Linear Algebra – Lesson 2

Matrices and linear transformations

1. Linear transformations

This lesson focuses on how we can manipulate vectors. This will lead us to the concept of matrices. In the next lesson we will see how matrices are important for concepts like "change of basis" or the solution to a system of linear equations.

1.1 Transformations

Linear transformations can be thought of as functions that take a vector as input and give a vector as output. As you know, a function is a mapping of numbers from one space, called "the domain", to another space, called "the codomain".

For example, $f(x) = 2x$ is a function from the real numbers to the real numbers $f: \mathbb{R} \to \mathbb{R}$.
The codomain can also be different. For example, $f(x) = floor(x)$ is a function from the real numbers to the integers ($f: \mathbb{R} \to \mathbb{Z}$).

Similarly, a transformation T takes as input a vector and returns a vector.
For example, $T(\bar{v}) = 2\bar{v}$ is a transformation that scales every vector by 2.
Question: What are the domain and codomain of T?
Answer: Well, we didn't define it. But what is clear is that they are identical. If we start with a vector $\bar{v} \in \mathbb{R}^2$ we get another vector in $\mathbb{R}^2$. So we could write $T: \mathbb{R}^2 \to \mathbb{R}^2$.

Generally, we write: $T: V \to W$. This reads: "T is a transformation from vectors in V to vectors in W".

1.2 Linear transformations

A special kind of transformation is called a linear transformation. As you saw in the video, we can intuitively think of linear transformation as transformations that keep the origin in place, and keep parallel lines parallel to each other.
More formally, a transformation T is called a linear transformation iff for any two vectors $\bar{u}, \bar{v}$ and any scalar a:

$$T(\bar{v} + \bar{u}) = T(\bar{v}) + T(\bar{u})$$
$$T(a\bar{v}) = aT(\bar{v})$$

In other words, it doesn't matter if you first scale (or add) vectors and then apply the transform, or first apply the transform and then scale (or add) them.

Example 1 – is this a linear transform?

Let T be the projection transform $T: \mathbb{R}^3 \to \mathbb{R}^3$, such that $T\left(\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}\right) = \begin{bmatrix} v_1 \\ v_2 \\ 0 \end{bmatrix}$.

Is it linear? To check this, we have to verify the two conditions:

$$T\left(\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}\right) = T\left(\begin{bmatrix} v_1 + u_1 \\ v_2 + u_2 \\ v_3 + u_3 \end{bmatrix}\right) = \begin{bmatrix} v_1 + u_1 \\ v_2 + u_2 \\ 0 \end{bmatrix}$$

And on the other hand: $T\left(\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}\right) + T\left(\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}\right) = \begin{bmatrix} v_1 \\ v_2 \\ 0 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ 0 \end{bmatrix} = \begin{bmatrix} v_1 + u_1 \\ v_2 + u_2 \\ 0 \end{bmatrix}$
So the first condition is fulfilled.
The second condition you will prove in the exercise.

1.3 The effect of transforms on the basis vectors

Notice that to understand what a certain linear transform does, we don't have to go and see what it does to any given vector.
In fact, it's enough to know what it does to the basis vectors. Why? Because any vector $\bar{v}$ is just a linear combination of the basis vectors:

$$\bar{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = v_1\hat{\imath} + v_2\hat{\jmath}$$

And so, by definition of a linear transform:

$$T(\bar{v}) = T(v_1\hat{\imath} + v_2\hat{\jmath}) = v_1T(\hat{\imath}) + v_2T(\hat{\jmath})$$

In other words, to apply a linear transformation on $\bar{v}$, we can decompose $\bar{v}$ to its basis components, apply the transformation separately for each components, and then add them together.

As an example, let's look at the rotation transform, which rotates every vector by 90 degrees clockwise.
$T_{90}\left(\begin{bmatrix} 2 \\ 3 \end{bmatrix}\right) = ?$
$T_{90}(\hat{\imath}) = T_{90}\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$, $\quad T_{90}(\hat{\jmath}) = T_{90}\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$
So $T_{90}\left(\begin{bmatrix} 2 \\ 3 \end{bmatrix}\right) = 2\begin{bmatrix} 0 \\ -1 \end{bmatrix} + 3\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$

Now, imaging that we want to a quick "recipe" for calculating this transformation, so we don't have to go through these steps every time we get a new vector. To do so, we can generally write for any vector $\begin{bmatrix} x \\ y \end{bmatrix}$:

$$T_{90}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = x\begin{bmatrix} 0 \\ -1 \end{bmatrix} + y\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \cdot x + 1 \cdot y \\ -1 \cdot x + 0 \cdot y \end{bmatrix}$$

This is the most important point:
If $\overline{v}$ is a linear combination of some basis vectors, the transformed $\boldsymbol{T}(\overline{v})$ is the same linear combination, but of the transformed basis vectors[1].
This leads us the useful notation of matrix representation.

2. Matrix representation of linear transformations

2.1 Intuition to matrix representation

Now, let's say we're dealing with a different transformation, a general one, that takes $\hat{\imath}$ to $\begin{bmatrix} a \\ c \end{bmatrix}$ and $\hat{\jmath}$ to $\begin{bmatrix} b \\ d \end{bmatrix}$. What will be the result of applying this transformation to $\begin{bmatrix} x \\ y \end{bmatrix}$?

$$x\begin{bmatrix} a \\ c \end{bmatrix} + y\begin{bmatrix} b \\ d \end{bmatrix} = \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix}$$

We can represent our transformation with a matrix, an array of numbers, that tells us where each basis vector "lands":

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

This is the matrix representation of the transform.
The first column tells us where $\hat{\imath}$ lands, and the second column tells us where $\hat{\jmath}$ lands.

To apply the transform on a vector, we do what's called "matrix multiplication":

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = x\begin{bmatrix} a \\ c \end{bmatrix} + y\begin{bmatrix} b \\ d \end{bmatrix} = \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix}$$

So the result of matrix-vector multiplication is just the appropriate linear combination of the column vectors of the matrix.

In other words, multiplying a vector by a matrix is the same as applying the appropriate transformation to that vector. **A matrix is the coordinate-based description of a linear transformation** (which by itself need not have a coordinate system attached to it)[2].

Example – Matrix-vector multiplication as linear transformation
What does the following matrix "do" to a vector?

$$\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

We just have to imagine what it does to the basis vectors. So $\hat{\jmath}$ remains unchanged, but $\hat{\imath}$ moves to $\begin{bmatrix} 2 \\ 0 \end{bmatrix}$.
In other words, this transformation expands any vector on the x axis, and does nothing on the y axis.

Note: This matrix is called a **diagonal matrix**, because all its off-diagonal terms are zero ($\forall i \neq j$: $a_{ij} = 0$).

2.2 Matrix notation

---

[1] You can see this in video 3 in 3blue1brown, 3:45.
[2] This is discussed in Strang's lecture 31.

While we denote vectors by lowercase letter, we denote matrices by capital letters: A,B etc.
So multiplying a vector $\bar{v}$ by a matrix A is written as $A\bar{v}$.

Also, while for vectors we only needed one index to refer to different entries (v_1, v_2 $etc.$), matrices require two indices. By convention, the first index represents the row, and the second represents the column:

$$A = \begin{bmatrix} 2 & 3 \\ 0 & 5 \end{bmatrix} \rightarrow \quad a_{11} = 2, a_{12} = 3, a_{21} = 0, a_{22} = 5$$

As you can see, we usually use lower-case letters to refer to specific terms in the matrix.

2.3 Matrix vector multiplication using Σ notation

We can define the product of matrix-vector multiplication in the general case. Let $A_{m\times n}$ be a matrix with m rows and n columns, and $\bar{v}$ be a vector in $\mathbb{R}^n$. Then the ith entry of the new vector $A\bar{v}$ is:

$$[A\bar{v}]_i = \sum_{k=1}^{n} A_{ik}\bar{v}_k$$

3. Image (column space), rank, kernel (null space)

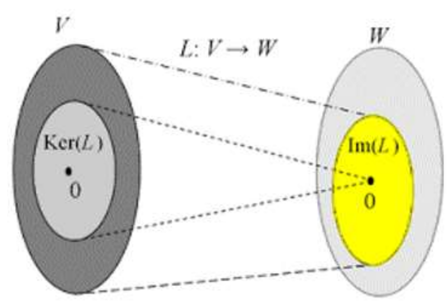
Given a matrix A which maps vectors from V to U:

$$A: V \rightarrow U$$

There are two interesting spaces we are interested in, called the Image of A ($Im(A)$) and the kernel of $(\ker(A))$.

3.1 $Im(A)$, the Image of a matrix, the column space of a matrix

A basic property we might want to know about a matrix is what its image is – what are the possible vectors it can give us. In other words – what is the **span** of the column vectors of the matrix? The span of the columns of a matrix is also called "the column space" of that matrix, the "image" of the matrix, or the "range" of the matrix. It is noted $Im(A)$. Naturally, $Im(A) \in U$.



3.1.2 The rank

The **dimension** of the $Im(A)$ is called the "rank of the A":

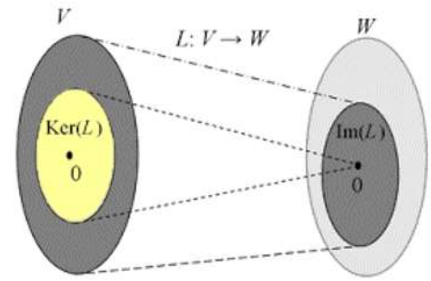
$$rank(A) = \dim\big(Im(A)\big)$$

This is also called the column rank of A.

Similarly, the row rank is defined as the dimension of the row space (the space spanned by the rows of the matrix). A theorem says that the column rank and row rank are equal. We can therefore talk about the rank of a matrix (knowing that it refers to both definitions).

3.2 Kernel, null space

A special case which will become very important when we use matrices to solve real-world problems, is the case in which a vector $\bar{v}$ is mapped to the 0 vector. The set of all vectors that A maps to the 0 vector is called the "null space" or "kernel".
Naturally, $ker(A) \in V$.
Given $A_{m\times n}$: $ker(A) = \{\bar{v} \in V | A\bar{v} = o\}$



3.2.1 Nullity of A

The **dimension** of $\ker(A)$ is called the "nullity of A":

$$nullity(A) = \dim(\ker(A))$$

Example – null space

For example, let's look at the matrix A:

$$A = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

Question: What linear transformation does it represent? What is its image? What is its kernel?
Answer: The matrix A represents a projection onto the y-axis (onto $\hat{\jmath}$).
Therefore, its image is the y-axis: $Im(A) = \left\{\bar{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} | v_1 = 0, v_2 \in \mathbb{R}\right\}$
What about its kernel? We can see that A will take any vector on the x-axis (i.e., along $\hat{\imath}$) to 0. So its null space is $Ker(A) = nullsp(A) = \left\{\bar{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} |, v_1 \in \mathbb{R}, v_2 = 0\right\}$

Example 2 – Rank of a matrix

What is the rank of $A = \begin{bmatrix} 1 & -3 \\ 2 & -6 \end{bmatrix}$?
Geometrically, this matrix transforms $\hat{\imath}$ to $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\hat{\jmath}$ to $\begin{bmatrix} -3 \\ -6 \end{bmatrix}$. We can see that any vector in the image of A will therefore fall on this line. The dimension of a line is 1, and therefore $rank(A) = 1$[3].
Note that we could also solve it in a different way. As you remember from the exercise, given a set of vectors, the dimension of their span is the maximum number of linearly independent vectors you can choose from them. It is easy to see that $\bar{a}_2 = 3\bar{a}_1$, and therefore $dim(SP(\{\bar{a}_1, \bar{a}_2\}) = 1$.

There is an easy algorithm for finding the rank of a matrix, we might talk about it later in the course. But in fact, sometimes you can just "see" what the rank is, if you can figure out linear dependencies between the columns.

3.3 Rank-nullity theorem

When we solve real-world problems, the rank and nullity have very important consequences. An important theorem that relates them is the rank-nullity theorem:

Left $A: V \to U$ be a linear transformation (V is finite). Then:

[3] You can see a similar example in video 3 of 3blue1brown, 9:20.

$$\dim(V) = \dim(\ker(A)) + \dim\big(Im(A)\big) = nullity(A) + rank(A)$$

We will not prove this theorem.

4. Composite transformations

4.1 Composite transformations as matrix multiplication

So far we've seen different transformations, like rotation, scaling and shear. But what happens if we want to apply two transformation one after the other?

Let A be a 2x2 rotation matrix of 90 degrees counterclockwise, and B be a 2x2 matrix that scales $\hat{\imath}$ by 3. First, let's find A and B (by thinking about where $\hat{\imath}$ and $\hat{\jmath}$ land:

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, B = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}$$

Now, given a vector $\bar{v}$, we want to first rotate, then scale. We first apply the rotation transformation:

$$A\bar{v}$$

This will give us a new vector. Now, we can apply the second transformation to this new vector:

$$B(A\bar{v}) = BA\bar{v}$$

Let's see a concrete example:

$$\bar{v} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

Read right to left ←

$$BA\bar{v} = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \cdots$$

In fact, we could simply compute one new matrix that combines the two transformations:

$$BA = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \cdots$$

Matrix multiplication is simply a series of matrix-vector multiplications.

Another way to look at matrix multiplication

Another way to look at matrix multiplication is using a formula[4].

$$AB = C$$

To find a formula for the general element C_{ij}, let's look at a specific example, say C_{34}.

$$C_{34} = (row\ 3\ of\ A) \cdot (column\ 4\ of\ B)$$

This is called a "dot product" between two vectors, and we will learn about it later in the course. In essence, it is the **sum of the element-by-element products**:

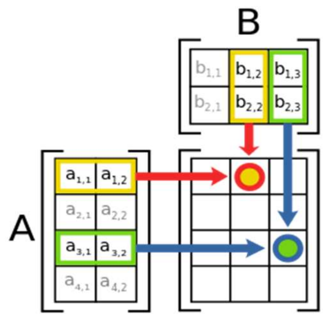
$$c_{34} = a_{31}b_{14} + a_{32}b_{24} + \cdots + a_{3n}b_{n4} = \sum_{k=1}^{n} a_{3k}b_{k4}$$

And in general:

[4] See examples and explanations at the beginning of video 3 from Gilbert Strang's videos.

$$c_{ij} = a_{i\rightarrow} b_{\downarrow j} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

The following diagram from [wikipedia]() also helps visualizing how each element in the product matrix is calculated:



Another visual explanation for matrix-matrix multiplication can be found in the book "No bullshit guide to linear algebra":

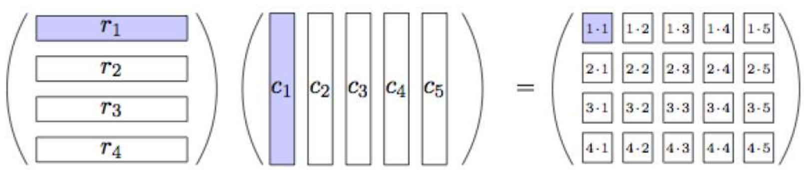


Figure 3.3: Matrix multiplication is performed rows-times-columns. The first-row, first-column entry of the product is the dot product of r_1 and c_1.

Example Consider the matrices $A \in \mathbb{R}^{2\times 3}$ and $B \in \mathbb{R}^{3\times 2}$. The product $AB = C \in \mathbb{R}^{2\times 2}$ is computed as

$$\underbrace{\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}}_{B} = \begin{bmatrix} 1+6+15 & 2+8+18 \\ 4+15+30 & 8+20+36 \end{bmatrix} = \underbrace{\begin{bmatrix} 22 & 28 \\ 49 & 64 \end{bmatrix}}_{C}.$$

We can also compute the product $BA = D \in \mathbb{R}^{3\times 3}$:

- $$\underbrace{\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}}_{B} \underbrace{\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}}_{A} = \begin{bmatrix} 1+8 & 2+10 & 3+12 \\ 3+16 & 6+20 & 9+24 \\ 5+24 & 10+30 & 15+36 \end{bmatrix} = \underbrace{\begin{bmatrix} 9 & 12 & 15 \\ 19 & 26 & 33 \\ 29 & 40 & 51 \end{bmatrix}}_{D}.$$

4.2 The order matters in matrix multiplication

Does the order matter? Looking at the previous example, we can easily see that in general:

$$AB \neq BA$$

4.3 Associativity

What if we have 3 matrices? What does it mean to have the product ABC?
Well, right now, it doesn't mean anything, because we only defined matrix multiplication for 2 matrices. This is where the property of associativity comes in. Associativity means that given 3 matrices A,B,C:

(AB)C=A(BC)

This simply means, that the order by which you compute the final matrix, doesn't matter (although the order of the matrices themselves, A,B,C, does matter, as we saw above).

How can we prove this?
Although annoying, you can simply calculate this product in two ways:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} h & i \\ j & k \end{pmatrix}\begin{pmatrix} w & x \\ y & z \end{pmatrix} = \begin{pmatrix} ah+bj & ai+bk \\ ch+dj & ci+dk \end{pmatrix}\begin{pmatrix} w & x \\ y & z \end{pmatrix} = \cdots$$

5. Matrix addition and scalar multiplication

5.1 Matrix addition

Matrix addition is very intuitively defined as the addition of the corresponding elements.
The sum of two matrices A and B of the same size (written A+B) is a matrix defined by adding the corresponding elements from A and B:

$$A+B = \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \dots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \dots & a_{2n}+b_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1}+b_{m1} & a_{m2}+b_{m2} & \dots & a_{mn}+b_{mn} \end{bmatrix}$$

5.2 Matrix scalar multiplication

The product of a matrix A and a scalar k (written as kA) is a matrix obtained by multiplying each element of A by k:

$$kA = \begin{bmatrix} ka_{11} & ka_{12} & \dots & ka_{1n} \\ ka_{21} & ka_{22} & \dots & ka_{2n} \\ \dots & \dots & \dots & \dots \\ ka_{m1} & ka_{m2} & \dots & ka_{mn} \end{bmatrix}$$

5.3 Left and right distributivity (of matrix product with respect to matrix addition)

Distributivity (הוצאת גורם משותף) of the matrix product means:

$$A(B+C) = AB + AC$$
$$(B+C)D = BD + CD$$

This is easy to show by thinking about the ij-th element of the resulting matrix:

$$\sum_k a_{ik}(b_{kj}+c_{kj}) = \sum_k a_{ik}b_{kj} + \sum_k a_{ik}c_{kj}$$

And similarly for left distributivity.
In other words, the distributivity property for matrices results from the distributivity property of scalars.

6. Non-square matrices

So far we saw only square matrices. But matrices don't have to be square, and in general we talk about matrices of size $m \times n$, where m is the number of rows and n is the number of columns.

What does a non-square matrix "do"? It takes vectors from n-dimensional space, and maps them to vectors in m-dimensional space:
$A_{m\times n}: \mathbb{R}^n \to \mathbb{R}^m$

Question: Given two matrices $A_{m\_\_}, B_{kn}$ What condition must be fulfilled for matrix multiplication AB to be defined?
Answer: The columns of A must be the same dimension as the rows of B.
Why? Well, the columns of B tell us where each basis vector ends up. Now, these vectors are described with k elements, meaning they live in k-dimensional space, with k basis vectors. Since we need to know where each one of them ends up after applying A, the matrix A must have k columns.

An easy way to remember this is that the adjacent subscripts must be the same: $A_{mk}B_{kn}$

7. Summary

(i) $(AB)C = A(BC)$ (associative law),

(ii) $A(B + C) = AB + AC$ (left distributive law),

(iii) $(B + C)A = BA + CA$ (right distributive law),

(iv) $k(AB) = (kA)B = A(kB)$, where k is a scalar.

A note about the identity matrix and scalar matrices (taken from Schaum):

Identity Matrix, Scalar Matrices

The n-square *identity* or *unit* matrix, denoted by I_n, or simply I, is the n-square matrix with 1's on the diagonal and 0's elsewhere. The identity matrix I is similar to the scalar 1 in that, for any n-square matrix A,

$$AI = IA = A$$

More generally, if B is an $m \times n$ matrix, then $BI_n = I_m B = B$.

For any scalar k, the matrix kI that contains k's on the diagonal and 0's elsewhere is called the *scalar matrix* corresponding to the scalar k. Observe that

$$(kI)A = k(IA) = kA$$

That is, multiplying a matrix A by the scalar matrix kI is equivalent to multiplying A by the scalar k.

EXAMPLE 2.8 The following are the identity matrices of orders 3 and 4 and the corresponding scalar matrices for $k = 5$:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix}, \quad \begin{bmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{bmatrix}, \quad \begin{bmatrix} 5 & & & \\ & 5 & & \\ & & 5 & \\ & & & 5 \end{bmatrix}$$

Remark 1: It is common practice to omit blocks or patterns of 0's when there is no ambiguity, as in the above second and fourth matrices.