1 & -6 & -2\\2 & 2 & 0\\2 & 4 & 1\end{bmatrix} = \begin{bmatrix}-3 & -6 & -1\\0 & -2 & -1\\5 & 12 & 3\end{bmatrix}.$$

From this, we have

$$S_{\mathcal{V}\mathcal{U}} = [S_{\mathcal{U}\mathcal{V}}]^{-1} == \begin{bmatrix}-3 & -6 & -1\\0 & -2 & -1\\5 & 12 & 3\end{bmatrix}^{-1} = \begin{bmatrix}3 & 3 & 2\\-\frac{5}{2} & -2 & -\frac{3}{2}\\5 & 3 & 3\end{bmatrix}.$$

For the second part, we're given the coordinates of $\mathbf{x}$ with respect to $\mathcal{U}$ and we want the coordinates with respect to $\mathcal{V}$. We can calculate these by

$$[\mathbf{x}]_{\mathcal{V}} = S_{\mathcal{V}\mathcal{U}}[\mathbf{x}]_{\mathcal{U}} = \begin{bmatrix}3 & 3 & 2\\-\frac{5}{2} & -2 & -\frac{3}{2}\\5 & 3 & 3\end{bmatrix}\begin{bmatrix}1\\2\\-1\end{bmatrix} = \begin{bmatrix}7\\-5\\8\end{bmatrix}.$$

To verify this, note that by saying $[\mathbf{x}]_{\mathcal{V}} = \begin{bmatrix} 7 & -5 & 8 \end{bmatrix}^T$, we're claiming that

$$
\begin{aligned}
\mathbf{x} &= \mathcal{V}[\mathbf{x}]_{\mathcal{V}} \\
&= \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix} \begin{bmatrix} 7 \\ -5 \\ 8 \end{bmatrix} \\
&= 7\mathbf{v}_1 - 5\mathbf{v}_2 + 8\mathbf{v}_3 \\
&= 7 \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix} - 5 \begin{bmatrix} -6 \\ 2 \\ 4 \end{bmatrix} + 8 \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix} \\
&= \begin{bmatrix} 7 \\ 4 \\ 2 \end{bmatrix}
\end{aligned}
$$

On the other hand, saying $[\mathbf{x}]_{\mathcal{U}} = \begin{bmatrix} 1 & 2 & -1 \end{bmatrix}^T$ is saying that

$$
\begin{aligned}
\mathbf{x} &= \mathcal{U}[\mathbf{x}]_{\mathcal{U}} \\
&= \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} \\
&= \mathbf{u}_1 + 2\mathbf{u}_2 - \mathbf{u}_3 \\
&= \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} + 2 \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \\
&= \begin{bmatrix} 7 \\ 4 \\ 2 \end{bmatrix}.
\end{aligned}
$$

$\square$

4. Matrix Representations of Linear Transformations

We want to study linear transformations between finite dimensional vector spaces.

So suppose that V is a vector space of dimension n, W is a vector space of dimension p and $L\colon V \to W$ is a linear transformation. Choose ordered bases $\mathcal{V}$ for V and $\mathcal{W}$ for W.

For each basis vector $\mathbf{v}_j$ in $\mathcal{V}$, the image $L(\mathbf{v}_j)$ is an element of W, and so can be expressed as a linear combination of the basis vectors in $\mathcal{W}$. Thus, we have

$$
\begin{aligned}
L(\mathbf{v}_1) &= a_{11}\mathbf{w}_1 + a_{21}\mathbf{w}_2 + \cdots + a_{p1}\mathbf{w}_p \\
L(\mathbf{v}_2) &= a_{12}\mathbf{w}_2 + a_{22}\mathbf{w}_2 + \cdots + a_{p2}\mathbf{w}_p \\
&\vdots \\
L(\mathbf{v}_n) &= a_{1n}\mathbf{w}_1 + a_{2n}\mathbf{w}_2 + \cdots + a_{pn}\mathbf{w}_p,
\end{aligned}
$$

for some scalars a_{ij}. We can rewrite this system of equations in matrix form as

$$\begin{bmatrix} L(\mathbf{v}_1) & L(\mathbf{v}_2) & \dots & L(\mathbf{v}_n) \end{bmatrix} = \begin{bmatrix} \mathbf{w}_1 & \mathbf{w}_2 & \dots & \mathbf{w}_p \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pn} \end{bmatrix} \tag{4.1}$$

If we introduce the notation $L(\mathcal{V}) = \begin{bmatrix} L(\mathbf{v}_1) & L(\mathbf{v}_2) & \dots & L(\mathbf{v}_n) \end{bmatrix}$, we can write (4.1) simply as

$$L(\mathcal{V}) = \mathcal{W}A \tag{4.2}$$

The matrix A is unique and we will denote it by $[L]_{\mathcal{W}\mathcal{V}}$ and call it *the matrix of L with respect to $\mathcal{V}$ and $\mathcal{W}$*. Thus, the defining equation of $[L]_{\mathcal{W}\mathcal{V}}$ is

$$\boxed{L(\mathcal{V}) = \mathcal{W}[L]_{\mathcal{W}\mathcal{V}}}.$$

Another way to look at the specification of this matrix is to note that (4.2) says that

$$L(\mathbf{v}_j) = \mathcal{W}\operatorname{col}_j(A),$$

and so $\operatorname{col}_j(A) = [L(\mathbf{v}_j)]_{\mathcal{W}}$. Thus, another we to interpret the definition of the matrix of L is

$$\text{Column } j \text{ of } [L]_{\mathcal{W}\mathcal{V}} = [L(\mathbf{v}_j)]_{\mathcal{W}},$$

in other words, column j of $[L]_{\mathcal{W}\mathcal{V}}$ is the coordinates with respect to the basis $\mathcal{W}$ in the target space of the image of the j-th basis vector in the source space under L.

We will need one fact about the way we've defined $L(\mathcal{V})$, which is just a restatement of the fact that L is linear.

Proposition 4.1. *Let $L\colon V \to W$ be a linear transformation, let $\mathcal{U}$ be any row of k vectors in V and let A be a $k \times \ell$ matrix. Then*

$$L(\mathcal{U}A) = L(\mathcal{U})A \tag{4.3}$$

Proof. Let $\mathcal{V} = \mathcal{U}A$. Then we have

$$\mathbf{v}_j = \mathcal{U}\operatorname{col}_j(A) = \sum_{i=1}^{k} \mathbf{u}_i a_{ij}.$$

Thus, we have

$$\begin{aligned} L(\mathbf{v}_i) &= L\left(\sum_{i=1}^{k} \mathbf{u}_i a_{ij}\right) \\ &= \sum_{i=1}^{k} L(\mathbf{u}_i) a_{ij}, && \text{since } L \text{ is linear,} \\ &= L(\mathcal{U})\operatorname{col}_j(A), \end{aligned}$$

Thus, the ith entry of the left-hand side of (4.3) is equal to the ith entry of the right-hand side, for all $i = 1, \dots, \ell$. ☐

Next, we want to describe the action of L on elements of V in terms of coordinates. To do this, let $\mathbf{x} \in V$ be arbitrary, so we have

$$\mathbf{x} = \mathcal{V}[\mathbf{x}]_{\mathcal{V}}.$$

Then, we have

$$\begin{aligned} L(\mathbf{x}) &= L(\mathcal{V}[\mathbf{x}]_{\mathcal{V}}) \\ &= L(\mathcal{V})[\mathbf{x}]_{\mathcal{V}} \\ &= (\mathcal{W}[L]_{\mathcal{W}\mathcal{V}})[\mathbf{x}]_{\mathcal{V}} \\ &= \mathcal{W}([L]_{\mathcal{W}\mathcal{V}}[\mathbf{x}]_{\mathcal{V}}) \end{aligned}$$

On the other hand, the defining equation of $[L(\mathbf{x})]_{\mathcal{W}}$ is

$$L(\mathbf{x}) = \mathcal{W}[L(\mathbf{x})]_{\mathcal{W}}.$$

Comparing with the last computation, we get the important equation

$$\boxed{[L(\mathbf{x})]_{\mathcal{W}} = [L]_{\mathcal{W}\mathcal{V}}[\mathbf{x}]_{\mathcal{V}}}. \tag{4.4}$$

Thus the matrix $[L]_{\mathcal{W}\mathcal{V}}$ tells you how to compute the *coordinates of* $L(\mathbf{x})$ from the *coordinates of* $\mathbf{x}$.

Another way to look at (4.4) is the following. Let n be the dimension of V and let p be the dimension of W. Then (4.4) is equivalent to the fact that the following diagram commutes.

$$\begin{tikzcd} V \arrow[r, "L"] \arrow[d, "{[\cdot]_{\mathcal{V}}}"'] & W \arrow[d, "{[\cdot]_{\mathcal{W}}}"] \\ \mathbb{R}^n \arrow[r, "{[L]_{\mathcal{W}\mathcal{V}}}"'] & \mathbb{R}^p \end{tikzcd} \tag{4.5}$$

where the arrow labeled $[L]_{\mathcal{W}\mathcal{V}}$ means the operation of left multiplication by that matrix. The vertical arrows are each a one-to-one correspondence, so the diagram says that L and multiplication by $[L]_{\mathcal{W}\mathcal{V}}$ are the same thing under the correspondence.

Example 4.2. Let $D\colon P_3 \to P_3$ be the derivative operator. Find the matrix of D with respect to the basis $\mathcal{U} = \begin{bmatrix} x^2 & x & 1 \end{bmatrix}$. Use this matrix to compute D of $p(x) = 5x^2 + 2x + 3$.

Solution. To find $[D]_{\mathcal{U}\mathcal{U}}$ we want to find the matrix that satisfies $D(\mathcal{U}) = \mathcal{U}[D]_{\mathcal{U}\mathcal{U}}$. But, we have

$$D(\mathcal{U}) = \begin{bmatrix} 2x & 1 & 0 \end{bmatrix} = \begin{bmatrix} x^2 & x & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

Thus, we've calculated that

$$[D]_{\mathcal{U}\mathcal{U}} = \begin{bmatrix} 0 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

Now consider $p(x)$. We have

$$p(x) = 5x^2 + 2x + 3 = \begin{bmatrix} x^2 & x & 1 \end{bmatrix} \begin{bmatrix} 5 \\ 2 \\ 3 \end{bmatrix}$$

Thus, $[p(x)]_{\mathcal{U}} = \begin{bmatrix} 5 & 2 & 3 \end{bmatrix}^T$. But then we have

$$\begin{aligned} [p'(x)]_{\mathcal{U}} &= [D(p(x))]_{\mathcal{U}} \\ &= [D]_{\mathcal{U}\mathcal{U}}[p(x)]_{\mathcal{U}} \\ &= \begin{bmatrix} 0 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 5 \\ 2 \\ 3 \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ 10 \\ 2 \end{bmatrix}, \end{aligned}$$

which tells us the coordinates of $p'(x)$. But the coordinates of $p'(x)$ tell us what $p'(x)$ is, namely,

$$p'(x) = \mathcal{U}[p'(x)]_{\mathcal{U}} = \begin{bmatrix} x^2 & x & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 10 \\ 2 \end{bmatrix} = 10x + 2,$$

which is, of course, what you get from taking the derivative of $p(x)$ using the rules from calculus. □

5. Composition of Linear Transformations

The next Proposition says that the composition of two linear transformations is a linear transformation.

Proposition 5.1. *Let $T: U \to V$ and $S: V \to W$ be linear transformations, where U, V and W are vector spaces. Let $L: U \to W$ be defined by $L = S \circ T$, i.e., $L(\mathbf{u}) = S(T(\mathbf{u}))$. Then L is a linear transformation.*

Proof. We have to show that L preserves addition and scalar multiplication.

To show L preserves addition, let $\mathbf{u}_1$ and $\mathbf{u}_2$ be vectors in U. Then we have

$$\begin{aligned} L(\mathbf{u}_1 + \mathbf{u}_2) &= S(T(\mathbf{u}_1 + \mathbf{u}_2)) && \text{by the definition of } L \\ &= S(T(\mathbf{u}_1) + T(\mathbf{u}_2)) && \text{since } T \text{ is linear} \\ &= S(T(\mathbf{u}_1)) + S(T(\mathbf{u}_2)) && \text{since } S \text{ is linear} \\ &= L(\mathbf{u}_1) + L(\mathbf{u}_2) && \text{by the definition of } L. \end{aligned}$$

Thus, $L(\mathbf{u}_1 + \mathbf{u}_2) = L(\mathbf{u}_1) + L(\mathbf{u}_2)$ and L preserves addition.

For scalar multiplication, let $\mathbf{u}$ be a vector in U and let α be a scalar. Then we have

$$\begin{aligned} L(\alpha\mathbf{u}) &= S(T(\alpha\mathbf{u})) && \text{by the definition of } L \\ &= S(\alpha T(\mathbf{u})) && \text{since } T \text{ is linear} \\ &\alpha S(T(\mathbf{u})) && \text{since } S \text{ is linear} \\ &= \alpha L(\mathbf{u}) && \text{by the definition of } L. \end{aligned}$$

Thus, $L(\alpha\mathbf{u}) = \alpha L(\mathbf{u})$, so L preserves scalar multiplication. This completes the proof that L is linear. □

The next Proposition says that the matrix of a composition of linear transformations is the product of the matrices of the transformations.

Proposition 5.2. *Let $T\colon U \to V$ and $S\colon V \to W$ be linear transformations, where U, V and W are vector spaces. Let $\mathcal{U}$ be an ordered basis for U, $\mathcal{V}$ an ordered basis for V and $\mathcal{W}$ an ordered basis for W. Then, we have*

$$\boxed{[S \circ T]_{\mathcal{W}\mathcal{U}} = [S]_{\mathcal{W}\mathcal{V}}[T]_{\mathcal{V}\mathcal{U}}}.$$

Proof. The defining equations are

$$(S \circ T)(\mathcal{U}) = \mathcal{W}[S \circ T]_{\mathcal{W}\mathcal{U}} \tag{5.1}$$

$$T(\mathcal{U}) = \mathcal{V}[T]_{\mathcal{V}\mathcal{U}} \tag{5.2}$$

$$S(\mathcal{V}) = \mathcal{W}[S]_{\mathcal{W}\mathcal{V}} \tag{5.3}$$

To derive (5.2), begin by applying S to both sides of (5.2). This gives

$$\begin{aligned}
(S \circ T)(\mathcal{U}) &= S(T(\mathcal{U})) \\
&= S(\mathcal{V}[T]_{\mathcal{V}\mathcal{U}}) && \text{by (5.2)} \\
&= S(\mathcal{V})[T]_{\mathcal{V}\mathcal{U}} \\
&= (\mathcal{W}[S]_{\mathcal{W}\mathcal{V}})[T]_{\mathcal{V}\mathcal{U}} && \text{by (5.3)} \\
&= \mathcal{W}([S]_{\mathcal{W}\mathcal{V}}[T]_{\mathcal{V}\mathcal{U}}).
\end{aligned}$$

Thus, we have

$$(S \circ T)(\mathcal{U}) = \mathcal{W}([S]_{\mathcal{W}\mathcal{V}}[T]_{\mathcal{V}\mathcal{U}}),$$

and comparing this with the defining equation (5.1) show that

$$[S \circ T]_{\mathcal{W}\mathcal{U}} = [S]_{\mathcal{W}\mathcal{V}}[T]_{\mathcal{V}\mathcal{U}}.$$

□

6. Change of Basis for Linear Transformation

This computation address the question of how the matrix of a linear transformation changes when we change the bases we're using.

Proposition 6.1. *Let $L\colon V \to W$ be a linear transformation, where V and W are vector spaces of dimension n and m respectively. Let $\mathcal{U}$ and $\mathcal{V}$ be ordered bases for V and let $\mathcal{X}$ and $\mathcal{Y}$ be ordered bases for W. Then*

$$\boxed{[L]_{\mathcal{Y}\mathcal{V}} = S_{\mathcal{Y}\mathcal{X}}[L]_{\mathcal{X}\mathcal{U}}S_{\mathcal{U}\mathcal{V}}}. \tag{6.1}$$

So, if we know $[L]_{\mathcal{U}\mathcal{X}}$ and then decide to use new bases $\mathcal{V}$ for V and $\mathcal{Y}$ for W, this formula tells us how to compute the matrix of L with respect to the new bases from the matrix with respect to the old bases.

Proof. The defining equations of the two matrices of L are

$$L(\mathcal{U}) = \mathcal{X}[L]_{\mathcal{X}\mathcal{U}} \tag{6.2}$$

$$L(\mathcal{V}) = \mathcal{Y}[L]_{\mathcal{Y}\mathcal{V}}. \tag{6.3}$$

The defining equations for the two transition matrices in (6.1) are

$$\mathcal{X} = \mathcal{Y}S_{\mathcal{Y}\mathcal{X}} \tag{6.4}$$

$$\mathcal{V} = \mathcal{U}S_{\mathcal{U}\mathcal{V}}. \tag{6.5}$$

Then, we have

$$\begin{aligned}
L(\mathcal{V}) &= L(\mathcal{U}S_{\mathcal{U}\mathcal{V}}) && \text{by (6.5)} \\
&= L(\mathcal{U})S_{\mathcal{U}\mathcal{V}} \\
&= (\mathcal{X}[L]_{\mathcal{X}\mathcal{U}})S_{\mathcal{U}\mathcal{V}} && \text{by (6.2)} \\
&= \mathcal{X}([L]_{\mathcal{X}\mathcal{U}}S_{\mathcal{U}\mathcal{V}}) \\
&= (\mathcal{Y}S_{\mathcal{Y}\mathcal{X}})([L]_{\mathcal{X}\mathcal{U}}S_{\mathcal{U}\mathcal{V}}) && \text{by (6.4)} \\
&= \mathcal{Y}(S_{\mathcal{Y}\mathcal{X}}[L]_{\mathcal{X}\mathcal{U}}S_{\mathcal{U}\mathcal{V}})
\end{aligned}$$

Thus, we have

$$L(\mathcal{V}) = \mathcal{Y}(S_{\mathcal{Y}\mathcal{X}}[L]_{\mathcal{X}\mathcal{U}}S_{\mathcal{U}\mathcal{V}}).$$

Comparing with the defining equation (6.3), we see that

$$[L]_{\mathcal{Y}\mathcal{V}} = S_{\mathcal{Y}\mathcal{X}}[L]_{\mathcal{X}\mathcal{U}}S_{\mathcal{U}\mathcal{V}}.$$

□

This equation, and much of our work above, can be brought together in the commutative "pup tent" diagram. We use the same notation as in the last Proposition. Then, the following diagram commutes:

$$\begin{tikzcd}
& V \arrow[rrr, "L"] \arrow[dl, "{[\cdot]_{\mathcal{U}}}"'] \arrow[ddr, "{[\cdot]_{\mathcal{V}}}"] & & & W \arrow[dl, "{[\cdot]_{\mathcal{X}}}"'] \arrow[ddr, "{[\cdot]_{\mathcal{Y}}}"] & \\
\mathbb{R}^n \arrow[rrr, "{[L]_{\mathcal{X}\mathcal{U}}}"] \arrow[drr, "S_{\mathcal{V}\mathcal{U}}"'] & & & \mathbb{R}^m \arrow[drr, "S_{\mathcal{Y}\mathcal{X}}"] & & \\
& & \mathbb{R}^n \arrow[rrr, "{[L]_{\mathcal{Y}\mathcal{V}}}"] & & & \mathbb{R}^m
\end{tikzcd}$$

In this diagram, the triangular ends of the tent are the diagrams corresponding to (3.2), the rectangular sides of the tent are (4.5), and the fact that the rectangular floor commutes is (6.1).

If we consider a linear operator $L\colon V \to V$ from a vector space to itself, we usually use the same basis for both sides. Thus, if $\mathcal{U}$ is a basis for V, the matrix $[L]_{\mathcal{U}\mathcal{U}}$ of L with respect to $\mathcal{U}$ is defined by

$$L(\mathcal{U}) = \mathcal{U}[L]_{\mathcal{U}\mathcal{U}}$$

and the action of L in coordinates with respect to this basis is given by

$$[L(\mathbf{v})]_{\mathcal{U}} = [L]_{\mathcal{U}\mathcal{U}}[\mathbf{v}]_{\mathcal{U}}.$$

If we have another basis $\mathcal{V}$ of V, we can use (6.1) to find

$$[L]_{\mathcal{V}\mathcal{V}} = S_{\mathcal{V}\mathcal{U}}[L]_{\mathcal{U}\mathcal{U}}S_{\mathcal{U}\mathcal{V}}. \tag{6.6}$$

If we've been working with the basis $\mathcal{U}$, we're most likely to know the transition matrix $S_{\mathcal{U}\mathcal{V}}$, since it is defined by

$$\mathcal{V} = \mathcal{U}S_{\mathcal{U}\mathcal{V}},$$

i.e., it tells us how to express the new basis elements in $\mathcal{V}$ as linear combinations of the old basis elements in $\mathcal{U}$. But then we also know $S_{\mathcal{V}\mathcal{U}}$, since $S_{\mathcal{V}\mathcal{U}} = S_{\mathcal{U}\mathcal{V}}^{-1}$. Thus, we can rewrite (6.6) as

$$\boxed{[L]_{\mathcal{V}\mathcal{V}} = S_{\mathcal{U}\mathcal{V}}^{-1}[L]_{\mathcal{U}\mathcal{U}}S_{\mathcal{U}\mathcal{V}}}.$$

Example 6.2. Recall the operator $D\colon P_3 \to P_3$ from Example 4.2. Let $\mathcal{U}$ be the basis in Example 4.2 and let

$$\mathcal{V} = \begin{bmatrix} x^2+x+2 & 4x^2+5x+6 & x^2+x+1 \end{bmatrix}$$

Show that $\mathcal{V}$ is a basis of P_3. Find the matrix of D with respect to $\mathcal{V}$. Let $p(x)$ be the polynomial such that $[p(x)]_{\mathcal{V}} = \begin{bmatrix} 2 & 1 & -1 \end{bmatrix}^T$. Find the coordinates with respect to $\mathcal{V}$ of $p'(x)$. Verify the computation.

Solution. The basis $\mathcal{U}$ from Example 4.2 was

$$\mathcal{U} = \begin{bmatrix} x^2 & x & 1 \end{bmatrix}$$

To see if $\mathcal{V}$ is a basis, we can write $\mathcal{V} = \mathcal{U}A$ for some matrix A. In fact, we have

$$\begin{bmatrix} x^2+x+2 & 4x^2+5x+6 & x^2+x+1 \end{bmatrix} = \begin{bmatrix} x^2 & x & 1 \end{bmatrix} \begin{bmatrix} 1 & 4 & 1 \\ 1 & 5 & 1 \\ 2 & 6 & 1 \end{bmatrix},$$

Thus, we see that

$$A = \begin{bmatrix} 1 & 4 & 1 \\ 1 & 5 & 1 \\ 2 & 6 & 1 \end{bmatrix}.$$

Punching it into the calculator shows that A is invertible. This implies that $\mathcal{V}$ is a basis, and we have $S_{\mathcal{U}\mathcal{V}} = A$.

From Example 4.2, we recall that

$$[D]_{\mathcal{U}\mathcal{U}} = \begin{bmatrix} 0 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

From our work above, we find that

$$\begin{aligned} [D]_{\mathcal{V}\mathcal{V}} &= S_{\mathcal{V}\mathcal{U}}[D]_{\mathcal{U}\mathcal{U}}S_{\mathcal{U}\mathcal{V}} \\ &= S_{\mathcal{U}\mathcal{V}}^{-1}[D]_{\mathcal{U}\mathcal{U}}S_{\mathcal{U}\mathcal{V}} \\ &= \begin{bmatrix} 1 & 4 & 1 \\ 1 & 5 & 1 \\ 2 & 6 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 4 & 1 \\ 1 & 5 & 1 \\ 2 & 6 & 1 \end{bmatrix} \\ &= \begin{bmatrix} -3 & -11 & -3 \\ 2 & 8 & 2 \\ -5 & -21 & -5 \end{bmatrix} \end{aligned}$$

Next, we compute $[p'(x)]_{\mathcal{V}}$. We're given that $[p(x)]_{\mathcal{V}} = \begin{bmatrix} 2 & 1 & -1 \end{bmatrix}^T$ Thus,

$$\begin{aligned}
[p'(x)]_{\mathcal{V}} &= [D(p(x))]_{\mathcal{V}} \\
&= [D]_{\mathcal{V}\mathcal{V}}[p(x)]_{\mathcal{V}} \\
&= \begin{bmatrix} -3 & -11 & -3 \\ 2 & 8 & 2 \\ -5 & -21 & -5 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix} \\
&= \begin{bmatrix} -14 \\ 10 \\ -26 \end{bmatrix},
\end{aligned}$$

so we have

$$[p'(x)]_{\mathcal{V}} = \begin{bmatrix} -14 \\ 10 \\ -26 \end{bmatrix}.$$

To verify this computation, note that our computation of $[p'(x)]_{\mathcal{V}}$ implies that

$$\begin{aligned}
p'(x) &= \mathcal{V}[p'(x)]_{\mathcal{V}} \\
&= \begin{bmatrix} x^2+x+2 & 4x^2+5x+6 & x^2+x+1 \end{bmatrix} \begin{bmatrix} -14 \\ 10 \\ -26 \end{bmatrix} \\
&= -14(x^2+x+2) + 10(4x^2+5x+6) - 26(x^2+x+1) \\
&= 10x + 6
\end{aligned}$$

(the reader is invited to check the algebra on the last step).

On the other hand, we know $[p(x)]_{\mathcal{V}}$, so

$$\begin{aligned}
p(x) &= \mathcal{V}[p(x)]_{\mathcal{V}} \\
&= \begin{bmatrix} x^2+x+2 & 4x^2+5x+6 & x^2+x+1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix} \\
&= 2(x^2+x+2) + (4x^2+5x+6) - (x^2+x+1) \\
&= 5x^2 + 6x + 6
\end{aligned}$$

(as the reader is invited to check). Thus, we see that we have computed $p'(x)$ correctly. □

Before doing an example in $\mathbb{R}^n$, some remarks are in order. Suppose that we have a linear transformation $L\colon \mathbb{R}^n \to \mathbb{R}^m$. Let $\mathcal{E}_n$ be the standard basis of $\mathbb{R}^n$ and $\mathcal{E}_m$ be the standard basis of $\mathbb{R}^m$ The matrix $[L]_{\mathcal{E}_m\mathcal{E}_n}$ of L with respect to the standard bases is defined by

$$L(\mathcal{E}_n) = \mathcal{E}_m[L]_{\mathcal{E}_m\mathcal{E}_n}$$

The equivalent matrix equation is

$$\operatorname{mat}(L(\mathcal{E}_n)) = \operatorname{mat}(\mathcal{E}_m)[L]_{\mathcal{E}_m\mathcal{E}_n} = I[L]_{\mathcal{E}_m\mathcal{E}_n} = [L]_{\mathcal{E}_m\mathcal{E}_n}$$

Thus, $[L]_{\mathcal{E}_m\mathcal{E}_n}$ is the matrix

$$\operatorname{mat}(L(\mathcal{E}_n)) = \operatorname{mat}(\begin{bmatrix} L(\mathbf{e}_1) & L(\mathbf{e}_2) & \dots & L(\mathbf{e}_n) \end{bmatrix}),$$

i.e., the matrix whose columns are the column vectors $L(\mathbf{e}_1), L(\mathbf{e}_2), \dots, L(\mathbf{e}_n)$. This is just the way we've previously described the standard matrix of L. Since $[\mathbf{x}]_{\mathcal{E}_n} = \mathbf{x}$ for $\mathbf{x} \in \mathbb{R}^n$ and and $[\mathbf{y}]_{\mathcal{E}_m} = \mathbf{y}$ for $\mathbf{y} \in \mathbb{R}^m$, the equation

$$[L(\mathbf{x})]_{\mathcal{E}_m} = [L]_{\mathcal{E}_m\mathcal{E}_n}[\mathbf{x}]_{\mathcal{E}_n}$$

for the coordinate action of L just becomes

$$L(\mathbf{x}) = [L]_{\mathcal{E}_m\mathcal{E}_n}\mathbf{x},$$

again agreeing with our discussion of the standard matrix of a linear transformation.

Example 6.3. Let $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation whose standard matrix is

$$A = \begin{bmatrix} 14/5 & 2/5 \\ 2/5 & 11/5 \end{bmatrix}.$$

Let $\mathcal{U}$ be the basis of $\mathbb{R}^2$ with

$$\mathbf{u}_1 = \begin{bmatrix} -1 \\ 2 \end{bmatrix}, \qquad \mathbf{u}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

Find the matrix of T with respect to $\mathcal{U}$. Discuss the significance of this matrix.

Solution. Letting $\mathcal{E}$ denote the standard basis of $\mathbb{R}^2$, we have $[T]_{\mathcal{E}\mathcal{E}} = A$, where A is given above. As usual in $\mathbb{R}^n$, we have

$$S_{\mathcal{E}\mathcal{U}} = \operatorname{mat}(\mathcal{U}) = \begin{bmatrix} -1 & 2 \\ 2 & 1 \end{bmatrix}$$

Thus, we have

$$\begin{aligned}
[T]_{\mathcal{U}\mathcal{U}} &= S_{\mathcal{U}\mathcal{E}}[T]_{\mathcal{E}\mathcal{E}}S_{\mathcal{E}\mathcal{U}} \\
&= S_{\mathcal{E}\mathcal{U}}^{-1}[T]_{\mathcal{E}\mathcal{E}}S_{\mathcal{E}\mathcal{U}} \\
&= \begin{bmatrix} -1 & 2 \\ 2 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 14/5 & 2/5 \\ 2/5 & 11/5 \end{bmatrix} \begin{bmatrix} -1 & 2 \\ 2 & 1 \end{bmatrix} \\
&= \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix},
\end{aligned}$$

so we have

$$[T]_{\mathcal{U}\mathcal{U}} = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}.$$

This is a diagonal matrix, i.e., the only nonzero entries are on the main diagonal. This has a special significance. The defining equation $T(\mathcal{U}) = \mathcal{U}[T]_{\mathcal{U}\mathcal{U}}$ becomes

$$\begin{bmatrix} T(\mathbf{u}_1) & T(\mathbf{u}_2) \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 2\mathbf{u}_1 & 3\mathbf{u}_2 \end{bmatrix}$$

Thus, $T(\mathbf{u}_1) = 2\mathbf{u}_1$ and $T(\mathbf{u}_2) = 3\mathbf{u}_2$. Thus, geometrically, T is a dilation by a factor of 2 in the $\mathbf{u}_1$ direction and dilation by a factor of 3 in the $\mathbf{u}_2$ direction. Also, if $\mathbf{c} = [\mathbf{x}]_{\mathcal{U}}$, then

$$[T(\mathbf{x})]_{\mathcal{U}} = [T]_{\mathcal{U}\mathcal{U}}[\mathbf{x}]_{\mathcal{U}} = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 2c_1 \\ 3c_2 \end{bmatrix}$$

Thus, T is easier to understand in this coordinate system than in the standard coordinate system. □

The same sort of reasoning leads to an important concept. Let A and B be $n \times n$ matrices. We say A and B are *similar* if there is a nonsingular matrix P such that $B = P^{-1}AP$. To see the significance of this concept, let $T\colon \mathbb{R}^n \to \mathbb{R}^n$ be the linear transformation whose *standard* matrix is A. Let $\mathcal{P}$ the the row of vectors in $\mathbb{R}^n$ formed by the columns of P. Then we have $\mathcal{P} = \mathcal{E}P$ (why?). Since P is invertible, $\mathcal{P}$ is a basis and $S_{\mathcal{E}\mathcal{P}} = P$. The matrix of T with respect to $\mathcal{P}$ is given by

$$\begin{aligned} [T]_{\mathcal{P}\mathcal{P}} &= S_{\mathcal{P}\mathcal{E}}[T]_{\mathcal{E}\mathcal{E}}S_{\mathcal{E}\mathcal{P}} \\ &= S_{\mathcal{E}\mathcal{P}}^{-1}[T]_{\mathcal{E}\mathcal{E}}S_{\mathcal{E}\mathcal{P}} \\ &= P^{-1}AP \\ &= B \end{aligned}$$

Thus, $[T]_{\mathcal{P}\mathcal{P}} = B$. The conclusion is that if $B = P^{-1}AP$, the matrices A and B represent the same linear transformation with respect to different bases. Indeed, if A is the matrix of the transformation with respect to the standard basis, then B is the matrix of the same transformation with respect to the basis formed by the columns of P.

If we can find P so that $P^{-1}AP = D$, where D is a diagonal matrix, then in the new basis formed by the columns of P, the corresponding transformation is just a dilation along each of the basis directions, and so is easier to understand in the new coordinate system.

Department of Mathematics, Texas Tech University, Lubbock, TX 79409-1042
E-mail address: drager@math.ttu.edu