Differentiability

Alex Nita

1 Matrices and Linear Functions

1.1 Algebraic Properties of Matrices

Let us define a **real** $m \times n$ **matrix** as an array of mn real numbers a_{ij} consisting of m rows and n columns,

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \tag{1.1}$$

The index i in a_{ij} indicates the ith row and the index j indicates the jth column. For example, $a_{2,3}$ is the real number in the second row, third column of A. The word "real" in the phrase "real matrix" refers to the fact that its entries a_i are real numbers. We could analogously define complex matrices, quaternionic matrices, integer matrices, etc.

Example 1.1 *The following is a real* 3×2 *matrix:*

$$\begin{pmatrix} 1 & -1 \\ 2 & -4 \\ 0 & \sqrt{2} \end{pmatrix}$$

while the following is a real 1×4 *matrix:*

$$\begin{pmatrix} -1 & 2 & 45 & -\frac{1}{\sqrt{3}} \end{pmatrix}$$

■

We can define **addition** of two matrices A and B of the same dimension $m \times n$ as we did with vectors, componentwise:

$$A + B = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} + \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{m1} & \cdots & b_{mn} \end{pmatrix} = \begin{pmatrix} a_{11} + b_{11} & \cdots & a_{1n} + b_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & \cdots & a_{mn} + b_{mn} \end{pmatrix} \tag{1.2}$$

We can also define **scalar multiplication** of an $m \times n$ real matrix A by a real number a:

$$aA = a\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} = \begin{pmatrix} aa_{11} & \cdots & aa_{1n} \\ \vdots & \ddots & \vdots \\ aa_{m1} & \cdots & aa_{mn} \end{pmatrix} \tag{1.3}$$

Example 1.2 *For example, here a sum of* 3×2 *matrices:*

$$\begin{pmatrix} 1 & -1 \\ 2 & -4 \\ 0 & \sqrt{2} \end{pmatrix} + \begin{pmatrix} 20 & 9 \\ 5 & 3 \\ \sqrt{5} & 2 \end{pmatrix} = \begin{pmatrix} 1+20 & -1+9 \\ 2+5 & -4+3 \\ 0+\sqrt{5} & \sqrt{2}+2 \end{pmatrix} = \begin{pmatrix} 21 & 8 \\ 7 & -1 \\ \sqrt{5} & 2+\sqrt{2} \end{pmatrix}$$

and here is a scalar multiple of a 3×2 *matrix.*

$$5 \begin{pmatrix} 1 & -1 \\ 2 & -4 \\ 0 & \sqrt{2} \end{pmatrix} = \begin{pmatrix} 5 \cdot 1 & 5 \cdot (-1) \\ 5 \cdot 2 & 5 \cdot (-4) \\ 5 \cdot 0 & 5 \cdot \sqrt{2} \end{pmatrix} = \begin{pmatrix} 5 & -5 \\ 10 & -20 \\ 0 & 5\sqrt{2} \end{pmatrix} \qquad \blacksquare$$

We can also **multiply matrices**. If A is an $m \times n$ matrix and B is an $n \times p$ matrix (m, n and p may be different here, but note that n appears in the dimensions of both A and B, in specific locations!), then the product of A and B, written AB, is an $m \times p$ matrix whose entries c_{ij} are the dot products of the ith row of A and the jth column of B. That is, if we write A in terms of *row vectors*,

$$A = \begin{pmatrix} \vec{a}_1 \\ \vec{a}_2 \\ \vdots \\ \vec{a}_m \end{pmatrix}, \qquad \text{where} \qquad \begin{array}{rcl} \vec{a}_1 & = & \langle a_{11}, \ldots, a_{1n} \rangle \\ \vec{a}_2 & = & \langle a_{21}, \ldots, a_{2n} \rangle \\ & \vdots & \\ \vec{a}_m & = & \langle a_{m1}, \ldots, a_{mn} \rangle \end{array} \tag{1.4}$$

and if we write B in terms of *column vectors*,

$$B = \begin{pmatrix} \vec{b}_1 & \vec{b}_2 & \cdots & \vec{b}_p \end{pmatrix} \qquad \text{where} \qquad \vec{b}_1 = \begin{pmatrix} b_{11} \\ \vdots \\ b_{n1} \end{pmatrix}, \quad \vec{b}_2 = \begin{pmatrix} b_{12} \\ \vdots \\ b_{n2} \end{pmatrix}, \quad \cdots \quad \vec{b}_1 = \begin{pmatrix} b_{1p} \\ \vdots \\ b_{np} \end{pmatrix} \tag{1.5}$$

then

$$AB = \begin{pmatrix} \vec{a}_1 \cdot \vec{b}_1 & \vec{a}_1 \cdot \vec{b}_2 & \cdots & \vec{a}_1 \cdot \vec{b}_p \\ \vec{a}_2 \cdot \vec{b}_1 & \vec{a}_2 \cdot \vec{b}_2 & \cdots & \vec{a}_1 \cdot \vec{b}_p \\ \vdots & \vdots & \ddots & \vdots \\ \vec{a}_m \cdot \vec{b}_1 & \vec{a}_m \cdot \vec{b}_2 & \cdots & \vec{a}_m \cdot \vec{b}_p \end{pmatrix} \tag{1.6}$$

Example 1.3 *Let us see how this works when we multiply a* 3×2 *matrix with a* 2×5*. We should, of course, get a* 3×5 *matrix. Let*

$$A = \begin{pmatrix} 1 & -1 \\ 0 & \sqrt{2} \\ 2 & -4 \end{pmatrix}, \qquad B = \begin{pmatrix} \sqrt{7} & 1 & 2 & 1 & 0 \\ 2 & -4 & 1 & 0 & 3 \end{pmatrix}$$

Then A *and* B *are written in terms of their row and column vectors as*

$$A = \begin{pmatrix} \vec{a}_1 \\ \vec{a}_2 \\ \vec{a}_3 \end{pmatrix} = \begin{pmatrix} \langle 1, -1 \rangle \\ \langle 0, \sqrt{2} \rangle \\ \langle 2, -4 \rangle \end{pmatrix}$$

$$B = \begin{pmatrix} \vec{b}_1 & \vec{b}_2 & \vec{b}_3 & \vec{b}_4 & \vec{b}_5 \end{pmatrix} = \left(\begin{pmatrix} \sqrt{7} \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ -4 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 3 \end{pmatrix} \right)$$

We really think of the $\vec{b}_j$ as vectors, of course, so for example $\vec{b}_1 = \langle\sqrt{7}, 2\rangle$. Then, we have

$$\begin{aligned}
AB &= \begin{pmatrix} \vec{a}_1\cdot\vec{b}_1 & \vec{a}_1\cdot\vec{b}_2 & \vec{a}_1\cdot\vec{b}_3 & \vec{a}_1\cdot\vec{b}_4 & \vec{a}_1\cdot\vec{b}_5 \\ \vec{a}_2\cdot\vec{b}_1 & \vec{a}_2\cdot\vec{b}_2 & \vec{a}_2\cdot\vec{b}_3 & \vec{a}_2\cdot\vec{b}_4 & \vec{a}_2\cdot\vec{b}_5 \\ \vec{a}_3\cdot\vec{b}_1 & \vec{a}_3\cdot\vec{b}_2 & \vec{a}_3\cdot\vec{b}_3 & \vec{a}_3\cdot\vec{b}_4 & \vec{a}_3\cdot\vec{b}_5 \end{pmatrix} \\
&= \begin{pmatrix} \langle 1,-1\rangle\cdot\langle\sqrt{7},2\rangle & \langle 1,-1\rangle\cdot\langle 1,-4\rangle & \langle 1,-1\rangle\cdot\langle 2,1\rangle & \langle 1,-1\rangle\cdot\langle 1,0\rangle & \langle 1,-1\rangle\cdot\langle 0,3\rangle \\ \langle 0,\sqrt{2}\rangle\cdot\langle\sqrt{7},2\rangle & \langle 0,\sqrt{2}\rangle\cdot\langle 1,-4\rangle & \langle 0,\sqrt{2}\rangle\cdot\langle 2,1\rangle & \langle 0,\sqrt{2}\rangle\cdot\langle 1,0\rangle & \langle 0,\sqrt{2}\rangle\cdot\langle 0,3\rangle \\ \langle 2,-4\rangle\cdot\langle\sqrt{7},2\rangle & \langle 2,-4\rangle\cdot\langle 1,-4\rangle & \langle 2,-4\rangle\cdot\langle 2,1\rangle & \langle 2,-4\rangle\cdot\langle 1,0\rangle & \langle 2,-4\rangle\cdot\langle 0,3\rangle \end{pmatrix} \\
&= \begin{pmatrix} \sqrt{7}-2 & 5 & 1 & 1 & -3 \\ 2\sqrt{2} & -4\sqrt{2} & \sqrt{2} & 0 & 3\sqrt{2} \\ 2\sqrt{7}-8 & 18 & 0 & 2 & -12 \end{pmatrix}
\end{aligned}$$

■

1.2 Geometric Properties of Matrices

The significance of matrices is that we can use them to define **linear functions**, which, as the name suggests, are functions that generalize functions of a single variable whose graphs are lines,

$$f(x) = mx + b \tag{1.7}$$

We have already seen an example of this in the "linear" equation for a plane, $z = mx + ny + b$, which is the graph of the function $z = g(x, y) = mx + ny + b$. This is a very geometric idea. Let us explain the geometric content of this now.

Recall the *equation of a plane*,

$$ax + by + cz + d = 0 \tag{1.8}$$

We can consider this plane either as the level surface of a function $f : \mathbb{R}^3 \to \mathbb{R}$, namely $f(x, y, z) = ax + by + cz + d$, or else, solving for z (which gives $z = -\frac{a}{c}x - \frac{b}{c}y - \frac{d}{c}$, assuming we can do this, i.e. assuming $c \neq 0$), we can consider it as the graph of a function $g : \mathbb{R}^2 \to \mathbb{R}$, namely $g(x, y) = -\frac{a}{c}x - \frac{b}{c}y - \frac{d}{c}$. Let us take the latter approach for the moment. In general, the graph of a function of the type

$$\begin{aligned} &g : \mathbb{R}^2 \to \mathbb{R} \\ &g(x, y) = mx + ny + b \end{aligned} \tag{1.9}$$

will be a plane. If $b \neq 0$, that plane will not pass through the origin $(0, 0, 0)$ (it will pass instead through $(0, 0, b)$, of course). Our goal is to take a complicated function $f : \mathbb{R}^2 \to \mathbb{R}$, whose graph may be curved and twisted, and locally approximate its graph by the graph of a **linear function**, that is a function of the type (1.9). g is called "linear" because it is a generalization of the one-dimensional line, $y = mx + b$. Now we have two dimensions, so the analog is $z = mx + ny + b$.

More generally, a function $f : \mathbb{R}^n \to \mathbb{R}$ will be called **linear** if it is of the form

$$f(\mathbf{x}) = f(x_1, x_2, \ldots, x_n) = m_1x_1 + m_2x_2 + \cdots + m_nx_n + b \tag{1.10}$$

What if the function is *vector-valued*, that is $f : \mathbb{R}^n \to \mathbb{R}^m$? Then f can be described in terms of m component functions $f_i : \mathbb{R}^n \to \mathbb{R}$,

$$f(\mathbf{x}) = \Big(f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x})\Big) \tag{1.11}$$

and if *each* of these component functions f_i is *linear*, i.e. if each f_i is of the form (1.10), then we get a system of linear expressions,

$$\begin{array}{rcccl} f_1(\mathbf{x}) & = & f_1(x_1, x_2, \ldots, x_n) & = & m_{11}x_1 + m_{12}x_2 + \cdots + m_{1n}x_n + b_1 \\ f_2(\mathbf{x}) & = & f_1(x_1, x_2, \ldots, x_n) & = & m_{21}x_1 + m_{22}x_2 + \cdots + m_{2n}x_n + b_2 \\ & \vdots & & & \\ f_m(\mathbf{x}) & = & f_1(x_1, x_2, \ldots, x_n) & = & m_{m1}x_1 + m_{m2}x_2 + \cdots + m_{mn}x_n + b_m \end{array} \tag{1.12}$$

With matrix and vector notation, we can neatly write the system (1.12) as

$$\begin{pmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_1(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} & \cdots & m_{1n} \\ m_{21} & m_{22} & \cdots & m_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ m_{m1} & m_{m2} & \cdots & m_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \tag{1.13}$$

If we write the component expression of f in (1.11) as a column vector instead of a row vector, then we can even more neatly write (1.12) as

$$f(\mathbf{x}) = A\mathbf{x} + \mathbf{b} \tag{1.14}$$

where

$$f(\mathbf{x}) = \begin{pmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_1(\mathbf{x}) \end{pmatrix}, \qquad A = \begin{pmatrix} m_{11} & m_{12} & \cdots & m_{1n} \\ m_{21} & m_{22} & \cdots & m_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ m_{m1} & m_{m2} & \cdots & m_{mn} \end{pmatrix}, \qquad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

A function $f : \mathbb{R}^n \to \mathbb{R}^m$ of the form (1.14) will be called a **linear function**. (Note, however, that in many other books only a function of the type $f(\mathbf{x}) = A\mathbf{x}$ is called linear, while a function of the type $f(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$ is called *affine*.)

Exercise 1.4 *Show that if A is an $m \times n$ matrix, $\mathbf{x}$ and $\mathbf{y}$ are (column) vectors in $\mathbb{R}^n$ (thought of as $n \times 1$ matrices), and a and b are real numbers, then we have the following distributivity law: $A(a\mathbf{x} + b\mathbf{y}) = a(A\mathbf{x}) + b(A\mathbf{y})$.* ■

1.3 Examples of Linear Functions

Example 1.5 *A very simple, though not very interesting, example of a linear function is the unction $f : \mathbb{R}^3 \to \mathbb{R}^2$,*

$$f(x, y, z) = \Big(x + 5y - z - 4,\ 4x + y + 3z + 2\Big)$$

The two component functions of f are

$$\begin{aligned} f_1(x, y, z) &= x + 5y - z - 4 \\ f_2(x, y, x) &= 4x + y + 3z + 2 \end{aligned}$$

and they are both linear, so the function f is linear. As we did in the general case, we can write the system of expressions above as a single matrix expression:

$$f(x, y, z) = \begin{pmatrix} f_1(x, y, z) \\ f_2(x, y, z) \end{pmatrix} = \begin{pmatrix} 1 & 5 & -1 \\ 4 & 1 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} -4 \\ 2 \end{pmatrix}$$ ■

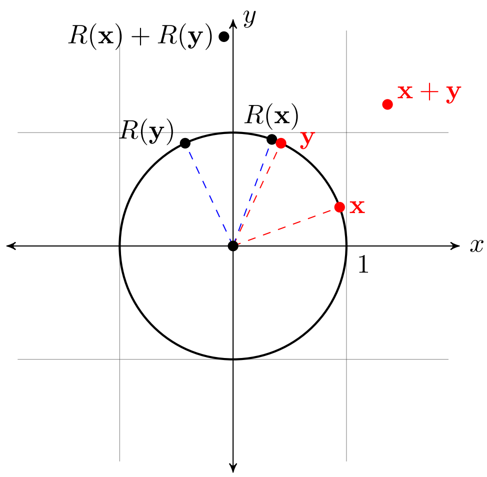
Example 1.6 *Let* $f : \mathbb{R}^2 \to \mathbb{R}^2$ *be the* **rotation-followed-by-a-translation function** *in the plane,*

$$f(x,y) = \Big(x\cos\theta - y\sin\theta + 5,\ x\sin\theta + y\cos\theta - 4\Big)$$

This is a rotation of the plane through an angle θ *followed by a translation by the vector/point* $(5,-4)$. *We can write this in matrix notation as*

$$f(x,y) = \begin{pmatrix} f_1(x,y) \\ f_2(x,y) \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 5 \\ -4 \end{pmatrix}$$

Let us briefly explain the geometric reasoning behind this expression. Let's start with a rotation through an angle θ, *call it* R. *In the following picture of the unit circle, we rotate the two points* $\mathbf{x}$ *and* $\mathbf{y}$ *lying on the circle, as well as the sum* $\mathbf{x}+\mathbf{y}$ *lying outside of it, by an angle* θ, *that is we apply a rotation* R *to them.*



Now, the first thing to notice is that every point $\mathbf{x} = (x,y)$ *in the plane can be written as a sum of scalar multiples of the coordinate vectors/points* $\mathbf{i} = (1,0)$ *and* $\mathbf{j} = (0,1)$, *namely by*

$$(x,y) = (x,0) + (0,y) = x(1,0) + y(0,1) = x\mathbf{i} + y\mathbf{j}$$

so if we can manage to characterize what the rotation does to $\mathbf{i}$ *and* $\mathbf{j}$, *then we'll understand what it does to any point* $\mathbf{x}$, *provided of course that* R *is* **linear**, *that is it satisfies*

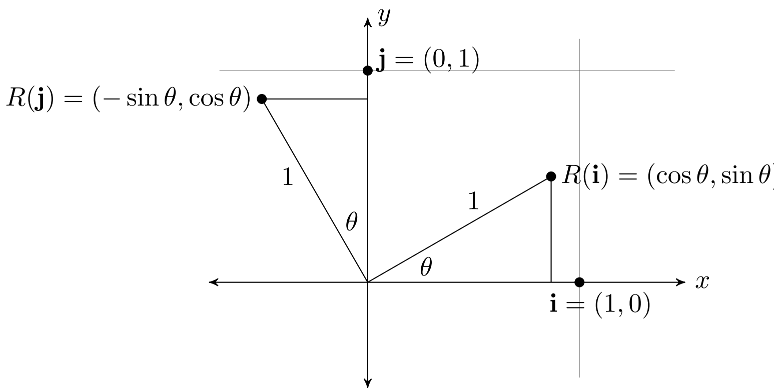
$$R(\mathbf{x}+\mathbf{y}) = R(\mathbf{x}) + R(\mathbf{y})$$
$$R(a\mathbf{x}) = aR(\mathbf{x})$$

for all points/vectors $\mathbf{x}$ *and* $\mathbf{y}$ *and all real numbers* a *(see Exercise 1.4 above for the* algebraic *reason for linearity). The intuitive* geometric *reasoning for linearity is contained in the above diagram: we'd like to be able to rotate the points* $\mathbf{x}$ *and* $\mathbf{y}$ *to* $R(\mathbf{x})$ *and* $R(\mathbf{y})$, *then add them to get* $R(\mathbf{x}) + R(\mathbf{y})$, *and we should get the same point/vector as when we* first *add* $\mathbf{x}$ *and* $\mathbf{y}$ *and* then *rotate the result to* $R(\mathbf{x}+\mathbf{y})$. *If* R *were linear, then knowing how* R *treats* $\mathbf{i}$ *and* $\mathbf{j}$ *means we know how it treats* $(x,y) = x\mathbf{i} + y\mathbf{j}$, *because*

$$R(x,y) = R(x\mathbf{i} + y\mathbf{j}) = xR(\mathbf{i}) + yR(\mathbf{j})$$

Let's now characterize R *on* $\mathbf{i}$ *and* $\mathbf{j}$. *Clearly, a rotation through a positive angle* θ *sends the point* $\mathbf{i} = (1,0)$ *to the point* $(\cos\theta, \sin\theta)$ *and it sends the point* $\mathbf{j} = (0,1)$ *to the point*

$(-\sin\theta, \cos\theta)$, as can be seen in the following diagram:



That is,

$$R(\mathbf{i}) = R(1,0) = (\cos\theta, \sin\theta)$$
$$R(\mathbf{j}) = R(0,1) = (-\sin\theta, \cos\theta)$$

Therefore, assuming the linearity of R we have that

$$\begin{aligned} R(x,y) &= R(x\mathbf{i} + y\mathbf{j}) \\ &= xR(\mathbf{i}) + yR(\mathbf{j}) \\ &= x(\cos\theta, \sin\theta) + y(-\sin\theta, \cos\theta) \\ &= \Big(\underbrace{x\cos\theta - y\sin\theta}_{R_1(x,y)}, \underbrace{x\sin\theta + y\cos\theta}_{R_2(x,y)}\Big) \end{aligned}$$

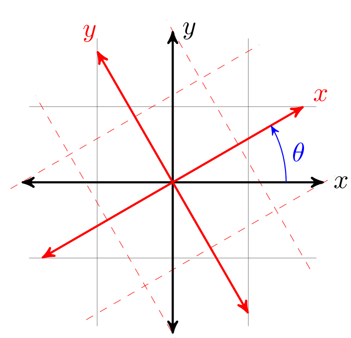
for all points $\mathbf{x} = (x,y)$ in $\mathbb{R}^2$. Putting this in matrix notation, R can be described as a matrix

$$R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \tag{1.15}$$

and its effect on a point/vector $\mathbf{x} = (x,y)$ in $\mathbb{R}^2$ is to left-multiply it, as a column vector, by the matrix R:

$$R\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$$

The cumulative effect of the rotation R, say through $\theta = \pi/6$, is to rotate the entire plane:



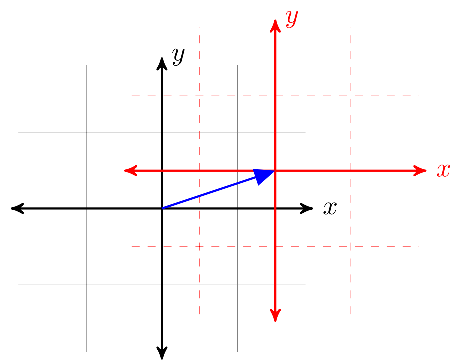
Next, consider a **translation** T *by a point/vector* $\mathbf{v} = (a, b)$. *This is much simpler to understand. A translation simply adds the point/vector* $\mathbf{v}$ *to any given point* $\mathbf{x} = (x, y)$*,*

$$T(\mathbf{x}) = \mathbf{x} + \mathbf{v} = (x, y) + (a, b) = (x + a, y + b)$$

or, if we're going to write everything in terms of column vectors,

$$\boxed{T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} a \\ b \end{pmatrix}}$$

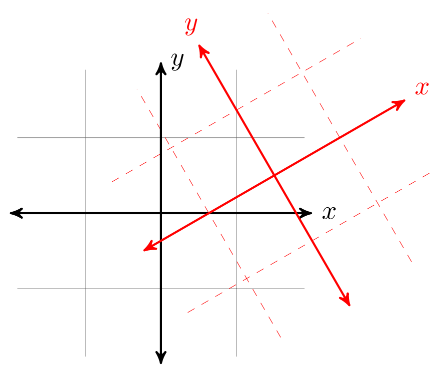
The effect of T *is thus to shift then entire plane* $\mathbb{R}^2$ *by the vector* $\mathbf{v}$. *For example, if* $\mathbf{v} = (3/2, 1/2)$*, the entire plane is shifted* $3/2$ *units to the right and* $1/2$ *a unit up:*



Lastly, let's combine the two. The way to do this is to compose the two, first apply the rotation, then apply the translation,

$$f(\mathbf{x}) = (T \circ R)(\mathbf{x}), \qquad i.e. \;\; f\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} a \\ b \end{pmatrix}$$

The picture is the following:



■

Example 1.7 *If ℓ is a line in the plane $\mathbb{R}^2$ passing through the origin, then a* **reflection** *about ℓ is a function $\mathcal{R} : \mathbb{R}^2 \to \mathbb{R}^2$, which obviously satisfies*

$$\mathcal{R}(\mathbf{x}) = \begin{cases} \mathbf{x} & \text{if } \mathbf{x} \text{ lies on } \ell \\ -\mathbf{x} & \text{if } \mathbf{x} \text{ is orthogonal to } \ell \end{cases}$$

Note that for $\mathbf{x}$ to be orthogonal to ℓ we must have the entire line containing $\mathbf{x}$ be perpendicular to ℓ (for if that is the case, then any point on the line will be orthogonal to any point in the other line!). What about every other point in the plane? Well, if $\mathbf{x} = (x, y)$ is any arbitrary point in $\mathbb{R}^2$, then we can decompose it into parallel and perpendicular components to ℓ:

$$\mathbf{x} = \mathbf{x}_{\|} + \mathbf{x}_{\perp}$$

In this case, we must have that

$$\mathcal{R}(\mathbf{x}) = \mathcal{R}(\mathbf{x}_{\|} + \mathbf{x}_{\perp}) = \mathcal{R}(\mathbf{x}_{\|}) + \mathcal{R}(\mathbf{x}_{\perp}) = \mathbf{x}_{\|} - \mathbf{x}_{\perp}$$

if we suppose, as we must, that $\mathcal{R}$ is linear. *Now, we know what $\mathbf{x}_{\|}$ and $\mathbf{x}_{\perp}$ are. Of course, ℓ be entirely described by a unit vector $\mathbf{v} = (a, b)$, since every other point on ℓ is a scalar multiple of $\mathbf{v}$, that is*

$$\ell = \{\lambda\mathbf{v} = (\lambda a, \lambda b) \mid \lambda \text{ is a real number}\}$$

As a consequence, $\mathbf{x}_{\|} = (\mathbf{x} \cdot \mathbf{v})\mathbf{v}$ and $\mathbf{x}_{\perp} = \mathbf{x} - \mathbf{x}_{\perp} = \mathbf{x} - (\mathbf{x} \cdot \mathbf{v})\mathbf{v}$, which means

$$\begin{aligned} \mathcal{R}(\mathbf{x}) &= \mathbf{x}_{\|} - \mathbf{x}_{\perp} \\ &= (\mathbf{x} \cdot \mathbf{v})\mathbf{v} - \big(\mathbf{x} - (\mathbf{x} \cdot \mathbf{v})\mathbf{v}\big) \\ &= 2(\mathbf{x} \cdot \mathbf{v})\mathbf{v} - \mathbf{x} \end{aligned}$$

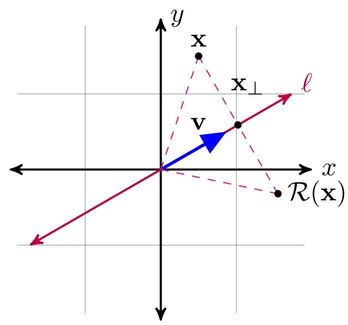
We can describe this operation as a matrix, by noting that $2\mathbf{x} \cdot \mathbf{v} = 2ax + 2by$, so that writing $\mathbf{v}$ and $\mathbf{x}$ as column vectors, we get

$$\begin{aligned} \mathcal{R}\begin{pmatrix} x \\ y \end{pmatrix} = \mathcal{R}(\mathbf{x}) &= 2(\mathbf{x} \cdot \mathbf{v})\mathbf{v} - \mathbf{x} \\ &= (2ax + 2by)\begin{pmatrix} a \\ b \end{pmatrix} - \begin{pmatrix} x \\ y \end{pmatrix} \\ &= \begin{pmatrix} 2a^2x + 2aby - x \\ 2abx + 2b^2y - y \end{pmatrix} \\ &= \begin{pmatrix} (2a^2 - 1)x + (2ab)y \\ (2ab)x + (2b^2 - 1)y \end{pmatrix} \\ &= \begin{pmatrix} 2a^2 - 1 & 2ab \\ 2ab & 2b^2 - 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} \end{aligned}$$

For example, if ℓ is the line containing the unit vector $\mathbf{v} = (\sqrt{3}/2, 1/2)$, i.e. $(\cos\pi/6, \sin\pi/6)$, then if we pick the point, say, $\mathbf{x} = (1/2, 3/2)$, it's reflection about ℓ will be

$$\mathcal{R}\begin{pmatrix} \frac{1}{2} \\ \frac{3}{2} \end{pmatrix} = \begin{pmatrix} 2 \cdot \frac{3}{4} - 1 & 2 \cdot \frac{\sqrt{3}}{2} \cdot \frac{1}{2} \\ 2 \cdot \frac{\sqrt{3}}{2} \cdot \frac{1}{2} & 2 \cdot \frac{1}{4} - 1 \end{pmatrix}\begin{pmatrix} \frac{1}{2} \\ \frac{3}{2} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}\begin{pmatrix} \frac{1}{2} \\ \frac{3}{2} \end{pmatrix} = \begin{pmatrix} \frac{1+3\sqrt{3}}{4} \\ \frac{\sqrt{3}-3}{4} \end{pmatrix}$$

so the picture of $\mathcal{R}$



In sum, then, if we're going to reflect all points in $\mathbb{R}^2$ *about a given line* ℓ *containing all scalar mutliples of a specific unit vector* $\mathbf{v} = (a, b)$*, then the reflection* $\mathcal{R}$ *is a linear function given in matrix terms by*

$$\mathcal{R}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2a^2 - 1 & 2ab \\ 2ab & 2b^2 - 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$$

■

2 Differentiability

2.1 The Total Derivative

The thing we want to do now is to locally appproximate a complicated function $f : \mathbb{R}^n \to \mathbb{R}^m$ by a much simpler *linear* function. This is the idea behind the total derivative of f at a point $\mathbf{a}$ in $\mathbb{R}^n$. Formally, we say that f is **differentiable at a point** $\mathbf{a} \in \mathbb{R}^n$ if there exists an $m \times n$ matrix (m and n here depend on the domain and range of f!)

$$Df(\mathbf{a}) = \begin{pmatrix} m_{11} & \cdots & m_{1n} \\ \vdots & \ddots & \vdots \\ m_{m1} & \cdots & m_{mn} \end{pmatrix} \qquad \text{(a real } m \times n \text{ matrix)} \tag{2.1}$$

called the **total derivative** (or the **Jacobi matrix**), which satisfies the following limit condition:

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{\|f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) - Df(\mathbf{a})\mathbf{h}\|}{\|\mathbf{h}\|} = 0 \tag{2.2}$$

Equivalently, f must locally be approximated by a linear function, that is

$$f(\mathbf{a} + \mathbf{h}) \approx Df(\mathbf{a})\mathbf{h} + f(\mathbf{a}) \tag{2.3}$$

where the **error** in the approximation

$$\begin{aligned} E(\mathbf{h}) &= f(\mathbf{a} + \mathbf{h}) - \big(\text{linear approximation of the value of } f \text{ at } \mathbf{x} = \mathbf{a} + \mathbf{h}\big) \\ &= f(\mathbf{a} + \mathbf{h}) - \Big(Df(\mathbf{a})\mathbf{h} + f(\mathbf{a})\Big) \end{aligned}$$

satisfies

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{\|E(\mathbf{h})\|}{\|\mathbf{h}\|} = 0 \tag{2.4}$$

This last statement, (2.4), is obviously the same statement as (2.2).

Remark 2.1 *The expression $Df(\mathbf{a})\mathbf{h}$ denotes matrix multiplication. Here, $\mathbf{h} = (h_1, \ldots, h_n)$ is a vector in $\mathbb{R}^n$ thought of as an $n \times 1$ column vector:*

$$Df(\mathbf{a})\mathbf{h} = \begin{pmatrix} m_{11} & \cdots & m_{1n} \\ \vdots & \ddots & \vdots \\ m_{m1} & \cdots & m_{mn} \end{pmatrix} \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} = \begin{pmatrix} m_{11}h_1 + \cdots + m_{1n}h_n \\ \vdots \\ m_{m1}h_1 + \cdots + m_{mn}h_1 \end{pmatrix}$$ ■

Remark 2.2 *Thus to say that a function f is differentiable at a point is equivalent to saying that f has a total derivative there. We shall see that f may be* partially *differentiable, and to have directional derivatives in all directions, yet not be differentiable. We will explain this further below.* ■

2.2 The Directional Derivative

Suppose $\mathbf{v}$ is a 'vector' in $\mathbb{R}^n$ and $\mathbf{a}$ is a 'point' in $\mathbb{R}^n$, and let $T : \mathbb{R} \to \mathbb{R}^n$ be the translation-by-$t\mathbf{v}$ function

$$T(t) = \mathbf{x} + t\mathbf{v}$$

We say that $f : \mathbb{R}^n \to \mathbb{R}^m$ has a **directional derivative** at $\mathbf{a}$ in the direction of $\mathbf{v}$ if the composition $f \circ T : \mathbb{R} \to \mathbb{R}^m$, $(f \circ T)(t) = f(\mathbf{a} + t\mathbf{v})$, is "differentiable at 0", in the sense that the limit

$$\begin{aligned} f_{\mathbf{v}}(\mathbf{a}) \quad \text{or} \quad D_{\mathbf{v}}f(\mathbf{a}) &= \left.\frac{d}{dt}\right|_{t=0} f(\mathbf{a} + t\mathbf{v}) \\ &= \lim_{t \to 0} \frac{f(\mathbf{a} + t\mathbf{v}) - f(\mathbf{a})}{t} \end{aligned} \tag{2.5}$$

exists in $\mathbb{R}^m$.

Remark 2.3 *Notice that for each fixed nonzero t the difference $f(\mathbf{a} + t\mathbf{v}) - f(\mathbf{a})$ in the numerator is a vector in $\mathbb{R}^m$, while $1/t$ is a real number, so the quotient $\frac{f(\mathbf{a}+t\mathbf{v})-f(\mathbf{a})}{t}$ is actually scalar multiplication of the vector $f(\mathbf{a} + t\mathbf{v}) - f(\mathbf{a})$ by $1/t$.*

Another thing to notice is that in the sum $\mathbf{a} + t\mathbf{v}$ we added a point to a vector. Since we have emphasized blurring the lines between points and vectors in $\mathbb{R}^n$, on account of algebraically *they are indistinguishable, at least in $\mathbb{R}^n$, the sum makes sense.* ■

2.3 The Partial Derivative

Now suppose f is a real-valued function, $f : \mathbb{R}^n \to \mathbb{R}$. The i**th partial derivative** of f at $\mathbf{a}$ is the directional derivative of f at $\mathbf{a}$ in a coordinate direction, that is in the direction of a unit coordinate vector $\mathbf{e}_i = (0, \ldots, i, \ldots, 0)$,

$$\begin{aligned} \left.\frac{\partial f}{\partial x_i}\right|_{\mathbf{x}=\mathbf{a}} \quad \text{or} \quad f_{x_i}(\mathbf{a}) \quad \text{or} \quad D_i f(\mathbf{a}) &= \left.\frac{d}{dt}\right|_{t=0} f(\mathbf{a} + t\mathbf{e}_i) \\ &= \lim_{t \to 0} \frac{f(\mathbf{a} + t\mathbf{e}_i) - f(\mathbf{a})}{t} \\ &= \lim_{t \to 0} \frac{f(a_1, \ldots, a_{i-1}, a_i + t, a_{i+1}, \ldots, a_n) - f(a_1, \ldots, a_i, \ldots, a_n)}{t} \end{aligned} \tag{2.6}$$

Remark 2.4 *The practical import of this definition will become clear in a minute. For now, notice that the derivative* $\frac{d}{dt}\big|_{t=0} f(\mathbf{a}+t\mathbf{e}_i)$ *is an ordinary derivative from Calc 1. It's the derivative of the real-valued function of a real variable* $f \circ T : \mathbb{R} \to \mathbb{R}$, $(f \circ T)(t) = f(\mathbf{a}+t\mathbf{e}_i)$. *This means that all the other coordinates, which we normally treat as variables, since they may vary, are treated here as constants. Thus, if we label the variables* $x_1, \ldots, x_i, \ldots, x_n$, *all the other* x_j *for* $j \neq i$ *are treated as constants in any expression for* f. *For example, if* $f(x, y, z) = xyz + x^2y + z^2$, *in the partial derivative* $\frac{\partial f}{\partial x}$ *with respect to* x *the 'variables'* y *and* z *are treated as constants, so we can do what we normally do when computing a Calc 1 derivative, pull the constants out. Here, for example, we'd have* $\frac{\partial f}{\partial x} = yz + 2xy$. ■

2.4 The Relationship Between the Total and Directional and Partial Derivatives

Nobody wants to compute an actual limit, though the limit idea is extremely important theoretically. Luckily, we don't have to here. The directional derivative, though defined in terms of a limit, is in fact computable in terms of a matrix product!

Theorem 2.5 *If* $f : \mathbb{R}^n \to \mathbb{R}^m$ *is differentiable at* $\mathbf{a}$ *in* $\mathbb{R}^n$, *then all of its directional derivatives at* $\mathbf{a}$ *exist, and for any choice of vector* $\mathbf{v}$ *in* $\mathbb{R}^n$ *we have*

$$D_{\mathbf{v}} f(\mathbf{x}) = Df(\mathbf{x})\mathbf{v} \tag{2.7}$$

The left-hand side is a limit, while the right-hand side is a matrix product, with $\mathbf{v}$ *treated as a column vector.*

Proof: Since f is differentiable at $\mathbf{a}$, fix $\mathbf{v}$ and consider $\mathbf{h} = t\mathbf{v}$ for some sufficiently small $t \in \mathbb{R}$. Applying the linear approximation (2.3) and the linearity of the derivative $Df(\mathbf{a})$ (i.e. $Df(\mathbf{a})(a\mathbf{x}+b\mathbf{y}) = aDf(\mathbf{a})\mathbf{x} + bDf(\mathbf{a})\mathbf{y}$, cf. Exercise 1.4 above) we get

$$\begin{aligned} f(\mathbf{x}+t\mathbf{v}) - f(\mathbf{x}) - tDf(\mathbf{x})\mathbf{v} &= f(\mathbf{x}+t\mathbf{v}) - f(\mathbf{x}) - Df(\mathbf{x})(t\mathbf{v}) \\ &= f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x}) - Df(\mathbf{x})\mathbf{h} \\ &= E(\mathbf{h}) \\ &= E(t\mathbf{v}) \end{aligned} \tag{2.8}$$

and applying the limit (2.4)

$$\lim_{t\to 0} \frac{\|E(t\mathbf{v})\|}{|t|} = \lim_{t\to 0} \frac{\|E(t\mathbf{v})\|}{|t|\|\mathbf{v}\|} \cdot \|\mathbf{v}\| = \lim_{t\to 0} \frac{\|E(t\mathbf{v})\|}{\|t\mathbf{v}\|} \cdot \|\mathbf{v}\| = \lim_{\mathbf{h}\to \mathbf{0}} \frac{\|E(\mathbf{h})\|}{\|\mathbf{h}\|} \cdot \|\mathbf{v}\| = 0 \cdot \|\mathbf{v}\| = 0$$

By (2.8) this means

$$\lim_{t\to 0} \frac{\|f(\mathbf{x}+t\mathbf{v}) - f(\mathbf{x}) - tDf(\mathbf{x})(\mathbf{v})\|}{t} = 0$$

and hence

$$\begin{aligned} \lim_{t\to 0} \frac{f(\mathbf{x}+t\mathbf{v}) - f(\mathbf{x})}{t} - Df(\mathbf{x})(\mathbf{v}) &= \lim_{t\to 0} \frac{f(\mathbf{x}+t\mathbf{v}) - f(\mathbf{x})}{t} - \lim_{t\to 0} \frac{tDf(\mathbf{x})(\mathbf{v})}{t} \\ &= \lim_{t\to 0} \frac{f(\mathbf{x}+t\mathbf{v}) - f(\mathbf{x}) - tDf(\mathbf{x})(\mathbf{v})}{t} \\ &= \mathbf{0} \end{aligned}$$

i.e.

$$D_{\mathbf{v}} f(\mathbf{x}) = \lim_{t\to 0} \frac{f(\mathbf{x}+t\mathbf{v}) - f(\mathbf{x})}{t} = Df(\mathbf{x})(\mathbf{v})$$ ■

Example 2.6 *Suppose $f : \mathbb{R}^3 \to \mathbb{R}^2$ is differentiable at the point $\mathbf{a} = (1,1,2)$ and $\mathbf{v} = (-1,4,2)$ is a vector in $\mathbb{R}^3$. If we already knew the total derivative of f, say*

$$Df(\mathbf{a}) = \begin{pmatrix} 2 & 3 & 2 \\ 1 & 0 & 5 \end{pmatrix}$$

then computing the directional derivative of f at $\mathbf{a}$ in the direction of $\mathbf{v}$ would be easy, namely

$$f_{\mathbf{v}}(\mathbf{a}) = Df(\mathbf{a})\mathbf{v} = \begin{pmatrix} 2 & 3 & 2 \\ 1 & 0 & 5 \end{pmatrix} \begin{pmatrix} -1 \\ 4 \\ 2 \end{pmatrix} = \begin{pmatrix} 14 \\ 9 \end{pmatrix} \qquad \blacksquare$$

Thus, if f is differentiable at a, the task is to find a way to compute $Df(\mathbf{a})$. For then we can compute all directional derivatives by simple matrix multiplication. Luckily, we can do this, using the previous theorem, provided we know the that the total derivative $Df(\mathbf{a})$ has rows consisting of the total derivatives of the component functions, that is

$$\underbrace{D\begin{pmatrix} f_1(\mathbf{a}) \\ \vdots \\ f_m(\mathbf{a}) \end{pmatrix}}_{Df(\mathbf{a})} = \begin{pmatrix} Df_1(\mathbf{a}) \\ \vdots \\ Df_m(\mathbf{a}) \end{pmatrix}$$

In this case I claim that **the total derivative $Df(\mathbf{a})$ is the matrix of partial derivatives of the component functions f_i of f,**

$$Df(\mathbf{a}) = \begin{pmatrix} \left.\frac{\partial f_1}{\partial x_1}\right|_{\mathbf{a}} & \cdots & \left.\frac{\partial f_1}{\partial x_n}\right|_{\mathbf{a}} \\ \vdots & \ddots & \vdots \\ \left.\frac{\partial f_m}{\partial x_1}\right|_{\mathbf{a}} & \cdots & \left.\frac{\partial f_m}{\partial x_n}\right|_{\mathbf{a}} \end{pmatrix} \tag{2.9}$$

To see this, take a real-valued function first, $f : \mathbb{R}^n \to \mathbb{R}$ (for any vector-valued function as above is made up of its m real-valued component functions) and look at the directional derivative in the ith coordinate direction $\mathbf{e}_i = (0, \ldots, 1, \ldots, 0)$ (it has a 1 in the ith slot and 0 everywhere else). Letting

$$Df(\mathbf{a}) = \begin{pmatrix} m_1 & m_2 & \cdots & m_n \end{pmatrix}$$

be the $1 \times n$ matrix defining the total derivative of f, and noting that the ith partial derivative is the directional derivative in the ith coordinate direction, we have

$$\begin{aligned} \left.\frac{\partial f}{\partial x_i}\right|_{\mathbf{a}} = D_{\mathbf{e}_i} f(\mathbf{a}) = Df(\mathbf{a})\mathbf{e}_i &= \begin{pmatrix} m_1 & m_2 & \cdots & m_n \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \\ &= m_1 \cdot 0 + m_2 \cdot 0 + \cdots m_i \cdot 1 + \cdots m_n \cdot 0 \\ &= m_i \end{aligned}$$

This is true for each $i = 1, \ldots, n$, so

$$Df(\mathbf{a}) = \begin{pmatrix} m_1 & m_2 & \cdots & m_n \end{pmatrix} = \begin{pmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \end{pmatrix}$$

Now consider a vector-valued function $f : \mathbb{R}^n \to \mathbb{R}^m$, $f = (f_1, \ldots, f_m)$. Then, each of its component functions f_i is a real-valued function, and the above result applies separately to each, from which we get our result,

$$Df(\mathbf{a}) = D\begin{pmatrix} f_1(\mathbf{a}) \\ \vdots \\ f_m(\mathbf{a}) \end{pmatrix} = \begin{pmatrix} Df_1(\mathbf{a}) \\ \vdots \\ Df_m(\mathbf{a}) \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}$$

The only questionable thing about this is the legality of the second equality, where we "pulled the D inside the column vector." It was, in fact, legal, and moreover our ability to do this gives us another useful way to decide the differentiability of a function. Let us state and prove this result carefully.

Theorem 2.7 *A function $f : \mathbb{R}^n \to \mathbb{R}^m$, $f = (f_1, \ldots, f_n)$, is differentiable at $\mathbf{a}$ if and only if each of its component functions $f_i : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\mathbf{a}$. In that case, we have*

$$Df(\mathbf{a}) = D\begin{pmatrix} f_1(\mathbf{a}) \\ \vdots \\ f_m(\mathbf{a}) \end{pmatrix} = \begin{pmatrix} Df_1(\mathbf{a}) \\ \vdots \\ Df_m(\mathbf{a}) \end{pmatrix} \tag{2.10}$$

That is, to compute $Df(\mathbf{a})$ we can just compute the $1 \times n$ derivative matrices of the f_i first, which we know are of the form $Df_i(\mathbf{a}) = \left(\frac{\partial f_i}{\partial x_1}\big|_{\mathbf{a}} \quad \cdots \quad \frac{\partial f_i}{\partial x_n}\big|_{\mathbf{a}} \right)$, and enter the result into the ith row of the larger matrix.

Proof: This follows from the inequalities

$$|a_i| \leq \|\mathbf{a}\| \leq \sqrt{n} \max_{1 \leq i \leq n} |a_i|$$

for all i, since if f is differentiable at $\mathbf{a}$, then the limit (2.2) exists, so the first inequality above implies that the limit of zero exists in each of the coordinates, and so for each of the coordinate functions. Indeed, by that limit we must have that $Df_i(\mathbf{a})$ is the ith component function of $Df(\mathbf{a})$. Conversely, if the component functions are differentiable at $\mathbf{a}$, then multiplying the limit (2.2) for f_i by $\sqrt{n}$ and using the second inequality above we have that the limit (2.2) for f holds as well (just choose the f_i with maximum absolute value), and moreover we must have that $Df_i(\mathbf{x})$ are the coordinate linear functionals of $Df(\mathbf{a})$ by the first inequality. ■

Now that we know how to compute $Df(\mathbf{a})$ if we know that $Df(\mathbf{a})$ exists, we have to answer the question, **"How do we determine the <u>existence</u> of $Df(\mathbf{a})$?"**. Well, we have seen that it boils down do determining the existence of the m separate total derivatives of the component functions $Df_i(\mathbf{a})$. The remaining question, therefore, is, **"How do we determine the existence of the m separate total derivatives $Df_i(\mathbf{a})$ of the component functions f_i?"** The naïve answer is, "Well, just compute the partials $\frac{\partial f_i}{\partial x_j}$ at $\mathbf{a}$ of each f_i and put them in a matrix," unfortunately, is not entirely correct. It *would* be if we knew that the partials were also *continuous* on a neighborhood of the point $\mathbf{a}$, but not otherwise. Here is an **example of why the existence of the partials $\frac{\partial f_i}{\partial x_j}$ at a alone is not enough to conclude the existence of $Df(\mathbf{a})$ (we must also have their continuity)**:

Example 2.8 *Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ given by*

$$f(x, y) = \begin{cases} \dfrac{x^2 y}{x^4 + y^2}, & \text{if } (x, y) \neq (0, 0) \\ 0, & \text{if } (x, y) = (0, 0) \end{cases}$$

First, notice that all its directional derivatives exist at the origin, for if $\mathbf{v} = (h,k)$ *is any vector in* $\mathbb{R}^2$*, then the directional derivative* $D_\mathbf{v} f(\mathbf{0})$ *is computable directly:*

$$D_\mathbf{v} f(\mathbf{0}) = \lim_{t\to 0} \frac{f(\mathbf{0}+t\mathbf{v}) - f(\mathbf{0})}{t} = \lim_{t\to 0} \frac{(th)^2(tk) - 0}{(th)^4 + (tk)^2} \cdot \frac{1}{t} = \lim_{t\to 0} \frac{\cancel{t^3} h^2 k}{\cancel{t^3}(t^2h^4 + k^2)} = \begin{cases} \dfrac{h^2}{k} & \textit{if } k \neq 0 \\ 0 & \textit{if } k = 0 \end{cases}$$

In particular, choosing $\mathbf{v} = \mathbf{e}_1 = (1,0)$ *and* $\mathbf{v} = \mathbf{e}_2 = (0,1)$ *shows that it has partial derivatives* $\frac{\partial f}{\partial x}|_{(0,0)} = \frac{\partial f}{\partial y}|_{(0,0)} = 0$ *at the origin. Outside the origin it is easily seen to be partially differentiable, and its partial derivatives exist everywhere on* $\mathbb{R}^2$*, and are given by*

$$\frac{\partial f}{\partial x} = \begin{cases} \dfrac{(x^4+y^2)2xy - 4x^5y}{(x^4+y^2)^2}, & \textit{if } (x,y) \neq (0,0) \\ (0,0), & \textit{if } (x,y) = (0,0) \end{cases}$$

$$\frac{\partial f}{\partial y} = \begin{cases} \dfrac{(x^4+y^2)x^2 - 2x^2y^2}{(x^4+y^2)^2}, & \textit{if } (x,y) \neq (0,0) \\ (0,0), & \textit{if } (x,y) = (0,0) \end{cases}$$

Thus, f *is partially differentiable everywhere in* $\mathbb{R}^2$*. However,* f *is not differentiable at the origin, in fact it is not even continuous there. For notice that on the parabola* $y = x^2$ *the function is constant with value* $1/2$*:*

$$f(h,h^2) = \frac{h^4}{2h^4} = \frac{1}{2}$$

so that arbitrarily close to the origin there are points for which $f(x,y) = 1/2$*, while* $f(0,0) = 0$*. On the other hand, along any straight line* $y = mx$ *the function satisfies*

$$f(x,mx) = \frac{mx^3}{x^2(x^2+m^2)} = \frac{mx}{x^2+m^2}$$

so f *approaches* 0 *along straight lines. By one of your homework problems, however, all differentiable functions must be continuous, so we conclude that* f *is not differentiable at the origin. (We prove that differentiability implies continuity below!)*

The problem here, of course, is that the partials $\frac{\partial f}{\partial x}$ *and* $\frac{\partial f}{\partial y}$ *are not continuous at the origin. For example,* $\frac{\partial f}{\partial x}$ *approaches* 0 *along the parabola* $y = x^2$ *while it diverges to* $-\infty$ *along the line* $y = x$*. (Check this!)* ■

Remark 2.9 *The problem point* $(0,0)$ *isn't special. We could make any point a problem point, for example* $(1,5)$*, by translating the above example function by* $(1,5)$*, i.e. by considering* $f(x,y) = \frac{(x-1)^2(y-5)}{(x-1)^4+(y-5)^2}$ *when* $(x,y) \neq (1,5)$ *and* $f(0,0) = (0,0)$*.* ■

OK, so now we know that the mere existence of the partials $\frac{\partial f_i}{\partial x_j}\big|_\mathbf{a}$ of $f = (f_1, \dots, f_m)$ isn't enough to ensure the existence of $Df(\mathbf{a})$. What we need is the *continuity of the partials* $\frac{\partial f_i}{\partial x_j}$ *on a neighborhood of* $\mathbf{a}$. Let us prove this!

Theorem 2.10 *Let* $f : \mathbb{R}^n \to \mathbb{R}^m$*. If all the partial derivatives* $\frac{\partial f_i}{\partial x_j}\big|_\mathbf{a}$ *of* f *exist and are continuous at* $\mathbf{a}$*, then* f *is differentiable at* $\mathbf{a}$*.*

Proof: By Proposition 2.7 it is enough to prove this for the component functions f_i of f. Indeed, let f_i be a component function of f, and suppose it's partial derivatives all exist and are continuous in a neighborhood of $\mathbf{a}$. Then, since $\frac{\partial f_i}{\partial x_j}\big|_{\mathbf{a}}$ moves only in the jth coordinate direction, we need only $\mathbf{h}_j = (0, \ldots, h_j, \ldots, 0)$ in those directions. By the definition of continuity of $\frac{\partial f_i}{\partial x_j}\big|_{\mathbf{a}}$, for any $\varepsilon > 0$ we choose there is a $\delta > 0$ such that if $\|\mathbf{h}\| = |h_j| < \delta$ then

$$\frac{\left\| \frac{\partial f_i}{\partial x_j}\Big|_{\mathbf{a}+\mathbf{h}} - \frac{\partial f_i}{\partial x_j}\Big|_{\mathbf{a}} \right\|}{|h_j|} < \frac{\varepsilon}{n}$$

Let $\mathbf{h}$ be a point in $\mathbb{R}^n$, so that $\mathbf{h} = \mathbf{h}_1 + \cdots + \mathbf{h}_n$ using our notation above. By the Mean Value Theorem from Calc 1, the continuity of f and the existence of the jth partial implies the existence of a point $\mathbf{a} + \mathbf{h}_j + t_j\mathbf{e}_j$ between $\mathbf{a} + \mathbf{h}_j$ and $\mathbf{a} + \mathbf{h}_j + \mathbf{e}_j$ such that

$$f(\mathbf{a} + \mathbf{h}_j) - f(\mathbf{a}) = \left(\frac{\partial f_i}{\partial x_j}\Big|_{\mathbf{x}+\mathbf{h}_i+t_i\mathbf{e}_i} \right) h_j \tag{2.11}$$

(Note: in the jth coordinate, keeping all other coordinates fixed, f_j is a real-valued function of a single variable, so this works. Recall the MVT: If f is continuous on $[a, b]$ and differentiable on (a, b) then there is a point c between a and b such that $f(b) - f(a) = f'(c)(b - a)$!) As a consequence, we have

$$\begin{aligned}
\left\| f_i(\mathbf{a}+\mathbf{h}) - f_i(\mathbf{a}) - \begin{pmatrix} \frac{\partial f_i}{\partial x_1}\Big|_{\mathbf{a}} & \cdots & \frac{\partial f_i}{\partial x_j}\Big|_{\mathbf{a}} \end{pmatrix} \mathbf{h} \right\| &= \left\| f_i(\mathbf{a}+\mathbf{h}) - f_i(\mathbf{a}) - \sum_{j=1}^{n} \left(\frac{\partial f_i}{\partial x_j}\Big|_{\mathbf{a}} \right) h_j \right\| \\
&= \left\| \sum_{j=1}^{n} \left(\frac{\partial f_i}{\partial x_j}\Big|_{\mathbf{a}+\mathbf{h}+t_j\mathbf{e}_i} \right) h_j - \sum_{j=1}^{n} \left(\frac{\partial f_i}{\partial x_j}\Big|_{\mathbf{a}} \right) h_i \right\| \\
&\leq \sum_{j=1}^{n} \left\| \frac{\partial f_i}{\partial x_j}\Big|_{\mathbf{a}+\mathbf{h}+t_j\mathbf{e}_i} - \frac{\partial f_i}{\partial x_j}\Big|_{\mathbf{a}} \right\| |h_j| \\
&< \sum_{j=1}^{n} \frac{|h_j|\varepsilon}{n} \\
&\leq \|\mathbf{h}\|\varepsilon
\end{aligned}$$

where the first inequality is from factoring out $|h_j|$ and then using the triangle inequality, the second is by application of (2.11) for each j, and the third by observing that $|h_1| + \cdots + |h_n| \leq \sqrt{h_1^2 + \cdots + h_n^2} + \cdots + \sqrt{h_1^2 + \cdots + h_n^2} = n\|\mathbf{h}\|$. Dividing the above inequality through by $\|\mathbf{h}\|$ gives our desired inequality,

$$\frac{\left\| f_i(\mathbf{a}+\mathbf{h}) - f_i(\mathbf{a}) - \begin{pmatrix} \frac{\partial f_i}{\partial x_1}\Big|_{\mathbf{a}} & \cdots & \frac{\partial f_i}{\partial x_j}\Big|_{\mathbf{a}} \end{pmatrix} \mathbf{h} \right\|}{\|\mathbf{h}\|} < \varepsilon$$

We have thus demonstrated the limit

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{\left\| f_i(\mathbf{a}+\mathbf{h}) - f_i(\mathbf{a}) - \begin{pmatrix} \frac{\partial f_i}{\partial x_1}\Big|_{\mathbf{a}} & \cdots & \frac{\partial f_i}{\partial x_j}\Big|_{\mathbf{a}} \end{pmatrix} \mathbf{h} \right\|}{\|\mathbf{h}\|} = 0$$

which is the definition of differentiability, and moreover, in the course of the proof, we have also shown that $Df(\mathbf{a}) = \begin{pmatrix} \frac{\partial f_i}{\partial x_1}\Big|_{\mathbf{a}} & \cdots & \frac{\partial f_i}{\partial x_j}\Big|_{\mathbf{a}} \end{pmatrix}$ as well! ■

Conclusion: If we know that the component functions of $f = (f_1, \dots, f_m)$ are each *continuously* differentiable on a neighborhood of our point a in $\mathbb{R}^n$, then we know that the f_i, and therefore f itself, are differentiable, and the total derivative $Df(\mathrm{a})$ is in fact the $m \times n$ matrix of partial derivatives $\frac{\partial f_i}{\partial x_j}\big|_{\mathbf{a}}$!

Example 2.11 *Let $f : \mathbb{R}^3 \to \mathbb{R}^2$ be given by $f(x, y, z) = (x^2 + y - z,\ e^{xy} \sin z + xz)$. Then $f_1(x, y, z) = x^2 + y - z$ and $f_2(x, y, z) = e^{xy} \sin z + xz$ are each clearly continuously differentiable in each partial derivative (for example, $\frac{\partial f_1}{\partial x} = 2x$ is continuous on all of $\mathbb{R}^n$). Therefore, f is differentiable and, say at $\mathbf{a} = (1, 1, 2)$, we have*

$$\begin{aligned} Df(1,1,2) &= \begin{pmatrix} \frac{\partial f_1}{\partial x}\Big|_{(1,1,2)} & \frac{\partial f_1}{\partial y}\Big|_{(1,1,2)} & \frac{\partial f_1}{\partial z}\Big|_{(1,1,2)} \\ \frac{\partial f_2}{\partial x}\Big|_{(1,1,2)} & \frac{\partial f_2}{\partial y}\Big|_{(1,1,2)} & \frac{\partial f_2}{\partial z}\Big|_{(1,1,2)} \end{pmatrix} \\ &= \begin{pmatrix} 2x\big|_{(1,1,2)} & 1\big|_{(1,1,2)} & -1\big|_{(1,1,2)} \\ ye^{xy}\sin z + z\big|_{(1,1,2)} & xe^{xy}\sin z\big|_{(1,1,2)} & e^{xy}\cos z + x\big|_{(1,1,2)} \end{pmatrix} \\ &= \begin{pmatrix} 2 & 1 & -1 \\ e\sin 2 + 2 & e\sin 2 & e\cos 2 + 1 \end{pmatrix} \end{aligned}$$

Moreover, if $\mathbf{v} = (-1, 4, 2)$ is a vector in $\mathbb{R}^3$, then we can compute the directional derivative of f at $(1, 1, 2)$ in the direction of $\mathbf{v}$ by simple matrix multiplication:

$$\begin{aligned} D_{(-1,4,2)}f(1,1,2) &= Df(1,1,2)\begin{pmatrix} -1 \\ 4 \\ 2 \end{pmatrix} \\ &= \begin{pmatrix} 2 & 1 & -1 \\ e\sin 2 + 2 & e\sin 2 & e\cos 2 + 1 \end{pmatrix}\begin{pmatrix} -1 \\ 4 \\ 2 \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ 3e\sin 2 + 2e\cos 2 \end{pmatrix} \end{aligned}$$

■

2.5 Further Properties of the Total and Partial Derivative

Theorem 2.12 (Chain Rule I) *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^m \to \mathbb{R}^p$ be functions such that $g \circ f$ is defined (i.e. the image of f is contained in the domain of g). If f differentiable at $\mathbf{a} \in \mathbb{R}^n$ and g is differentiable at $\mathbf{b} = f(\mathbf{a})$, then their composite $g \circ f : \mathbb{R}^n \to \mathbb{R}^p$ is differentiable at $\mathbf{a}$ and their derivative is a matrix product, namely the product of their two respective total derivatives,*

$$D(g \circ f)(\mathbf{a}) = Dg\big(f(\mathbf{a})\big) \cdot Df(\mathbf{a}) \tag{2.12}$$

The components of the matrix $D(g \circ f)(\mathbf{a})$ in (2.12) may explicitly be given by the formulas:

$$\frac{\partial (g \circ f)_i}{\partial x_j}\bigg|_{\mathbf{a}} = \frac{\partial g_i}{\partial y_1}\bigg|_{\mathbf{b}} \frac{\partial f_1}{\partial x_j}\bigg|_{\mathbf{a}} + \cdots + \frac{\partial g_i}{\partial y_m}\bigg|_{\mathbf{b}} \frac{\partial f_m}{\partial x_j}\bigg|_{\mathbf{a}} \tag{2.13}$$

or, if we let $z_i := g_i(y_1, \dots, y_m)$ and $y_k := f_k(x_1, \dots, x_n)$,

$$\frac{\partial z_i}{\partial x_j} = \frac{\partial z_i}{\partial y_1}\frac{\partial y_1}{\partial x_j} + \cdots + \frac{\partial z_i}{\partial y_m}\frac{\partial y_m}{\partial x_j} \tag{2.14}$$

Theorem 2.13 (Clairaut: Equality of Mixed Partial Derivatives) *If $f : \mathbb{R}^n \to \mathbb{R}^m$ has twice continuously differentiable partial derivatives or equivalently if for all $1 \le i, j \le n$ the partial derivatives $\frac{\partial^2 f}{\partial x_i \partial x_j}$ and $\frac{\partial^2 f}{\partial x_j \partial x_i}$ exist on a neighborhood of a point $\mathbf{a}$ and are continuous at $\mathbf{a}$, then*

$$\left.\frac{\partial^2 f}{\partial x_i \partial x_j}\right|_{\mathbf{a}} = \left.\frac{\partial^2 f}{\partial x_i \partial x_j}\right|_{\mathbf{a}} \tag{2.15}$$

for all $1 \le i, j \le n$.

Remark 2.14 *Failure of continuity at $\mathbf{a}$ may lead to inequality of the mixed partials at $\mathbf{a}$. Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ given by*

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

Then,

$$\begin{aligned}
\frac{\partial f}{\partial x} &= \begin{cases} \dfrac{(x^2+y^2)(3x^2y - y^3) - 2x(x^3y - xy^3)}{(x^2+y^2)^4}, & \textit{if } (x,y) \neq (0,0) \\ 0, & \textit{if } (x,y) = (0,0) \end{cases} \\
&= \begin{cases} \dfrac{3x^4y - x^2y^3 + 3x^2y^3 - y^5 - 2x^4y + 2x^2y^3}{(x^2+y^2)^4}, & \textit{if } (x,y) \neq (0,0) \\ 0, & \textit{if } (x,y) = (0,0) \end{cases} \\
&= \begin{cases} \dfrac{x^4y + 4x^2y^3 - y^5}{(x^2+y^2)^4}, & \textit{if } (x,y) \neq (0,0) \\ 0, & \textit{if } (x,y) = (0,0) \end{cases}
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial f}{\partial y} &= \begin{cases} \dfrac{(x^2+y^2)(x^3 - 3xy^2) - 2y(x^3y - xy^3)}{(x^2+y^2)^4}, & \textit{if } (x,y) \neq (0,0) \\ 0, & \textit{if } (x,y) = (0,0) \end{cases} \\
&= \begin{cases} \dfrac{x^5 - 3x^3y^2 + x^3y^2 - 3xy^4}{(x^2+y^2)^4}, & \textit{if } (x,y) \neq (0,0) \\ 0, & \textit{if } (x,y) = (0,0) \end{cases} \\
&= \begin{cases} \dfrac{x^5 - 4x^3y^2 - xy^4}{(x^2+y^2)^4}, & \textit{if } (x,y) \neq (0,0) \\ 0, & \textit{if } (x,y) = (0,0) \end{cases}
\end{aligned}$$

Therefore, for $a, b \neq 0$ we have

$$\left.\frac{\partial f}{\partial x}\right|_{(0,b)} = \frac{-5b^5}{b^4} = -b \qquad\qquad \left.\frac{\partial f}{\partial y}\right|_{(a,0)} = \frac{a^5}{a^4} = a$$

and consequently

$$\begin{aligned}
\left.\frac{\partial^2 f}{\partial y \partial x}\right|_{(0,0)} &= \lim_{t\to 0} \frac{\left.\frac{\partial f}{\partial x}\right|_{(0,t)} - \left.\frac{\partial f}{\partial x}\right|_{(0,0)}}{t} = \lim_{t\to 0} \frac{-t-0}{t} = -1 \\
\left.\frac{\partial^2 f}{\partial x \partial y}\right|_{(0,0)} &= \lim_{t\to 0} \frac{\left.\frac{\partial f}{\partial y}\right|_{(t,0)} - \left.\frac{\partial f}{\partial y}\right|_{(0,0)}}{t} = \lim_{t\to 0} \frac{t-0}{t} = 1
\end{aligned}$$

and so $\left.\frac{\partial^2 f}{\partial y \partial x}\right|_{(0,0)} \neq \left.\frac{\partial^2 f}{\partial x \partial y}\right|_{(0,0)}$. *The problem, of course, is the discontinuity of the second derivatives at* $(0,0)$*:*

$$\frac{\partial^2 f}{\partial x \partial y} = \begin{cases} \frac{(x^2+y^2)^2(5x^4-12x^2y^2-y^4)-2(x^2y^2)2x(x^5-4x^3y^2-xy^4)}{(x^2+y^2)^4}, & \textit{if } (x,y) \neq (0,0) \\ 0, & \textit{if } (x,y) = (0,0) \end{cases}$$

$$\frac{\partial^2 f}{\partial y \partial x} = \begin{cases} \frac{(x^2+y^2)^2(x^4+12x^2y^2-5y^4)-2(x^2y^2)2y(x^4y+4x^2y^3-y^5)}{(x^2+y^2)^4}, & \textit{if } (x,y) \neq (0,0) \\ 0, & \textit{if } (x,y) = (0,0) \end{cases}$$

For example, along the line $x = y$ *we have* $\frac{\partial^2 f}{\partial x \partial y} = 2(1-x)$*, so it approaches a value of* 2*, while along the line* $x = 0$ *it stays constant at* 1*, as noted above.* ■