Math S-21b – Lecture #3 Notes

Today's lecture features a continuation of geometrically-defined linear transformations – specifically projections and reflections, conditions for invertibility of a matrix and how to find an inverse matrix, and the basic rules of matrix algebra. We use the understanding of a matrix as a linear transformation to define the product of two appropriately-sized matrices as the matrix of the composition of the respective functions. This approach makes a number of the facts about matrix algebra remarkably simple to prove and to understand.

Notes on the dot product and orthogonal projection

An important tool for working with vectors in $\mathbf{R}^n$ (and in abstract vector spaces) is the **dot product** (or, more generally, the inner product). The algebraic definition of the dot product in $\mathbf{R}^n$ is quite simple: Just multiply corresponding components and add.

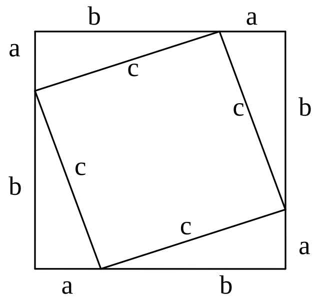
$$\mathbf{u}\cdot\mathbf{v}=\left\langle u_1,u_2,\cdots,u_n\right\rangle\cdot\left\langle v_1,v_2,\cdots,v_n\right\rangle=u_1v_1+u_2v_2+\cdots+u_nv_n$$

However, the true value of the dot product is realized when you relate this to the measurement of angles using trigonometry and the Law of Cosines.

Here are a couple of classic facts:

I. The <u>Pythagorean Theorem</u>: If a right triangle has legs of length a and b and the hypotenuse has length c, then $a^2+b^2=c^2$.

Proof of the Pythagorean Theorem – Perhaps the easiest way to prove this is with areas:

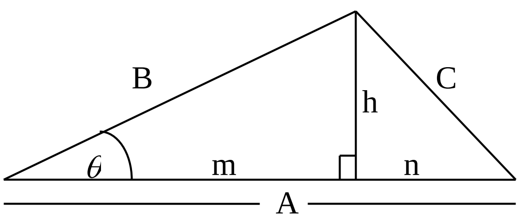


The area of the larger square is the sum of the areas of the smaller square and the four right triangles. This gives us:

$$(a+b)^2 = c^2 + 4\left(\tfrac{1}{2}ab\right)$$
$$a^2 + 2ab + b^2 = c^2 + 2ab$$
$$a^2 + b^2 = c^2$$

II. The <u>Law of Cosines</u>: Given any triangle with sides of length A and B adjacent to an angle θ and with the side opposite this angle of length C, then $\mathrm{C}^2 = \mathrm{A}^2 + \mathrm{B}^2 - 2\mathrm{AB}\cos\theta$.

Proof of the Law of Cosines – Referring to the variables in the diagram, this is a straightforward application of the Pythagorean Theorem and basic trigonometry. The case for an acute angle shown. The proof is similar for an obtuse angle.



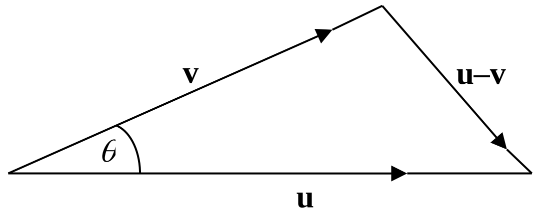
$\mathrm{B}^2 = \mathrm{m}^2 + \mathrm{h}^2$ and $\mathrm{C}^2 = \mathrm{h}^2 + \mathrm{n}^2$

$\mathrm{A} = \mathrm{m} + \mathrm{n}$, so we can write $\mathrm{n} = \mathrm{A} - \mathrm{m}$.

If we substitute this, we get:

$$\mathrm{C}^2 = \mathrm{h}^2 + (\mathrm{A}-\mathrm{m})^2 = \mathrm{h}^2 + \mathrm{A}^2 - 2\mathrm{Am} + \mathrm{m}^2$$
$$= \mathrm{B}^2 + \mathrm{A}^2 - 2\mathrm{A}(\mathrm{B}\cos\theta)$$
$$= \mathrm{A}^2 + \mathrm{B}^2 - 2\mathrm{AB}\cos\theta$$

Measuring angles using the dot product:
Referring to the "vectorized" diagram to the right, we can restate the Law of Cosines in terms of the lengths of the respective vectors as:



$$\|\mathbf{u}-\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\|\mathbf{u}\|\|\mathbf{v}\|\cos\theta .$$

In order to relate this to the dot product, we need to use a few easy-to-show facts about the dot product, namely: Given vectors $\mathbf{u},\mathbf{v},\mathbf{w}\in\mathbf{R}^n$ and any scalar r:

a) $\mathbf{v}\cdot\mathbf{u}=\mathbf{u}\cdot\mathbf{v}$ (commutative law)
b) $(r\mathbf{u})\cdot\mathbf{v}=r(\mathbf{u}\cdot\mathbf{v})=\mathbf{u}\cdot(r\mathbf{v})$
c) $\mathbf{u}\cdot(\mathbf{v}+\mathbf{w})=\mathbf{u}\cdot\mathbf{v}+\mathbf{u}\cdot\mathbf{w}$ and $(\mathbf{u}+\mathbf{v})\cdot\mathbf{w}=\mathbf{u}\cdot\mathbf{w}+\mathbf{v}\cdot\mathbf{w}$ (left and right distributive laws).
d) $\mathbf{u}\cdot\mathbf{u}=\|\mathbf{u}\|^2\ge 0$ for all $\mathbf{u}$, and $\mathbf{u}\cdot\mathbf{u}=\|\mathbf{u}\|^2=0$ only if $\mathbf{u}=\mathbf{0}$ (the zero vector). Here $\|\mathbf{u}\|$ denotes the length of the vector $\mathbf{u}$.

Using these facts, the left-hand side of our vectorized Law of Cosines reads:

$$\|\mathbf{u}-\mathbf{v}\|^2=(\mathbf{u}-\mathbf{v})\cdot(\mathbf{u}-\mathbf{v})=\mathbf{u}\cdot\mathbf{u}-\mathbf{u}\cdot\mathbf{v}-\mathbf{v}\cdot\mathbf{u}+\mathbf{v}\cdot\mathbf{v}=\|\mathbf{u}\|^2+\|\mathbf{v}\|^2-2\mathbf{u}\cdot\mathbf{v} .$$

Comparing this to the original expression, we get the all-important property that
$\boxed{\mathbf{u}\cdot\mathbf{v}=\|\mathbf{u}\|\|\mathbf{v}\|\cos\theta}$ where θ is the angle between the two vectors **u** and **v**.

The significance of this property is that the left-hand side is purely algebraic and the right-hand side is purely geometric. This opens the possibility that we can use basic algebraic operations to calculate geometric quantities like lengths and angles. For example, we can rewrite this result as:

$$\boxed{\cos\theta = \frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|}}.$$

The relation $\mathbf{u}\cdot\mathbf{v} = \|\mathbf{u}\|\|\mathbf{v}\|\cos\theta$ can also be used to provide a simple way of calculating the **scalar projection** of one vector in the direction of another. If we let l denote the orthogonal (perpendicular) projection of **v** in the direction of a another vector **u**, then from the diagram we see that $l = \|\mathbf{v}\|\cos\theta$. We can solve for this in the previous relation to get:

$$l = \|\mathbf{v}\|\cos\theta = \frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|} = \mathbf{v}\cdot\frac{\mathbf{u}}{\|\mathbf{u}\|} = \mathbf{v}\cdot(\text{unit vector in the direction of } \mathbf{u}).$$

In other words, if you want to find out "how much" of a vector **v** is in a given direction, you "dot **v** with a unit vector in that direction".

We can further adapt this to find an expression for the **vector projection** of **v** in the direction of **u**. Simply take a unit vector in the direction of **u** and scale it by the scalar projection of **v** in the **u**-direction to construct this vector projection, a vector in the same direction as **u**, but with length equal to the scalar projection. That is:

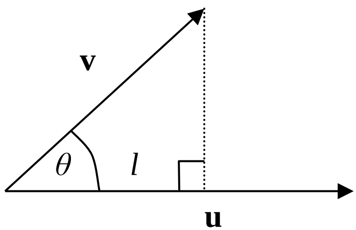
$$\text{Proj}_{\mathbf{u}}\mathbf{v} = \left(\mathbf{v}\cdot\frac{\mathbf{u}}{\|\mathbf{u}\|}\right)\frac{\mathbf{u}}{\|\mathbf{u}\|} = \left(\frac{\mathbf{v}\cdot\mathbf{u}}{\|\mathbf{u}\|^2}\right)\mathbf{u}.$$

In the special case where the vector **u** is a unit vector, i.e. where $\|\mathbf{u}\| = 1$, this simplifies to:

$$\boxed{\text{Proj}_{\mathbf{u}}\mathbf{v} = (\mathbf{v}\cdot\mathbf{u})\mathbf{u}}.$$

The scalar projection of a vector in a given direction is also known as the **component** of the vector in the given direction. It's easy to see that this coincides with the usual *x*, *y*, and *z* components in the case of a vector in $\mathbf{R}^3$. Simply calculate the dot product of the vector $\langle x, y, z \rangle$ with unit vectors in these respective directions. However, with the dot product you can now easily calculate the component of a vector in <u>any</u> direction.

The ability to decompose a vector into its component parts is a fundamental theme in linear algebra. In the case of a more abstract vector space such as a space of functions, this will form the basis of Fourier analysis and other methods for deconstructing functions. These methods play significant roles in fields such as quantum mechanics and digital audio and video recording.



Orthogonal projection onto a line *L* (through the origin) in $\mathbf{R}^n$: We can use the above results to calculate the orthogonal projection of any vector onto a line in $\mathbf{R}^n$. This is, in fact, a linear transformation defined by an $n \times n$ matrix. If the direction of the line L is determined by a unit vector **u**, then for any vector $\mathbf{x} \in \mathbf{R}^n$, we'll have $\text{Proj}_L \mathbf{x} = \text{Proj}_{\mathbf{u}} \mathbf{x} = (\mathbf{x} \cdot \mathbf{u})\mathbf{u}$.

Example: Suppose we want to find the 2×2 matrix for orthogonal projection in $\mathbf{R}^2$ onto the 30° line. Looking at a standard 30°-60°-90° triangle, we see that this line is in the direction of the vector $\mathbf{w} = \langle \sqrt{3}, 1 \rangle = \begin{bmatrix} \sqrt{3} \\ 1 \end{bmatrix}$. We scale to get the unit vector $\mathbf{u} = \frac{1}{2}\begin{bmatrix} \sqrt{3} \\ 1 \end{bmatrix}$. We can proceed in two different ways here.

We can find the columns of the matrix by calculating

$$T(\mathbf{e}_1) = (\mathbf{e}_1 \cdot \mathbf{u})\mathbf{u} = \tfrac{1}{2}(\sqrt{3})\tfrac{1}{2}\begin{bmatrix} \sqrt{3} \\ 1 \end{bmatrix} = \begin{bmatrix} 3/4 \\ \sqrt{3}/4 \end{bmatrix} = \mathbf{v}_1 \text{ and}$$

$$T(\mathbf{e}_2) = (\mathbf{e}_2 \cdot \mathbf{u})\mathbf{u} = \tfrac{1}{2}(1)\tfrac{1}{2}\begin{bmatrix} \sqrt{3} \\ 1 \end{bmatrix} = \begin{bmatrix} \sqrt{3}/4 \\ 1/4 \end{bmatrix} = \mathbf{v}_2 .$$

So the matrix is $\mathbf{A} = \begin{bmatrix} 3/4 & \sqrt{3}/4 \\ \sqrt{3}/4 & 1/4 \end{bmatrix}$.

We can also find the rows of the matrix by calculating $T(\mathbf{x})$ for any $\mathbf{x}$ and then just read off the coefficients to get the entries in each row of the matrix. That is:

$$T(\mathbf{x}) = \mathrm{Proj}_{\mathbf{u}}\mathbf{x} = (\mathbf{x} \cdot \mathbf{u})\mathbf{u} = \left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \cdot \tfrac{1}{2}\begin{bmatrix} \sqrt{3} \\ 1 \end{bmatrix}\right)\tfrac{1}{2}\begin{bmatrix} \sqrt{3} \\ 1 \end{bmatrix} = \tfrac{1}{4}(\sqrt{3}x_1 + x_2)\begin{bmatrix} \sqrt{3} \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{3}{4}x_1 + \frac{\sqrt{3}}{4}x_2 \\ \frac{\sqrt{3}}{4}x_1 + \frac{1}{4}x_2 \end{bmatrix} = \begin{bmatrix} \frac{3}{4} & \frac{\sqrt{3}}{4} \\ \frac{\sqrt{3}}{4} & \frac{1}{4} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{A}\mathbf{x}.$$

So we again see that $\mathbf{A} = \begin{bmatrix} 3/4 & \sqrt{3}/4 \\ \sqrt{3}/4 & 1/4 \end{bmatrix}$.

Reflection through a line (through the origin) in $\mathbf{R}^n$: This builds on orthogonal projection. It's easy to see (if you draw a picture) that if we denote by $\mathbf{p} = \mathrm{Proj}_L\mathbf{x} = (\mathbf{x} \cdot \mathbf{u})\mathbf{u}$ for the orthogonal projection of a vector $\mathbf{x}$ onto the line L (in the direction of the unit vector $\mathbf{u}$), then the reflection of $\mathbf{x}$ through this line will be:

$$\mathrm{Ref}(\mathbf{x}) = \mathbf{x} + 2(\mathbf{p} - \mathbf{x}) = 2\mathbf{p} - \mathbf{x} = 2\mathrm{Proj}_{\mathbf{u}}\mathbf{x} - \mathbf{x} = 2\mathbf{A}\mathbf{x} - \mathbf{x} = 2\mathbf{A}\mathbf{x} - \mathbf{I}\mathbf{x} = (2\mathbf{A} - \mathbf{I})\mathbf{x}$$

The last part of this calculation uses a little matrix algebra that we'll discuss momentarily, but the point is that if you can find the matrix **A** representing orthogonal projection onto a line L, you can then calculate the matrix for reflection (which is also a linear transformation) as $2\mathbf{A}-\mathbf{I}$. [This same fact will hold true when we talk about subspaces and orthogonal projection onto subspaces.]

In our 30° line example above, we found the matrix $\mathbf{A}=\begin{bmatrix} 3/4 & \sqrt{3}/4 \\ \sqrt{3}/4 & 1/4 \end{bmatrix}$ for orthogonal projection onto the line, so the matrix for reflection across this same line will be:

$$2\mathbf{A}-\mathbf{I}=2\begin{bmatrix} 3/4 & \sqrt{3}/4 \\ \sqrt{3}/4 & 1/4 \end{bmatrix}-\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}=\begin{bmatrix} 1/2 & \sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{bmatrix}.$$

These calculations for matrices of rotations, projections, and reflections are relatively easy in $\mathbf{R}^2$. They're not so much more difficult in $\mathbf{R}^3$, especially if you think geometrically. Several homework problems will require you to do this.

Another important geometric linear transformation is called a **shear**. You should <u>read about this in the text</u>.

Inverse of a linear transformation

<u>Definition</u>: We call a linear transformation $T:\mathbf{R}^n \to \mathbf{R}^n$ ***invertible*** (also called nonsingular) if it is both <u>one-to-one</u> (if $T(\mathbf{x})=T(\mathbf{y})$ then necessarily $\mathbf{x}=\mathbf{y}$) and <u>onto</u> the codomain (for every vector $\mathbf{z}\in\mathbf{R}^n$ there is a (unique) $\mathbf{x}\in\mathbf{R}^n$ such that $T(\mathbf{x})=\mathbf{z}$).

It's relatively easy to see why invertibility will only make sense for linear transformations $T:\mathbf{R}^n \to \mathbf{R}^n$ given by (square) $n \times n$ matrices, $T(\mathbf{x}) = \mathbf{Ax}$; and certainly not all such transformations will have inverses.

This is the same notion of invertibility we have for functions elsewhere. However, in the context of linear transformations given by $T(\mathbf{x}) = \mathbf{Ax}$ we have a simple algorithmic way of not only determining if this linear transformation is invertible, but also for determining the matrix of this inverse function (referred to as its **inverse matrix** $\mathbf{A}^{-1}$) if it exists. It all comes down to an enhanced view of row reduction and what invertibility means in terms of rank and the reduced row-echelon form of an associated $n \times 2n$ matrix.

> Some of you may already know about inverse matrices and may be tempted to use them to solve arbitrary systems of linear equations. **This is a very bad idea!** Linear systems can be inconsistent, and they can also have infinitely many solutions. If you restrict yourself to using inverse matrices for solving all linear systems, you will very soon come to regret this. Row reduction is universally valid.

Consider a simple example like $T(\mathbf{x}) = \mathbf{Ax}$ where $\mathbf{A} = \begin{bmatrix} 3 & 1 \\ 2 & -1 \end{bmatrix}$. Given any input vector

$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, this transformation will give the output vector $T(\mathbf{x}) = \mathbf{Ax} = \begin{bmatrix} 3 & 1 \\ 2 & -1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \mathbf{y}$.

To be invertible, given any vector $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$, we would have to be able to solve uniquely for

$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ in terms of the components of $\mathbf{y}$.

What does this mean in terms of algebra?

$$T(\mathbf{x}) = \mathbf{A}\mathbf{x} = \begin{bmatrix} 3 & 1 \\ 2 & -1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3x_1 + x_2 \\ 2x_1 - x_2 \end{bmatrix} = \begin{bmatrix} y_1 & \\ & y_2 \end{bmatrix}.$$

I've staggered the right-hand-sides a bit to suggest the approach. All we have to do is augment the matrix a little more and represent these two equations by entering the coefficients on both the left-hand-side <u>and</u> the right-hand side. This gives $\left[\begin{array}{cc|cc} 3 & 1 & 1 & 0 \\ 2 & -1 & 0 & 1 \end{array}\right] = [\mathbf{A} \mid \mathbf{I}]$ where **I** is the appropriate Identity matrix. If it's possible to solve uniquely for $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, we'll discover this by row reduction. That is:

$$\left[\begin{array}{cc|cc} 3 & 1 & 1 & 0 \\ 2 & -1 & 0 & 1 \end{array}\right] \to \left[\begin{array}{cc|cc} 1 & 2 & 1 & -1 \\ 2 & -1 & 0 & 1 \end{array}\right] \to \left[\begin{array}{cc|cc} 1 & 2 & 1 & -1 \\ 0 & -5 & -2 & 3 \end{array}\right]$$

$$\to \left[\begin{array}{cc|cc} 1 & 2 & 1 & -1 \\ 0 & 1 & 2/5 & -3/5 \end{array}\right] \to \left[\begin{array}{cc|cc} 1 & 0 & 1/5 & 1/5 \\ 0 & 1 & 2/5 & -3/5 \end{array}\right] = [\mathbf{I} \mid \mathbf{A}^{-1}].$$

This last array reads: $\begin{Bmatrix} x_1 = \frac{1}{5}y_1 + \frac{1}{5}y_2 \\ x_2 = \frac{2}{5}y_1 - \frac{3}{5}y_2 \end{Bmatrix}$

We discover two things from this example:

(1) If the matrix **A** has full rank, then we will be able to solve uniquely for $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

(2) If the matrix **A** has full rank, the matrix of its inverse will appear in the right half of rref $[\mathbf{A} \mid \mathbf{I}_n] = [\mathbf{I}_n \mid \mathbf{A}^{-1}]$.

The situation in general is no different. If $T:\mathbf{R}^n \to \mathbf{R}^n$ is given by a (square) $n \times n$ matrix, $T(\mathbf{x}) = \mathbf{Ax}$, and if we write $T(\mathbf{x}) = \mathbf{Ax} = \mathbf{y}$, then we write the $n \times 2n$ matrix $\left[\mathbf{A} \mid \mathbf{I}_n\right]$ and carry out row reduction to determine whether this has full rank *n*. If it doesn't have full rank, then we can't solve uniquely for $\mathbf{x}$, and the transformation (and its matrix) is not invertible. However,

(1) If the matrix **A** has full rank *n*, then we will be able to solve uniquely for

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \text{ in terms of } \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}.$$

(2) If the matrix **A** has full rank *n*, the matrix of its inverse will appear in the right half of $\operatorname{rref}\left[\mathbf{A} \mid \mathbf{I}_n\right] = \left[\mathbf{I}_n \mid \mathbf{A}^{-1}\right]$.

Note: This is generally the simplest way to find the inverse of a matrix by hand. There is a formulaic way of doing this using determinants (based on Cramer's Rule), but it's impractical for matrices larger than 3×3.

There is an ever-so-simple way to find the inverse of a 2×2 matrix $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. First, calculate its determinant $\det(\mathbf{A}) = ad - bc$. You can easily show using our row reduction method that if $\det(\mathbf{A}) = ad - bc = 0$, then the matrix **A** will not have full rank and will not be invertible. If $\det(\mathbf{A}) = ad - bc \neq 0$, then **A** will have full rank and will be invertible, and

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

For example, if $\mathbf{A} = \begin{bmatrix} 3 & 1 \\ 2 & -1 \end{bmatrix}$, then $\det(\mathbf{A}) = (3)(-1) - (1)(2) = -5 \neq 0$ and

$$\mathbf{A} = -\tfrac{1}{5}\begin{bmatrix} -1 & -1 \\ -2 & 3 \end{bmatrix} = \begin{bmatrix} 1/5 & 1/5 \\ 2/5 & -3/5 \end{bmatrix}.$$

The corresponding method for 3×3 matrices has similar elements to this, but involves far more calculation.

Matrix algebra

Definition: Given any scalar $k \in \mathbf{R}$ and an $m \times n$ matrix $\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \cdots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$, we define the scalar multiple of the matrix as $k\mathbf{A} = \begin{bmatrix} ka_{11} & \cdots & ka_{1n} \\ \vdots & \cdots & \vdots \\ ka_{m1} & \cdots & ka_{mn} \end{bmatrix}$. In the case of an $m \times 1$ matrix (a column vector) or a $1 \times n$ matrix (a row vector), this is the same as the ordinary scaling of a vector.

Example: $3\begin{bmatrix} 2 & 1 \\ -1 & 5 \end{bmatrix} = \begin{bmatrix} 6 & 3 \\ -3 & 15 \end{bmatrix}$

Definition: Given two $m \times n$ matrices $\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \cdots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \cdots & \vdots \\ b_{m1} & \cdots & b_{mn} \end{bmatrix}$, we define the sum of these two like matrices by adding their respective entries. That is:

$$\mathbf{A}+\mathbf{B}=\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \cdots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}+\begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \cdots & \vdots \\ b_{m1} & \cdots & b_{mn} \end{bmatrix}=\begin{bmatrix} (a_{11}+b_{11}) & \cdots & (a_{1n}+b_{1n}) \\ \vdots & \cdots & \vdots \\ (a_{m1}+b_{m1}) & \cdots & (a_{mn}+b_{mn}) \end{bmatrix}.$$

Example: $3\begin{bmatrix} 2 & 1 \\ -1 & 5 \end{bmatrix}+5\begin{bmatrix} 1 & 0 \\ 3 & -1 \end{bmatrix}=\begin{bmatrix} 6 & 3 \\ -3 & 15 \end{bmatrix}+\begin{bmatrix} 5 & 0 \\ 15 & -5 \end{bmatrix}=\begin{bmatrix} 11 & 3 \\ 12 & 10 \end{bmatrix}$

Proposition: For any $m \times n$ matrix $\mathbf{A}$, any scalar k, and any $1 \times n$ column vector **x**, $\boxed{(k\mathbf{A})\mathbf{x} = k(\mathbf{A}\mathbf{x})}$.

Proof: This is a straightforward calculation. Writing $\mathbf{A}=\begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ \downarrow & & \downarrow \end{bmatrix}$ in terms of its column vectors, we have

$$(k\mathbf{A})\mathbf{x}=\left(k\begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ \downarrow & & \downarrow \end{bmatrix}\right)\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}=\begin{bmatrix} \uparrow & & \uparrow \\ k\mathbf{v}_1 & \cdots & k\mathbf{v}_n \\ \downarrow & & \downarrow \end{bmatrix}\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}=x_1(k\mathbf{v}_1)+\cdots+x_n(k\mathbf{v}_n)$$

$$=k(x_1\mathbf{v}_1+\cdots+x_n\mathbf{v}_n)=k\left(\begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ \downarrow & & \downarrow \end{bmatrix}\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}\right)=k(\mathbf{A}\mathbf{x}).$$

Proposition: For any $m \times n$ matrices **A** and **B** and any $1 \times n$ column vector **x,** $\boxed{(\mathbf{A}+\mathbf{B})\mathbf{x} = \mathbf{A}\mathbf{x}+\mathbf{B}\mathbf{x}}$.

Proof: If we write $\mathbf{A} = \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ \downarrow & & \downarrow \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{w}_1 & \cdots & \mathbf{w}_n \\ \downarrow & & \downarrow \end{bmatrix}$, and $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$, then

$$(\mathbf{A}+\mathbf{B})\mathbf{x} = \left(\begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ \downarrow & & \downarrow \end{bmatrix} + \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{w}_1 & \cdots & \mathbf{w}_n \\ \downarrow & & \downarrow \end{bmatrix} \right) \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \uparrow & & \uparrow \\ (\mathbf{v}_1+\mathbf{w}_1) & \cdots & (\mathbf{v}_n+\mathbf{w}_n) \\ \downarrow & & \downarrow \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$= x_1(\mathbf{v}_1+\mathbf{w}_1)+\cdots+x_n(\mathbf{v}_n+\mathbf{w}_n) = x_1\mathbf{v}_1 + x_1\mathbf{w}_1 + \cdots + x_n\mathbf{v}_n + x_n\mathbf{w}_n$$

$$= x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n + x_1\mathbf{w}_1 + \cdots + x_n\mathbf{w}_n = \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ \downarrow & & \downarrow \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{w}_1 & \cdots & \mathbf{w}_n \\ \downarrow & & \downarrow \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \mathbf{Ax}+\mathbf{Bx}.$$

Matrix products

Though it's possible to take a formulaic approach to the multiplication of matrices, it's much better to think of each matrix as representing a linear transformation and to define matrix product by considering the composition of these linear transformations.

Proposition: Where defined, the composition of linear transformations is a linear transformation.

Proof: Suppose **A** is an $m \times n$ matrix that corresponds to a linear transformation $T_{\mathbf{A}} : \mathbf{R}^n \to \mathbf{R}^m$, i.e. $T_{\mathbf{A}}(\mathbf{y}) = \mathbf{Ay}$. Also, suppose **B** is an $n \times p$ matrix that corresponds to a linear transformation $T_{\mathbf{B}} : \mathbf{R}^p \to \mathbf{R}^n$, i.e. $T_{\mathbf{B}}(\mathbf{x}) = \mathbf{Bx}$. We can then define the composition $T_{\mathbf{A}} \circ T_{\mathbf{B}} : \mathbf{R}^p \to \mathbf{R}^m$ by

$\left(T_{\mathbf{A}} \circ T_{\mathbf{B}}\right)(\mathbf{x}) = T_{\mathbf{A}}\left(T_{\mathbf{B}}(\mathbf{x})\right)$. Since both of these functions are linear, for any scalars c_1, c_2 and vectors $\mathbf{v}_1, \mathbf{v}_2 \in \mathbf{R}^p$, we have:

$$\left(T_{\mathbf{A}} \circ T_{\mathbf{B}}\right)(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) = T_{\mathbf{A}}\left(T_{\mathbf{B}}(c_1\mathbf{v}_1 + c_2\mathbf{v}_2)\right) = T_{\mathbf{A}}\left(c_1 T_{\mathbf{B}}(\mathbf{v}_1) + c_2 T_{\mathbf{B}}(\mathbf{v}_2)\right)$$
$$= c_1 T_{\mathbf{A}}\left(T_{\mathbf{B}}(\mathbf{v}_1)\right) + c_2 T_{\mathbf{A}}\left(T_{\mathbf{B}}(\mathbf{v}_2)\right) = c_1\left(T_{\mathbf{A}} \circ T_{\mathbf{B}}\right)\mathbf{v}_1 + c_2\left(T_{\mathbf{A}} \circ T_{\mathbf{B}}\right)\mathbf{v}_2$$

So $T_{\mathbf{A}} \circ T_{\mathbf{B}} : \mathbf{R}^p \to \mathbf{R}^m$ is also linear and is represented by an $m \times p$ matrix. Call this matrix **AB**.

Corollary (really a restatement of the definition): For any vector $\mathbf{x} \in \mathbf{R}^p$, $\boxed{(\mathbf{AB})\mathbf{x} = \mathbf{A}(\mathbf{Bx})}$.

This statement look very much like an associative law for multiplication, but it’s really just the statement that **AB** is <u>defined</u> to be the matrix of the composition.

Calculation of the matrix product

How do we actually <u>calculate</u> the matrix product **AB** (where defined)? Perhaps the simplest way to do this is to recall the meaning of the columns of any matrix. The columns tell us where the corresponding linear function takes the elementary vectors $\{\mathbf{e}_1, \cdots, \mathbf{e}_p\}$, so

$$\mathbf{AB} = \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{AB}(\mathbf{e}_1) & \cdots & \mathbf{AB}(\mathbf{e}_p) \\ \downarrow & & \downarrow \end{bmatrix} = \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{A}(\mathbf{Be}_1) & \cdots & \mathbf{A}(\mathbf{Be}_p) \\ \downarrow & & \downarrow \end{bmatrix}.$$

But if we write $\mathbf{B} = \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{B}(\mathbf{e}_1) & \cdots & \mathbf{B}(\mathbf{e}_p) \\ \downarrow & & \downarrow \end{bmatrix} = \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{v}_1 & \cdots & \mathbf{v}_p \\ \downarrow & & \downarrow \end{bmatrix}$, we then see that:

$$\mathbf{AB} = \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{A}(\mathbf{v}_1) & \cdots & \mathbf{A}(\mathbf{v}_p) \\ \downarrow & & \downarrow \end{bmatrix}. \qquad \text{That is, } \mathbf{AB} = \mathbf{A}\begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{v}_1 & \cdots & \mathbf{v}_p \\ \downarrow & & \downarrow \end{bmatrix} = \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{Av}_1 & \cdots & \mathbf{Av}_p \\ \downarrow & & \downarrow \end{bmatrix}.$$

In other words, the matrix **A** simply individually multiplies each of the column vectors of **B**.

Example: If $\mathbf{A} = \begin{bmatrix} 2 & -1 \\ 1 & 5 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 1 & 0 & 4 \\ -1 & 2 & 1 \end{bmatrix}$, then the product **AB** is defined (though **BA** is not), and $\mathbf{AB} = \begin{bmatrix} 2 & -1 \\ 1 & 5 \end{bmatrix}\begin{bmatrix} 1 & 0 & 4 \\ -1 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 3 & -2 & 7 \\ -4 & 10 & 9 \end{bmatrix}$.

It should be relatively easy to see that each entry is calculated as $(\mathbf{AB})_{ij} = (i\text{th row of } \mathbf{A}) \bullet (j\text{th column of } \mathbf{B})$. This dot product is only defined when the number of columns of **A** matches the number of rows of **B**.

Matrix multiplication (where defined) is not commutative: $\boxed{\mathbf{AB} \neq \mathbf{BA}}$

It's easy to understand why matrix multiplication cannot be commutative even in the case where both products are defined. Matrix product is just the composition of functions, and composing functions in reverse order does not generally give the same functions, i.e. $f \circ g \neq g \circ f$. This is most simply understood by thinking about it in less mathematical terms. For example, if you put on your socks and then put on your shoes, this is clearly different than first putting on your shoes and then putting on your socks. Sometimes you can get the same result, just as it is the case that there are some matrices **A** and **B** such that $\mathbf{AB} = \mathbf{BA}$, but this will not generally be the case.

Matrix multiplication (where defined) is associative: $\boxed{(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})}$

This follows from the corresponding fact about composition of functions, namely that $(f \circ g) \circ h = f \circ (g \circ h)$.

The Identity matrix acts as a multiplicative identity: For an $m \times n$ matrix $\mathbf{A}$,

$$\boxed{\mathbf{I}_m \mathbf{A} = \mathbf{A}} \text{ and } \boxed{\mathbf{A}\mathbf{I}_n = \mathbf{A}}.$$

Though this is easy to see by calculation, it follows from the general fact about functions that $Id \circ f = f$ and $f \circ Id = f$, i.e. for any x in the domain of f, we have $(Id \circ f)(x) = Id(f(x)) = f(x)$ and $(f \circ Id)(x) = f(Id(x)) = f(x)$.

Proposition: If $\mathbf{A}$ is an invertible $n \times n$ matrix with inverse matrix $\mathbf{A}^{-1}$, then

$$\boxed{\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}} \text{ and } \boxed{\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}}.$$

Proof: These follow directly from the fact that matrix product represents the composition of linear functions and the fact that a function composed with its inverse (in either order) yields the identity function.

Proposition: If both $\mathbf{A}$ and $\mathbf{B}$ are invertible $n \times n$ matrices, then $\mathbf{AB}$ is also invertible and

$$\boxed{(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}}.$$

Proof: This also follows directly from the general fact about functions, i.e. $(f \circ g)^{-1} = g^{-1} \circ f^{-1}$.

In nonmathematical terms, if you first put on your socks and then put on your shoes, the inverse of this is to first take off your shoes and then take off your socks.

More easy-to-prove matrix algebra facts: For any scalar k and appropriate sized matrices

$\mathbf{A}(\mathbf{C}+\mathbf{D}) = \mathbf{AC}+\mathbf{AD}$ (left-hand distributive law)

$(\mathbf{A}+\mathbf{B})\mathbf{C} = \mathbf{AC}+\mathbf{BC}$ (right-hand distributive law)

$(k\mathbf{A})\mathbf{C} = k(\mathbf{AC})$

These and the facts previously stated are not meant to be exhaustive. Except for the fact that matrix multiplication is not commutative, most of the familiar algebraic rules are also true for matrices.

An application to trigonometry: Sum of angle formulas for sine and cosine

We previously showed that counterclockwise rotation in $\mathbf{R}^2$ through an angle θ is a linear transformation represented by the rotation matrix $\mathbf{R}_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$. It is geometrically clear that the composition of rotation through angle ϕ and rotation through angle θ is just rotation through the angle $(\theta+\phi)$, so $\mathbf{R}_\theta\mathbf{R}_\phi = \mathbf{R}_{\theta+\phi}$. Therefore:

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} = \begin{bmatrix} \cos\theta\cos\phi - \sin\theta\sin\phi & -(\sin\theta\cos\phi + \cos\theta\sin\phi) \\ \sin\theta\cos\phi + \cos\theta\sin\phi & \cos\theta\cos\phi - \sin\theta\sin\phi \end{bmatrix}$$

$$= \begin{bmatrix} \cos(\theta+\phi) & -\sin(\theta+\phi) \\ \sin(\theta+\phi) & \cos(\theta+\phi) \end{bmatrix} \Rightarrow \boxed{\begin{array}{l} \cos(\theta+\phi) = \cos\theta\cos\phi - \sin\theta\sin\phi \\ \sin(\theta+\phi) = \sin\theta\cos\phi + \cos\theta\sin\phi \end{array}}$$

Notes by Robert Winters