LINEAR MAPS, THE TOTAL DERIVATIVE AND THE CHAIN RULE

ROBERT LIPSHITZ

ABSTRACT. We will discuss the notion of linear maps and introduce the total derivative of a function $f\colon \mathbb{R}^n \to \mathbb{R}^m$ as a linear map. We will then discuss composition of linear maps and the chain rule for derivatives.

CONTENTS

1. MAPS $\mathbb{R}^n \to \mathbb{R}^m$

So far in this course, we have talked about:

- Functions $\mathbb{R} \to \mathbb{R}$; you worked with these a lot in Calculus 1.
- Parametric curves, i.e., functions $\mathbb{R} \to \mathbb{R}^2$ and $\mathbb{R} \to \mathbb{R}^3$, and
- Functions of several variables, i.e., functions $\mathbb{R}^2 \to \mathbb{R}$ and $\mathbb{R}^3 \to \mathbb{R}$.

(We've probably also seen some examples of maps $\mathbb{R}^n \to \mathbb{R}$ for some $n > 3$, but we haven't worked with these so much.)

What we haven't talked about much are functions $\mathbb{R}^3 \to \mathbb{R}^3$, say, or $\mathbb{R}^3 \to \mathbb{R}^2$, or $\mathbb{R}^2 \to \mathbb{R}^2$.

Definition 1.1. A function $f\colon \mathbb{R}^3 \to \mathbb{R}^3$ is something which takes as input a vector in $\mathbb{R}^3$ and gives as output a vector in $\mathbb{R}^3$. Similarly, a map $\mathbb{R}^3 \to \mathbb{R}^2$ takes as input a vector in $\mathbb{R}^3$ and gives as output a vector in $\mathbb{R}^2$; and so on.

Example 1.2. Rotation by $\pi/6$ counter-clockwise around the z-axis is a function $\mathbb{R}^3 \to \mathbb{R}^3$: it takes as input a vector in $\mathbb{R}^3$ and gives as output a vector in $\mathbb{R}^3$. Let's let $R(\vec{v})$ denote rotation of $\vec{v}$ by $\pi/6$ counter-clockwise around the z-axis. Then, doing some trigonometry (see Figure 1 for the two-dimensional analogue), we have for example

$$\begin{aligned}
R(1,0,0) &= (\cos(\pi/6), \sin(\pi/6), 0) = (\sqrt{3}/2, 1/2, 0)\\
R(0,1,0) &= (-\sin(\pi/6), \cos(\pi/6), 0)(-1/2, \sqrt{3}/2, 0)\\
R(0,0,1) &= (0,0,1)\\
R(2,3,1) &= (2\cos(\pi/6) - 3\sin(\pi/6), 2\sin(\pi/6) + 3\cos(\pi/6), 1).
\end{aligned}$$

Example 1.3. Translation by the vector $(1,2,3)$ is a map $\mathbb{R}^3 \to \mathbb{R}^3$: it takes as input any vector $\vec{v}$ and gives as output $\vec{v} + (1,2,3)$. Let $T(\vec{v})$ denote translation of $\vec{v}$ by $(1,2,3)$. Then, for example

$$\begin{aligned}
T(1,0,0) &= (2,2,3)\\
T(0,1,0) &= (1,3,3)\\
T(0,0,1) &= (1,2,4)\\
T(2,3,1) &= (3,5,4).
\end{aligned}$$

Example 1.4. Orthogonal projection from $\mathbb{R}^3$ to the xy-plane can be viewed as a function $P\colon \mathbb{R}^3 \to \mathbb{R}^2$. Then, for example

$$\begin{aligned}
P(1,0,0) &= (1,0)\\
P(0,1,0) &= (0,1)\\
P(0,0,1) &= (0,0)\\
P(2,3,1) &= (2,3).
\end{aligned}$$

Example 1.5. Rotation by an angle θ around the origin gives a map $R_\theta\colon \mathbb{R}^2 \to \mathbb{R}^2$. For example,

$$\begin{aligned}
R_\theta(1,0) &= (\cos(\theta), \sin(\theta))\\
R_\theta(0,1) &= (-\sin(\theta), \cos(\theta))\\
R_\theta(2,3) &= (2\cos(\theta) - 3\sin(\theta), 2\sin(\theta) + 3\cos(\theta)).
\end{aligned}$$

(See Figure 1 for the trigonometry leading to the computation of $R_\theta(2,3)$.)

Just like a function $\mathbb{R} \to \mathbb{R}^3$ corresponds to three functions $\mathbb{R} \to \mathbb{R}$, a function $\mathbb{R}^3 \to \mathbb{R}^3$ corresponds to three functions $\mathbb{R}^3 \to \mathbb{R}$. That is, any function $F\colon \mathbb{R}^3 \to \mathbb{R}^3$ is given by

$$F(x,y,z) = (u(x,y,z), v(x,y,z), w(x,y,z)).$$

Example 1.6. The function R from Example 1.2 has the form

$$R(x,y,z) = (x\cos(\pi/6) - y\sin(\pi/6), x\sin(\pi/6) + y\cos(\pi/6), z).$$

Example 1.7. The function T from Example 1.3 has the form

$$T(x,y,z) = (x+1, y+2, z+3).$$

Example 1.8. There is a function which takes a point written in cylindrical coordinates and rewrites it in rectangular coordinates, which we can view as a map $\mathbb{R}^3 \to \mathbb{R}^3$ given by

$$F(r,\theta,z) = (r\cos(\theta), r\sin(\theta), z).$$

Similarly, there is a function which takes a point written in spherical coordinates and rewrites it in rectangular coordinates; viewed as a function $\mathbb{R}^3 \to \mathbb{R}^3$ it is given by

$$G(\rho,\theta,\phi) = (\rho\cos(\theta)\sin(\phi), \rho\sin(\theta)\sin(\phi), \rho\cos(\phi)).$$

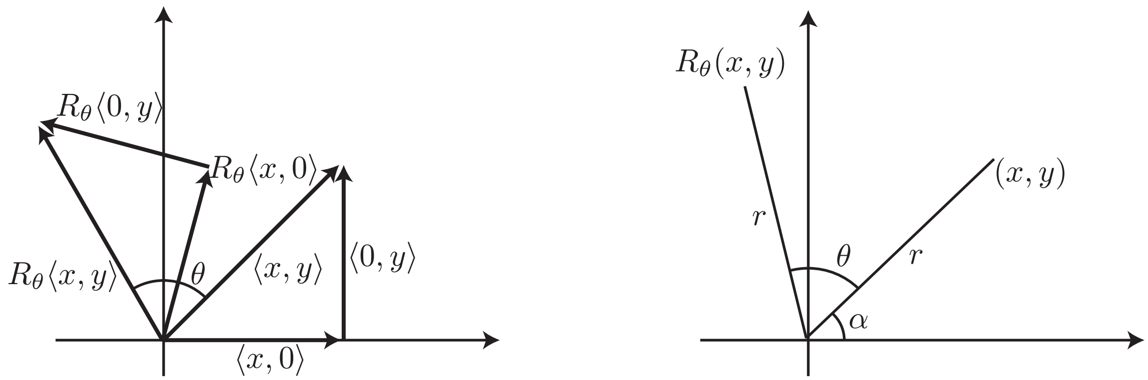

FIGURE 1. **Two derivations of the formula for rotation of $\mathbb{R}^2$ by an angle θ.** Left: Notice that $(x, y) = (x, 0) + (0, y)$. So, rotating the whole picture, $R_\theta(x, y) = R_\theta(x, 0)+R_\theta(0, y) = (x\cos(\theta), x\sin(\theta))+(-y\sin(\theta), y\cos(\theta))$, which is exactly $(x\cos(\theta) - y\sin(\theta), y\cos(\theta) + x\sin(\theta))$. Right: the point (x, y) is $(r\cos(\alpha), r\sin(\alpha))$ and $R_\theta(x, y) = (r\cos(\alpha + \theta), \sin(\alpha + \theta))$. So, by addition formulas for sine and cosine, $R_\theta(x, y) = (r\cos(\alpha)\cos(\theta) - r\sin(\alpha)\sin(\theta), r\sin(\alpha)\cos(\theta) + r\cos(\alpha)\sin(\theta))$ which is exactly $(x\cos(\theta) - y\sin(\theta), y\cos(\theta) + x\sin(\theta))$.

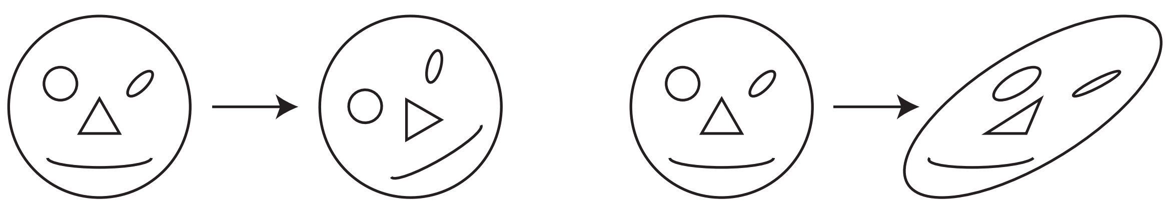

FIGURE 2. **Visualizing maps $\mathbb{R}^2 \to \mathbb{R}^2$.** Left: the effect of the map $R_{\pi/6}$ from Example 1.5. Right: the effect of the function $F(x, y) = (x + y, y)$.

Example 1.9. The function R_θ from Example 1.5 is given in coordinates by:

$$R_\theta(x, y) = (x\cos(\theta) - y\sin(\theta), x\sin(\theta) + y\cos(\theta)).$$

See Figure 1 for the trigonometry leading to this formula.

Functions $\mathbb{R}^2 \to \mathbb{R}^2$ are a little easier to visualize than functions $\mathbb{R}^3 \to \mathbb{R}^3$ or $\mathbb{R}^3 \to \mathbb{R}^2$ (or. . .). One way to visualize them is to think of how they transform a picture. A couple of examples are illustrated in Figure 2.

Given functions $F\colon \mathbb{R}^p \to \mathbb{R}^n$ and $G\colon \mathbb{R}^n \to \mathbb{R}^m$ you can compose F and G to get a new function $G \circ F\colon \mathbb{R}^p \to \mathbb{R}^m$. That is, $G \circ F$ is the function which takes a vector $\vec{v}$, does F to $\vec{v}$, and then does G to the result; symbolically, $G \circ F(\vec{v}) = G(F(v))$.

Note that $G \circ F$ only makes sense if the target of F is the same as the source of G.

Example 1.10. For R_θ the rotation map from Example 1.5, $R_{\pi/2}\circ R_{\pi/6}$ means you first rotate by $\pi/6$ and then by another $\pi/2$. So, this is just a rotation by $\pi/2+\pi/6=2\pi/3$, i.e.,

$$R_{\pi/2}\circ R_{\pi/6}=R_{2\pi/3}.$$

Example 1.11. If you apply the rotation map R from Example 1.2 and then the projection map P from Example 1.4 the result is the same as applying P first and then rotating in the plane by $\pi/6$. That is,

$$P\circ R=R_{\pi/6}\circ P.$$

In terms of coordinates, composition of maps corresponds to substituting variables, as the following example illustrates.

Example 1.12. Let R be the function from Example 1.2 and T the function from Example 1.3; that is,

$$\begin{aligned}R(x,y,z)&=(x\cos(\pi/6)-y\sin(\pi/6),x\sin(\pi/6)+y\cos(\pi/6),z)\\T(x,y,z)&=(x+1,y+2,z+3).\end{aligned}$$

Then,

$$\begin{aligned}T\circ R(x,y,z)&=T(x\cos(\pi/6)-y\sin(\pi/6),x\sin(\pi/6)+y\cos(\pi/6),z)\\&=(x\cos(\pi/6)-y\sin(\pi/6)+1,x\sin(\pi/6)+y\cos(\pi/6)+2,z+3).\end{aligned}$$

Similarly,

$$\begin{aligned}R\circ T(x,y,z)&=R(x+1,y+2,z+3)\\&=((x+1)\cos(\pi/6)-(y+2)\sin(\pi/6),\\&\qquad(x+1)\sin(\pi/6)+(y+2)\cos(\pi/6),z+3).\end{aligned}$$

Notice that composition of maps is *not* commutative: $R\circ T$ is not the same as $T\circ R$. Geometrically, this says that translating and then rotating is not the same as rotating and then translating; if you think about it, that makes sense.

As another example of composition as substitution of variables, let's do Example 1.11 in terms of coordinates:

Example 1.13. Let's compute $P\circ R$ in coordinates. We had

$$\begin{aligned}P(x,y,z)&=(x,y)\\R(x,y,z)&=(x\cos(\pi/6)-y\sin(\pi/6),x\sin(\pi/6)+y\cos(\pi/6),z).\end{aligned}$$

So,

$$\begin{aligned}P\circ R(x,y,z)&=P(x\cos(\pi/6)-y\sin(\pi/6),x\sin(\pi/6)+y\cos(\pi/6),z)\\&=(x\cos(\pi/6)-y\sin(\pi/6),x\sin(\pi/6)+y\cos(\pi/6)).\end{aligned}$$

On the other hand,

$$R_{\pi/6}(x,y)=(x\cos(\pi/6)-y\sin(\pi/6),x\sin(\pi/6)+y\cos(\pi/6)).$$

So,

$$R_{\pi/6}\circ P(x,y,z)=R_{\pi/6}(x,y)=(x\cos(\pi/6)-y\sin(\pi/6),x\sin(\pi/6)+y\cos(\pi/6)).$$

So, again we see that $P\circ R=R_{\pi/6}\circ P$.

As one last example, we'll do Example 1.10 again in terms of coordinates:

Example 1.14. We have

$$\begin{aligned} R_{\pi/2}(x,y) &= (x\cos(\pi/2) - y\sin(\pi/2), x\sin(\pi/2) + y\cos(\pi/2)) = (-y, x) \\ R_{\pi/6}(x,y) &= (x\cos(\pi/6) - y\sin(\pi/6), x\sin(\pi/6) + y\cos(\pi/6)) \\ &= (x\sqrt{3}/2 - y/2, x/2 + y\sqrt{3}/2) \\ R_{2\pi/3}(x,y) &= (x\cos(2\pi/3) - y\sin(2\pi/3), x\sin(2\pi/3) + y\cos(2\pi/3)) \\ &= (-x/2 - y\sqrt{3}/2, x\sqrt{3}/2 - y/2). \end{aligned}$$

Composing $R_{\pi/2}$ and $R_{\pi/6}$ we get

$$R_{\pi/2} \circ R_{\pi/6}(x,y) = R_{\pi/2}(x\sqrt{3}/2 - y/2, x/2 + y\sqrt{3}/2) = (-x/2 - y\sqrt{3}/2, x\sqrt{3}/2 - y/2).$$

So, again we see that $R_{\pi/2} \circ R_{\pi/6} = R_{2\pi/3}$.

Exercise 1.15. Let T be the map from Example 1.3. What does the map $T \circ T$ mean geometrically? What is it in coordinates?

Exercise 1.16. Let F be the "write cylindrical coordinates in rectangular coordinates" map from Example 1.8. Let $H(r,\theta,z) = (r, \theta + \pi/6, z)$. Compute $F \circ H$ in coordinates.

Exercise 1.17. Let R_θ be the rotation map from Example 1.5. Compose R_θ and R_ϕ by substituting, like in Example 1.14.

We also know that $R_\phi \circ R_\theta = R_{\theta+\phi}$. Deduce the formulas for $\sin(\theta+\phi)$ and $\cos(\theta+\phi)$.

2. Linear maps

We'll start with a special case:

Definition 2.1. A map $F\colon \mathbb{R}^2 \to \mathbb{R}^2$ is *linear* if F can be written in the form

$$F(x,y) = (ax + by, cx + dy)$$

for some real numbers a, b, c, d.

That is, a linear map is one given by homogeneous linear equations:

$$\begin{aligned} x_{\text{new}} &= a x_{\text{old}} + b y_{\text{old}} \\ y_{\text{new}} &= c x_{\text{old}} + d y_{\text{old}}. \end{aligned}$$

(The word homogeneous means that there are no constant terms.)

Example 2.2. The map $f(x,y) = (2x + 3y, x + y)$ is a linear map $\mathbb{R}^2 \to \mathbb{R}^2$.

Example 2.3. The map $R_{\pi/2}\colon \mathbb{R}^2 \to \mathbb{R}^2$ which is rotation by the angle $\pi/2$ around the origin is a linear map: it is given in coordinates as

$$R_{\pi/2}(x,y) = (-y, x) = (0x + (-1)y, 1x + 0y).$$

Example 2.4. More generally, the map $R_\theta\colon \mathbb{R}^2 \to \mathbb{R}^2$ which is rotation by the angle θ around the origin is a linear map: it is given in coordinates as

$$R_\theta(x,y) = (\cos(\theta)x - \sin(\theta)y, \sin(\theta)x + \cos(\theta)y).$$

(Notice that $\cos(\theta)$ and $\sin(\theta)$ are constants, so it's fine for them to be coefficients in a linear map.)

More generally:

Definition 2.5. A map $F\colon \mathbb{R}^n \to \mathbb{R}^m$ is *linear* if F can be written in the form

$$\begin{aligned}F(x_1,&\dots,x_n)\\ &= (a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,n}x_n, a_{2,1}x_1 + a_{2,2}x_2 + \cdots + a_{2,n}x_n, \dots,\\ &\qquad a_{m,1}x_1 + a_{m,2}x_2 + \cdots + a_{m,n}x_n)\end{aligned}$$

for some constants $a_{1,1}, \dots, a_{m,n}$.

Example 2.6. The projection map P from Example 1.4 is a linear map $\mathbb{R}^3 \to \mathbb{R}^2$. If we take $a_{1,1} = 1$, $a_{1,2} = 0$, $a_{1,3} = 0$, $a_{2,1} = 0$, $a_{2,2} = 1$ and $a_{2,3} = 0$ then

$$P(x, y, z) = (a_{1,1}x + a_{1,2}y + a_{1,3}z, a_{2,1}x + a_{2,2}y + a_{2,3}z) = (x, y).$$

Example 2.7. The rotation map R from Example 1.2 is a linear map $\mathbb{R}^3 \to \mathbb{R}^3$: we wrote R in the desired form in Example 1.6.

Lemma 2.8. *The composition of two linear maps is a linear map. That is, if $F\colon \mathbb{R}^p \to \mathbb{R}^n$ and $G\colon \mathbb{R}^n \to \mathbb{R}^m$ are linear maps then $G \circ F\colon \mathbb{R}^p \to \mathbb{R}^m$ is a linear map.*

Proof. We will just check the two-dimensional case; the general cases is similar but the notation is more complicated. Suppose

$$\begin{aligned}F(x, y) &= (ax + by, cx + dy)\\ G(x, y) &= (ex + fy, gx + hy).\end{aligned}$$

Then

$$\begin{aligned}G \circ F(x, y) &= G(ax + by, cx + dy)\\ &= (e(ax + by) + f(cx + dy), g(ax + by) + h(cx + dy))\\ &= ((ae + cf)x + (be + df)y, (ag + ch)x + (bg + dh)y),\end{aligned}$$

which is, indeed, a linear map. □

Notice that the proof of Lemma 2.8 actually gave us a formula for $G \circ F$ in terms of the formulas for F and G.

Example 2.9. In Example 1.14, we composed the maps $R_{\pi/2}$ and $R_{\pi/6}$. We had:

$$\begin{aligned}R_{\pi/2}(x, y) &= (-y, x)\\ R_{\pi/6}(x, y) &= (x\sqrt{3}/2 - y/2, x/2 + y\sqrt{3}/2).\end{aligned}$$

So, if we set $a = 0$, $b = -1$, $c = 1$, $d = 0$; and $e = \sqrt{3}/2$, $f = -1/2$, $g = 1/2$ and $h = \sqrt{3}/2$ then

$$\begin{aligned}R_{\pi/2}(x, y) &= (ax + by, cx + dy)\\ R_{\pi/6}(x, y) &= (ex + fy, gx + hy).\end{aligned}$$

So, by the computation from Lemma 2.8,

$$\begin{aligned}R_{\pi/2} \circ R_{\pi/6} &= ((ae + cf)x + (be + df)y, (ag + ch)x + (bg + dh)y)\\ &= ((1)(-1/2)x + (-1)(\sqrt{3}/2)y, (1)(\sqrt{3}/2)x + (-1)(1/2)y)\\ &= (-x/2 - y\sqrt{3}/2, x\sqrt{3}/2 - y/2),\end{aligned}$$

which is exactly what we found in Example 1.14.

Lemma 2.10. *Let $F\colon \mathbb{R}^n \to \mathbb{R}^m$ be a linear map. Then for any $\vec{v}, \vec{w}$ in $\mathbb{R}^n$ and λ in $\mathbb{R}$,*

- $F(\vec{v}+\vec{w}) = F(\vec{v}) + F(\vec{w})$ *and*
- $F(\lambda\vec{v}) = \lambda F(\vec{v})$.

Proof. Again, to keep notation simple, we will just prove the lemma for maps $\mathbb{R}^2 \to \mathbb{R}^2$. Suppose $F(x,y) = (ax+by, cx+dy)$. Let $\vec{v} = (r,s)$ and $\vec{w} = (t,u)$. Then

$$\begin{aligned} F(\vec{v}+\vec{w}) &= F(r+t, s+u) \\ &= (a(r+t)+b(s+u), c(r+t)+d(s+u)) \\ &= (ar+bs, cr+ds) + (at+bu, ct+du) \\ &= F(\vec{v}) + F(\vec{w}) \\ F(\lambda\vec{v}) &= F(\lambda r, \lambda s) \\ &= (a\lambda r + b\lambda s, c\lambda r + d\lambda s) \\ &= \lambda(ar+bs, cr+ds) \\ &= \lambda F(\vec{v}), \end{aligned}$$

as desired. □

Example 2.11. The map F from Example 1.8 is *not* linear. The form we wrote it in is certainly not that of Definition 2.5. But this doesn't necessarily mean F *can not* be written in the form of Definition 2.5. To see that F can not be written in that form, we use Lemma 2.10. If we take $\vec{v} = (1, \pi/2, 0)$ and $\lambda = 2$ then

$$\begin{aligned} F(\lambda\vec{v}) &= F(2,\pi,0) = (2\cos(\pi), 2\sin(\pi), 0) = (-2,0,0) \\ \lambda F(\vec{v}) &= 2F(1,\pi/2,0) = 2(\cos(\pi/2), \sin(\pi/2), 0) = 2(0,1,0) = (0,2,0). \end{aligned}$$

So, $F(\lambda\vec{v}) \neq \lambda F(\vec{v})$ so F is not linear.

Example 2.12. The function $f(x) = x^2\colon \mathbb{R} \to \mathbb{R}$ is not linear: taking $x = 5$ and $\lambda = 2$ we have $f(\lambda x) = 100$ but $\lambda f(x) = 50$.

Example 2.13. The function $f(x) = |x|\colon \mathbb{R} \to \mathbb{R}$ is not linear: taking $\vec{v} = (1)$ and $\vec{w} = (-1)$, $f(\vec{v}+\vec{w}) = f(0) = 0$ but $f(\vec{v}) + f(\vec{w}) = 1+1 = 2$.

Remark 2.14. The converse to Lemma 2.10 is also true; the proof is slightly (but not much) harder. We outline the argument in Challenge Problem 2.17, below.

Exercise 2.15. Which of the following maps are linear? Justify your answers.

(1) $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ defined by $f(x,y) = (-5x-7y, x+y)$.
(2) $f\colon \mathbb{R}^3 \to \mathbb{R}^2$ defined by $f(x,y,z) = (xy+yz, xz+yz)$.
(3) $f\colon \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x,y) = |x| - |y|$.
(4) $f\colon \mathbb{R}^3 \to \mathbb{R}^2$ defined by $f(x,y,z) = (\sin(\pi/7)x + \cos(\pi/7)y, -e^3 z)$.
(5) $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$f(x,y) = \begin{cases} (\frac{x^3+xy^2-yx^2-y^3}{x^2+y^2}, x+y) & \text{if } (x,y) \neq (0,0) \\ (0,0) & \text{if } (x,y) = (0,0). \end{cases}$$

Exercise 2.16. Use the formula from Lemma 2.8 to compute $R_{\pi/6} \circ R_{\pi/6}$.

Challenge Problem 2.17. In this problem we will prove the converse of Lemma 2.10, in the $m = n = 2$ case. Suppose that $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ satisfies $f(\vec{v}+\vec{w}) = f(\vec{v}) + f(\vec{w})$ and $f(\lambda\vec{v}) = \lambda f(\vec{v})$ for any $\vec{v}$, $\vec{w}$ in $\mathbb{R}^2$ and λ in $\mathbb{R}$. Let $(a,c) = f(1,0)$ and $(b,d) = f(0,1)$. Prove that for any (x,y), $f(x,y) = (ax+by, cx+dy)$.

3. Matrices

Let's look again at the form of a linear map F from Definition 2.1:

$$F(x,y) = (ax+by, cx+dy).$$

If we want to specify such a map, all we need to do is specify a, b, c and d: the x and the y are just placeholders. So, we could record this data by writing:

$$[F] = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

This is a *matrix*, i.e., a rectangular array of numbers. The main point to keep in mind is that **a matrix is just a shorthand for a linear map**.[1]

If we are not just interested in linear maps $\mathbb{R}^2 \to \mathbb{R}^2$, our matrices will not be 2×2. For a linear map $F\colon \mathbb{R}^n \to \mathbb{R}^m$ given by

$$\begin{aligned} &F(x_1,\dots,x_n) \\ &\qquad = (a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,n}x_n, a_{2,1}x_1 + a_{2,2}x_2 + \cdots + a_{2,n}x_n, \dots, \\ &\qquad\qquad a_{m,1}x_1 + a_{m,2}x_2 + \cdots + a_{m,n}x_n), \end{aligned}$$

as in Definition 2.5, the corresponding matrix is

$$[F] = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{pmatrix}.$$

Notice that this matrix has m rows and n columns. Again: the matrix for a linear map $F\colon \mathbb{R}^n \to \mathbb{R}^m$ is an $m \times n$ matrix. (Notice also that I am writing $[F]$ to denote the matrix for F.)

Example 3.1. Suppose P is the projection map from Example 1.4. Then the matrix for P is

$$[P] = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Notice that $[P]$ is a 2×3 matrix, and P maps $\mathbb{R}^3 \to \mathbb{R}^2$.

Example 3.2. If R is the rotation map from Example 1.2 then the matrix for R is

$$[R] = \begin{pmatrix} \cos(\pi/6) & -\sin(\pi/6) & 0 \\ \sin(\pi/6) & \cos(\pi/6) & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \sqrt{3}/2 & -1/2 & 0 \\ 1/2 & \sqrt{3}/2 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Notice that $[R]$ is a 3×3 matrix and R maps $\mathbb{R}^3 \to \mathbb{R}^3$.

[1]There are also other uses of matrices, but their use as stand-ins for linear maps is their most important use, and the only way we will use them.

For this matrix shorthand to be useful, we should find a way to recover $F(\vec{v})$ from the matrix for F. You do this as follows: the i^{th} entry of $F(\vec{v})$ is the dot product of the i^{th} row of $[F]$ with the vector $\vec{v}$.

Example 3.3. Suppose $F(x, y) = (x + 2y, 3x + 4y)$. So, the matrix for F is

$$[F] = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}.$$

To compute $F(5, 6)$, the first entry is the dot product $(1, 2) \cdot (5, 6)$ of the first row of $[F]$ with the vector $(5, 6)$. The second entry is $(3, 4) \cdot (5, 6)$. That is,

$$F(5, 6) = ((1, 2) \cdot (5, 6), (3, 4) \cdot (5, 6)) = (5 + 12, 15 + 24) = (17, 39).$$

Another way of saying this will be helpful later on (Section 5). Write $\vec{v}$ as a *column vector*. That is, instead of writing $(1, 2, 3)$ write $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$. Now, the i^{th} entry of $F(\vec{v})$ is obtained by running your left index finger along the i^{th} row of $[F]$ while running your right index finger down $\vec{v}$, multiplying each pair of numbers your fingers touch, and adding up the results. See Figure 3 for the examples

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 5 \\ 6 \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} \sqrt{3}/2 & -1/2 & 0 \\ 1/2 & \sqrt{3}/2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}.$$

(Advice: when doing this, use your fingers; don't try to do it in your head until you have had a lot of practice.)

Exercise 3.4. Write the matrices corresponding to the following linear maps:

(a) $f \colon \mathbb{R}^2 \to \mathbb{R}^2$ defined by $f(x, y) = (2x + 7y, x + 8y)$.
(b) $f \colon \mathbb{R}^2 \to \mathbb{R}^3$ defined by $f(x, y) = (3x + y, 4x + y, x + 5y)$.
(c) $f \colon \mathbb{R}^3 \to \mathbb{R}^3$ defined by $f(x, y, z) = (x + y + z, y + z, z)$.
(d) $f \colon \mathbb{R} \to \mathbb{R}^3$ defined by $f(x) = (3x, 7x, 13x)$.
(e) $f \colon \mathbb{R}^3 \to \mathbb{R}$ defined by $f(x, y, z) = 3x + 7y + 13z$.

Exercise 3.5. Write the linear maps $F \colon \mathbb{R}^n \to \mathbb{R}^m$ corresponding to the following matrices. For each, say what n and m are:

$$\underset{(a)}{\begin{pmatrix} 1 & 2 \\ 4 & 3 \end{pmatrix}} \quad \underset{(b)}{\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}} \quad \underset{(c)}{\begin{pmatrix} -1 & 3 & -2 & 4 \\ 2 & 1 & 5 & 13 \end{pmatrix}} \quad \underset{(d)}{\begin{pmatrix} -1 & 2 \\ 3 & 1 \\ -2 & 5 \\ 4 & 13 \end{pmatrix}} \quad \underset{(e)}{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}}$$

Exercise 3.6. For the given $[F]$ and $\vec{v}$'s, compute $F(\vec{v})$. (Do this using your fingers.)

(1) $[F] = \begin{pmatrix} 3 & 1 \\ 4 & 1 \end{pmatrix}$, $\vec{v} = \begin{pmatrix} 2 \\ 5 \end{pmatrix}$.

(2) $[F] = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$, $\vec{v} = \begin{pmatrix} -1 \\ 0 \\ 2 \end{pmatrix}$.

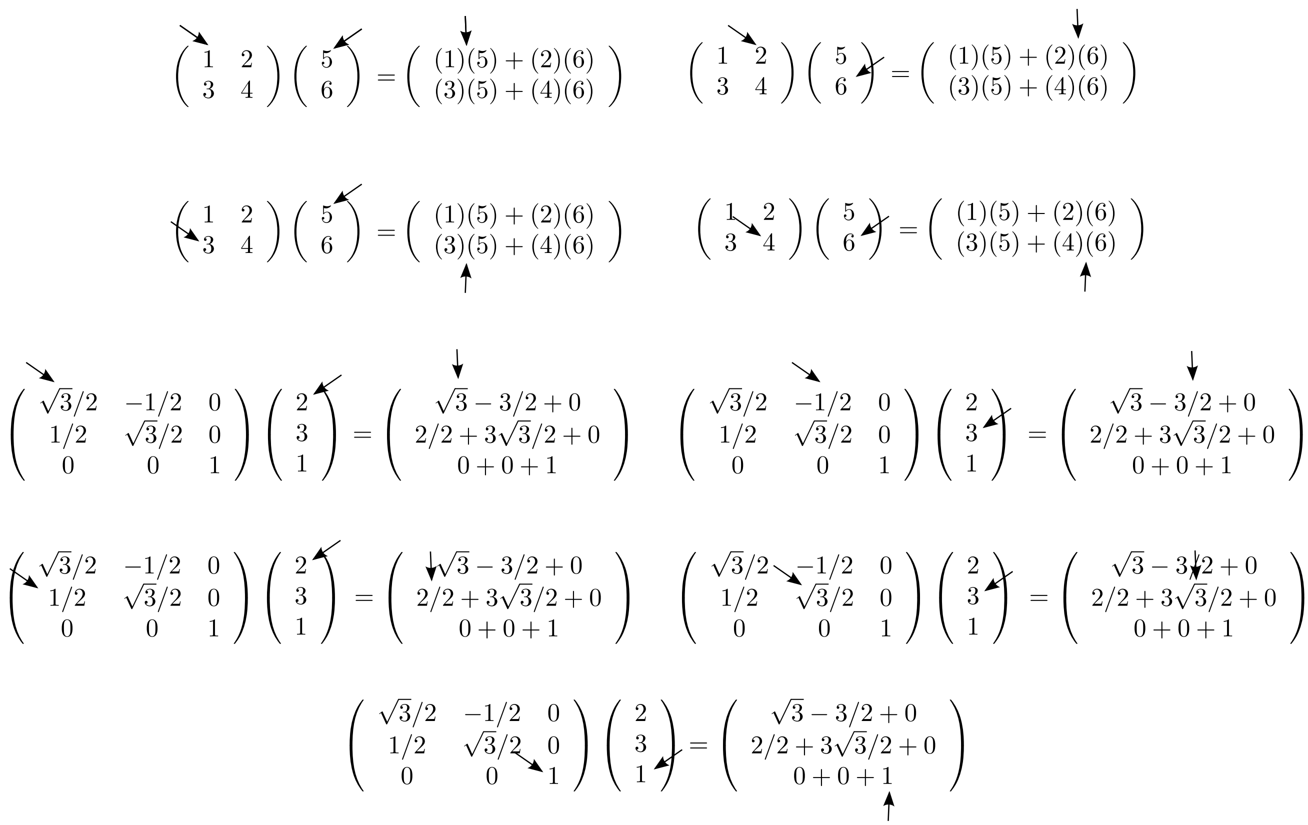

FIGURE 3. **Multiplying a matrix by a vector.** Two examples are shown, with arrows indicating your fingers during the computation. In the second example, only the nonzero terms are marked.

(3) $[F]=\begin{pmatrix}1&2&3\\3&2&1\\1&1&1\end{pmatrix}$, $\vec{v}=\begin{pmatrix}-1\\0\\2\end{pmatrix}$.

(4) $[F]=\begin{pmatrix}1&2\\3&4\\5&6\end{pmatrix}$, $\vec{v}=\begin{pmatrix}-1\\-2\end{pmatrix}$.

(5) $[F]=\begin{pmatrix}0&-1\\1&0\end{pmatrix}$, $\vec{v}=\begin{pmatrix}3\\5\end{pmatrix}$

4. The total derivative and the Jacobian matrix

The total derivative of a map $F\colon \mathbb{R}^n \to \mathbb{R}^m$ at a point $\vec{p}$ in $\mathbb{R}^n$ is the best linear approximation to F near $\vec{p}$. We will make this precise in Section 4.2. First, though, we review the various kinds of linear approximations we have seen already.

4.1. **Review of the derivative as linear approximation.** Suppose that $f\colon \mathbb{R} \to \mathbb{R}$ is a differentiable function. Then near any point $a \in \mathbb{R}$ we can approximate $f(x)$ using $f(a)$ and

$f'(a)$:

$$f(a+h) \approx f(a) + f'(a)h. \tag{4.1}$$

This is a good approximation in the sense that if we let

$$\epsilon(h) = f(a+h) - f(a) - f'(a)h$$

denote the error in the approximation (4.1) then $\epsilon(h)$ goes to zero faster than linearly:

$$\lim_{h\to 0} \epsilon(h)/h = 0.$$

Similarly, suppose $F\colon \mathbb{R} \to \mathbb{R}^n$ is a parametric curve in $\mathbb{R}^n$. If F is differentiable at some a in $\mathbb{R}$ then

$$F(a+h) \approx F(a) + F'(a)h.$$

This is now a vector equation; the $+$ denotes vector addition and $F'(a)h$ is scalar multiplication of the vector $F'(a)$ by the real number h. The sense in which this is a good approximation is the same as before: the error goes to 0 faster than linearly. In symbols:

$$\lim_{h\to 0} \frac{F(a+h) - F(a) - F'(a)h}{h} = \vec{0}.$$

(Again, this is a vector equation now.)

As a final special case, consider a function $F\colon \mathbb{R}^n \to \mathbb{R}$. To keep notation simple, let's consider the case $n = 2$. Then for any $\vec{a} = (a_1, a_2)$ in $\mathbb{R}^2$,

$$F(\vec{a} + (h_1, h_2)) \approx F(\vec{a}) + \frac{\partial F}{\partial x}h_1 + \frac{\partial F}{\partial y}h_2. \tag{4.2}$$

Here, the $\approx$ means that

$$\lim_{(h_1,h_2)\to(0,0)} \frac{F(\vec{a} - (h_1, h_2)) - F(\vec{a}) - \frac{\partial F}{\partial x}h_1 - \frac{\partial F}{\partial y}h_2}{\sqrt{h_1^2 + h_2^2}} = 0.$$

Notice that we could rewrite Equation (4.2) as

$$F(\vec{a} + \vec{h}) \approx F(\vec{a}) + (\nabla F) \cdot \vec{h}.$$

The dot product $(\nabla F) \cdot \vec{h}$ term looks a lot like matrix multiplication. Indeed, if we define $D_{\vec{a}}F$ to be the linear map with matrix

$$[D_{\vec{a}}F] = \begin{pmatrix} \frac{\partial F}{\partial x} & \frac{\partial F}{\partial y} \end{pmatrix}$$

then Equation (4.2) becomes

$$F(\vec{a} + \vec{h}) \approx F(\vec{a}) + (D_{\vec{a}}F)(\vec{h}).$$

Thus inspired...

4.2. **The total derivative of a function $\mathbb{R}^n \to \mathbb{R}^m$.**

Definition 4.3. Let $F\colon \mathbb{R}^n \to \mathbb{R}^m$ be a map and let $\vec{a}$ be a point in $\mathbb{R}^n$. We say F is *differentiable at* $\vec{a}$ if there is a linear map $L\colon \mathbb{R}^2 \to \mathbb{R}^2$ so that

$$F(\vec{a}+\vec{h}) \approx F(\vec{a}) + L(\vec{h})$$

in the sense that

$$\lim_{\vec{h}\to\vec{0}} \frac{F(\vec{a}+\vec{h}) - F(\vec{a}) - L(\vec{h})}{\|\vec{h}\|} = \vec{0}.$$

In this case, we say that L is the *total derivative of* F *at* $\vec{a}$ and write $DF(\vec{a})$ to denote L.

Example 4.4. If $n = m = 1$, so F is a function $\mathbb{R} \to \mathbb{R}$, then Definition 4.3 agrees with the usual notion, and $DF(a)$ is the map $(DF(a))(h) = F'(a)h$. This follows from the discussion in Section 4.1.

Example 4.5. If $n = 1$, so F is a function $\mathbb{R} \to \mathbb{R}^m$ then Definition 4.3 agrees with the usual notion, and $DF(a)$ is the map $(DF(a))(h) = hF'(a)$ (which is a vector in $\mathbb{R}^m$). Again, this follows from the discussion in Section 4.1.

Example 4.6. If $m = 1$, so F is a function $\mathbb{R}^n \to \mathbb{R}$ then Definition 4.3 agrees with the usual notion (Stewart, Section 14.4), and the map $DF(\vec{a})$ is given by

$$DF(\vec{a})(h_1,\dots,h_n) = \frac{\partial F}{\partial x_1}(\vec{a})h_1 + \cdots + \frac{\partial F}{\partial x_n}(\vec{a})h_n.$$

Again, this follows from the discussion in Section 4.1 (or, in other words, immediately from the definition).

Example 4.7. Consider the function $F\colon \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$F(x,y) = (x+y^2, x^3+5y).$$

Then the derivative at $(1,1)$ of F is the map

$$L(h_1,h_2) = (h_1+2h_2, 3h_1+5h_2).$$

To see this, we must verify that

$$\lim_{(h_1,h_2)\to(0,0)} \frac{F(1+h_1,1+h_2) - F(1,1) - L(h_1,h_2)}{\sqrt{h_1^2+h_2^2}} = \vec{0}.$$

But:

$$\begin{aligned}
&\lim_{(h_1,h_2)\to(0,0)} \frac{F(1+h_1,1+h_2) - F(1,1) - L(h_1,h_2)}{\sqrt{h_1^2+h_2^2}}\\
&\quad= \lim_{(h_1,h_2)\to(0,0)} \frac{\big((1+h_1)+(1+h_2)^2, (1+h_1)^3+5(1+h_2)\big) - (2,6) - (h_1+2h_2, 3h_1+5h_2)}{\sqrt{h_1^2+h_2^2}}\\
&\quad= \lim_{(h_1,h_2)\to(0,0)} \frac{(h_2^2, 3h_1^2+h_1^3)}{\sqrt{h_1^2+h_2^2}}\\
&\quad= \left(\lim_{(h_1,h_2)\to(0,0)} \frac{h_2^2}{\sqrt{h_1^2+h_2^2}}, \lim_{(h_1,h_2)\to(0,0)} \frac{3h_1^2+h_1^3}{\sqrt{h_1^2+h_2^2}}\right)\\
&\quad= (0,0),
\end{aligned}$$

as we wanted.

Example 4.8. For $F(x,y)=(x+2y,3x+4y)$ and any $\vec{a}\in\mathbb{R}^2$,

$$DF(\vec{a})(h_1,h_2)=(h_1+2h_2,3h_1+4h_2).$$

To see this, write $\vec{a}=(a_1,a_2)$. Then,

$$\begin{aligned}
\lim_{(h_1,h_2)\to(0,0)} & \frac{F(a_1+h_1,a_2+h_2)-F(a_1,a_2)-(h_1+2h_2,3h_1+4h_2)}{\sqrt{h_1^2+h_2^2}} \\
&= \lim_{(h_1,h_2)\to(0,0)} \big[(a_1+h_1+2a_2+2h_2,3a_1+3h_1+4a_2+4h_2)-(a_1+2a_2,3a_1+4a_2) \\
&\qquad -(h_1+2h_2,3h_1+4h_2)\big]/\sqrt{h_1^2+h_2^2} \\
&= \lim_{(h_1,h_2)\to(0,0)} \frac{(0,0)}{\sqrt{h_1^2+h_2^2}} \\
&= (0,0),
\end{aligned}$$

as desired.

More generally, if F is a linear map then for any vector $\vec{a}$, $DF(\vec{a})=F$. This makes sense: if F is linear then the best linear approximation to F is F itself.

We have been calling $DF(\vec{a})$ *the* derivative of F at $\vec{a}$, but *a priori* there might be more than one. As you might suspect, this is not the case:

Lemma 4.9. *Let $F\colon\mathbb{R}^n\to\mathbb{R}^m$ be a map and let $\vec{a}$ be a point in $\mathbb{R}^n$. Suppose that L and M are both linear maps such that*

$$\lim_{\vec{h}\to\vec{0}}\frac{F(\vec{a}+\vec{h})-F(\vec{a})-L(\vec{h})}{\|\vec{h}\|}=\vec{0} \qquad \textit{and}$$
$$\lim_{\vec{h}\to\vec{0}}\frac{F(\vec{a}+\vec{h})-F(\vec{a})-M(\vec{h})}{\|\vec{h}\|}=\vec{0}.$$

Then $L=M$.

Proof. If L and M are different linear maps then there is some vector $\vec{k}$ so that $L(\vec{k})\neq M(\vec{k})$. Let's consider taking the limit $\vec{h}\to 0$ along the line $\vec{h}=t\vec{k}$. We have

$$\begin{aligned}
0 &= \lim_{t\to 0}\frac{F(\vec{a}+t\vec{k})-F(\vec{a})-L(t\vec{k})}{\|t\vec{k}\|} \\
&= \lim_{t\to 0}\left(\frac{F(\vec{a}+t\vec{k})-F(\vec{a})-M(t\vec{k})}{\|t\vec{k}\|}+\frac{L(t\vec{k})-M(t\vec{k})}{\|t\vec{k}\|}\right) \\
&= 0+\lim_{t\to 0}\frac{tL(\vec{k})-tM(\vec{k})}{|t|\|\vec{k}\|}.
\end{aligned}$$

But the limit on the right hand side does not exist: if $t>0$ then we get $(L(\vec{k})-M(\vec{k}))/\|\vec{k}\|$ while if $t<0$ we get $-(L(\vec{k})-M(\vec{k}))/\|\vec{k}\|$.

So, our supposition that $L(\vec{k})\neq M(\vec{k})$ must have been false. ☐

4.3. **The Jacobian matrix.** Our definition of the total derivative should seem like a useful one, and it generalizes the cases we already had, but some questions remain. Chief among them:

(1) How do you tell if a function $\mathbb{R}^n \to \mathbb{R}^m$ is differentiable? (And, are many of the functions one encounters differentiable?)
(2) How do you compute the total derivative of a function $\mathbb{R}^n \to \mathbb{R}^m$?

The next theorem answers both questions; we will state it and then unpack it in some examples.

Theorem 1. *Suppose that* $F\colon \mathbb{R}^n \to \mathbb{R}^m$. *Write* $F = (f_1, \dots, f_m)$, *where* $f_i\colon \mathbb{R}^n \to \mathbb{R}$. *If for all* i *and* j, $\partial f_i/\partial x_j$ *is (defined and) continuous near* $\vec{a}$ *then* F *is differentiable at* $\vec{a}$, *and the matrix for* $DF(\vec{a})$ *is given by*

$$[DF(\vec{a})] = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}.$$

To avoid being sidetracked, we wont prove this theorem; its proof is similar to the proof of Stewart's Theorem 8 in Section 14.4 (proved in Appendix F). (It is fairly easy to see that if F is differentiable at $\vec{a}$ then $[DF(\vec{a})]$ has the specified form. Slightly harder is to show that if all of the partial derivatives of the components of F are continuously differentiable then F is differentiable.)

The matrix $[DF(\vec{a})]$ is called the *total derivative matrix* or *Jacobian matrix* of F at $\vec{a}$.

Example 4.10. For the function $F(x, y) = (x + y^2, x^3 + 5y)$ from Example 4.7, the Jacobian matrix at $(1, 1)$ is

$$[DF((1,1))] = \begin{pmatrix} 1 & 2 \\ 3 & 5 \end{pmatrix}.$$

(Why?) This is, indeed, the matrix associated to the linear map from Example 4.7.

Example 4.11. Suppose F is the "turn cylindrical coordinates into rectangular coordinates" map from Example 1.8. Then

$$[DF(5, \pi/3, 0)] = \begin{pmatrix} \cos(\theta) & -r\sin(\theta) & 0 \\ \sin(\theta) & r\cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix}_{r=5,\theta=\pi/3,z=0} = \begin{pmatrix} 1/2 & -5\sqrt{3}/2 & 0 \\ \sqrt{3}/2 & 5/2 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Example 4.12. *Problem.* Suppose $F\colon \mathbb{R}^2 \to \mathbb{R}^2$ is a differentiable map, $F(1, 1) = (5, 8)$ and

$$[DF(1,1)] = \begin{pmatrix} 1 & 1 \\ 2 & 3 \end{pmatrix}.$$

Estimate $F(1.1, 1.2)$.

Solution. Since F is differentiable,

$$F((1,1) + (.1,.2)) \approx F(1,1) + DF(1,1)(.1,.2).$$

So,

$$F((1,1)+(.1,.2)) \approx \begin{pmatrix}5\\8\end{pmatrix}+\begin{pmatrix}1&1\\2&3\end{pmatrix}\begin{pmatrix}.1\\.2\end{pmatrix}$$
$$=\begin{pmatrix}5\\8\end{pmatrix}+\begin{pmatrix}.3\\.8\end{pmatrix}=\begin{pmatrix}5.3\\8.8\end{pmatrix}.$$

Of course, we don't know whether this is a good estimate or not: it depends whether $(.1,.2)$ is a big or a small change for F.

Exercise 4.13. Compute the Jacobian matrices of the given maps at the given points:

(1) The map $F\colon \mathbb{R}^2 \to \mathbb{R}^2$ defined by $F(x,y)=(x^2+y^3,xy)$ at $(-1,1)$.
(2) The map $F\colon \mathbb{R}^2 \to \mathbb{R}^3$ defined by $F(x,y)=(xe^y,ye^{2x},x^2+y^2)$ at $(1,1)$.
(3) The map $F\colon \mathbb{R}^3 \to \mathbb{R}$ defined by $F(x,y,z)=x^2+y^2+z^2$ at $(1,2,3)$.
(4) The map G from Example 1.8,
$$G(\rho,\theta,\phi)=(\rho\cos(\theta)\sin(\phi),\rho\sin(\theta)\sin(\phi),\rho\cos(\phi)),$$
at the point $(1,\pi/2,\pi/2)$.
(5) The map G from Example 1.8,
$$G(\rho,\theta,\phi)=(\rho\cos(\theta)\sin(\phi),\rho\sin(\theta)\sin(\phi),\rho\cos(\phi)),$$
at the point $(1,\pi/2,0)$.

Exercise 4.14. Let $F(r,\theta)=(r\cos(\theta),r\sin(\theta))$. (This map takes a vector in polar coordinates and rewrites it in rectangular coordinates.) Compute $F(1,\pi/4)$ and $[DF(1,\pi/4)]$. Use your computation to estimate $F(1.01,\pi/4+.02)$. Check your answer is reasonable using a calculator.

Exercise 4.15. Consider the map $F\colon \mathbb{C}\to\mathbb{C}$ defined by $F(z)=z^2$. We can identify $\mathbb{C}$ with $\mathbb{R}^2$ by identifying $z=x+iy$ in $\mathbb{C}$ with (x,y) in $\mathbb{R}^2$. Then, F is a map $\mathbb{R}^2\to\mathbb{R}^2$ given by

$$F(x,y)=(x^2-y^2,2xy).$$

Compute $[DF]$, $[DF(1,2)]$ and $[DF(0,0)]$.

Exercise 4.16. Consider the map $F\colon \mathbb{C}\to\mathbb{C}$ given by $F(z)=z^3+2z$. As in Exercise 4.15, we can view F as a map $\mathbb{R}^2\to\mathbb{R}^2$. Write F as a function $F(x,y)$ and compute its Jacobian matrix. Do you notice any symmetry in the matrix?

Exercise 4.17. Consider the map $F\colon \mathbb{C}\to\mathbb{C}$ given by $F(z)=e^z$. As in Exercise 4.15, we can view F as a map $\mathbb{R}^2\to\mathbb{R}^2$. Write F as a function $F(x,y)$ and compute its Jacobian matrix. (You will want to use Euler's formula.) Do you notice any symmetry in the matrix?

5. Composition of linear maps and matrix multiplication

We showed in Lemma 2.8 that the composition of linear maps is a linear map. How does this relate to matrices? That is, if $F\colon\mathbb{R}^p\to\mathbb{R}^n$ and $G\colon\mathbb{R}^n\to\mathbb{R}^m$ are linear maps, how does $[G\circ F]$ relate to $[G]$ and $[F]$? The answer is that $[G\circ F]$ is the product of $[G]$ and $[F]$.

Given a $m\times n$ and $n\times p$ matrices

$$A=\begin{pmatrix}a_{1,1}&a_{1,2}&\cdots&a_{1,n}\\a_{2,1}&a_{2,2}&\cdots&a_{2,n}\\\vdots&\vdots&\ddots&\vdots\\a_{m,1}&a_{m,2}&\cdots&a_{m,n}\end{pmatrix}\qquad\text{and}\qquad B=\begin{pmatrix}b_{1,1}&b_{1,2}&\cdots&b_{1,p}\\b_{2,1}&b_{2,2}&\cdots&b_{2,p}\\\vdots&\vdots&\ddots&\vdots\\b_{n,1}&b_{n,2}&\cdots&b_{n,p}\end{pmatrix},$$

$$\begin{pmatrix}1&2&3\\5&6&7\end{pmatrix}\begin{pmatrix}1&1\\2&3\\5&8\end{pmatrix}=\begin{pmatrix}1\cdot1+2\cdot2+3\cdot5 & 1\cdot1+2\cdot3+3\cdot8\\5\cdot1+6\cdot2+7\cdot5 & 5\cdot1+6\cdot3+7\cdot8\end{pmatrix}=\begin{pmatrix}20&31\\52&79\end{pmatrix}$$

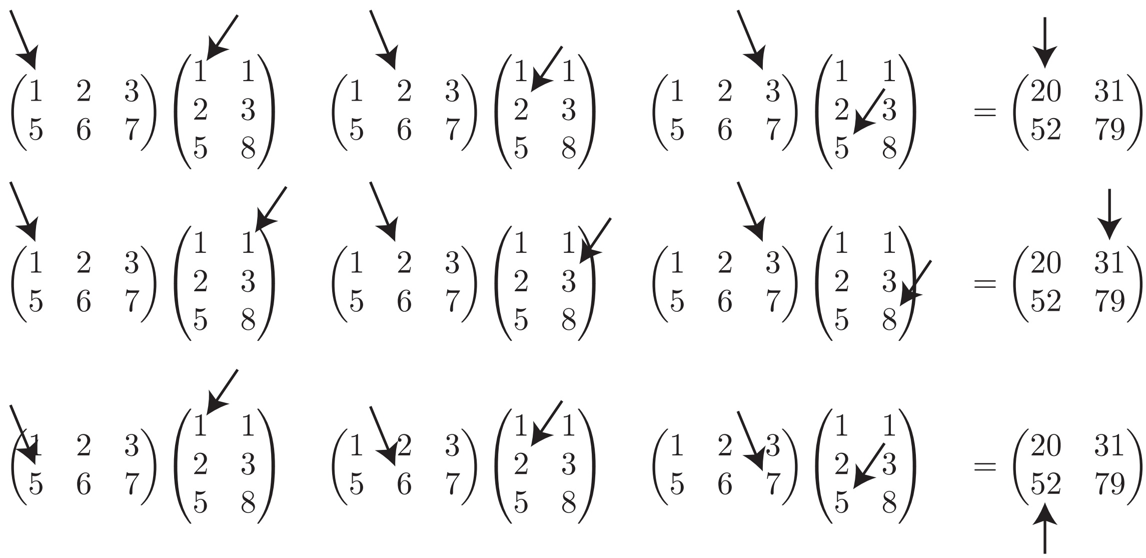

FIGURE 4. **Multiplying matrices.** Arrows indicate what your fingers should do when computing three of the four entries.

the matrix product AB of A and B is the $m \times p$ matrix whose (i, j) entry[2] is

$$c_{i,j} = a_{i,1}b_{1,j} + a_{i,2}b_{2,j} + \cdots + a_{i,n}b_{n,j}.$$

That is, to find the (i, j) entry of AB you run your left index finger across the i^{th} row of A and your right index finger down the j^{th} column of B. Multiply the pairs of numbers your fingers touch simultaneously and add the results. See Figure 4.

Example 5.1. A few example of matrix products:

$$\begin{pmatrix}2&3\\0&4\end{pmatrix}\begin{pmatrix}-1&1\\-2&2\end{pmatrix}=\begin{pmatrix}-2-6&2+6\\0-8&0+8\end{pmatrix}=\begin{pmatrix}-8&8\\-8&8\end{pmatrix}.$$

$$\begin{pmatrix}1&2&3\\4&5&6\end{pmatrix}\begin{pmatrix}1&3\\5&7\\9&11\end{pmatrix}=\begin{pmatrix}1+10+27&3+14+33\\4+25+54&12+35+66\end{pmatrix}=\begin{pmatrix}38&50\\83&113\end{pmatrix}.$$

$$\begin{pmatrix}1&3\\5&7\\9&11\end{pmatrix}\begin{pmatrix}1&2&3\\4&5&6\end{pmatrix}=\begin{pmatrix}1+12&2+15&3+18\\5+28&10+35&15+42\\9+44&18+55&27+66\end{pmatrix}=\begin{pmatrix}13&17&21\\33&45&57\\53&73&93\end{pmatrix}.$$

(By the way, I did in fact use my fingers when computing these products.)

Notice that the product of a matrix with a column vector as in Section 3 is a special case of multiplication of two matrices.

[2] i.e., the entry in the i^{th} row and j^{th} column

Theorem 2. *If* $F\colon \mathbb{R}^p \to \mathbb{R}^n$ *and* $G\colon \mathbb{R}^n \to \mathbb{R}^m$ *are linear maps then* $[G \circ F] = [G][F]$*.*

Proof. We will only do the case $m = n = p = 2$; the general case is similar (but with indices). Write

$$\begin{aligned} F(x,y) &= (ax+by, cx+dy) \\ G(x,y) &= (ex+fy, gx+hy), \end{aligned}$$

so on the one hand we have

$$\begin{aligned} G \circ F(x,y) &= ((ae+cf)x + (be+df)y, (ag+ch)x + (bg+dh)y), \\ [G \circ F] &= \begin{pmatrix} ae+cf & be+df \\ ag+ch & bg+dh \end{pmatrix}, \end{aligned}$$

and on the other hand we have

$$\begin{aligned} [G] &= \begin{pmatrix} e & f \\ g & h \end{pmatrix} \\ [F] &= \begin{pmatrix} a & b \\ c & d \end{pmatrix} \\ [G][F] &= \begin{pmatrix} ae+cf & be+df \\ ag+ch & bg+dh \end{pmatrix}. \end{aligned}$$

So, indeed, $[G][F] = [G \circ F]$, as desired. ☐

Corollary 5.2. *The matrix product is associative. That is, if* A *is an* $m \times n$ *matrix,* B *is an* $n \times p$ *matrix and* C *is a* $p \times r$ *matrix then* $(AB)C = A(BC)$*.*

Proof. The matrix product corresponds to composition of functions, and composition of functions is (obviously) associative.

(We could also verify this directly, by writing out all of the sums—but they become painfully complicated.) ☐

Notice that we have only defined the matrix product AB when the width of A is the same as the height of B. If the width of A is not the same as the height of B then AB is simply not defined. This corresponds to the fact that given maps $F\colon \mathbb{R}^q \to \mathbb{R}^p$ and $G\colon \mathbb{R}^n \to \mathbb{R}^m$, $G \circ F$ only makes sense if $p = n$.

Example 5.3. The product

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 7 & 8 & 9 \\ 10 & 11 & 12 \end{pmatrix}$$

is not defined: the first matrix has 3 columns but the second has only 2 rows. This corresponds to the fact that you can't compose a map $\mathbb{R}^3 \to \mathbb{R}^2$ with another map $\mathbb{R}^3 \to \mathbb{R}^2$.

If you use your fingers, it's obvious when a matrix product isn't defined: one finger runs out of entries before the other.

Example 5.4. Even when the products in both orders are defined, matrix multiplication is typically non-commutative. For example:

$$\begin{aligned} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} &= \begin{pmatrix} -2 & 1 \\ -4 & 3 \end{pmatrix} \\ \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} &= \begin{pmatrix} 3 & 4 \\ -1 & -2 \end{pmatrix}. \end{aligned}$$

(Actually, we had another example of this phenomenon in Example 5.1.)

Example 5.5. The dot product can be thought of as a special case of matrix multiplication. For example:

$$\begin{pmatrix} 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 2 \\ 2 \\ 4 \end{pmatrix} = (2 + 4 + 12)$$

is the dot product of the vectors $\vec{v} = \langle 1, 2, 3\rangle$ and $\vec{w} = \langle 2, 2, 4\rangle$. The trick is to write the first vector as a row vector and the second as a column vector.

Exercise 5.6. For each of the following, say whether the matrix product AB is defined, and if so compute it. Say also whether BA is defined, and if so compute it.

(1) $A = \begin{pmatrix} 3 & 2 \\ 1 & 0 \end{pmatrix}$, $B = \begin{pmatrix} 2 & 4 \\ 6 & -2 \end{pmatrix}$.
(2) $A = \begin{pmatrix} 3 & 2 \\ 1 & 0 \end{pmatrix}$, $B = \begin{pmatrix} 5 & 3 & 1 \\ 2 & 2 & 4 \end{pmatrix}$.
(3) $A = B = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$
(4) $A = \begin{pmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 4 \end{pmatrix}$, $B = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$
(5) $A = \begin{pmatrix} 1 & 2 & 3 \end{pmatrix}$, $B = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$

5.1. **Matrix arithmetic.** As something of an aside, we should mention that you can also add matrices: if A and B are both $n \times m$ matrices then $A + B$ is obtained by adding corresponding entries of A and B. For example:

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} + \begin{pmatrix} 2 & 2 & 3 \\ 3 & 4 & 4 \end{pmatrix} = \begin{pmatrix} 3 & 4 & 6 \\ 7 & 9 & 10 \end{pmatrix}.$$

There are two special (families of) matrices for matrix arithmetic. The $n \times n$ identity matrix is the matrix with 1's on it's diagonal:

$$I_{n\times n} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}.$$

The matrix $I_{n\times n}$ represents the identity map $\mathbb{I}\colon \mathbb{R}^n \to \mathbb{R}^n$ given by $\mathbb{I}(\vec{x}) = \vec{x}$. (This takes a little checking.)

The other special family are the $m \times n$ zero matrices, $0_{m\times n}$, all of whose entries are 0:

$$0_{m\times n} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}.$$

We might drop the subscripts $n \times n$ or $m \times n$ from $I_{n\times n}$ or $0_{m\times n}$ when the dimensions are either obvious or irrelevant.

The properties of matrix arithmetic are summarized as follows.

Lemma 5.7. *Suppose A, B, C and D are $n\times m$, $n\times m$, $m\times l$ and $p\times n$ matrices, respectively. Then:*

(1) *(Addition is commutative:)* $A + B = B + A$.
(2) *(Additive identity:)* $0_{n\times m} + A = A$.
(3) *(Multiplicative identity:)* $I_{n\times n}A = A = AI_{m\times m}$.
(4) *(Multiplicative zero:)* $A0_{m\times k} = 0_{n\times k}$. $0_{p\times n}A = 0_{p\times m}$.
(5) *(Distributivity:)* $(A + B)C = AC + BC$ *and* $D(A + B) = DA + DB$.

(All of the properties are easy to check directly, and also follow from corresponding properties of the arithmetic of linear maps.)

Exercise 5.8. Square the matrix $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, i.e., multiply A by itself. Explain geometrically why the answer makes sense. (Hint: what linear map does A correspond to? What happens if you do that map twice?)

Exercise 5.9. Square the matrix $B = \begin{pmatrix} 2 & -4 \\ 1 & -2 \end{pmatrix}$. Does anything surprise you about the answer?

Challenge Problem 5.10. Find a matrix A so that $A \neq I$ but $A^3 = I$. (Hint: think geometrically.)

6. The chain rule for total derivatives

With the terminology we have developed, the chain rule is very succinct:

$$D(G \circ F) = (DG) \circ (DF).$$

More precisely:

Theorem 3. *Let $F\colon \mathbb{R}^p \to \mathbb{R}^n$ and $G\colon \mathbb{R}^n \to \mathbb{R}^m$ be maps so that F is differentiable at $\vec{a}$ and G is differentiable at $F(\vec{a})$. Then*

$$D(G \circ F)(\vec{a}) = DG(F(\vec{a})) \circ DF(\vec{a}). \tag{6.1}$$

In particular, the Jacobian matrices satisfy

$$[D(G \circ F)(\vec{a})] = [DG(F(\vec{a}))][DF(\vec{a})].$$

We will prove the theorem a little later. First, some examples and remarks.

Notice that Equation (6.1) looks a lot like the chain rule for functions of one variable, $(g \circ f)'(x) = f'(x)g'(f(x))$. Of course, the one-variable chain rule is a (very) special case of Theorem 3, so this isn't a total surprise. But it's nice that we have phrased things, and chosen notation, so that the similarity manifests itself.

Example 6.2. *Question.* Let $F(x, y) = (x + y^2, x^3 + 5y)$ and $G(x, y) = (y^2, x^2)$. Compute $[D(G \circ F)]$ at $(1, 1)$.

Solution. We already computed in Example 4.10 that

$$[DF(1, 1)] = \begin{pmatrix} 1 & 2 \\ 3 & 5 \end{pmatrix}.$$

Computing. . .

$$\begin{aligned} F(1,1) &= (2,6) \\ [DG] &= \begin{pmatrix} 0 & 2y \\ 2x & 0 \end{pmatrix} \\ [DG(2,6)] &= \begin{pmatrix} 0 & 12 \\ 4 & 0 \end{pmatrix}. \end{aligned}$$

So, by the chain rule,

$$[D(G \circ F)(1,1)] = \begin{pmatrix} 0 & 12 \\ 4 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 5 \end{pmatrix} = \begin{pmatrix} 36 & 60 \\ 4 & 8. \end{pmatrix}$$

As a sanity check, notice that the dimensions are right: $G \circ F$ is a map $\mathbb{R}^2 \to \mathbb{R}^2$, so its Jacobian should be a 2×2 matrix.

Note also that we could compute $[D(G \circ F)]$ by composing G and F first and then differentiating. We should get the same answer; this would be a good check to do.

Example 6.3. *Question.* Let $F(x,y) = (x+y^2, x^3+5y)$ and $G(x,y) = 2xy$. Compute $[D(G \circ F)(1,1)]$.

Solution. As in the previous example, $F(1,1) = (2,6)$ and

$$[DF(1,1)] = \begin{pmatrix} 1 & 2 \\ 3 & 5 \end{pmatrix}.$$

Computing,

$$\begin{aligned} [DG] &= \begin{pmatrix} 2y & 2x \end{pmatrix} \\ [DG(2,6)] &= \begin{pmatrix} 4 & 12 \end{pmatrix}. \end{aligned}$$

So, by the chain rule,

$$[D(G \circ F)(1,1)] = \begin{pmatrix} 4 & 12 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 5 \end{pmatrix} = \begin{pmatrix} 40 & 68 \end{pmatrix}.$$

Example 6.4. *Question.* An ant is crawling on a surface. The temperature at a point (x,y) is $T(x,y) = 70 + 5\sin(x)\cos(y)$. The ant's position at time t is $\gamma(t) = (t\cos(\pi t), t\sin(\pi t))$. How fast is the temperature under the ant changing at $t = 5$?

Solution. We are interested in the derivative of $T \circ \gamma(t)$ at $t = 5$. Computing,

$$\begin{aligned} \gamma(5) &= (5\cos(5\pi), 5\sin(5\pi)) = (-5, 0) \\ [D\gamma] &= \begin{pmatrix} \cos(\pi t) - \pi t \sin(\pi t) \\ \sin(\pi t) + \pi t \cos(\pi t) \end{pmatrix} \\ [D\gamma(5)] &= \begin{pmatrix} -1 \\ -5\pi \end{pmatrix} \\ [DT] &= \begin{pmatrix} 5\cos(x)\cos(y) & -5\sin(x)\sin(y) \end{pmatrix} \\ [DT(-5,0)] &= \begin{pmatrix} 5\cos(5) & 0 \end{pmatrix}. \end{aligned}$$

So, by the chain rule,

$$[D(T \circ \gamma)] = \begin{pmatrix} 5\cos(5) & 0 \end{pmatrix} \begin{pmatrix} -1 \\ -5\pi \end{pmatrix} = (-5\cos(5)).$$

So, the answer is $-5\cos(5)$ (degrees per second, or whatever).

Again, this could be checked directly, though it would be a bit of a pain.

Example 6.5. *Question.* The height of a mountain is given by $H(x,y) = \frac{1}{1+x^2+2y^2}$. Compute $\frac{\partial H}{\partial r}$, the rate of change of H directly away from the origin.

Solution.. Let $F(r,\theta) = (r\cos(\theta), r\sin(\theta))$ be the "turn polar coordinates into rectangular coordinates" map. Then the total derivative of $H \circ F$ is

$$\begin{pmatrix} \frac{\partial(H\circ F)}{\partial r} & \frac{\partial(H\circ F)}{\partial \theta} \end{pmatrix}.$$

The first entry, $\frac{\partial(H\circ F)}{\partial r}$, is what we want to compute. Now, working directly,

$$\begin{aligned}
[DH] &= \begin{pmatrix} \frac{-2x}{(1+x^2+2y^2)^2} & \frac{-4y}{(1+x^2+2y^2)^2} \end{pmatrix} \\
[DH(F(r,\theta))] &= \begin{pmatrix} \frac{-2r\cos(\theta)}{(1+r^2\cos^2(\theta)+2r^2\sin^2(\theta))^2} & \frac{-4r\sin(\theta)}{(1+r^2\cos^2(\theta)+2r^2\sin^2(\theta))^2} \end{pmatrix} \\
[DF] &= \begin{pmatrix} \cos(\theta) & -r\sin(\theta) \\ \sin(\theta) & r\cos(\theta) \end{pmatrix}.
\end{aligned}$$

So,

$$\begin{aligned}
[D(H\circ F)] &= \begin{pmatrix} \frac{-2r\cos(\theta)}{(1+r^2\cos^2(\theta)+2r^2\sin^2(\theta))^2} & \frac{-4r\sin(\theta)}{(1+r^2\cos^2(\theta)+2r^2\sin^2(\theta))^2} \end{pmatrix} \begin{pmatrix} \cos(\theta) & -r\sin(\theta) \\ \sin(\theta) & r\cos(\theta) \end{pmatrix} \\
&= \begin{pmatrix} \frac{-2r\cos^2(\theta)-4r\sin^2(\theta)}{(1+r^2\cos^2(\theta)+2r^2\sin^2(\theta))^2} & \frac{-2r^2\sin(\theta)\cos(\theta)}{(1+r^2\cos^2(\theta)+2r^2\sin^2(\theta))^2} \end{pmatrix}.
\end{aligned}$$

Again, you could check this directly.

We turn now to the proof of Theorem 3. It's a little bit complicated, and you can safely view it as optional.

Proof of Theorem 3. To keep notation simple, let $\vec{b} = F(\vec{a})$, $L = DF(\vec{a})$ and $M = DG(\vec{b})$ (so L and M are linear maps). We must show that $M \circ L$ is a good linear approximation to $G \circ F$ near $\vec{a}$, i.e., that

$$\lim_{\vec{h}\to\vec{0}} \left(G(F(\vec{a}+\vec{h})) - G(\vec{b}) - M(L(\vec{h})) \right) / \|\vec{h}\| = 0.$$

Let

$$\begin{aligned}
\epsilon(\vec{h}) &= F(\vec{a}+\vec{h}) - F(\vec{a}) - L(\vec{h}) \\
\eta(\vec{k}) &= G(\vec{b}+\vec{k}) - G(\vec{b}) - M(\vec{k})
\end{aligned}$$

denote the errors in the linear approximations to F and G. Then

$$\begin{aligned}
G(F(\vec{a}+\vec{h})) - G(\vec{b}) - M(L(\vec{h})) &= G(F(\vec{a}) + L(\vec{h}) + \epsilon(\vec{h})) - G(\vec{b}) - M(L(\vec{h})) \\
&= G(\vec{b} + L(\vec{h}) + \epsilon(\vec{h})) - G(\vec{b}) - M(L(\vec{h})) \\
&= G(\vec{b}) + M(L(\vec{h}) + \epsilon(\vec{h})) + \eta(L(\vec{h}) + \epsilon(\vec{h})) - G(\vec{b}) - M(L(\vec{h})) \\
&= M(\epsilon(\vec{h})) + \eta(L(\vec{h}) + \epsilon(\vec{h})),
\end{aligned}$$

where we used the fact that M is linear to get the cancellation in the last equality.

So,

$$\lim_{\vec{h}\to\vec{0}} \left(G(F(\vec{a}+\vec{h})) - G(\vec{b}) - M(L(\vec{h})) \right) / \|\vec{h}\| = \lim_{\vec{h}\to\vec{0}} \left(M(\epsilon(\vec{h})) + \eta(L(\vec{h}) + \epsilon(\vec{h})) \right) / \|\vec{h}\|.$$

Thus, it's enough to show that

$$\lim_{\vec{h}\to\vec{0}} \frac{M(\epsilon(\vec{h}))}{\|\vec{h}\|} = 0 \qquad \text{and}$$

$$\lim_{\vec{h}\to\vec{0}} \frac{\eta(L(\vec{h})+\epsilon(\vec{h}))}{\|\vec{h}\|} = 0.$$

For the first, notice that there is some constant C, depending on M, so that for any vector $\vec{v}$, $\|M(\vec{v})\| \leq C\|\vec{v}\|$. So,

$$\left\| \frac{M(\epsilon(\vec{h}))}{\|\vec{h}\|} \right\| \leq C \left\| \frac{\epsilon(\vec{h})}{\|\vec{h}\|} \right\| \to 0,$$

as desired.

For the second, if we let $\vec{k} = L(\vec{h}) + \epsilon(\vec{h})$ then there is a constant D so that for any $\vec{h}$, $\|\vec{h}\| \geq D\|\vec{k}\|$. Moreover, as $\vec{h} \to 0$, $\vec{k} \to 0$ as well. So, for $\vec{h}$ so that $\vec{k} \neq 0$,

$$\left\| \frac{\eta(L(\vec{h})+\epsilon(\vec{h}))}{\|\vec{h}\|} \right\| \leq \frac{1}{D} \left\| \frac{\eta(\vec{k})}{\|\vec{k}\|} \right\| \to 0$$

as $\vec{k} \to 0$. For $\vec{h}$ so that $\vec{k} = \vec{0}$, $\frac{\eta(L(\vec{h})+\epsilon(\vec{h}))}{\|\vec{h}\|}$ is exactly zero. So, as $\vec{h} \to 0$, $\eta(L(\vec{h})+\epsilon(\vec{h}))/\|\vec{h}\| \to 0$, as desired.

This proves the result. □

Exercise 6.6. Let $F(x,y) = (x^2y, xy^2)$ and $G(x,y) = (x/y, y/x)$. Use the chain rule to compute $[D(G\circ F)(1,2)]$. Check your answer by composing G and F and then differentiating.

6.1. **Comparison with the treatment in Stewart's *Calculus*.** We already saw several special cases of Theorem 3, in a somewhat disguised form, earlier in the semester. The following is quoted from Stewart, Section 14.5 (box 4):

Theorem 4. *Suppose u is a differentiable function of n variables $x_1, x_2, \dots, x_n$ and each x_j is a differentiable function of m variables $t_1, t_2, \dots, t_m$. Then u is a function of $t_1, t_2, \dots, t_m$ and*

$$\frac{\partial u}{\partial t_i} = \frac{\partial u}{\partial x_1}\frac{\partial x_1}{\partial t_i} + \frac{\partial u}{\partial x_2}\frac{\partial x_2}{\partial t_i} + \cdots + \frac{\partial u}{\partial x_n}\frac{\partial x_n}{\partial t_i}$$

for each $i = 1, 2, \dots, m$.

As a special case, he noted that if z is a function of x and y and x and y are functions of s and t then

$$\frac{\partial z}{\partial s} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s} \qquad \text{and} \qquad \frac{\partial z}{\partial t} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t}.$$

Let's see how the special case fits into our framework, and then how Stewart's "general" case does.

In the special case, we have $z = z(x,y)$. Let $F(s,t) = (x(s,t), y(s,t))$. We are interested in the partial derivatives of $z \circ F$; for example, $\frac{\partial z}{\partial s}$ really means $\frac{\partial(z\circ F)}{\partial s}$. On the one hand,

$$[D(z\circ F)] = \begin{pmatrix} \frac{\partial(z\circ F)}{\partial s} & \frac{\partial(z\circ F)}{\partial t} \end{pmatrix}.$$

On the other hand,

$$[Dz] = \begin{pmatrix} \frac{\partial z}{\partial x} & \frac{\partial z}{\partial y} \end{pmatrix}$$

$$[DF] = \begin{pmatrix} \frac{\partial x}{\partial s} & \frac{\partial x}{\partial t} \\ \frac{\partial y}{\partial s} & \frac{\partial y}{\partial t} \end{pmatrix} \qquad \text{so}$$

$$[Dz][DF] = \begin{pmatrix} \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s} & \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t} \end{pmatrix}.$$

By the chain rule, $[D(z \circ F)] = [Dz] \circ [DF]$, so

$$\frac{\partial(z \circ F)}{\partial s} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s}$$

$$\frac{\partial(z \circ F)}{\partial t} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t},$$

exactly as in Stewart.

The general case is similar. Let $F(t_1, \dots, t_m) = (x_1(t_1, \dots, t_m), \dots, x_n(t_1, \dots, t_m))$. Then,

$$[Du] = \begin{pmatrix} \frac{\partial u}{\partial x_1} & \cdots & \frac{\partial u}{\partial x_n} \end{pmatrix}$$

$$[DF] = \begin{pmatrix} \frac{\partial x_1}{\partial t_1} & \cdots & \frac{\partial x_1}{\partial t_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial t_1} & \cdots & \frac{\partial x_n}{\partial t_m} \end{pmatrix}.$$

So,

$$[Du][DF] = \begin{pmatrix} \frac{\partial u}{\partial x_1}\frac{\partial x_1}{\partial t_1} + \cdots + \frac{\partial u}{\partial x_n}\frac{\partial x_n}{\partial t_1} & \cdots & \frac{\partial u}{\partial x_1}\frac{\partial x_1}{\partial t_m} + \cdots + \frac{\partial u}{\partial x_n}\frac{\partial x_n}{\partial t_m} \end{pmatrix}.$$

Equating this with

$$D(u \circ F) = \begin{pmatrix} \frac{\partial(u \circ F)}{\partial t_1} & \cdots & \frac{\partial(u \circ x)}{\partial t_m} \end{pmatrix}$$

gives the "general" form of the chain rule from Stewart.

Exercise 6.7. Stewart also stated a special case of the chain rule when $z = z(x, y)$ and $x = x(t)$, $y = y(t)$. Explain how this special case follows from Theorem 3.

Department of Mathematics, Columbia University, New York, NY 10027
E-mail address: rl2327@columbia.edu