33A
Discussion Worksheet 2

1. *Proof.* Here's a short answer:

$$T\begin{pmatrix}x_1\\ \vdots\\ x_m\end{pmatrix} = T(x_1\vec{e_1} + \cdots + x_m\vec{e_m})$$
$$= x_1T(\vec{e_1}) + \cdots + x_mT(\vec{e_m})$$

There are two key realizations in this problem, corresponding to the two steps in the above chain of equalities.

(a) Every vector is a linear combination of the standard basis vectors $\vec{e_i}$.

(b) Linear transformations "preserve" linear combinations.

I'll provide some discussion on those two facts below.

(a) Written concisely, all we're saying here is that

$$\begin{pmatrix}x_1\\ \vdots\\ x_m\end{pmatrix} = x_1\vec{e_1} + \cdots + x_m\vec{e_m} \qquad (*)$$

This comes down to the very definition of adding and scaling vectors. Let's start with a more concrete example first:

$$\begin{pmatrix}1\\2\end{pmatrix}$$

If we interpret our above equation $(*)$ in this case, it would say that

$$\begin{pmatrix}1\\2\end{pmatrix} = 1\vec{e_1} + 2\vec{e_2}$$

So let's verify this directly. First things first, we'll expand the right hand side using the definitions of $\vec{e_i}$. Indeed, we get

$$1\vec{e_1} + 2\vec{e_2} = 1\begin{pmatrix}1\\0\end{pmatrix} + 2\begin{pmatrix}0\\1\end{pmatrix}$$

Now let's simplify the right hand side here. How do we do this? We have to use the definitions of scalar multiplication and vector addition. These operations have geometric meaning that's important to know, but for now let's just consider it algebraically. Indeed, addition and scalar multiplication of vectors are done component-wise. So for instance, we get the equation

$$1\begin{pmatrix}1\\0\end{pmatrix} + 2\begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}1\\0\end{pmatrix} + \begin{pmatrix}0\\2\end{pmatrix}$$

Addition is also done component-wise, so that

$$\begin{pmatrix}1\\0\end{pmatrix}+\begin{pmatrix}0\\2\end{pmatrix}=\begin{pmatrix}1\\2\end{pmatrix}$$

In short, we showed

$$\begin{aligned}1\vec{e_1}+2\vec{e_2}&=1\begin{pmatrix}1\\0\end{pmatrix}+2\begin{pmatrix}0\\1\end{pmatrix}\\&=\begin{pmatrix}1\\0\end{pmatrix}+\begin{pmatrix}0\\2\end{pmatrix}=\begin{pmatrix}1\\2\end{pmatrix}\\&=\begin{pmatrix}1\\2\end{pmatrix}\end{aligned}$$

Which is exactly what equation $(*)$ predicted in this case.Every step here was an application of the definition of addition and scalar multiplication of vectors.

The general argument goes exactly the same as this special case. If we have an expression like

$$x_1\vec{e_1}+\cdots+x_m\vec{e_m}$$

then the $\vec{e_i}$ term can only affect the i^{th} component of this vector. And indeed, the i^{th} component of this vector must then be x_i. Put more formally, we can manipulate this expression as follows:

$$\begin{aligned}x_1\vec{e_1}+\cdots+x_m\vec{e_m}&=x_1\begin{pmatrix}1\\0\\0\\\vdots\\\vdots\\0\\0\end{pmatrix}+x_2\begin{pmatrix}0\\1\\0\\\vdots\\\vdots\\0\\0\end{pmatrix}+\cdots+x_i\begin{pmatrix}0\\\vdots\\0\\1\\0\\\vdots\\0\end{pmatrix}+\cdots+x_m\begin{pmatrix}0\\0\\0\\\vdots\\\vdots\\0\\1\end{pmatrix}\\&=\begin{pmatrix}x_1\\0\\0\\\vdots\\\vdots\\0\\0\end{pmatrix}+\begin{pmatrix}0\\x_2\\0\\\vdots\\\vdots\\0\\0\end{pmatrix}+\cdots+\begin{pmatrix}0\\\vdots\\0\\x_i\\0\\\vdots\\0\end{pmatrix}+\cdots+\begin{pmatrix}0\\0\\0\\\vdots\\\vdots\\0\\x_m\end{pmatrix}\\&=\begin{pmatrix}x_1\\x_2\\\vdots\\x_i\\\vdots\\x_m\end{pmatrix}\end{aligned}$$

...yikes

I hope this also explains to some extent why we choose to write vectors as linear combinations like $x_1\vec{e_1} + \cdots + x_m\vec{e_m}$. Writing everything down with these super long column vectors is tedious and takes a lot of space. In contrast, $x_1\vec{e_1} + \cdots + x_m\vec{e_m}$ is very concise. Brevity of notation is no small thing! It's meaningful to be able to express more with less. If you're comfortable with it, things become even faster if we use Σ notation $\sum_{i=1}^{m} x_i\vec{e_i}$.

(b) First, I should explain what exactly is meant by "preserve" here. Formally, if $T : \mathbb{R}^m \longrightarrow \mathbb{R}^n$ is linear, I mean that the following two equations hold:

$$T(\vec{x} + \vec{y}) = T(\vec{x}) + T(\vec{y})$$
$$T(c\vec{x}) = cT(\vec{x})$$

for any scalar $c \in \mathbb{R}$ and any vectors $\vec{x}, \vec{y} \in \mathbb{R}^m$. It may be helpful to think about these algebraic formulas in a geometric fashion. The idea is that these equations essentially say that a linear transformation takes an evenly space grid of parallelograms to another evenly space grid of parallelograms. I'll leave this as a bit of a puzzle, but you can see this 3Blue1Brown video for some great exposition on this topic.

Anyways, these two algebraic identities are quite poweful. They say that T preserves addition and scalar multiplication. You can also think of this as saying doing addition or scalar multiplication inside of T is the same as doing it outside. A linear combination of vectors is some combination of addition and scalar multiplication. Hence, these two equations tell us that T preserves linear combinations. Explicitly, we have that

$$T(c_1\vec{v_1} + c_2\vec{v_2} + \cdots + c_k\vec{v_k}) = c_1T(\vec{v_1}) + c_2T(\vec{v_2}) + \cdots + c_kT(\vec{v_k})$$

for any collection of scalars $c_1, \ldots, c_k$ and vectors $\vec{v_1}, \ldots, \vec{v_k}$. This arises by taking those two equations I had above and iterating them. For example, for $k = 3$ we can prove this follows:

$$\begin{aligned}
T(c_1\vec{v_1} + c_2\vec{v_2} + c_3\vec{v_3}) &= T(c_1\vec{v_1} + (c_2\vec{v_2} + c_3\vec{v_3})) \\
&= T(c_1\vec{v_1}) + T(c_2\vec{v_2} + c_3\vec{v_3}) \\
&= T(c_1\vec{v_1}) + T(c_2\vec{v_2}) + T(c_3\vec{v_3}) \\
&= c_1T(\vec{v_1}) + c_2T(\vec{v_2}) + c_3T(\vec{v_3})
\end{aligned}$$

A rigorous proof for the general case of k many vectors and scalars isn't in the scope of what we're doing here, but the pattern is the same. If you don't believe this, try it for 4 vectors, then for 5 vectors, and keep going to convince yourself it'll work no matter how many vectors you tried it with. If you don't feel convinced that these particular cases can prove the general case (which I completely understand!), the relevant thing to look up is "proof by induction", which we won't discuss in this class.

With these two facts together, we can use the short answer I provided at the start. By the way, this result yields one of the most incredible facts of linear algebra – to specify a linear transformation, you only need to tell me what it does on $\vec{e_1}, \dots, \vec{e_m}$, which is a finite number of points! Indeed, look at both sides of the equation in this problem. The left hand side is about T on an *arbitrary* input. Whereas on the right hand side, we only evaluate T at the vectors $\vec{e_i}$! $\square$

2. *Proof.* Short answer: They are of the form $T(x) = ax$, and their graphs are non-vertical lines through the origin.

   Let $T : \mathbb{R} \longrightarrow \mathbb{R}$ be linear. Then T is represented by a 1×1 matrix A. This means that for $\vec{v} \in \mathbb{R}$, $T(\vec{v}) = A\vec{v}$. But hang on, a 1×1 matrix A? And a 1×1 column vector $\vec{v}$? These are just dressed up real numbers! If we were to write down a 1×1 matrix A, we'd say $A = (a)$ for some real number a. And out vector $\vec{v}$ would be $\vec{v} = (x)$ for some real number x. So the equation $T(\vec{v}) = A\vec{v}$ is a fancy way of saying $T(x) = ax$.

   Another way to think of this above argument is as follows. As T is linear, we must have that $T(x) = xT(1)$, so we just let $a = T(1)$.

   In any case, we have classified every linear map $T : \mathbb{R} \longrightarrow \mathbb{R}$. They're functions of the form $T(x) = ax$ for some $a \in \mathbb{R}$.

   So what do their graphs look like? Let's remember that the graph of a function $T : \mathbb{R} \longrightarrow \mathbb{R}$ is a subset of the plane $\mathbb{R}^2$ so that at every point on the x - axis, we draw a point with y - coordinate equal to $T(x)$. In other words, it's the set of points in $\mathbb{R}^2$ with coordinates (x, y) such that $T(x) = y$. In formal set builder notation, it's given by $\{(x, y) \in \mathbb{R}^2 : y = T(x)\}$. Any way we say it, it's the same notion of graphs in, say, precalculus.

   And indeed, you may recall from a class like precalculus that the graph of a function like $T(x) = ax$ is a line with y - intercept 0 and slope a. Hence, the graphs look like lines through the origin. But there's one important qualification - these lines cannot be vertical! A vertical line is thought of as having infinite slope, so it's not the graph of $T(x) = ax$, which has a slope a, a real number. $\square$

3. *Proof.* Short answer: They are of the form

$$T\begin{pmatrix} x \\ y \end{pmatrix} = ax + by$$

   for some real numbers a and b. Their graphs are planes in $\mathbb{R}^3$ through the origin which do not contain the z - axis.

   This question is certainly more difficult than the previous one, but the method will be similar. We can write down an explicit form of these linear maps using matrices. In problem 2, I appealed to previous knowledge about graphs of common single variable functions, but this time I'll discuss in greater detail how we arrive at the shape of the graph.

First thing's first, let $T : \mathbb{R}^2 \longrightarrow \mathbb{R}$ be linear. By definition, this means that there is some 1×2 matrix A so that $T(\vec{x}) = A\vec{x}$. Let's write A as

$$A = \begin{pmatrix} a & b \end{pmatrix}$$

for some real numbers a, b. Let's also name our variables

$$\vec{x} = \begin{pmatrix} x \\ y \end{pmatrix}$$

Then we can compute their product as

$$A\vec{x} = \begin{pmatrix} a & b \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = ax + by$$

Therefore, we have that

$$T \begin{pmatrix} x \\ y \end{pmatrix} = ax + by$$

This gives an explicit formula for T. Now what about its graph? First, we should determine what we mean by the graph of a function $T : \mathbb{R}^2 \longrightarrow \mathbb{R}$. It's similar to the notion of graphs of functions $\mathbb{R} \longrightarrow \mathbb{R}$ we used in the previous problem, and that you likely saw in courses like precalculus. There, the input was viewed as the x - axis and the output of the function on an x value was represented as the height over that point. So if we had, for example, $f(5) = 7$ then we'd draw a point at the coordinates $(5, 7)$ in $\mathbb{R}^2$. When we do this for all inputs, we get the graph. Formally then, the graph of a function $f : \mathbb{R} \longrightarrow \mathbb{R}$ is given as

$$\{(x, f(x)) : x \in \mathbb{R}\}$$

This means the set of all points of the form $(x, f(x))$ where x ranges over all real numbers. We can equivalently think of this as the set of points (x, y) satisfying $y = f(x)$.

Now, how about the graph of a function $T : \mathbb{R}^2 \longrightarrow \mathbb{R}$? Here, the input will consist of length 2 vectors of the form

$$\begin{pmatrix} x \\ y \end{pmatrix}$$

So we can think of the inputs to this function T as being the points on the xy - plane, similarly to how in the single variable case we thought of the inputs as being on the x - axis. Now, how do we represent the output? We'll again think of this as the height over the point in the xy - plane. Try visualizing this as, say, having the xy - plane be your desk and the z - axis as the height above (and below for negative values!) your desk. For example, if we had

$$T \begin{pmatrix} 1 \\ 2 \end{pmatrix} = 5$$

Then we'd draw a point at height 5 above $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ on the xy - plane. That is, we'd draw the point

$$\begin{pmatrix} 1 \\ 2 \\ 5 \end{pmatrix}$$

In general, over the point

$$\begin{pmatrix} x \\ y \end{pmatrix}$$

on the xy - plane, we draw the point

$$\begin{pmatrix} x \\ y \\ T\begin{pmatrix} x \\ y \end{pmatrix} \end{pmatrix}$$

The graph is then the set of all such points, as $\begin{pmatrix} x \\ y \end{pmatrix}$ ranges over all points in the xy - plane.

So back to linear functions, recall that we showed any linear function $T : \mathbb{R}^2 \longrightarrow \mathbb{R}$ is of the form

$$T\begin{pmatrix} x \\ y \end{pmatrix} = ax + by$$

for some real numbers a, b. By definition, its graph will consist of points of the form

$$\begin{pmatrix} x \\ y \\ ax + by \end{pmatrix}$$

ranging over all real numbers x, y. In other words, it's the set of points in $\mathbb{R}^3$ satisfying the equation $z = ax + by$. You may recognize this already as the equation defining a plane in three dimensional space. In, say, a multivariable calculus class, you may have also learned that this plane has normal vector

$$\begin{pmatrix} a \\ b \\ -1 \end{pmatrix}$$

Without this prior knowledge, we can still see a way to realize this graph as a plane. Indeed, points on the graph look like

$$\begin{pmatrix} x \\ y \\ ax + by \end{pmatrix}$$

Notice that there are two free parameters here, x and y, which can be taken to be any real numbers we please. This suggests that the graph may also be some sort of two dimensional shape. To better understand this, we split the x and y contributions apart. That is, we write

$$\begin{pmatrix} x \\ y \\ ax+by \end{pmatrix} = x \begin{pmatrix} 1 \\ 0 \\ a \end{pmatrix} + y \begin{pmatrix} 0 \\ 1 \\ b \end{pmatrix}$$

The fancy word here is that the graph of T is therefore the *span* of the vectors

$$\begin{pmatrix} 1 \\ 0 \\ a \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ b \end{pmatrix}$$

Let's try to reason some more about why this must be a plane. Suppose we hold y constant. Then letting x vary, we get a line parallel to the vector

$$\begin{pmatrix} 1 \\ 0 \\ a \end{pmatrix}$$

and going through the point

$$\begin{pmatrix} 0 \\ y \\ yb \end{pmatrix}$$

Holding y - fixed like this yields a slice of the graph, which we recognize as a line. Now, if we change y we get another parallel line, but through a different point now. By doing so, we can recognize that this graph looks like a union of lines parallel to

$$\begin{pmatrix} 1 \\ 0 \\ a \end{pmatrix}$$

We can apply the same reasoning by first fixing x and realizing that the resulting slice is another line, this time parallel to

$$\begin{pmatrix} 0 \\ 1 \\ b \end{pmatrix}$$

See the Figure 1 below. In this particular example I took $a = 1$ and $b = 1.5$. The red part is the xy - plane. The blue lines arise by holding y constant. Hence, the blue lines on the red plane are parallel to the x - axis. The black lines arise by holding x constant. So similarly, the black lines on the red plane are parallel to the y - axis. The lines that are off of the xy - plane represent the slices of the graph I was talking about above. They each correspond to the part of the graph of T lying over one of the lines drawn on the xy - plane. So the grid of lines on the xy - plane correspond to a grid on the graph of T.

Now imagine I made the grid even finer, like in Figure 2. Then the corresponding

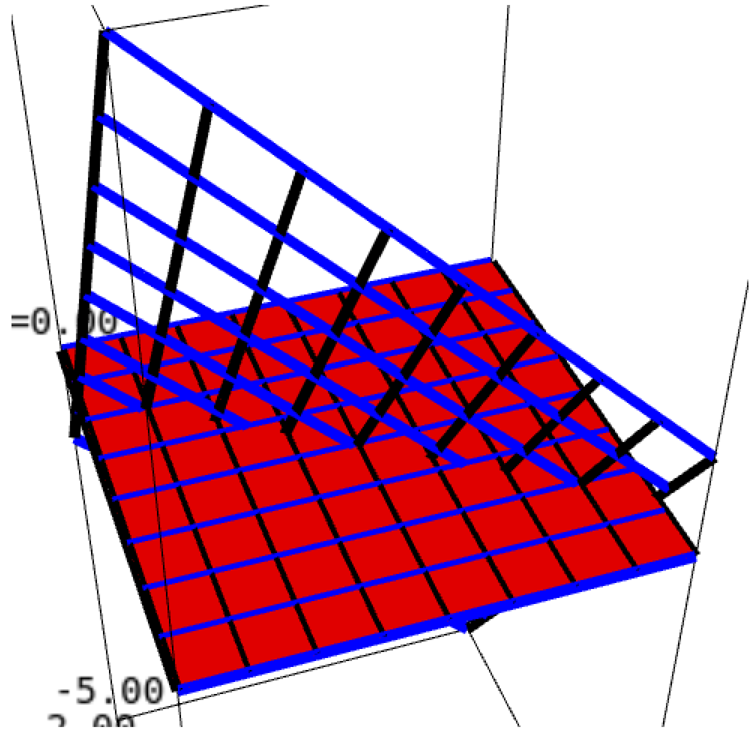

Figure 1: planes

grid on the graph of T gets even finer too. If we make the grid on the xy - plane extremely fine, like in Figure 3 the it'll look like we're covering the whole xy - plane. The corresponding grid on the graph will then grow even finer, and perhaps now we can believe that the graph will indeed be a plane.

Let's now return to the question at hand. We said above in the short answer that the graphs are planes which do not contain the z - axis. Certainly, the pictures above don't contain the z - axis. But why is this impossible. This is analogous to the vertical line test from classes like precalculus. Over any point in the xy - plane, there can be only one output of the function T and therefore only one point on the graph. The z - axis, however, consists of infinitely many points above the origin in the xy - plane, so the graph of a function cannot contain the z - axis. □

4. *Proof.* Short answer: yeah

The easiest way to do this is to recall that a function T is linear if and only if the following two equations hold

$$T(\vec{x}+\vec{y}) = T(\vec{x}) + T(\vec{y})$$
$$T(c\vec{x}) = cT(\vec{x})$$

for all vectors $\vec{x}, \vec{y}$ and scalars c. This is theorem 2.1.3 in Bretscher. The officially given definition of T being linear was that $T(\vec{x}) = A\vec{x}$ for some matrix A. To briefly explain this, if we only had the above two equations, how do we find this matrix A so

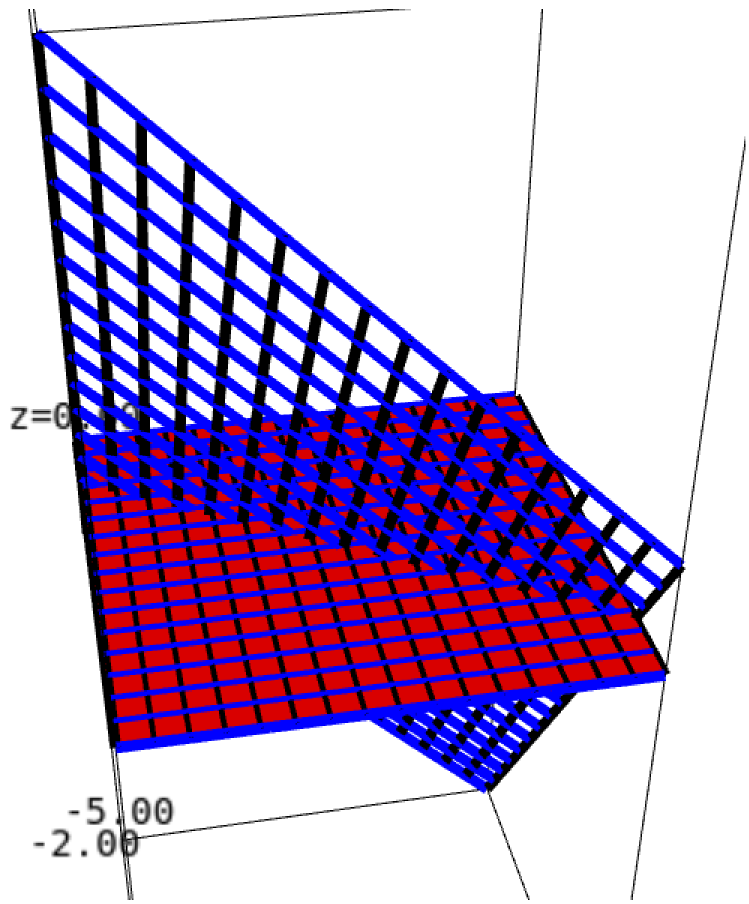

Figure 2: finer grid

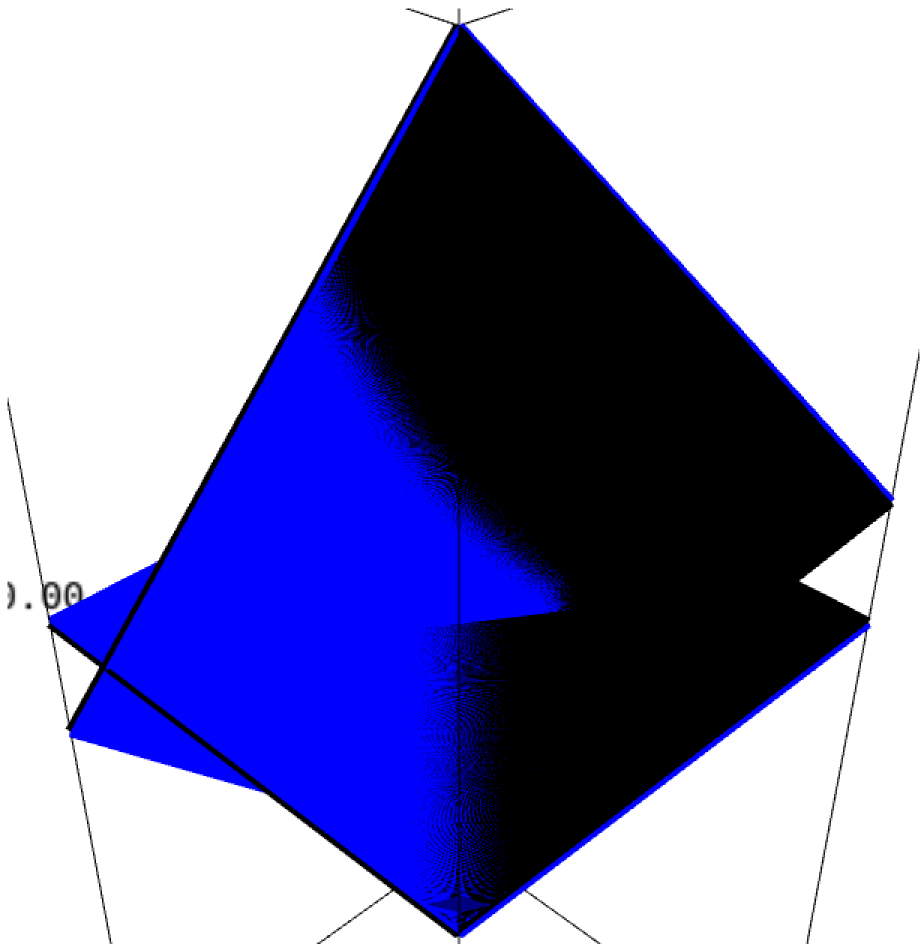

Figure 3: very fine grid

that $T(x) = A\vec{x}$? We let A be the matrix with i^{th} column equal to $T(\vec{e_i})$. Problem 1 is an essential part of understanding why we can therefore conclude that $T(\vec{x}) = A\vec{x}$.

With this in mind, let's show that the transformation $\vec{z} = L(T(\vec{x}))$ satisfies these two equations. In other words, that the composition of linear transformations L and T is also linear.

Indeed, let's start with addition. We want to show that

$$L(T(\vec{x}+\vec{y})) \stackrel{?}{=} L(T(\vec{x})) + L(T(\vec{y}))$$

for any vectors $\vec{x}, \vec{y}$ in $\mathbb{R}^m$. We know that L and T preserve addition, because we are assuming they are linear. So we know

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

This lets us simplify the left hand side of our desired equation as

$$L(T(\vec{x}+\vec{y})) = L(T(\vec{x}) + T(\vec{y}))$$

Aha, but L also preserves addition, as it is also assumed to be linear! So this simplifies further to become

$$L(T(\vec{x}) + T(\vec{y})) = L(T(\vec{x})) + L(T(\vec{y}))$$

Then we conclude

$$L(T(\vec{x}+\vec{y})) = L(T(\vec{x})) + L(T(\vec{y}))$$

as desired.

We'll do the same thing now for scalar multiplication, but written more quickly as it's the same idea. We apply the hypothesis that T is linear to pull out the scalar once. We apply the hypothesis that L is linear to pull out the scalar again, then we win.

Indeed, our goal now is to show

$$L(T(c\vec{x})) \stackrel{?}{=} cL(T(\vec{x}))$$

We manipulate the left hand side as follows:

$$\begin{aligned} L(T(c\vec{x})) &= L(cT(\vec{x})) \\ &= cL(T(\vec{x})) \end{aligned}$$

which is our desired formula,

Hence, the composition $L \circ T$ satisfies the equationswe wrote at the start, namely that it preserves addition and scalar multiplication. So we conclude that $L \circ T$ is linear. $\square$

5. *Proof.* Short answer: BA

So in problem 4 above, we were able to show that the composition of linear functions is linear, but we did so by sidestepping the matrices and instead using the more "intrinsic" notion of a linear transformation as being a function which preserves addition and scalar multiplication. But the correspondence between linear transformations and matrices is a very fruitful one, so here we want to explore how composition works in terms of matrices.

We'll explicitly compute $T(\vec{x})$ in terms of the formula the problem gives us. Indeed, let's write

$$\vec{x} = \begin{pmatrix} x \\ y \end{pmatrix}$$

and do some multiplying.

$$\begin{aligned} T(\vec{x}) &= B(A\vec{x}) \\ &= \begin{pmatrix} p & q \\ r & s \end{pmatrix} \left(\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \right) \end{aligned}$$

We'll first multiply the expression within the parentheses

$$\begin{pmatrix} p & q \\ r & s \end{pmatrix} \left(\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \right) = \begin{pmatrix} p & q \\ r & s \end{pmatrix} \begin{pmatrix} ax+by \\ cx+dy \end{pmatrix}$$

And now, we multiply again

$$\begin{aligned} \begin{pmatrix} p & q \\ r & s \end{pmatrix} \begin{pmatrix} ax+by \\ cx+dy \end{pmatrix} &= \begin{pmatrix} p(ax+by)+q(cx+dy) \\ r(ax+by)+s(cx+dy) \end{pmatrix} \\ &= \begin{pmatrix} pax+pby+qcx+qdy \\ rax+rby+scx+sdy \end{pmatrix} \\ &= \begin{pmatrix} (pa+qc)x+(pb+qd)y \\ (ra+sc)x+(rb+sd)y \end{pmatrix} \end{aligned}$$

We can recognize this last expression as looking like a matrix times the vector $\begin{pmatrix} x \\ y \end{pmatrix}$ Indeed, we have

$$\begin{pmatrix} (pa+qc)x+(pb+qd)y \\ (ra+sc)x+(rb+sd)y \end{pmatrix} = \begin{pmatrix} pa+qc & pb+qd \\ ra+sc & rb+sd \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

This new matrix looks pretty gross, but we can note that this is nothing more than the product BA.

There's a subtle point here that T is *not* originally defined with BA, though it may look like it is. Notice where the parentheses are! It's defined as $T(\vec{x}) = B(A\vec{x})$, so we first multiply $A\vec{x}$ and then multiply the resulting vector on the left by B. We showed here that this is the same as $T(\vec{x}) = (BA)\vec{x}$. In other words, we showed

$$(BA)\vec{x} = B(A\vec{x})$$

This is called "associativity", but the word itself is not so important.

By the way, we could have made this computation a bit easier by realizing that the columns of the matrix of T are given by $T(\vec{e_1})$ and $T(\vec{e_2})$. To be precise, applying this fact would require us to know that T is actually a linear transformation. How do we know that this is true without already knowing that T is represented by a matrix? Well multiplying by A and multiplying by B are both linear transformations, and T is the composition of these two. More formally, if we let $S(\vec{x}) = A\vec{x}$ and $R(\vec{x}) = B\vec{x}$, then T is defined as $T(\vec{x}) = R(S(\vec{x}))$. This is exactly the scenario of problem 4, so we know then that T must be a linear transformation without knowing yet what its matrix is.

Finally, I'd like to say that this is an important fact to note. We have a correspondence between linear transformations and matrices. This leads us to ask how this correspondence acts of preexisting operations on linear transformations or on matrices. An operation that exists of linear transformations is composition - we take two linear transformations and get another one. An operation on matrices is multiplication - we take two matrices and get another one. This problem is a special case of the fact that these operations correspond to one another! $\square$