This chapter is about the derivative of a function \( f : \mathbb{R}^n \to \mathbb{R}^m \). First, though, we review some linear algebra.

## 1 Linear transformations and their matrix representations

**Definition 1** A linear transformation from \( \mathbb{R}^n \) to \( \mathbb{R}^m \) is a function \( L : \mathbb{R}^n \to \mathbb{R}^m \) such that for any \( x \) and \( y \) in \( \mathbb{R}^n \), and any scalars \( \alpha \) and \( \beta \),

\[
L (\alpha x + \beta y) = \alpha L (x) + \beta L (y).
\]

We can associate \( L \) with a matrix once we have chosen bases for \( \mathbb{R}^n \) and \( \mathbb{R}^m \). Recall that a basis for \( \mathbb{R}^n \) space is a set \( \{v_1, \ldots, v_n\} \) of vectors in \( \mathbb{R}^n \) which is linearly independent and spans the space. Most commonly we use the standard bases,

\[
\begin{pmatrix} 1 \\ 0 \\ \cdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ \cdots \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} 0 \\ 0 \\ \cdots \\ 1 \end{pmatrix}
\]

for each space. Of course the \( n \) basis vectors for \( \mathbb{R}^n \) have \( n \) components and the \( m \) basis vectors for \( \mathbb{R}^m \) have \( m \) components.

As an example, consider the linear transformation from \( \mathbb{R}^3 \) to \( \mathbb{R}^2 \) defined by

\[
L \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x + 2y + 3z \\ 4x + 5y + 6z \end{pmatrix}.
\]

The standard basis for \( \mathbb{R}^3 \) is \( \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\} \) and the standard basis for \( \mathbb{R}^2 \) is \( \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\} \) and the matrix for \( L \) with respect to these bases is the “obvious” matrix

\[
A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}.
\]
But we should keep in mind that $L$ is not the same thing as $A$. In order to emphasize this we will compute that matrix of $L$ with respect to another pair of bases. For $R^3$ we choose the basis
\[
\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \\ \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \\ \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \\ \end{pmatrix} \right\}.
\]

For $R^2$ we choose the basis
\[
\left\{ \begin{pmatrix} 1 \\ 0 \\ \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ \end{pmatrix} \right\}.
\]

Suppose that $x \in R^3$ and $x = c_1v_1 + c_2v_2 + c_3v_3$, where $\{v_1, v_2, v_3\}$ is the basis chosen for $R^3$. And suppose that $Lx = d_1w_1 + d_2w_2$, where $\{w_1, w_2\}$ is the basis chosen for $R^2$. Then the matrix for $L$ with respect to these two bases is the matrix $B$ such that
\[
\begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = B \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}.
\]

In other words, the matrix finds the coefficients of $Lx$ with respect to the chosen basis for $R^2$ in terms of the coefficients of $x$ with respect to the chosen basis for $R^3$.

In your linear algebra text you can probably find the formula:
\[
B = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^{-1} A \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}
= \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 4 & 9 & 15 \\ -3 & -6 & -9 \end{pmatrix}.
\]

We can check this on an example: Let $x = 2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \end{pmatrix} + 3 \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 2 \end{pmatrix}$. Then
\[
Lx = \begin{pmatrix} 17 \\ 44 \end{pmatrix} = 44 \begin{pmatrix} 1 \\ 1 \end{pmatrix} - 27 \begin{pmatrix} 1 \\ 0 \end{pmatrix}.
\]
And
\[
\begin{pmatrix} 4 & 9 & 15 \\ -3 & -6 & -9 \end{pmatrix} \begin{pmatrix} 2 \\ -1 \\ 3 \end{pmatrix} = \begin{pmatrix} 44 \\ -27 \end{pmatrix}
\]

We will encounter linear transformations that are not defined originally by a matrix; the matrix comes later after the basis is specified.
2 The definition of Derivative

Derivatives have not appeared previously in this book, but from earlier classes you should be familiar with the formula

\[ f'(x) = \lim_{h \to 0} \frac{f(x + h) - f(x)}{h}, \]  

(1)

though perhaps you used different symbols, such as “\( \Delta x \)” for “\( h \)”. This is for a function \( f : \mathbb{R} \to \mathbb{R} \). Another version of this is probably also familiar:

\[ f'(x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}. \]

But either of these formulas is rather misleading when it comes to considering \( f : \mathbb{R}^n \to \mathbb{R}^m \). Whichever formula we use, the message that comes across from the \( n = m = 1 \) case is that the derivative of \( f \) at a point \( x_0 \) is a number. True, \( f' \) is a function, but for each \( x_0 \), \( f'(x_0) \) is a number. As an example, if \( f(x) = x^2 \) then \( f'(2) = 4 \).

Now, however, consider the following function from \( \mathbb{R}^2 \) to \( \mathbb{R}^1 \):

\[ f(x_1, x_2) = x_1^2 + 2x_2^2. \]  

(2)

Both of the formulas above are difficult to interpret, because now \( x \in \mathbb{R}^2 \), and so to make sense of the terms in these formula, such as \( x + h \), both \( h \) and \( x_0 \) must be also be in \( \mathbb{R}^2 \). And we don’t know how to divide by an element of \( \mathbb{R}^2 \).

To get around this difficulty, we recall that one important interpretation of the derivative in \( \mathbb{R}^1 \) is as the slope of a tangent line. If \( f(x) = x^2 \), then \( f'(2) \) is the slope of the line which is tangent to the graph of \( f \) at the point \((2, 4)\). We can easily write down the equation of this tangent line as \( y = 4x - 4 \). The \(-4\) insures that this line passes through the point \((2, 4)\).

So now we look at the graph of the function \( f \) in equation (2). This is a surface, \( x_3 = x_1^2 + 2x_2^2 \). One point on this surface is \((1, 2, 9)\). We can ask: what is the equation of the plane which is tangent to the surface at this point?

You probably saw this in Calc 3. To answer it we use partial derivatives. We have \( \frac{\partial x_3}{\partial x_1} = 2x_1 \) and \( \frac{\partial x_3}{\partial x_2} = 4x_2 \), and substituting \( x_1 = 1, x_2 = 2 \) we get the “gradient”:

\[ \left( \frac{\partial x_3}{\partial x_1}, \frac{\partial x_3}{\partial x_2} \right) = (2, 8). \]

The equation of the tangent plane is then

\[ x_3 = 2(x_1 - 1) + 8(x_2 - 2) + 9, \]

which is the equation of a plane which passes through the point \((1, 2, 9)\).
So we ask: what is the derivative of the function \( f(x_1, x_2) = x_1^2 + 2x_2^2 \) at the point \((1, 2)\). And the answer is: The derivative is the linear function of two new variables \( u \) and \( v \) defined by the formula

\[
Df(1, 2)(u, v) = 2u + 8v.
\]

You can think of \( u \) and \( v \) as representing \((x_1 - 1)\) and \((x_2 - 2)\), and aside from adding 9 to get the point \((1, 2, 9)\), the derivative is the function defining the tangent plane at this point.

And now we come to the formula for derivative (Definition 6.1.1 in the text). But we need a little more motivation. We will rewrite the one dimensional formula above to get

\[
\lim_{x \to x_0} \left( \frac{f(x) - f(x_0)}{x - x_0} - f'(x_0) \right) = 0,
\]

and rewriting again, as

\[
\lim_{x \to x_0} \frac{f(x) - f(x_0) - f'(x_0)(x - x_0)}{x - x_0} = 0.
\]

Still, this wouldn’t extend easily to higher dimensions, because we are dividing by \( x - x_0 \), and we don’t know how to do this when \( x \) and \( x_0 \) are vectors. But the limit above is zero if and only if

\[
\lim_{x \to x_0} \frac{|f(x) - f(x_0) - f'(x_0)(x - x_0)|}{|x - x_0|} = 0.
\]

Now we need only a small change in order to interpret this in higher dimensions:

**Definition 2** The derivative of a function \( f : \mathbb{R}^n \to \mathbb{R}^m \) at a point \( x_0 \) is a linear transformation \( Df(x_0) : \mathbb{R}^n \to \mathbb{R}^m \) such that

\[
\lim_{x \to x_0} \frac{||f(x) - f(x_0) - Df(x_0)(x - x_0)||}{||x - x_0||} = 0. \tag{3}
\]

To make sense of this, notice that \( Df(x_0) \) is a function defined on \( \mathbb{R}^n \), so by \( Df(x_0)(x - x_0) \) we mean this function evaluated at the vector \( x - x_0 \). It is a linear function, or a linear transformation, from \( \mathbb{R}^n \) to \( \mathbb{R}^m \), so after agreeing on a basis for each space, usually the standard bases, we can think of it as a matrix. You can probably guess the matrix with respect to the standard basis. If \( x = (x_1, \ldots, x_n) \) and \( f = (f_1, \ldots, f_m) \), then the matrix is

\[
Df(x_0) = \left( \frac{\partial f_i}{\partial x_j}(x_0) \right),
\]

4
which we call the “Jacobian of \( f \) at \( x_0 \).” It is an \( m \times n \) matrix, as befits a linear transformation from \( \mathbb{R}^n \) to \( \mathbb{R}^m \).

We should test this out on the example above, namely \( f(x_1, x_2) = x_1^2 + 2x_2^2 \). Then \( n = 2, m = 1 \). It is convenient here to let \( x_0 = (x_1, x_2) \) and to use \( (u, v) \) as the argument to which we apply the function \( Df(x_1, x_2) \). We switch to vector and matrix notation in the 2nd expression below:

\[
Df(x_1, x_2)(u, v) = (2x_1, 4x_2) \left( \begin{array}{c} u \\ v \end{array} \right) = 2x_1u + 4x_2v.
\]

This shows that when \( m = 1 \), the matrix for \( Df(x_0) \) is the gradient of \( f \) at \( x_0 \). With \( x_0 = (1, 2) \), we get

\[
Df(x_0) = (2, 8),
\]

which is a \( 1 \times 2 \) matrix. Now we want to check the equation (3).

We are lucky here, because the norm \( || \cdot || \) in the numerator is fairly simple. That is because \( m = 1 \), so the term inside \( || \cdot || \) on the top is a scalar. To make sense of the second term we have to write “\( x - x_0 \)” as \( \left( \begin{array}{c} x_1 - 1 \\ x_2 - 2 \end{array} \right) \). We get

\[
x_1^2 + 2x_2^2 - 1 - 8 - 2(x_1 - 1) - 8(x_2 - 2)
\]

which simplifies to \( (x_1 - 1)^2 + 2(x_2 - 2)^2 \). If we let \( u = x_1 - 1, v = x_2 - 2 \), then the ratio in (3) becomes

\[
\frac{u^2 + 2v^2}{\sqrt{u^2 + v^2}} \leq 2\sqrt{u^2 + v^2}
\]

which obviously tends to zero as \( (u, v) \to (0, 0) \). Thus we have verified that \( Df(1, 2) = (2, 8) \).

### 3 6.2; Matrix representation

As discussed earlier, each linear transformation from \( \mathbb{R}^n \) to \( \mathbb{R}^m \) has a matrix representation relative to a choice of bases for the two spaces. By far the most frequent case is when the standard basis is used for both \( \mathbb{R}^n \) and \( \mathbb{R}^m \). In this case the matrix representation is easy. Theorem 6.2.2 tells us that if \( Df(x) \) exists and is continuous in an open set, then the matrix representation of \( Df(x) \) in this set is the Jacobian matrix

\[
\begin{pmatrix}
\frac{\partial f_1}{\partial x_1}(x) & \frac{\partial f_1}{\partial x_2}(x) & \cdots & \frac{\partial f_1}{\partial x_n}(x) \\
\frac{\partial f_2}{\partial x_1}(x) & \frac{\partial f_2}{\partial x_2}(x) & \cdots & \frac{\partial f_2}{\partial x_n}(x) \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial f_m}{\partial x_1}(x) & \cdots & \cdots & \frac{\partial f_m}{\partial x_n}(x)
\end{pmatrix}
\]
As an example, let \( f(x_1, x_2, x_3) = (2x_1x_3^2, x_2 - x_1) \). Thus, \( f : \mathbb{R}^3 \to \mathbb{R}^2 \). Then
\[
Df(1, 1, 1) (u, v, w) = \left( \begin{array}{ccc}
2x_3^2 & 0 & 4x_3x_1 \\
-1 & 1 & 0
\end{array} \right) \left( \begin{array}{c}
u \\
v \\
w
\end{array} \right) 
= \left( \begin{array}{c}
2u + 4w \\
-u + v
\end{array} \right).
\]

\[\text{(5)}\]

4 6.3; Differentiability implies continuity.

This is obvious in one dimension, since \( \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x-x_0} \) cannot approach a limit unless the numerator approaches zero. For the case of \( f : \mathbb{R}^n \to \mathbb{R}^m \) it is just as obvious; recall that \( f \) is differentiable at \( x_0 \) with derivative \( Df(x_0) \) if
\[
\lim_{x \to x_0} \frac{||f(x) - f(x_0) - Df(x_0)(x-x_0)||}{||x-x_0||} = 0.
\]
Assuming this is true, the second term in the numerator tends to zero since \( Df(x_0) \) is a linear function and by definition of linearity, \( Df(x_0)(0) = 0 \). Therefore, the limit of the ratio cannot be zero unless \( \lim_{x \to x_0} f(x) - f(x_0) = 0 \), which is the definition of continuity at \( x_0 \).

The text then discusses the case \( c : \mathbb{R} \to \mathbb{R}^m \). In this case, \( c \) is called a curve. Often, the image of \( c \), which is a set of points in \( \mathbb{R}^n \) is also called a curve, with it being understood that there is some function \( c \) involved. (The text says that \( c \) “represents” a curve, but what “represents” means is not made clear. I prefer to say that \( c \) is the curve, though this has the drawback that different functions can have the same image set, which sometimes leads to confusion.)

For the case of a curve \( c : \mathbb{R} \to \mathbb{R}^m \), differentiability is simpler to discuss. There are \( n \) coordinate functions, so
\[
c(t) = \left( \begin{array}{c}
c_1(t) \\
c_2(t) \\
\vdots \\
c_m(t)
\end{array} \right),
\]
and it is easy to show that \( c \) is differentiable to \( t_0 \) if and only if \( c_1, \ldots, c_m \) are all differentiable at \( t_0 \). If \( c \) is differentiable at \( t_0 \), then we can find the value in \( \mathbb{R}^m \) of

1Indeed, more often, but not in analysis except informally.
the linear function $Dc(t_0)$, which is

$$Dc(t_0)(u) = \begin{pmatrix} c_1'(t_0) u \\ c_2'(t_0) u \\ \vdots \\ c_m'(t_0) u \end{pmatrix} = u \begin{pmatrix} c_1'(t_0) \\ c_2'(t_0) \\ \vdots \\ c_m'(t_0) \end{pmatrix}.$$ 

In the text it is stated that $Dc(t)$ is “represented by” the column vector $\begin{pmatrix} c_1'(t) \\ c_2'(t) \\ \vdots \\ c_m'(t) \end{pmatrix}$.

Interpreting all this in the image space $R^m$, this vector is tangent to the image of $c$ at $c(t)$. In fact, we usually say, a bit sloppily, that the vector $Dc(t)$ is tangent to $c$ at the point $c(t)$.

**Definition 3** A curve $c$ is “smooth” at $t_0$ if it is differentiable at $t_0$ and $Dc(t_0)$ is not the zero vector.

It is useful to look at the image of a non-smooth curve, such as

$$c(t) = (t^2, t^3).$$

For this example, $Dc(0) = (0, 0)$ (using a row vector to save space), and so $c$ is not smooth at $t_0 = 0$. And a plot shows that the image of the curve doesn’t look smooth at $c(0) = (0,0)$:

![Graph of a curve](image)

On the other hand, $c(t) = (t^3, t^3)$ is also not smooth at $(0,0)$, and yet its image looks perfectly smooth, being a straight line.

**4.1 Functions which have partial derivatives at a point but are not differentiable, or even continuous.**

The text gives a number of these. One of the simplest is:

$$f(x,y) = \begin{cases} 
  x & \text{if } y = 0 \\
  y & \text{if } x = 0 \\
  1 & \text{if } x \neq 0 \text{ and } y \neq 0.
\end{cases}$$
The partial derivatives exist because \( f(x,0) = 0 \) and \( f(0,y) = 0 \). But this function isn’t even continuous at \((0,0)\).

Another example is related to the idea of a “directional derivative”. This is defined for a function from \( \mathbb{R}^n \) to \( \mathbb{R} \).

**Definition 4** If \( f : \Omega \to \mathbb{R} \) where \( \Omega \) is an open subset of \( \mathbb{R}^n \), and \( x_0 \in \Omega \), then the directional derivative of \( f \) in the direction of a unit vector \( e \) is given by

\[
\lim_{t \to 0} \frac{f(x_0 + te) - f(x_0)}{t}.
\]

Here is a picture when \( n = 2 \), \( x_0 = (0,0) \), and \( e = \frac{1}{\sqrt{5}} (2, 1) \). The line is in the direction of \( e \). If \( g(t) = f(x_0 + te) \), then \( g \) is the value of \( f \) as a function of distance along the line, and the directional derivative in the direction of \( e \) is \( g'(0) \).

It is shown in the text that if \( f \) is differentiable at \( x_0 \), then \( g'(t) = Df(x_0) e \). But all of the directional derivatives (i.e. in every direction) can exist even if \( f \) is not differentiable. Examples are given on page 343 and will be discussed in class.

## 5 Homework, due January 19 at the beginning of class

1. pg. 330, # 3.

2. pg. 334, #4, But in the first part find \( Df(1,2)(u,v) \) and use \((x,y) = (1,2)\) in the second part as well. It is assumed that you are using the given basis for both \( \mathbb{R}^n \) and \( \mathbb{R}^m \), since in this example \( n = m = 2 \).

3. pg. 344 # 2. Also investigate the existence of the directional derivatives of \( f \) at \((0,0)\) in every direction.
4. pg. 384, #5 b,f.
5. pg. 389, #40 a.
6. pg. 388, #36.