EECS 16A Designing Information Devices and Systems I
Fall 2020 Lecture Notes Note 2B

2.1 Matrix-Matrix Multiplication (i.e., Transformation of Spaces)

Matrix-matrix multiplication is another powerful tool for modeling linear systems, which we will discuss further in later notes. As an example, two matrices A and B in $\mathbb{R}^{2\times 2}$ can be multiplied as follows:

$$\underbrace{\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}}_{\text{A}} \underbrace{\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}}_{\text{B}} = \underbrace{\begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{bmatrix}}_{\text{AB}}$$

Computationally, matrix-matrix multiplication involves multiplying each row vector in A with each column vector in B, starting from the top row of matrix A and leftmost column of matrix B. Effectively, the left matrix is multiplied by each column vector in the second matrix to produce a new column of AB. Why columns and not rows? That's just convention. But this does lead to an important point about the dimensions of matrix-matrix multiplication.

To left-multiply a matrix B by another matrix A, the number of columns in A must equal the number of rows in B. Otherwise, the product $A \times B$ cannot be calculated. Moreover, if A is an $m \times n$ matrix and B is $n \times p$, the product $A \times B$ will have dimensions $m \times p$. A visual illustration of this can be seen here, where the left matrix is broken up into m row vectors and the right matrix is represented as p column vectors:

$$\begin{bmatrix} - - - - \vec{r}_1^T - - - - \\ - - - - \vec{r}_2^T - - - - \\ \vdots \\ - - - - \vec{r}_m^T - - - - \end{bmatrix} \begin{bmatrix} | & | & & | \\ | & | & & | \\ \vec{c}_1 & \vec{c}_2 & \cdots & \vec{c}_p \\ | & | & & | \\ | & | & & | \end{bmatrix} = \begin{bmatrix} \vec{r}_1^T\vec{c}_1 & \vec{r}_1^T\vec{c}_2 & \cdots & \vec{r}_1^T\vec{c}_p \\ \vec{r}_2^T\vec{c}_1 & \vec{r}_2^T\vec{c}_2 & \cdots & \vec{r}_2^T\vec{c}_p \\ \vdots & \vdots & & \vdots \\ \vec{r}_m^T\vec{c}_1 & \vec{r}_m^T\vec{c}_2 & \cdots & \vec{r}_m^T\vec{c}_p \end{bmatrix}$$

In order for the inner product $\vec{r}_i^T\vec{c}_j$ to be defined, each row vector ($\vec{r}_i^T$) must have the same number of entries as each column vector ($\vec{c}_j$). As a result, matrix-matrix multiplication is typically not commutative — $A \times B$ does not necessarily equal $B \times A$. In fact, both quantities can only be calculated if the number of rows in A equals the number of columns in B **and** the number of rows in B equals the number of columns in A.

To illustrate this, consider the following example of taking the product of two 2×2 matrices.

Example 2.1 (Matrix Multiplication):

$$\begin{bmatrix} 2 & 4 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} (2)(1)+(4)(3) & (2)(2)+(4)(4) \\ (3)(1)+(1)(3) & (3)(2)+(1)(4) \end{bmatrix} = \begin{bmatrix} 14 & 20 \\ 6 & 10 \end{bmatrix}$$

Example 2.2 (Matrix Multiplication is Not Commutative!): Above, we mentioned that matrix multiplication does not commute - that is to say, there exist matrices A and B such that $AB \neq BA$. Let's see if we can come up with such an example to verify that assertion.

A natural approach would be to take the matrices from the above example, multiply them in the other order, and see if we get the same answer. Let's try it out! Swapping the order of the matrices, we obtain

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 2 & 4 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} (1)(2)+(2)(3) & (1)(4)+(2)(1) \\ (3)(2)+(4)(3) & (3)(4)+(4)(1) \end{bmatrix} = \begin{bmatrix} 8 & 6 \\ 18 & 16 \end{bmatrix}$$

As expected, we did not end up with the same result as we did before. Having produced a counterexample, we have therefore proven that matrix multiplication is not generally commutative.

Be aware, however, that there still might (and indeed do!) exist pairs of matrices whose product *is* commutative. All we have shown here is that *not all* pairs of matrices produce the same product when multiplied in the opposite order.

Example 2.3 (Matrix Multiplication is Associative!): Having seen above that matrix multiplication is not commutative, we might start asking questions about associativity, as well. In particular, is it true that given three matrices A, B, and C, that $(AB)C = A(BC)$? Put differently, does the *grouping* of matrices in a product not matter, if the order is kept the same throughout?

As it turns out, this is true. Unfortunately, a general proof of associativity is tedious and relies on just repeatedly applying the component-wise definition of matrix multiplication. To gain some intuition about associativity, it's better to simply consider an example, such as the following:

$$\begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} 7 & 8 \\ 9 & 10 \end{bmatrix} \begin{bmatrix} 11 & 12 \\ 13 & 14 \end{bmatrix}.$$

The above product can be evaluated in two different ways - we will do both, and verify that we get the same answer either way.

Let's first multiply the first two matrices together, before multiplying their product with the third:

$$\begin{aligned}
\left(\begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} 7 & 8 \\ 9 & 10 \end{bmatrix} \right) \begin{bmatrix} 11 & 12 \\ 13 & 14 \end{bmatrix} &= \begin{bmatrix} (3)(7)+(4)(9) & (3)(8)+(4)(10) \\ (5)(7)+(6)(9) & (5)(8)+(6)(10) \end{bmatrix} \begin{bmatrix} 11 & 12 \\ 13 & 14 \end{bmatrix} \\
&= \begin{bmatrix} 57 & 64 \\ 89 & 100 \end{bmatrix} \begin{bmatrix} 11 & 12 \\ 13 & 14 \end{bmatrix} \\
&= \begin{bmatrix} (57)(11)+(64)(13) & (57)(12)+(64)(14) \\ (89)(11)+(100)(13) & (89)(12)+(100)(14) \end{bmatrix} \\
&= \begin{bmatrix} 1459 & 1580 \\ 2279 & 2468 \end{bmatrix}.
\end{aligned}$$

Then, let's try multiplying the last two matrices together first, before multiplying the first matrix with that product:

$$\begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix} \left(\begin{bmatrix} 7 & 8 \\ 9 & 10 \end{bmatrix} \begin{bmatrix} 11 & 12 \\ 13 & 14 \end{bmatrix} \right) = \begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} (7)(11)+(8)(13) & (7)(12)+(8)(14) \\ (9)(11)+(10)(13) & (9)(12)+(10)(14) \end{bmatrix}$$
$$= \begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} 181 & 196 \\ 229 & 248 \end{bmatrix}$$
$$= \begin{bmatrix} (3)(181)+(4)(229) & (3)(196)+(4)(248) \\ (5)(181)+(6)(229) & (5)(196)+(6)(248) \end{bmatrix}$$
$$= \begin{bmatrix} 1459 & 1580 \\ 2279 & 2468 \end{bmatrix},$$

which is the same as what we got before!

The fact that three fairly arbitrary matrices exhibit associativity when being multiplied should be a strong hint that matrix multiplication is probably associative - however, it is important to understand that *this is not a proof* of the associativity of matrix multiplication. To prove that matrix multiplication is associative, we'd have to show that *any* triplet of matrices can be multiplied in either order without changing the final answer - showing that it seems to work for particular examples is not sufficient.

Example 2.4 (Matrices as Functions): In a single-variable situation, we might have a function f that takes in a number x and outputs a number $f(x)$. If we want functions of multiple variables, we can use vectors. The input $\vec{x}$ is now a list of variables. The output is another list of numbers. If f is **linear**, then it acts on a list of variables by multiplying them by scalars and adding them together. In this case, we can represent f as a matrix. Therefore, matrices are also called **linear maps** or **linear transformations**.

As an example, recall the water reservoir, where applying the matrix to the current distribution of water gives us the next day's distribution:

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{3} \\ \frac{1}{2} & 0 & \frac{1}{3} \\ 0 & \frac{3}{4} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} x_A \\ x_B \\ x_C \end{bmatrix} = \begin{bmatrix} x'_A \\ x'_B \\ x'_C \end{bmatrix}$$

Here, we have three variables, one for each reservoir. We want a function that takes in a water distribution and gives us the water distribution one day later, which is represented by the matrix A.

What if we want the water distribution two days later? We could apply A twice, giving us $A(A\vec{x})$. Alternatively, a key property of matrix multiplication is **associativity**, or $(AB)C = A(BC)$, so we know that $A(A\vec{x}) = (AA)\vec{x}$. Therefore, we can use matrix-matrix multiplication to produce AA:

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{3} \\ \frac{1}{2} & 0 & \frac{1}{3} \\ 0 & \frac{3}{4} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{3} \\ \frac{1}{2} & 0 & \frac{1}{3} \\ 0 & \frac{3}{4} & \frac{1}{3} \end{bmatrix} = \begin{bmatrix} \frac{3}{8} & \frac{3}{8} & \frac{13}{36} \\ \frac{1}{4} & \frac{3}{8} & \frac{5}{18} \\ \frac{3}{8} & \frac{1}{4} & \frac{13}{36} \end{bmatrix}$$

This operation gives us a single matrix representing two days of water flow. In other words, matrix multiplication implements function composition, and AA represents applying the function A twice.

In algebra, we learned how to manipulate functions of one variable. Linear algebra teaches us how to manipulate linear functions of multiple variables.

In a later note, we will further explore how matrix-matrix multiplication applies to linear transformations.

Additional Resources For more on matrix-matrix multiplication, read *Strang* pages 61-62, and try Problem Set 2.3.

In *Schuam's*, read pages 30-33 and try Problems 2.4 to 2.11, 2.39 to 2.40, 2.42, 2.44 - 2.49, 2.12 to 2.16, 2.41, 2.43, and 2.72. *Extra: Understand Polynomials in Matrices.*

2.2 Linear Transformations

2.2.1 A Natural Generalization

In an earlier note, we looked at linear functions over the reals - specifically, we defined a scalar function $f(x)$ to be linear if, for any scalar k,

$$f(kx) = k \cdot f(x).$$

We will now work to generalize this definition to functions acting on vectors. The most natural generalization would simply be to replace x with $\vec{x}$ in the above definition - in other words, we might define a function f to be linear if and only

$$f(k\vec{x}) = kf(\vec{x}).$$

Unfortunately, the above definition isn't quite sufficient. Why? Well, in the scalar ("one-dimensional") case, knowing $f(x_0)$ for a single nonzero scalar x_0 was sufficient to define $f(x)$ over all the reals, since we could write

$$x = (x/x_0)x_0 \implies f(x) = (x/x_0)f(x_0).$$

Can we do something similar now, working over vectors? Imagine working in two dimensional space, where the span of two vectors $\vec{x}_0$ and $\vec{x}_1$ is $\mathbb{R}^2$. By definition, we know any vector $\vec{x} \in \mathbb{R}^2$ can be expressed as a linear combination

$$\vec{x} = \alpha\vec{x}_0 + \beta\vec{x}_1$$

of the two vectors whose span we are considering. Thus, a natural analog of our result over scalars would be to say that

$$f(\vec{x}) = f(\alpha\vec{x}_0 + \beta\vec{x}_1) = \alpha f(\vec{x}_0) + \beta f(\vec{x}_1).$$

More generally, given the output of a linear function for a given set of vectors, we'd like to be able to evaluate the function at any point in the span of the given set of vectors. Unfortunately, our proposed definition of linearity doesn't let us do this. Why? Consider the following function $f : \mathbb{R}^2 \to \mathbb{R}$:

$$f(\vec{x}) = f\left(\begin{bmatrix} x_0 \\ x_1 \end{bmatrix}\right) = \left\{\begin{array}{lr} 2x_0, & \text{for } x_0 = x_1 \\ x_1, & \text{for } x_0 = -x_1 \\ 0, & \text{otherwise} \end{array}\right\}.$$

Some inspection of $f(\vec{x})$ will show that $f(k\vec{x}) = kf(\vec{x})$ for all $\vec{x}$. But observe that while

$$\begin{bmatrix} 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ -1 \end{bmatrix},$$

$$f\left(\begin{bmatrix} 2 \\ 0 \end{bmatrix}\right) = 0 \neq 2 - 1 = f\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) + f\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right),$$

so our desired generalization doesn't hold. Clearly, if we want our generalized result to hold, we need to strengthen our definition of linearity.

2.2.2 Additivity

One way to do so is to introduce one further requirement, known as **additivity** - specifically, that

$$f(\vec{x}+\vec{y}) = f(\vec{x}) + f(\vec{y})$$

for all $\vec{x}$ and $\vec{y}$. Observe now that, by applying additivity as well as our previous requirement (known as **homogeneity**), we can directly show that

$$\begin{aligned} f(\alpha\vec{x}_0 + \beta\vec{x}_1) &= f(\alpha\vec{x}_0) + f(\beta\vec{x}_1) \\ &= \alpha f(\vec{x}_0) + \beta f(\vec{x}_1), \end{aligned}$$

as desired! As it turns out, these two requirements are all that are needed to generalize linear functions to act over vectors, where they are known as **linear transformations**. One interesting thing to note is that additivity also holds for scalar linear functions, and can be derived from homogeneity - it's only when working with vectors that additivity starts to give us something new.[1]

2.2.3 Matrices as Linear Transformations

So far, we've established the requirements that a linear transformation must satisfy. But what *is* a linear transformation, really? As it turns out, multiplying a matrix with a column vector is a linear transformation - specifically, the function

$$f_A(\vec{x}) = A\vec{x}$$

is a linear transformation for any matrix A. Typically, we simplify this statement by stating that the matrix A *itself* is a linear transformation, with the matrix used to represent the transformation f_A.

But why is this true? To check if a function is a linear transformation, we simply need to verify that it satisfies the requirements of homogeneity and additivity. Observe that, by the rules of matrix-vector multiplication,

$$\begin{aligned} f_A(\vec{x}+\vec{y}) &= A(\vec{x}+\vec{y}) = A\vec{x} + A\vec{y} = f_A(\vec{x}) + f_A(\vec{y}) \\ f_A(k\vec{x}) &= A(k\vec{x}) = k(A\vec{x}) = k f_A(k\vec{x}), \end{aligned}$$

where $\vec{x}$ and $\vec{y}$ are arbitrary vectors, A is a matrix with the appropriate dimensions, and k is an arbitrary real scalar, so both additivity and homogeneity are satisfied by matrix multiplication. Thus, matrix multiplication is a linear transformation, as we claimed earlier.

One final piece of jargon remains to be introduced - when a linear transformation yields vectors of the same dimension as its input (i.e. if $f(\vec{x})$ has the same dimension as $\vec{x}$) then it is sometimes called a **linear operator**.

[1] A good question to ask at this point would be: does additivity imply homogeneity? As it turns out, the answer is no - try to produce a function that satisfies additivity but not homogeneity! It's fairly easy to do so when working over the field of complex numbers, but *much* harder to do so when working over the reals, like we do here.

2.3 Practice Problems

These practice problems are also available in an interactive form on the course website.

1. Multiply $\begin{bmatrix} 1 & 5 & 0 \\ 10 & 3 & 7 \\ 6 & 4 & 11 \end{bmatrix}$ with $\begin{bmatrix} 2 & 12 & 3 \\ 1 & 8 & 0 \\ 9 & 1 & 2 \end{bmatrix}$. What is the first row of the resulting matrix?

   (a) $\begin{bmatrix} 16 & 52 & 5 \end{bmatrix}$

   (b) $\begin{bmatrix} 3 & 40 & 7 \end{bmatrix}$

   (c) $\begin{bmatrix} 7 & 52 & 3 \end{bmatrix}$

   (d) $\begin{bmatrix} 14 & 13 & 2 \end{bmatrix}$

2. **Matrix Multiplication**

   Consider the following matrices:

   $\mathbf{A_1} = \begin{bmatrix} 1 & 4 \end{bmatrix}$ $\mathbf{B_1} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$ $\mathbf{A} = \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}$ $\mathbf{B} = \begin{bmatrix} 3 & 2 \\ 2 & 1 \end{bmatrix}$ $\mathbf{C} = \begin{bmatrix} 1 & 9 & 5 & 7 \\ 4 & 3 & 2 & 2 \end{bmatrix}$ $\mathbf{D} = \begin{bmatrix} 5 & 5 & 8 \\ 6 & 1 & 2 \\ 4 & 1 & 7 \\ 3 & 2 & 2 \end{bmatrix}$ $\mathbf{E} = \begin{bmatrix} 8 & 1 & 6 \\ 3 & 5 & 7 \\ 4 & 9 & 2 \end{bmatrix}$ $\mathbf{F} = \begin{bmatrix} 5 & 3 & 4 \\ 1 & 8 & 2 \\ 2 & 3 & 5 \end{bmatrix}$

   For each matrix multiplication problem, if the product exists, find the product by hand. Otherwise, explain why the product does not exist.

   (a) $\mathbf{A_1 B_1}$

   (b) $\mathbf{AB}$

   (c) $\mathbf{BA}$

   (d) $\mathbf{AC}$

   (e) $\mathbf{DC}$

   (f) $\mathbf{CD}$ (Write down the dimensions of the product if it exists. For practice, you can compute the product on your own)

   (g) $\mathbf{EF}$ (Practice on your own)

   (h) $\mathbf{FE}$ (Practice on your own)