6. Lecture 6: 2.3 Composition of linear maps and matrix product

One is often interested in **composing** transformations i.e. one first perform one transformations and then another to the result of the first, e.g. if we first rotate a vector and then scale the resulting vector or rotate it again. If the transformations involved are linear then the composite transformation is also linear and one would like to calculate the matrix for the composite transformation from those of the involved transformations.

Suppose that B is an $m \times n$ matrix and A is an $n \times p$ matrix. Then multiplication by A defines a map $\mathbf{R}^p \ni \mathbf{x} \to A\mathbf{x} = \mathbf{y} \in \mathbf{R}^n$ and multiplication by B defines a map $\mathbf{R}^n \ni \mathbf{y} \to B\mathbf{y} = \mathbf{z} \in \mathbf{R}^m$ and multiplication by first A and then B

$$\mathbf{x} \xrightarrow{\text{multiply by } A} A\mathbf{x} \xrightarrow{\text{multiply by } B} B(A\mathbf{x})$$

defines a map $\mathbf{R}^p \ni \mathbf{x} \to T(\mathbf{x}) = B(A\mathbf{x}) \in \mathbf{R}^m$.

We claim that this map is linear. In fact, this follows from that multiplication by A and by B are linear

$$T(\mathbf{x}+\mathbf{y}) = B\big(A(\mathbf{x}+\mathbf{y})\big) = B(\big(A\mathbf{x}+A\mathbf{y}\big) = B(A\mathbf{x}) + B(A\mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y}),$$

and similarly one proves that $T(\lambda\mathbf{x}) = \lambda T(\mathbf{x})$.

We want to define the matrix product BA to be the $m \times p$ matrix that represents this map so that $(BA)\mathbf{x} = B(A\mathbf{x})$:

$$\mathbf{x} \xrightarrow{\text{multiply by } BA} (BA)\mathbf{x} = B(A\mathbf{x})$$

Another way to formulate this is to say that we want the matrix multiplication to be defined so it is **associative**, i.e. $(BA)\mathbf{x} = B(A\mathbf{x})$, it shouldn't matter if you first calculate $A\mathbf{x}$ and then $B(A\mathbf{x})$ or if you first calculate BA and then $(BA)\mathbf{x}$.

Ex Find the matrix for the linear transformation obtained by first rotating the vector an angle θ and then multiplying the resulting vector by a the scalar r.

Sol The matrix for the rotation $T(\mathbf{x}) = A\mathbf{x}$, is $A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ and the matrix for the scaling $M(\mathbf{x}) = r\,\mathbf{x} = S\mathbf{x}$ is $S = \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix}$.

The combined map obtained by first rotating the vector an angle θ and then multiplying the resulting vector by a the scalar r, is by the definition of the multiplication of a matrix by a vector in the column picture

$$\mathbf{x} \to r\,T(\mathbf{x}) = r\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = r\Big(\begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}x_1 + \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}x_2\Big) = \begin{bmatrix} r\cos\theta & -r\sin\theta \\ r\sin\theta & r\cos\theta \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix},$$

which gives the matrix for the combined map.

Question: How can we find the matrix for the composition of linear maps if we know the matrices for the maps themselves? Let us calculate $B(A\mathbf{x})$. We have

$$A\mathbf{x} = \begin{bmatrix}\mathbf{a}_1 \cdots \mathbf{a}_p\end{bmatrix}\begin{bmatrix}x_1\\ \vdots\\ x_p\end{bmatrix} = x_1\mathbf{a}_1 + \cdots + x_p\mathbf{a}_p.$$

By linearity and the column picture of multiplication of a matrix by a vector

$$B(A\mathbf{x}) = B\big(x_1\mathbf{a}_1 + \cdots + x_p\mathbf{a}_p\big) = x_1 B\mathbf{a}_1 + \cdots + x_p B\mathbf{a}_p = \begin{bmatrix}B\mathbf{a}_1 \cdots B\mathbf{a}_p\end{bmatrix}\begin{bmatrix}x_1\\ \vdots\\ x_p\end{bmatrix}$$

can be interpreted as the matrix product of the $m \times p$ matrix with columns $B\mathbf{a}_1, \ldots, B\mathbf{a}_p$ and the column vector $\mathbf{x}$. Since we already know how to calculate $B\mathbf{a}_j$ where $\mathbf{a}_j$ is a column vector this allows us to define the **matrix multiplication** to be

$$BA = \begin{bmatrix}B\mathbf{a}_1 \cdots B\mathbf{a}_p\end{bmatrix}$$

and we have achieved that $(BA)\mathbf{x} = B(A\mathbf{x})$. (That the columns of the matrix of the transformation $\mathbf{x} \to B(A\mathbf{x})$ are $B(A\mathbf{e}_j) = B\mathbf{a}_j$ also follows from section 1.9.)

It is more efficient to use the alternative **row-column rule** to compute the (i,j)th entry of BA as the dot product between the ith row of $B = \begin{bmatrix} b_{ij} \end{bmatrix}$ and jth column of $A = \begin{bmatrix} a_{ij} \end{bmatrix}$:

$$(BA)_{ij} = b_{i1}a_{1j} + \cdots + b_{in}a_{nj} \tag{6.1}$$

$$i \text{ th } \mathbf{row} \quad \begin{bmatrix} \\ b_{i1} \cdots b_{in} \\ \\ \end{bmatrix}\begin{bmatrix} a_{1j} \\ \vdots \\ a_{nj} \end{bmatrix}_{j \text{ th } \mathbf{column}} = \begin{bmatrix} \vdots \\ \cdots (BA)_{ij} \cdots \\ \vdots \\ j \text{ th } \mathbf{column} \end{bmatrix} \quad i \text{ th } \mathbf{row}$$

This is because the jth column of BA is $B\mathbf{a}_j$ and the ith row of $B\mathbf{a}_j$ is the dot product of the ith row of B with $\mathbf{a}_j$ in the row picture of multiplication of a matrix by a vector.

Ex Let $B = \begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix}$, $A = \begin{bmatrix} -1 & 1 \\ 1 & 2 \end{bmatrix}$. Find BA

Sol $$B\mathbf{a}_1 = \begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \cdot (-1) + 2 \cdot 1 \\ 0 \cdot (-1) + (-1) \cdot 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix},$$

$$B\mathbf{a}_2 = \begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \cdot 1 + 2 \cdot 2 \\ 0 \cdot 1 + (-1) \cdot 2 \end{bmatrix} = \begin{bmatrix} 5 \\ -2 \end{bmatrix}$$

Hence

$$BA = B\begin{bmatrix}\mathbf{a}_1\mathbf{a}_2\end{bmatrix} = \begin{bmatrix}B\mathbf{a}_1 B\mathbf{a}_2\end{bmatrix} = \begin{bmatrix} 1 & 5 \\ -1 & -2 \end{bmatrix}$$

Alternatively using the row-column method

$$BA = \begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} -1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 1(-1) + 2 \cdot 1 & 1 \cdot 1 + 2 \cdot 2 \\ 0(-1) + (-1)1 & 0 \cdot 1 + (-1)2 \end{bmatrix} = \begin{bmatrix} 1 & 5 \\ -1 & -2 \end{bmatrix}$$

Alternatively one can also write this as

$$\begin{array}{cc} & \begin{bmatrix} -1 & 1 \\ 1 & 2 \end{bmatrix} \\ BA = \begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix} & \begin{bmatrix} 1 & 5 \\ -1 & -2 \end{bmatrix} \end{array}$$

Question Is matrix multiplication **commutative**, i.e. is $AB = BA$?

Why do people expect things to be commutative in math when they are not commutative in real life? It is not the same thing to first put on the shoes and then the socks as it is to first put on the socks and then the shoes?

Question What if A is a 2×3 and B is 3×4? Are AB and BA defined?

Question Is it the same to first rotate and then reflect as it is to first reflect and then rotate?

Ex Let $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ be the matrix of rotation $\frac{\pi}{2}$ counterclockwise and $B = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ be the matrix of reflection in the x_1 axis. Find AB and BA. Interpret geometrically.

Sol

$$AB = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{matrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \\ \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \end{matrix} \qquad BA = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{matrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \\ \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix} \end{matrix}$$

Matrix multiplication need not be commutative, i.e. in general $AB \neq BA$.

Question Given examples of nonzero matrices such that $AB = 0$.

If you first project on the x_1 axis and then on the x_2 axis the result is $\mathbf{0}$.

The identity matrix is $I = \begin{bmatrix} \delta_{ij} \end{bmatrix}$, where $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$:

$$I = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \qquad \text{in case } 4 \times 4.$$

We have $AI = IA = A$ for any matrix A if I has the right size.

Matrix multiplication is however **associative**, i.e. $(CB)A = C(AB)$. This just follows from that they both define the combined map

$$\mathbf{x} \xrightarrow{\text{multiply by } A} A\mathbf{x} \xrightarrow{\text{multiply by } B} B(A\mathbf{x}) \xrightarrow{\text{multiply by } C} C(B(A\mathbf{x}))$$

The transpose of an orthogonal matrix

The transpose A^T is the matrix with rows and columns interchanged, $(A^T)_{ij} = (A)_{ji}$

Ex If $A = \begin{bmatrix} 1 & 2 & 3 \\ -2 & 0 & -1 \\ 4 & 5 & 2 \end{bmatrix}$ then $A^T = \begin{bmatrix} 1 & -2 & 4 \\ 2 & 0 & 5 \\ 3 & -1 & 2 \end{bmatrix}$.

We have e.g. $(AB)^T = B^T A^T$, $(A+B)^T = A^T + B^T$.

Def An $n \times n$ matrix

$$Q = \begin{bmatrix} | & | & & | \\ \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \\ | & | & & | \end{bmatrix}$$

is called orthogonal if the column vectors are orthonormal, i.e. for all i, j

$$\mathbf{q}_i \cdot \mathbf{q}_j = \delta_{ij}, \qquad \text{where} \qquad \delta_{ij} = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j, \end{cases}.$$

An equivalent way to formulate this is that $Q^T Q = I$. This is because the columns of Q become the rows of Q^T and the matrix product is formed by taking the dot product of the rows of Q^T by the columns of Q, by the row column rule (6.1).

Powers of Transition Matrices and equilibrium

A **distribution vector** is a vector with all components positive or 0 and adding up to 1. A **transition matrix** is a matrix in which each column vector is a distribution vector.

Ex Let us consider the mini-web in Ex 4 in section 2.3 in the book, with transition matrix

$$A = \begin{bmatrix} 0 & 1/2 & 0 & 0 \\ 1/2 & 0 & 0 & 1 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

The point is that as $N \to \infty$

$$A^N \to B = \begin{bmatrix} 1/6 & 1/6 & 1/6 & 1/6 \\ 1/3 & 1/3 & 1/3 & 1/3 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{bmatrix}$$

Which means that if we start with any distribution vector it will converge to an equilibrium vector that satisfy $A\mathbf{x}_{equ} = \mathbf{x}_{equ}$, as seen in Ex 9 in section 2.1 in the book.

$$\mathbf{x}_{equ} = \begin{bmatrix} 1/6 \\ 1/3 \\ 1/4 \\ 1/4 \end{bmatrix}$$

Summary and Conceptual Questions

Suppose that B is an $m \times n$ matrix and A is an $n \times p$ matrix. Then multiplication by A defines a map $\mathbf{R}^p \ni \mathbf{x} \to A\mathbf{x} = \mathbf{y} \in \mathbf{R}^n$ and multiplication by B defines a map $\mathbf{R}^n \ni \mathbf{y} \to B\mathbf{y} = \mathbf{z} \in \mathbf{R}^m$ and multiplication by first A and then B

$$\mathbf{x} \xrightarrow{\text{multiply by } A} A\mathbf{x} \xrightarrow{\text{multiply by } B} B(A\mathbf{x})$$

defines a map $\mathbf{R}^p \ni \mathbf{x} \to T(\mathbf{x}) = B(A\mathbf{x}) \in \mathbf{R}^m$, which it is easy to see is linear. We define the matrix product BA to be the $m \times p$ matrix that represents this map so that $(BA)\mathbf{x} = B(A\mathbf{x})$:

$$\mathbf{x} \xrightarrow{\qquad \text{multiply by } BA \qquad} (BA)\mathbf{x} = B(A\mathbf{x})$$

We can calculate this by first using the column picture first for $A\mathbf{x}$

$$A\mathbf{x} = \begin{bmatrix}\mathbf{a}_1 \cdots \mathbf{a}_p\end{bmatrix}\begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix} = x_1\mathbf{a}_1 + \cdots + x_p\mathbf{a}_p.$$

and then using linearity and the column picture for B

$$B(A\mathbf{x}) = x_1 B\mathbf{a}_1 + \cdots + x_p B\mathbf{a}_p = \begin{bmatrix}B\mathbf{a}_1 \cdots B\mathbf{a}_p\end{bmatrix}\begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}$$

This can be interpreted as the matrix product of $\mathbf{x}$ with the $m \times p$ matrix

$$BA = \begin{bmatrix}B\mathbf{a}_1 \cdots B\mathbf{a}_p\end{bmatrix}$$

It is more efficient to use the alternative **row-column rule** to compute the (i,j)th entry of BA as the dot product between the ith row of $B = \begin{bmatrix} b_{ij} \end{bmatrix}$ and jth column of $A = \begin{bmatrix} a_{ij} \end{bmatrix}$:

$$(BA)_{ij} = b_{i1}a_{1j} + \cdots + b_{in}a_{nj}$$

$$i \text{ th } \textbf{row} \quad \begin{bmatrix} \\ b_{i1} \cdots b_{in} \\ \\ \end{bmatrix} \begin{bmatrix} a_{1j} \\ \vdots \\ a_{nj} \end{bmatrix}_{j \text{ th } \textbf{column}} = \begin{bmatrix} \vdots \\ \cdots (BA)_{ij} \cdots \\ \vdots \\ j \text{ th } \textbf{column} \end{bmatrix} \quad i \text{ th } \textbf{row}$$

This is because the jth column of BA is $B\mathbf{a}_j$ and the ith row of $B\mathbf{a}_j$ is the dot product of the ith row of B with $\mathbf{a}_j$ in the row picture of multiplication of a matrix by a vector.

Matrix multiplication need not be commutative, i.e. in general $AB \neq BA$.

Ex Let $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ be the matrix of rotation $\dfrac{\pi}{2}$ counterclockwise and $B = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ be the matrix of reflection in the x_1 axis. Find AB and BA. Interpret geometrically.
Sol

$$AB = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \overset{\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}}{\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}} \qquad BA = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \overset{\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}}{\begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}}$$

Question What does AB represent geometrically?
Question What does BA represent geometrically?