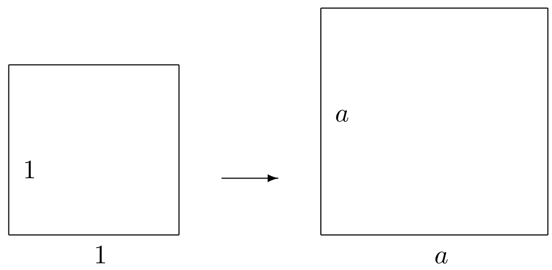
Matrices as Mappings

Scaling: The simplest sort of linear transformation of the plane we can write down is a rescaling (which is also called a dilation). If a is a positive number, we can send (x, y) to $a(x, y) = (ax, ay)$. Geometrically, we can imagine this transformation as taking the unit square $\{0 \leq x \leq 1, 0 \leq y \leq 1\}$ to a simliar square $\{0 \leq x \leq a, 0 \leq y \leq a\}$, which we represent in the following picture.



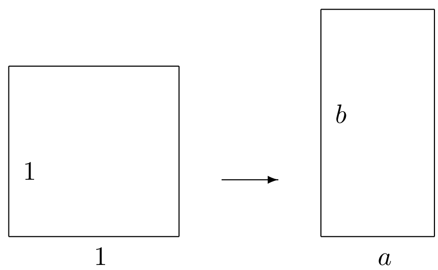
Actually, if we adopt the convention that we always start with the unit square $\{0 \leq x \leq 1, 0 \leq y \leq 1\}$, we really only need to draw the square on the right to have a geometric picture of or transformation. If we want to write this transformation in terms of matrices, we can write

$$\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ax \\ ay \end{bmatrix}.$$

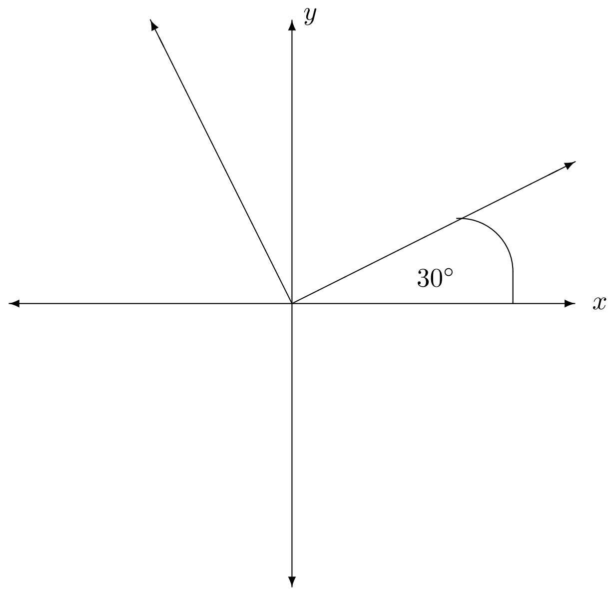
In the previous example, we scaled the horizontal and vertical axes by the same factor, but there's no reason we have to do this. More generally, we might scale the horizontal axis by $a > 0$ and the vertical axis by $b > 0$. This time, we can write the transformation as

$$\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ax \\ by \end{bmatrix}.$$

As before, we can represent this transformation with a picture.



Rotations and reflections: Let's suppose we want to write down a formula for a $30°$ counterclockwise rotation in the plane; call this rotation R_{30}. For instance, we may have a collection of data points we'd like to put into a database, but the coordinates of these datapoints are all rotated by $30°$ in the clockwise direction, and so we want to undo this rotation, by rotating through the same angle in the opposite direction. So let's find out how to write down a formula for the rotation. We start with a picture of the vectors $(1, 0)$ and $(0, 1)$ rotated by $30°$.



We see that $(1,0)$ gets mapped to the point

$$R_{30}\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} \cos(30°) \\ \sin(30°) \end{bmatrix} = \begin{bmatrix} \sqrt{3}/2 \\ 1/2 \end{bmatrix},$$

and that $(0,1)$ gets mapped to the point

$$R_{30}\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} -\sin(30°) \\ \cos(30°) \end{bmatrix} = \begin{bmatrix} -1/2 \\ \sqrt{3}/2 \end{bmatrix}.$$

Next observe that we can rescale the vectors $(1,0)$ and $(0,1)$, and, because rotations don't change lengths, the rotation will carry these scalings along:

$$R_{30}\left(\begin{bmatrix} x \\ 0 \end{bmatrix}\right) = \begin{bmatrix} \sqrt{3}x/2 \\ x/2 \end{bmatrix}, \qquad R_{30}\left(\begin{bmatrix} 0 \\ y \end{bmatrix}\right) = \begin{bmatrix} -y/2 \\ \sqrt{3}y/2 \end{bmatrix}.$$

Finally, we can put this all together, because acts independently on the vectors $(1,0)$ and $(0,1)$. This means R_{30} rotatates $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ independently, which we can write as

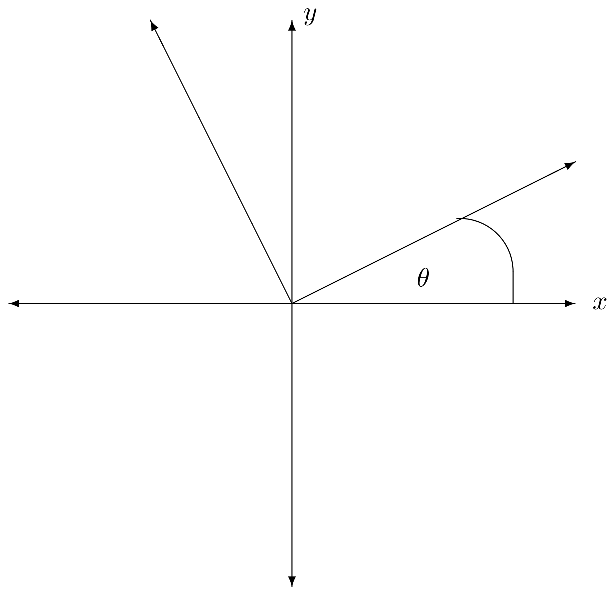
$$R_\theta\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = R_\theta\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) + R_\theta\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right).$$

(You can verify this formula geometrically, by seeing where the rotation carries the top right corner of the unit square.) Adding these two vectors together, we have

$$R_{30}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} \sqrt{3}x/2 - y/2 \\ x/2 + \sqrt{3}y/2 \end{bmatrix} = \begin{bmatrix} \sqrt{3}/2 & -1/2 \\ 1/2 & \sqrt{3}/2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

In this last step we used the rule for multiplying matrices we stated previously in the notes, which starts to explain why we defined matrix multiplication the way we did.

We can redo this whole discussion with a rotation through any angle. Let R_θ be the rotation through angle θ in the counterclockwise direction, whose action on the vectors $(1,0)$ and $(0,1)$ is drawn below.



Then, just as before, we have

$$R_\theta\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}, \qquad R_\theta\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix},$$

and, by the same argument we have above,

$$R_\theta\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x\cos\theta - y\sin\theta \\ x\sin\theta + y\cos\theta \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}.$$

In this way, we can say that the rotation R_θ is given by multiplication (on the left) by the matrix

$$[R_\theta] = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

Exercise: Verify that

$$[R_{-\theta}] = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}.$$

Exercise: Verify that $[R_\theta]^2 = [R_{2\theta}]$. (You'll need to remember double angle formulas from trigonometry.)

Exercise: Verify that $[R_\theta][R_\phi] = [R_{\theta+\phi}]$. (You'll need to remember the angle addition formulas from trigonometry.) Notice that rotation matrices commute! That is, $[R_\theta][R_\phi] = [R_\phi][R_\theta]$.

Now that we've figured out how to write any rotation as multiplication by a matrix, let's be a little ambitious and see wht else we can write. The next natural thing to consider is a reflection. The reflection through the x axis sends (x, y) to $(x, -y)$. We can write this as a matrix product by

$$\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} x \\ -y \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}.$$

Now that we know how to write reflection through the y axis and any rotation, we can write down any reflection through a line that intersections the origin $(0, 0)$. Indeed, let l be a line passing through the origin making an angle θ with the positive x axis, and let r_l be reflection through the line l. We build the matrix for r_l by performing three transformations in succession. We first rotate our coordinates by $-\theta$, then reflect through the x axis, and then rotate back by the angle θ. The result is a reflection fixing the line l, so it must be r_l, and it has the matrix representation

$$[r_l] = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} \cos^2\theta - \sin^2\theta & 2\cos\theta\sin\theta \\ 2\cos\theta\sin\theta & \sin^2\theta - \cos^2\theta \end{bmatrix}.$$

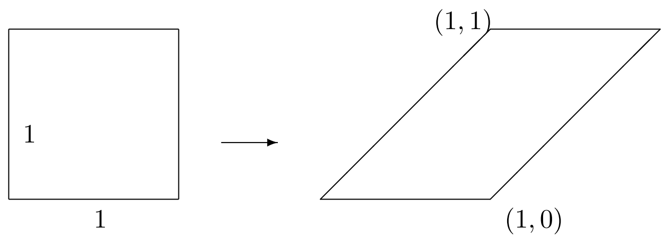
Let's check quickly that we have the right matrix for the reflection through the line l. This line l is uniquely determined by the two points $(0,0)$ and $(\cos\theta, \sin\theta)$, so we only need to check that $[r_l]$ fixes these two vectors. You can check for yourself that

$$[r_l]\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \cos^2\theta - \sin^2\theta & 2\cos\theta\sin\theta \\ 2\cos\theta\sin\theta & \sin^2\theta - \cos^2\theta \end{bmatrix}\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Now we check that $[r_l]$ fixes $(\cos\theta, sin\theta)$:

$$\begin{aligned} [r_l]\begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} &= \begin{bmatrix} \cos^2\theta - \sin^2\theta & 2\cos\theta\sin\theta \\ 2\cos\theta\sin\theta & \sin^2\theta - \cos^2\theta \end{bmatrix}\begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \\ &= \begin{bmatrix} \cos^3\theta - \sin^2\theta\cos\theta + 2\sin^2\theta\cos\theta \\ 2\cos^2\theta\sin\theta - \cos^2\theta\sin\theta + \sin^3\theta \end{bmatrix} \\ &= \begin{bmatrix} \cos\theta(\cos^2\theta + \sin^2\theta) \\ \sin\theta(\cos^2\theta + \sin^2\theta) \end{bmatrix} = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}. \end{aligned}$$

Sheers: The next sort of transformation we'll talk about is a *sheer*, which you can imagine as what happens to a deck of cards (as viewed from the side) when you push the top card to the side and hold the bottom card still. This means a sheer will fix one direction, say the direction of $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$, but it will move the other directions. We draw a picture of what this transformation does the unit square $\{0 \le x \le 1, 0 \le y \le 1\}$ below.



We'll call this sheer S.

We'll construct the matrix for this sheer mapping S by seeing what it does to the two coordinate vectors $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$, which, in a way, is how we constructed the rotation matrix. We can see from the picture that $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ keeps the same direction, so we can rescale in the horizontal direction to make

$$S\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

which tells us

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

Here the $*$'s can stand for any number, because we haven't figured out yet what these parts of $[S]$ are.

On the other hand, the vector $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ gets tilted in the clockwise direction, and it looks like it gets stretched as well. If we look a little more closely, we see

$$S\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

which tells us

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

We can check this last formula by separating out the components:

$$S\left(\begin{bmatrix} x \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ y \end{bmatrix}\right) = S\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}.$$

We've just constructed the matrix of a particular sheer which fixes the horizontal direction. In general, the vertical direction will go to some other direction, so that

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} \mapsto \begin{bmatrix} a \\ 1 \end{bmatrix},$$

where $a \neq 0$ is a number. Notice that we have the second component equal to 1, which we can arrange by rescaling if necessary. We always have the second component nonzero, because otherwise the sheer would collaps the unit square down to a (horizontal) line segment. Now, following the same reasoning as we did above, we find the matrix of this sheer is

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

Exercise: Notice that a can be negative in the formula just above. What does the paralelogram which is the image of the unit square $\{0 \leq x \leq 1, 0 \leq y \leq 1\}$ look like in this case? In particular, what can you say about the angle at the origin $(0,0)$?

Exercise: Show that the general sheer which fixes the y axis is given by a matrix of the form

$$[S] = \begin{bmatrix} 1 & 0 \\ a & 1 \end{bmatrix},$$

where $a \neq 0$ is a number.

Exercise: Construct the general sheer which fixes the $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ direction. Hint: you might want to apply a rotation.

General matrices as mappings: We just saw how to construct the matrix associated to a sheer by tracking where the sheer transformation sends the basis vectors $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$. In fact, this technique is exactly how we can produce the matrix associated to any linear transformation. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be any linear transformation, which means $T(v + w) = T(v) + T(w)$ for all vectors $v, w \in \mathbb{R}^2$ and $T(av) = aT(v)$ for all scalars a.

Exercise: Prove that $T\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ for any linear mapping. Hint: suppose otherwise; then what is $T\left(2\begin{bmatrix} 0 \\ 0 \end{bmatrix}\right)$?

We can construct a matrix associated to T, which we call $[T]$, as follows. The first column of $[T]$ is $T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right)$, and the second column of $[T]$ is $T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right)$.

Let's check this is actually the right matrix. Suppose we have a linear mapping $T : \mathbb{R}^2 \to \mathbb{R}^2$ with

$$T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} a \\ c \end{bmatrix}, \qquad T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} b \\ d \end{bmatrix}.$$

In this case we'd like to check that the matrix associated with T is

$$[T] = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Indeed,

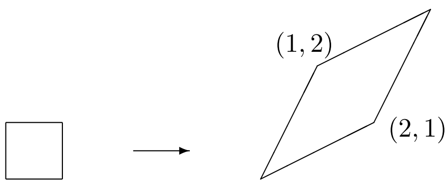
$$\begin{aligned}
T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) &= T\left(\begin{bmatrix} x \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) \\
&= xT\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) + yT\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) \\
&= x\begin{bmatrix} a \\ c \end{bmatrix} + y\begin{bmatrix} b \\ d \end{bmatrix} \\
&= \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix} \\
[T]\begin{bmatrix} x \\ y \end{bmatrix} &= \begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} \\
&= \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix}.
\end{aligned}$$

In both computations we end up with the same answer, regardless of which x and y we choose, so this matrix must be the correct choice.

Let's look at an example. Suppose we want to find the linear map which takes the unit square $\{0 \leq x \leq 1, 0 \leq y \leq 1\}$ to the parallelogram with the vertices

$$(0,0), \quad (2,1), \quad (1,2), \quad (3,3).$$

Here's a picture.



In fact, we have two choices for this linear mapping; we can either have

$$T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \qquad T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 2 \end{bmatrix},$$

or we can have

$$T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \qquad T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

In the first case we have

$$[T] = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix},$$

and in the second case we have

$$[T] = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}.$$

Notice that we can get from one of these matrices to the other by swapping the columns, which geometrically corresponds to the swapping $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$. We can write this in terms of matrix multiplication as

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

(You should check the matrix product.) This should not surprise you. The matrix $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ corresponds to the reflection through the line $y = x$, which maps our parallelogram to itself and interchanges the vectors $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Thus we see that **we represent the composition of linear mappings as matrix multiplication.** We will return to this important idea later on in these notes.

Exercise: Why can't we have $T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$? Hint: what is $\begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}$?

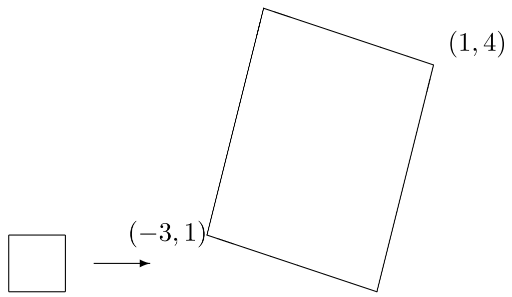
In fact, we can reverse this process. Suppose we have a matrix, let's say

$$[T] = \begin{bmatrix} 1 & -3 \\ 4 & 1 \end{bmatrix},$$

and we want to understand the linear transformation associated to this matrix. We can draw the parallelogram that T sends the unit square $\{0 \leq x \leq 1, 0 \leq y \leq 1\}$ onto, which gives us all the geometric information about T. We see from the matrix that

$$T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 4 \end{bmatrix}, \qquad T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} -3 \\ 1 \end{bmatrix}.$$

To draw the parallelogram, all we need to do is draw in these two edges starting at $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and connect them. We end up with the following picture.



Exercise: Can you explain why the image of a square is always a parallelogram? (Or a line segment, which is really a degenerate parallelogram, with one pair of opposite angles collapsed to 0 ...)

Beyond two dimensions: So far we've seen how to write down the matrix of a linear transformation taking the unit square to an arbitrary parallelogram, and how to draw the parallelogram which is the image of the unit square under an arbitrary linear mapping. However, nothing we've done so far is special to two dimensions, and everything works in higher dimensions. Let's suppose $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear mapping, which means

$$T(v+w) = T(v) + T(w), \qquad T(av) = aT(v),$$

where $v, w \in \mathbb{R}^n$ are vectors and $a \in \mathbb{R}$ is a scalar. Also let $\{e_1, e_2, e_3, \ldots, e_n\}$ be the vectors in $\mathbb{R}^n$ where e_i has a 1 in the ith component and 0 elsewhere. Then we can write down a matrix $[T]$, where the ith column of T is $T(e_i)$. We have

$$[T] = \begin{bmatrix} \vdots & & & \vdots \\ T(e_1) & T(e_2) & \cdots & T(e_n) \\ \vdots & & & \vdots \end{bmatrix}.$$

We can do a quick example, and write down the linear mapping $T : \mathbb{R}^3 \to \mathbb{R}^3$ taking the unit cube $\{0 \le x \le 1, 0 \le y \le 1, 0 \le z \le 1\}$ to the parallelpiped which has the vertices

$$\begin{array}{cccc} (0,0,0) & (3,1,1) & (1,3,1) & (1,1,3) \\ (4,4,2) & (4,2,4) & (2,4,4) & (5,5,5) \end{array}.$$

This time there are actually six such examples; we'll write one of them down, and leave the other five to you. If we want

$$T\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, \qquad T\left(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}, \qquad T\left(\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix},$$

then we must have

$$[T] = \begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{bmatrix}.$$

Exercise: Write down the matrices of the other five possible linear mappings which carry the unit cube onto this parallelpiped.

Composition of mappings and matrix multiplication: There is one very important thing we mentioned above, which we emphasize here. Namely, the composition of linear transformations is given by matrix multiplication. More precisely, if $T_1 : \mathbb{R}^n \to \mathbb{R}^k$ and $T_2 : \mathbb{R}^k \to \mathbb{R}^m$ are linear mappings, then the composition

$$T_2 \circ T_1 : \mathbb{R}^n \to \mathbb{R}^m, \qquad T_2 \circ T_1(v) = T_2(T_1(v))$$

is also linear, and the matrix associated to the composition is the product of the matrices:

$$[T_2 \circ T_1] = [T_2][T_1].$$

This explains why the definition of matrix multiplication is the way it is. As a quick check, it's good to see that the matrix product is well-defined. We have $T_1 : \mathbb{R}^n \to \mathbb{R}^k$, so that $[T_1]$ is a $k \times n$ matrix, and (similarly) $[T_2]$ is a $m \times k$ matrix. Then the product $[T_2][T_1]$ is well-defined, and it is an $m \times n$ matrix. Also, the composition $T_2 \circ T_1 : \mathbb{R}^m \to \mathbb{R}^n$ corresponds to an $m \times n$ matrix. And so everything fits together nicely.

In a later set of notes we'll concentrate on $n \times n$ matrices, and see how a linear mapping effects volume. This leads naturally to the idea of determinants, which we saw earlier in the special case of 2×2 matrices. But this is a story for another day ...