Math 108a	Professor: Padraic Bartlett
Lecture 15: Matrices	
Week 7	*UCSB 2013*

Today's talk: matrices!

1 Matrices: Definitions

We formally define a matrix as follows:

Definition. Take a linear map $T : \mathbb{R}^n \to \mathbb{R}^m$. Let the vectors $\vec{e_1}, \ldots \vec{e_n}$ denote the standard basis vectors for $\mathbb{R}^n$: i.e. $\vec{e_1} = (1, 0, \ldots 0), \vec{e_2} = (0, 1, 0 \ldots 0), \ldots \vec{e_n} = (0, 0 \ldots 0, 1)$.

For each of the vectors $T(\vec{e_i})$ in $\mathbb{R}^m$, write

$$T(\vec{e_i}) = (t_{1,i}, t_{2,i}, \ldots, t_{m,i}),$$

where the values $t_{i,j}$ are all real numbers

We can turn T into an $m \times n$ **matrix**, i.e. a $m \times n$ grid of real numbers, as follows:

$$T \longrightarrow T_{\text{matrix}} = \begin{bmatrix} t_{1,1} & t_{1,2} & \ldots & t_{1,n} \\ t_{2,1} & t_{2,2} & \ldots & t_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ t_{m,1} & t_{m,2} & \ldots & t_{m,n} \end{bmatrix}.$$

In other words,

$$T \longrightarrow T_{\text{matrix}} = \begin{bmatrix} \vdots & \vdots & \ldots & \vdots \\ T(\vec{e_1}) & T(\vec{e_2}) & \ldots & T(\vec{e_n}) \\ \vdots & \vdots & \ldots & \vdots \end{bmatrix},$$

Similarly, given some $m \times n$ matrix

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \ldots & a_{1,n} \\ a_{2,1} & a_{2,2} & \ldots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \ldots & a_{m,n} \end{bmatrix},$$

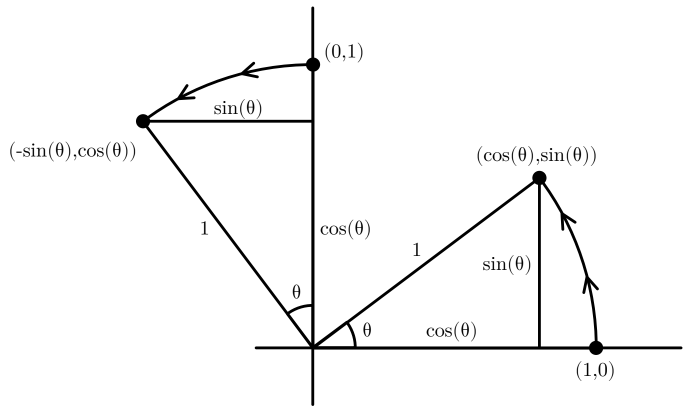
we can interpret A as a **linear map** $A_{\text{map}} : \mathbb{R}^n$ to $\mathbb{R}^m$ as follows:

- For any of the standard basis vectors $\vec{e_i}$, we define $A_{\text{map}}(\vec{e_i})$ to simply be the vector $(a_{1,i}, \ldots a_{m,i})$.
- For any other vector $(x_1, \ldots x_n) \in \mathbb{R}^n$, we define $A_{\text{map}}(x_1, \ldots x_n)$ to simply be the corrresponding linear combination of the $\vec{e_i}$'s: i.e.

$$A_{\text{map}} : (x_1, \ldots x_n) := x_1 \cdot A_{\text{map}}(\vec{e_1}) + \ldots + x_n A_{\text{map}}(\vec{e_n}).$$

In practice, we will usually not bother writing the subscripts "map" and "matrix" on these objects, and think of linear maps from $\mathbb{R}^n$ to $\mathbb{R}^m$ and $m \times n$ matrices as basically the same things.

For example, consider the map $T_\theta : \mathbb{R}^2 \to \mathbb{R}^2$, that we worked with on problem #6 of homework #6.



Because this map sends $(1,0)$ to $(\cos(\theta), \sin(\theta))$, and $(0,1)$ to $(-\sin(\theta), \cos(\theta))$, we would express this map as a matrix as follows:

$$T_\theta = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}.$$

Under this interpretation, we would say that $T_\theta(x,y) = (x\cos(\theta) - y\sin(\theta), x\sin(\theta + y\cos(\theta))$, i.e.

$$T_\theta(x,y) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = (x\cos(\theta) - y\sin(\theta), x\sin(\theta) + y\cos(\theta))$$

2 Matrix Properties

Many of you have used matrices in other classes; in these settings, you've probably applied matrices to vectors, and taken "products" of matrices. However, in these classes, people typically never justify **why** we take products of matrices in the way we do, or why applying a matrix to a vector is calculated in the way you're shown.

That's dumb. We're fixing this here.

Theorem. Take any linear map $A : \mathbb{R}^n \to \mathbb{R}^m$. Let A have the associated matrix

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} \end{bmatrix}.$$

Denote the row vectors $(a_{i,1}, \dots a_{i,n}) \in \mathbb{R}^n$ of our matrix A with the vectors $\vec{a_{r_i}}$, for shorthand.

We claim that for any $\vec{x} \in \mathbb{R}^n$, we have

$$A(\vec{x}) = (\vec{x} \cdot \vec{a_{r_1}}, \vec{x} \cdot \vec{a_{r_2}}, \ldots \vec{x} \cdot \vec{a_{r_m}}) .$$

In other words, to calculate what a matrix A does to a vector $\vec{x}$, we simply construct the vector given by dotting each of the **rows** of A with the vector $\vec{x}$.

Proof. We prove this fact by using the definition of A, along with linearity.

By definition, we know that A is the linear map $\mathbb{R}^n \to \mathbb{R}^m$, that sends the standard basis vectors to the columns of A: i.e.

$$\begin{aligned}
A(\vec{e_1}) &= (a_{1,1}, a_{2,1}, \ldots, a_{m,1}) \\
A(\vec{e_2}) &= (a_{1,2}, a_{2,2}, \ldots, a_{m,2}) \\
&\vdots \\
A(\vec{e_n}) &= (a_{1,n}, a_{2,n}, \ldots, a_{m,n})
\end{aligned}$$

Therefore, by linearity, we can see that

$$\begin{aligned}
A(x_1, \ldots x_n) &= A(x_1\vec{e_1} + x_2\vec{e_2} + \ldots + x_n\vec{e_n}) \\
&= x_1 A(\vec{e_1}) + x_2 A(\vec{e_2}) + \ldots + x_n A(\vec{e_n}) \\
&= x_1(a_{1,1}, a_{2,1}, \ldots, a_{m,1}) + x_2(a_{1,2}, a_{2,2}, \ldots, a_{m,2}) + \ldots + x_n(a_{1,n}, a_{2,n}, \ldots, a_{m,n}).
\end{aligned}$$

If we add all of these vectors together, we get that the first coördinate of the result $A(x_1, \ldots x_n)$ is just

$$x_1 a_{1,1} + x_2 a_{1,2} + x_3 a_{1,3} + \ldots + x_n a_{1,n},$$

the second coördinate is

$$x_1 a_{2,1} + x_2 a_{2,2} + x_3 a_{2,3} + \ldots + x_n a_{2,n},$$

and in general the k-th coördinate is

$$x_1 a_{k,1} + x_2 a_{k,2} + x_3 a_{k,3} + \ldots + x_n a_{k,n}.$$

Notice that this expression is simply the dot product of $\vec{x}$ with the k-th row of A! If we plug this observation in for every coördinate of $A(\vec{x})$, we get

$$A(\vec{x}) = (\vec{x} \cdot \vec{a_{r_1}}, \vec{x} \cdot \vec{a_{r_2}}, \ldots \vec{x} \cdot \vec{a_{r_m}}) ,$$

which is what we claimed was true. $\square$

This proof above shows that the process that most people learn when they first see matrices — take the rows of the matrix and dot them with the vector you're applying the matrix to — is in fact the only thing that applying a matrix to a vector could sensibly be! This is reassuring: it's nice to see that the definitions we've seen before in other classes weren't just made up, but were actually chosen because they're the only possible things that can be true. (This is kind of like the feeling of proving $1 + 1 = 2$; on one hand you already know that it's true, but it's cool seeing **why** it must be true!)

Example. Again, return to the rotation map T_θ. We showed above that it has matrix

$$T_\theta = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}.$$

If we just use the result above, we have that

$$\begin{aligned} T_\theta(x,y) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} &= ((x,y)\cdot(\cos(\theta), -\sin(\theta)), (x,y)\cdot(\sin(\theta)+\cos(\theta))) \\ &= (x\cos(\theta) - y\sin(\theta), x\sin(\theta) + y\cos(\theta)). \end{aligned}$$

Reassuringly, this is the same thing we got in our direct calculation for where this matrix sends (x,y)!

Another property people often know from earlier classes is how to "multiply" matrices. We interpret this in the sense of **composing linear maps corresponding to matrices** in the following theorem:

Theorem. Take any pair of linear maps $A : \mathbb{R}^n \to \mathbb{R}^m, B : \mathbb{R}^m \to \mathbb{R}^k$ with associated matrices

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} \end{bmatrix}, B = \begin{bmatrix} b_{1,1} & b_{1,2} & \dots & b_{1,m} \\ b_{2,1} & b_{2,2} & \dots & b_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{k,1} & b_{k,2} & \dots & b_{k,m} \end{bmatrix}.$$

Look at the linear map given by the composition of these two maps: i.e. consider the linear map $B \circ A : \mathbb{R}^n \to \mathbb{R}^k$. Denote the row vectors of B as $\vec{b_{r_i}}$'s and the column vectors of A as $\vec{a_{c_j}}$'s. We claim that this linear map corresponds to the $k \times n$ matrix

$$\begin{bmatrix} \vec{b_{r_1}} \cdot \vec{a_{c_1}} & \vec{b_{r_1}} \cdot \vec{a_{c_2}} & \dots & \vec{b_{r_1}} \cdot \vec{a_{c_n}} \\ \vec{b_{r_2}} \cdot \vec{a_{c_1}} & \vec{b_{r_2}} \cdot \vec{a_{c_2}} & \dots & \vec{b_{r_2}} \cdot \vec{a_{c_n}} \\ \dots & \dots & \ddots & \dots \\ \vec{b_{r_k}} \cdot \vec{a_{c_1}} & \vec{b_{r_k}} \cdot \vec{a_{c_2}} & \dots & \vec{b_{r_k}} \cdot \vec{a_{c_n}} \end{bmatrix}.$$

In other words, to get the matrix given by composing two matrices, we simply dot the rows of the first matrix with the columns of the second matrix in the manner described above. In other words, we do matrix multiplication.

Proof. So: we want to find the matrix corresponding to $B \circ A$. To do this, according to the definitions, we just need to find where $B \circ A$ sends the basis vectors $\vec{e_1}, \dots \vec{e_n}$ of $\mathbb{R}^n$!

This is not too hard. Take any basis vector $\vec{e_i}$. Apply A to this basis vector. By definition, we know that this yields the i-th column of A: i.e. $A(\vec{e_i}) = \vec{a_{c_i}}$.

To find $B \circ A$ as applied to $\vec{e_i}$, then, we can just calculate B of $A(\vec{e_i}) = \vec{a_{c_i}}$. In particular, we can calculate $B(\vec{a_{c_i}})$ by using the theorem we just proved earlier for how matrices apply to vectors: this tells us that

$$B(A(\vec{e_i})) = \left(\vec{b_{r_1}} \cdot \vec{a_{c_i}}, \vec{b_{r_2}} \cdot \vec{a_{c_i}}, \dots \vec{b_{r_k}} \cdot \vec{a_{c_i}}\right)$$

Then, by definition, the matrix corresponding to $B \circ A$ is the matrix that has the vectors $\left(\vec{b_{r_1}} \cdot \vec{a_{c_i}}, \vec{b_{r_2}} \cdot \vec{a_{c_i}}, \ldots \vec{b_{r_k}} \cdot \vec{a_{c_i}}\right)$ as its columns: i.e.

$$\begin{bmatrix} \vec{b_{r_1}} \cdot \vec{a_{c_1}} & \vec{b_{r_1}} \cdot \vec{a_{c_2}} & \ldots & \vec{b_{r_1}} \cdot \vec{a_{c_n}} \\ \vec{b_{r_2}} \cdot \vec{a_{c_1}} & \vec{b_{r_2}} \cdot \vec{a_{c_2}} & \ldots & \vec{b_{r_2}} \cdot \vec{a_{c_n}} \\ \ldots & \ldots & \ddots & \ldots \\ \vec{b_{r_k}} \cdot \vec{a_{c_1}} & \vec{b_{r_k}} \cdot \vec{a_{c_2}} & \ldots & \vec{b_{r_k}} \cdot \vec{a_{c_n}} \end{bmatrix}.$$

□

Example. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear map

$$T = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix},$$

and $S : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear map

$$S = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

What is the matrix given by the composition $S \circ (T \circ S)$?

Answer. If we just apply problem 2 from the above section, we have

$$\begin{aligned}(T \circ S) = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} \cdot \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} &= \begin{bmatrix} \left(\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right) \cdot (0,1) & \left(\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right) \cdot (1,0) \\ \left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right) \cdot (0,1) & \left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right) \cdot (1,0) \end{bmatrix} \\ &= \begin{bmatrix} \frac{-\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}.\end{aligned}$$

Therefore, we have

$$\begin{aligned}S \circ (T \circ S) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} \frac{-\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} &= \begin{bmatrix} (0,1) \cdot \left(-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right) & (0,1) \cdot \left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right) \\ (1,0) \cdot \left(-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right) & (1,0) \cdot \left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right) \end{bmatrix} \\ &= \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}.\end{aligned}$$

In other words, this is the map that sends $(1,0)$ to $\left(\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right)$ and $(0,1)$ to $\left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)$. If you recall our discussion from the past homework set, this is in fact the rotation matrix that rotates $\mathbb{R}^2$ by $-\pi/4$!

We can double-check this answer by thinking geometrically: the map T is just the matrix given by the linear map $T_{\pi/4}$ that rotates space by $\pi/4$ radians, while the map S is the matrix that flips the x and y-coördinates. Composing these maps as $S \circ T \circ S$, geometrically speaking, should give you a map that first switches the x and y coordinates, then rotates by $\pi/4$ in the "switched" space, then flips back — which is just rotation by $-\pi/4$!