MULTIVARIABLE AND VECTOR ANALYSIS

W W L CHEN

© W W L Chen, 1997, 2008.
This chapter is available free to all individuals, on the understanding that it is not to be used for financial gain, and may be downloaded and/or photocopied, with or without permission from the author.
However, this document may not be kept on any information storage and retrieval system without permission from the author, unless such system is not accessible to any individuals other than its owners.

Chapter 2

DIFFERENTIATION

2.1. Partial Derivatives

The notion of differentiability for vector valued functions of several variables is more complicated than one might expect. As a first step, we consider partial derivatives of real valued functions of several variables.

Definition. Consider a function of the form $f : A \to \mathbb{R}$, where $A \subseteq \mathbb{R}^n$ is an open set. For every $\mathbf{x} = (x_1, \ldots, x_n) \in A$ and every $j = 1, \ldots, n$, the limit

$$\frac{\partial f}{\partial x_j}(x_1, \ldots, x_n) = \lim_{h \to 0} \frac{f(x_1, \ldots, x_{j-1}, x_j + h, x_{j+1}, \ldots, x_n) - f(x_1, \ldots, x_n)}{h},$$

if it exists, is called the j-th partial derivative of f.

Remark. If for every $j = 1, \ldots, n$, we write

$$\mathbf{e}_j = (\underbrace{0, \ldots, 0}_{j-1}, 1, \underbrace{0, \ldots, 0}_{n-j}),$$

then

$$\frac{\partial f}{\partial x_j}(x_1, \ldots, x_n) = \lim_{h \to 0} \frac{f(\mathbf{x} + h\mathbf{e}_j) - f(\mathbf{x})}{h}.$$

Example 2.1.1. Suppose that $f : \mathbb{R}^2 \to \mathbb{R}$, where $f(x, y) = \sin xy + x \cos y$ for every $(x, y) \in \mathbb{R}^2$. Then

$$\frac{\partial f}{\partial x} = y \cos xy + \cos y \qquad \text{and} \qquad \frac{\partial f}{\partial y} = x \cos xy - x \sin y.$$

EXAMPLE 2.1.2. Suppose that $f : \mathbb{R}^2 \to \mathbb{R}$, where

$$f(x,y) = \frac{xy}{\sqrt{x^2+y^2}}$$

for every $(x,y) \in \mathbb{R}^2$. Then

$$\frac{\partial f}{\partial x} = \frac{y^3}{(x^2+y^2)^{3/2}} \qquad \text{and} \qquad \frac{\partial f}{\partial y} = \frac{x^3}{(x^2+y^2)^{3/2}}.$$

EXAMPLE 2.1.3. Suppose that $f : \mathbb{R}^2 \to \mathbb{R}$, where $f(x,y) = x^{1/5}y^{1/5}$ for every $(x,y) \in \mathbb{R}^2$. Then to obtain the partial derivatives at $(0,0)$, we cannot simply differentiate and substitute $(x,y) = (0,0)$. In fact, we can work from first definitions that

$$\frac{\partial f}{\partial x}(0,0) = \lim_{h\to 0} \frac{f(h,0)-f(0,0)}{h} = 0 \qquad \text{and} \qquad \frac{\partial f}{\partial y}(0,0) = \lim_{h\to 0} \frac{f(0,h)-f(0,0)}{h} = 0.$$

Let us study the restriction of the function to the plane $x = y$. Try to draw a picture to convince yourself that the function $f(x,x) = x^{2/5}$ is not differentiable at $x = 0$. It follows that the graph of the function $f(x,y) = x^{1/5}y^{1/5}$ does not have a tangent plane at $(x,y) = (0,0)$. This suggests that differentiability of a function at a point has to be more than just the existence of partial derivatives at that point.

2.2. Total Derivatives

Next, we turn to the idea of differentiability at a point. To motivate this, let us consider the following two examples.

EXAMPLE 2.2.1. Suppose that $f : A \to \mathbb{R}$, where $A \subseteq \mathbb{R}$, is differentiable at a point x_0. Consider the tangent to the curve of f at the point $(x_0, f(x_0))$. The slope of this tangent is $f'(x_0)$, and it is easy to see that the equation of the tangent is the line

$$y = f(x_0) + f'(x_0)(x - x_0).$$

It follows that as $x \to x_0$, the quantity $f(x_0) + f'(x_0)(x - x_0)$ is a good approximation of the function $f(x)$. On the other hand, note that

$$\lim_{x\to x_0} \frac{f(x)-f(x_0)}{x-x_0} = f'(x_0),$$

so that

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

whence

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

EXAMPLE 2.2.2. Suppose that $f : A \to \mathbb{R}$, where $A \subseteq \mathbb{R}^2$. Suppose further that $(x_0, y_0) \in A$ and that there is a tangent plane to the surface of f at $(x_0, y_0, f(x_0, y_0))$. The equation of a non-vertical plane is of the form $z = ax + by + c$, and it is not difficult to show that the tangent plane must be of the form

$$z = f(x_0,y_0) + \left(\frac{\partial f}{\partial x}(x_0,y_0)\right)(x-x_0) + \left(\frac{\partial f}{\partial y}(x_0,y_0)\right)(y-y_0).$$

It follows that as $(x, y) \to (x_0, y_0)$, the quantity

$$f(x_0, y_0) + \left(\frac{\partial f}{\partial x}(x_0, y_0)\right)(x - x_0) + \left(\frac{\partial f}{\partial y}(x_0, y_0)\right)(y - y_0)$$

is a good approximation of the function $f(x, y)$. If we take into consideration Example 2.2.1, we may perhaps wish to say that f is differentiable at (x_0, y_0) if $\partial f/\partial x$ and $\partial f/\partial y$ exist at (x_0, y_0) and if

$$\lim_{(x,y)\to(x_0,y_0)} \frac{\left| f(x, y) - f(x_0, y_0) - \left(\frac{\partial f}{\partial x}(x_0, y_0)\right)(x - x_0) - \left(\frac{\partial f}{\partial y}(x_0, y_0)\right)(y - y_0) \right|}{\|(x, y) - (x_0, y_0)\|} = 0. \qquad (1)$$

Now write

$$(\mathbf{D}f)(x_0, y_0) = \begin{pmatrix} \frac{\partial f}{\partial x}(x_0, y_0) & \frac{\partial f}{\partial y}(x_0, y_0) \end{pmatrix}$$

as a row matrix. Then with a slight abuse of notation, we have

$$f(x_0, y_0) + \left(\frac{\partial f}{\partial x}(x_0, y_0)\right)(x - x_0) + \left(\frac{\partial f}{\partial y}(x_0, y_0)\right)(y - y_0) = f(x_0, y_0) + (\mathbf{D}f)(x_0, y_0)\begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix},$$

and so (1) can be written in the form

$$\lim_{(x,y)\to(x_0,y_0)} \frac{\left| f(x, y) - f(x_0, y_0) - (\mathbf{D}f)(x_0, y_0)\begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix} \right|}{\|(x, y) - (x_0, y_0)\|} = 0.$$

We now try to generalize our discussion so far to the case of real valued functions of several real variables.

Definition. Consider a function of the form $f : A \to \mathbb{R}$, where $A \subseteq \mathbb{R}^n$ is an open set. We say that f is differentiable at $\mathbf{x}_0 \in A$ if all partial derivatives

$$\frac{\partial f}{\partial x_j}, \qquad \text{where } j = 1, \ldots, n,$$

exist and if

$$\lim_{\mathbf{x}\to\mathbf{x}_0} \frac{\|f(\mathbf{x}) - f(\mathbf{x}_0) - (\mathbf{D}f)(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)\|}{\|\mathbf{x} - \mathbf{x}_0\|} = 0,$$

where $(\mathbf{D}f)(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)$ denotes the matrix product of

$$(\mathbf{D}f)(\mathbf{x}_0) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}_0) & \ldots & \frac{\partial f}{\partial x_n}(\mathbf{x}_0) \end{pmatrix}$$

with the vector $\mathbf{x} - \mathbf{x}_0$ regarded as a column matrix.

The generalization to vector valued functions is now rather straightforward.

Definition. Consider a function of the form $f : A \to \mathbb{R}^m$, where $A \subseteq \mathbb{R}^n$ is an open set. Suppose that $f(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_m(\mathbf{x}))$ for every $\mathbf{x} \in A$. Then we say that f is differentiable at $\mathbf{x}_0 \in A$ if $f_i : A \to \mathbb{R}$ is differentiable at $\mathbf{x}_0 \in A$ for every $i = 1, \ldots, m$. We say that $f : A \to \mathbb{R}^m$ is differentiable if f is differentiable at every $\mathbf{x}_0 \in A$.

DEFINITION. Consider a function of the form $f : A \to \mathbb{R}^m$, where $A \subseteq \mathbb{R}^n$ is an open set. Suppose that $f(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_m(\mathbf{x}))$ for every $\mathbf{x} \in A$. Then the total derivative of f at $\mathbf{x}_0 \in A$ is defined to be the $m \times n$ matrix

$$\mathbf{T} = (\mathbf{D}f)(\mathbf{x}_0) = \begin{pmatrix} \dfrac{\partial f_1}{\partial x_1}(\mathbf{x}_0) & \ldots & \dfrac{\partial f_1}{\partial x_n}(\mathbf{x}_0) \\ \vdots & & \vdots \\ \dfrac{\partial f_m}{\partial x_1}(\mathbf{x}_0) & \ldots & \dfrac{\partial f_m}{\partial x_n}(\mathbf{x}_0) \end{pmatrix} \tag{2}$$

if all the partial derivatives exist. The matrix $\mathbf{T}$ is also called the matrix of partial derivatives.

REMARK. Consider a function of the form $f : A \to \mathbb{R}^m$, where $A \subseteq \mathbb{R}^n$ is an open set. It can be shown that f is differentiable at $\mathbf{x}_0 \in A$ if and only if all partial derivatives

$$\frac{\partial f_i}{\partial x_j}, \qquad \text{where } i = 1, \ldots, m \text{ and } j = 1, \ldots, n,$$

exist and

$$\lim_{\mathbf{x} \to \mathbf{x}_0} \frac{\|f(\mathbf{x}) - f(\mathbf{x}_0) - \mathbf{T}(\mathbf{x} - \mathbf{x}_0)\|}{\|\mathbf{x} - \mathbf{x}_0\|} = 0, \tag{3}$$

where $\mathbf{T}(\mathbf{x} - \mathbf{x}_0)$ denotes the matrix product of $\mathbf{T}$ given by (2) with the vector $\mathbf{x} - \mathbf{x}_0$ regarded as a column matrix. To see this, note first of all that the existence of the partial derivatives is obvious. Next, note that for any vector $\mathbf{y} = (y_1, \ldots, y_m) \in \mathbb{R}^m$, we have $|y_i| \le \|\mathbf{y}\|$ for every $i = 1, \ldots, m$, and $\|\mathbf{y}\| \le |y_1| + \ldots + |y_m|$. Note now that

$$f(\mathbf{x}) - f(\mathbf{x}_0) - \mathbf{T}(\mathbf{x} - \mathbf{x}_0) = \begin{pmatrix} f_1(\mathbf{x}) - f_1(\mathbf{x}_0) - (\mathbf{D}f_1)(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) \\ \vdots \\ f_m(\mathbf{x}) - f_m(\mathbf{x}_0) - (\mathbf{D}f_m)(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) \end{pmatrix},$$

so that

$$|f_i(\mathbf{x}) - f_i(\mathbf{x}_0) - (\mathbf{D}f_i)(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)| \le \|f(\mathbf{x}) - f(\mathbf{x}_0) - \mathbf{T}(\mathbf{x} - \mathbf{x}_0)\|$$

for every $i = 1, \ldots, m$, and

$$\|f(\mathbf{x}) - f(\mathbf{x}_0) - \mathbf{T}(\mathbf{x} - \mathbf{x}_0)\| \le \sum_{i=1}^{m} |f_i(\mathbf{x}) - f_i(\mathbf{x}_0) - (\mathbf{D}f_i)(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)|.$$

EXAMPLE 2.2.3. Consider the function $f : \mathbb{R}^3 \to \mathbb{R}^2 : (x, y, z) \mapsto (x \sin y, x + \sin z)$. In this case, we have $f(x, y, z) = (f_1(x, y, z), f_2(x, y, z))$, where $f_1(x, y, z) = x \sin y$ and $f_2(x, y, z) = x + \sin z$ for every $(x, y, z) \in \mathbb{R}^3$. It follows that

$$\mathbf{D}f = \begin{pmatrix} \dfrac{\partial f_1}{\partial x} & \dfrac{\partial f_1}{\partial y} & \dfrac{\partial f_1}{\partial z} \\ \dfrac{\partial f_2}{\partial x} & \dfrac{\partial f_2}{\partial y} & \dfrac{\partial f_2}{\partial z} \end{pmatrix} = \begin{pmatrix} \sin y & x \cos y & 0 \\ 1 & 0 & \cos z \end{pmatrix}.$$

EXAMPLE 2.2.4. Consider the function $f : A \to \mathbb{R}^2 : (r, \theta) \mapsto (r\cos\theta, r\sin\theta)$, where

$$A = \{(r, \theta) \in \mathbb{R}^2 : r > 0\}.$$

In this case, we have $f(r, \theta) = (f_1(r, \theta), f_2(r, \theta))$, where $f_1(r, \theta) = r\cos\theta$ and $f_2(r, \theta) = r\sin\theta$ for every $(r, \theta) \in A$. It follows that

$$\mathbf{D}f = \begin{pmatrix} \dfrac{\partial f_1}{\partial r} & \dfrac{\partial f_1}{\partial \theta} \\ \dfrac{\partial f_2}{\partial r} & \dfrac{\partial f_2}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}.$$

REMARK. Consider the special case $m = 1$. Then

$$\mathbf{T} = (\mathbf{D}f)(\mathbf{x}_0) = \begin{pmatrix} \dfrac{\partial f}{\partial x_1} & \dots & \dfrac{\partial f}{\partial x_n} \end{pmatrix}.$$

The corresponding vector

$$\left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right)$$

is called the gradient of f and denoted by ∇f or $\operatorname{grad} f$. For $f : \mathbb{R}^2 \to \mathbb{R}$ and $f : \mathbb{R}^3 \to \mathbb{R}$, we can use the special notation

$$\nabla f = \frac{\partial f}{\partial x}\mathbf{i} + \frac{\partial f}{\partial y}\mathbf{j} \qquad \text{and} \qquad \nabla f = \frac{\partial f}{\partial x}\mathbf{i} + \frac{\partial f}{\partial y}\mathbf{j} + \frac{\partial f}{\partial z}\mathbf{k}$$

respectively. Here $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$ denote, as usual, unit vectors in the x, y and z directions.

2.3. Consequences of Differentiability

We know that in the theory of real valued functions of one real variable, a function f is continuous whenever it is differentiable. The purpose of this section is to extend this result to vector valued functions of several variables.

THEOREM 2A. *Suppose that $f : A \to \mathbb{R}^m$, where $A \subseteq \mathbb{R}^n$ is an open set. Suppose further f is differentiable at $\mathbf{x}_0 \in A$. Then f is continuous at $\mathbf{x}_0$.*

† PROOF. Since (3) holds, it follows that there exists $\delta_1 > 0$ such that

$$\frac{\|f(\mathbf{x}) - f(\mathbf{x}_0) - \mathbf{T}(\mathbf{x} - \mathbf{x}_0)\|}{\|\mathbf{x} - \mathbf{x}_0\|} < 1,$$

and so

$$\|f(\mathbf{x}) - f(\mathbf{x}_0) - \mathbf{T}(\mathbf{x} - \mathbf{x}_0)\| < \|\mathbf{x} - \mathbf{x}_0\|$$

for every $\mathbf{x} \in A$ satisfying $0 < \|\mathbf{x} - \mathbf{x}_0\| < \delta_1$. It follows from the Triangle inequality that

$$\begin{aligned} \|f(\mathbf{x}) - f(\mathbf{x}_0)\| &= \|f(\mathbf{x}) - f(\mathbf{x}_0) - \mathbf{T}(\mathbf{x} - \mathbf{x}_0) + \mathbf{T}(\mathbf{x} - \mathbf{x}_0)\| \\ &\le \|f(\mathbf{x}) - f(\mathbf{x}_0) - \mathbf{T}(\mathbf{x} - \mathbf{x}_0)\| + \|\mathbf{T}(\mathbf{x} - \mathbf{x}_0)\| \\ &< \|\mathbf{x} - \mathbf{x}_0\| + \|\mathbf{T}(\mathbf{x} - \mathbf{x}_0)\| \end{aligned}$$

for every $\mathbf{x} \in A$ satisfying $0 < \|\mathbf{x} - \mathbf{x}_0\| < \delta_1$. Next, note that

$$\begin{aligned}
\|\mathbf{T}(\mathbf{x} - \mathbf{x}_0)\| &= \left\| \begin{pmatrix} \dfrac{\partial f_1}{\partial x_1}(\mathbf{x}_0) & \dots & \dfrac{\partial f_1}{\partial x_n}(\mathbf{x}_0) \\ \vdots & & \vdots \\ \dfrac{\partial f_m}{\partial x_1}(\mathbf{x}_0) & \dots & \dfrac{\partial f_m}{\partial x_n}(\mathbf{x}_0) \end{pmatrix} (\mathbf{x} - \mathbf{x}_0) \right\| = \left\| \begin{pmatrix} (\nabla f_1)(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0) \\ \vdots \\ (\nabla f_m)(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0) \end{pmatrix} \right\| \\
&= \left(\sum_{i=1}^m ((\nabla f_i)(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0))^2 \right)^{1/2} \le \left(\sum_{i=1}^m \|(\nabla f_i)(\mathbf{x}_0)\|^2 \|\mathbf{x} - \mathbf{x}_0\|^2 \right)^{1/2} \\
&= \left(\sum_{i=1}^m \|(\nabla f_i)(\mathbf{x}_0)\|^2 \right)^{1/2} \|\mathbf{x} - \mathbf{x}_0\|,
\end{aligned}$$

where the inequality follows from the Cauchy-Schwarz inequality in $\mathbb{R}^m$. Now let

$$M = \left(\sum_{i=1}^m \|(\nabla f_i)(\mathbf{x}_0)\|^2 \right)^{1/2} + 1.$$

Then

$$\|f(\mathbf{x}) - f(\mathbf{x}_0)\| < M\|\mathbf{x} - \mathbf{x}_0\|$$

for every $\mathbf{x} \in A$ satisfying $0 < \|\mathbf{x} - \mathbf{x}_0\| < \delta_1$. Suppose that $\epsilon > 0$ is given. Then let $\delta = \min\{\delta_1, \epsilon/M\}$. It is now easily seen that $\|f(\mathbf{x}) - f(\mathbf{x}_0)\| < \epsilon$ for every $\mathbf{x} \in A$ satisfying $\|\mathbf{x} - \mathbf{x}_0\| < \delta$. It follows that f is continuous at $\mathbf{x}_0$. ◯

Remark. The Cauchy-Schwarz inequality in $\mathbb{R}^m$ states that for every two vectors $\mathbf{x} = (x_1, \dots, x_m)$ and $\mathbf{y} = (y_1, \dots, y_m)$ in $\mathbb{R}^m$, we have $|\mathbf{x} \cdot \mathbf{y}| \le \|\mathbf{x}\| \, \|\mathbf{y}\|$, where $\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + \dots + x_m y_m$.

If we look at the proof of Theorem 2A carefully, then it is easy to see that we have also proved the following result.

THEOREM 2B. *Suppose that $f : A \to \mathbb{R}^m$, where $A \subseteq \mathbb{R}^n$ is an open set. Suppose further f is differentiable at $\mathbf{x}_0 \in A$. Then there exist positive real numbers M and δ_1 such that*

$$\|f(\mathbf{x}) - f(\mathbf{x}_0)\| \le M\|\mathbf{x} - \mathbf{x}_0\|$$

for every $\mathbf{x} \in A$ satisfying $\|\mathbf{x} - \mathbf{x}_0\| < \delta_1$.

2.4. Conditions for Differentiability

In the theory of real valued functions of one real variable, we have many examples of continuous functions that are not differentiable. On the other hand, the example

$$f(x, y) = \begin{cases} 1 & \text{if } x = 0 \text{ or } y = 0, \\ 0 & \text{otherwise,} \end{cases}$$

shows that while the partial derivatives $\partial f/\partial x$ and $\partial f/\partial y$ exist at $(0, 0)$, the function is not continuous at $(0, 0)$ and so not differentiable at $(0, 0)$.

On the other hand, the definition for differentiability is very difficult to use, since it is practically impossible to verify condition (3) in most instances. The following result eases the pain somewhat.

THEOREM 2C. *Suppose that $f : A \to \mathbb{R}^m$, where $A \subseteq \mathbb{R}^n$ is an open set. Suppose further that all partial derivatives*

$$\frac{\partial f_i}{\partial x_j}, \qquad \text{where } i = 1, \ldots, m \text{ and } j = 1, \ldots, n,$$

exist and are continuous in a neighbourhood of a point $\mathbf{x}_0 \in A$. Then f is differentiable at $\mathbf{x}_0$.

† PROOF. We need to establish (3). To do this, note first of all that

$$f(\mathbf{x}) - f(\mathbf{x}_0) - \mathbf{T}(\mathbf{x} - \mathbf{x}_0) = \begin{pmatrix} f_1(\mathbf{x}) - f_1(\mathbf{x}_0) - (\nabla f_1)(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0) \\ \vdots \\ f_m(\mathbf{x}) - f_m(\mathbf{x}_0) - (\nabla f_m)(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0) \end{pmatrix}.$$

It is then easy to see that

$$\|f(\mathbf{x}) - f(\mathbf{x}_0) - \mathbf{T}(\mathbf{x} - \mathbf{x}_0)\| \le \sum_{i=1}^m |f_i(\mathbf{x}) - f_i(\mathbf{x}_0) - (\nabla f_i)(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0)|.$$

It follows that to prove (3), it suffices to show that for every $i = 1, \ldots, m$,

$$\lim_{\mathbf{x} \to \mathbf{x}_0} \frac{|f_i(\mathbf{x}) - f_i(\mathbf{x}_0) - (\nabla f_i)(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0)|}{\|\mathbf{x} - \mathbf{x}_0\|} = 0. \tag{4}$$

Our proof will depend on the Mean value theorem for real valued functions of a real variable. Let $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{x}_0 = (X_1, \ldots, X_n)$. Then

$$\begin{aligned} f_i(\mathbf{x}) - f_i(\mathbf{x}_0) &= f_i(x_1, \ldots, x_n) - f_i(X_1, \ldots, X_n) \\ &= f_i(x_1, \ldots, x_n) - f_i(X_1, x_2, \ldots, x_n) \\ &\quad + f_i(X_1, x_2, \ldots, x_n) - f_i(X_1, X_2, x_3 \ldots, x_n) \\ &\quad + \ldots \\ &\quad + f_i(X_1, \ldots, X_{n-2}, x_{n-1}, x_n) - f_i(X_1, \ldots, X_{n-1}, x_n) \\ &\quad + f_i(X_1, \ldots, X_{n-1}, x_n) - f_i(X_1, \ldots, X_n) \\ &= \frac{\partial f_i}{\partial x_1}(\mathbf{y}_1)(x_1 - X_1) + \ldots + \frac{\partial f_i}{\partial x_n}(\mathbf{y}_n)(x_n - X_n) = \sum_{j=1}^n \frac{\partial f_i}{\partial x_j}(\mathbf{y}_j)(x_j - X_j) \end{aligned}$$

by the Mean value theorem, where for every $j = 1, \ldots, n$,

$$\mathbf{y}_j = (X_1, \ldots, X_{j-1}, y_j, x_{j+1}, \ldots, x_n)$$

for some y_j between x_j and X_j. On the other hand,

$$(\nabla f_i)(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0) = \sum_{j=1}^n \frac{\partial f_i}{\partial x_j}(\mathbf{x}_0)(x_j - X_j).$$

It follows that

$$\begin{aligned} |f_i(\mathbf{x}) - f_i(\mathbf{x}_0) - (\nabla f_i)(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0)| &= \left| \sum_{j=1}^n \left(\frac{\partial f_i}{\partial x_j}(\mathbf{y}_j) - \frac{\partial f_i}{\partial x_j}(\mathbf{x}_0) \right) (x_j - X_j) \right| \\ &\le \sum_{j=1}^n \left| \frac{\partial f_i}{\partial x_j}(\mathbf{y}_j) - \frac{\partial f_i}{\partial x_j}(\mathbf{x}_0) \right| |x_j - X_j| \end{aligned}$$

by the Triangle inequality. Note also that $|x_j - X_j| \le \|\mathbf{x} - \mathbf{x}_0\|$ for every $j = 1, \ldots, n$, so that

$$\frac{|f_i(\mathbf{x}) - f_i(\mathbf{x}_0) - (\nabla f_i)(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0)|}{\|\mathbf{x} - \mathbf{x}_0\|} \le \sum_{j=1}^{n} \left| \frac{\partial f_i}{\partial x_j}(\mathbf{y}_j) - \frac{\partial f_i}{\partial x_j}(\mathbf{x}_0) \right|.$$

Clearly the right hand side converges to zero as $\mathbf{x} \to \mathbf{x}_0$, in view of the continuity of the partial derivatives. (4) follows immediately. ◯

EXAMPLE 2.4.1. Consider the function $f : A \to \mathbb{R}^2$, given by

$$f(x, y) = \left(\frac{\sin x + \mathrm{e}^y}{x^2 + y^2}, \frac{1}{x^2 + y^2 - 1} \right)$$

for $(x, y) \in A$, where $A \subseteq \mathbb{R}^2$ is some suitable domain. We can write $f(x, y) = (f_1(x, y), f_2(x, y))$, where

$$f_1(x, y) = \frac{\sin x + \mathrm{e}^y}{x^2 + y^2} \qquad \text{and} \qquad f_2(x, y) = \frac{1}{x^2 + y^2 - 1}.$$

Now

$$\frac{\partial f_1}{\partial x} = \frac{(x^2 + y^2)\cos x - 2x(\sin x + \mathrm{e}^y)}{(x^2 + y^2)^2} \qquad \text{and} \qquad \frac{\partial f_1}{\partial y} = \frac{(x^2 + y^2)\mathrm{e}^y - 2y(\sin x + \mathrm{e}^y)}{(x^2 + y^2)^2}$$

are continuous except when $x = y = 0$. On the other hand,

$$\frac{\partial f_2}{\partial x} = -\frac{2x}{(x^2 + y^2 - 1)^2} \qquad \text{and} \qquad \frac{\partial f_2}{\partial y} = -\frac{2y}{(x^2 + y^2 - 1)^2}$$

are continuous except when $x^2 + y^2 = 1$. It follows from Theorem 2C that f is differentiable at every (x, y) such that $x^2 + y^2 \ne 0$ or 1.

2.5. Properties of the Derivative

The first two results in this section concern the arithmetic of derivatives, and are stated without proof. The interested reader is invited to write the proofs which proceed in a similar way as in the case of real valued functions of one real variable.

THEOREM 2D. *Suppose that $f : A \to \mathbb{R}^m$ and $g : A \to \mathbb{R}^m$ where $A \subseteq \mathbb{R}^n$ is an open set. Suppose further that $\mathbf{x}_0 \in A$.*

(a) If f is differentiable at $\mathbf{x}_0$, then cf is also differentiable at $\mathbf{x}_0$ for every $c \in \mathbb{R}$, and

$$(\mathbf{D}(cf))(\mathbf{x}_0) = c(\mathbf{D}f)(\mathbf{x}_0).$$

(b) If f and g are differentiable at $\mathbf{x}_0$, then $f + g$ is also differentiable at $\mathbf{x}_0$, and

$$(\mathbf{D}(f + g))(\mathbf{x}_0) = (\mathbf{D}f)(\mathbf{x}_0) + (\mathbf{D}g)(\mathbf{x}_0).$$

THEOREM 2E. *Suppose that $f : A \to \mathbb{R}$ and $g : A \to \mathbb{R}$, where $A \subseteq \mathbb{R}^n$ is an open set. Suppose further that f and g are differentiable at $\mathbf{x}_0 \in A$.*
(a) Then fg is also differentiable at $\mathbf{x}_0$, and

$$(\mathbf{D}(fg))(\mathbf{x}_0) = g(\mathbf{x}_0)(\mathbf{D}f)(\mathbf{x}_0) + f(\mathbf{x}_0)(\mathbf{D}g)(\mathbf{x}_0).$$

(b) If $g(\mathbf{x}_0) \neq 0$, then f/g is also differentiable at $\mathbf{x}_0$, and

$$\left(\mathbf{D}\left(\frac{f}{g}\right)\right)(\mathbf{x}_0) = \frac{g(\mathbf{x}_0)(\mathbf{D}f)(\mathbf{x}_0) - f(\mathbf{x}_0)(\mathbf{D}g)(\mathbf{x}_0)}{g^2(\mathbf{x}_0)}.$$

EXAMPLE 2.5.1. Suppose that $f : \mathbb{R}^2 \to \mathbb{R}$ and $g : \mathbb{R}^2 \to \mathbb{R}$ are given by $f(x, y) = x^2 + y^2$ and $g(x, y) = x + y$ for every $(x, y) \in \mathbb{R}^2$. Let

$$h(x, y) = \frac{f(x, y)}{g(x, y)} = \frac{x^2 + y^2}{x + y}.$$

Then

$$(\mathbf{D}h)(x, y) = \left(\frac{\partial h}{\partial x}, \frac{\partial h}{\partial y}\right) = \left(\frac{x^2 + 2xy - y^2}{(x + y)^2}, \frac{y^2 + 2xy - x^2}{(x + y)^2}\right)$$

whenever $x + y \neq 0$. On the other hand,

$$\begin{aligned}
\frac{g(x, y)(\mathbf{D}f)(x, y) - f(x, y)(\mathbf{D}g)(x, y)}{g^2(x, y)} &= \frac{g(x, y)\left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right) - f(x, y)\left(\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y}\right)}{g^2(x, y)} \\
= \frac{(x + y)(2x, 2y) - (x^2 + y^2)(1, 1)}{(x + y)^2} &= \frac{(2x(x + y), 2y(x + y)) - (x^2 + y^2, x^2 + y^2)}{(x + y)^2} \\
= \frac{(x^2 + 2xy - y^2, y^2 + 2xy - x^2)}{(x + y)^2} &= \left(\frac{x^2 + 2xy - y^2}{(x + y)^2}, \frac{y^2 + 2xy - x^2}{(x + y)^2}\right)
\end{aligned}$$

whenever $x + y \neq 0$. This verifies the Quotient rule in Theorem 2E.

It is in the Chain rule where there is substantial difference from the case of a real valued function of one real variable. Here, our notation helps to provide some visual resemblance at least.

THEOREM 2F. *Suppose that $f : A \to \mathbb{R}^m$ and $g : B \to \mathbb{R}^p$, where $A \subseteq \mathbb{R}^n$ and $B \subseteq \mathbb{R}^m$ are open sets. Suppose further that $f(A) \subseteq B$, so that $g \circ f : A \to \mathbb{R}^p$ is well defined. If f is differentiable at $\mathbf{x}_0 \in A$ and g is differentiable at $\mathbf{y}_0 = f(\mathbf{x}_0) \in B$, then $g \circ f$ is differentiable at $\mathbf{x}_0$, and*

$$(\mathbf{D}(g \circ f))(\mathbf{x}_0) = (\mathbf{D}g)(\mathbf{y}_0)(\mathbf{D}f)(\mathbf{x}_0),$$

where the right hand side represents the matrix product of $(\mathbf{D}g)(\mathbf{y}_0)$ and $(\mathbf{D}f)(\mathbf{x}_0)$.

EXAMPLE 2.5.2. Suppose that $f : \mathbb{R} \to \mathbb{R}^2 : t \mapsto (x(t), y(t))$ and $g : \mathbb{R}^2 \to \mathbb{R} : (x, y) \mapsto g(x, y)$ are both differentiable. Then

$$(\mathbf{D}f)(t) = \begin{pmatrix} x'(t) \\ y'(t) \end{pmatrix} \qquad \text{and} \qquad (\mathbf{D}g)(x, y) = \begin{pmatrix} \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{pmatrix},$$

so that we obtain the familiar formula

$$(\mathbf{D}(g \circ f))(t) = \begin{pmatrix} \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{pmatrix} \begin{pmatrix} x'(t) \\ y'(t) \end{pmatrix} = \frac{\partial g}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial g}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}t}.$$

EXAMPLE 2.5.3. Suppose that $f : \mathbb{R}^2 \to \mathbb{R}^2 : (s,t) \mapsto (x(s,t), y(s,t))$ and $g : \mathbb{R}^2 \to \mathbb{R} : (x,y) \mapsto g(x,y)$ are both differentiable. Then

$$(\mathbf{D}f)(s,t) = \begin{pmatrix} \dfrac{\partial x}{\partial s} & \dfrac{\partial x}{\partial t} \\ \dfrac{\partial y}{\partial s} & \dfrac{\partial y}{\partial t} \end{pmatrix} \qquad \text{and} \qquad (\mathbf{D}g)(x,y) = \begin{pmatrix} \dfrac{\partial g}{\partial x} & \dfrac{\partial g}{\partial y} \end{pmatrix},$$

so that

$$(\mathbf{D}(g \circ f))(s,t) = \begin{pmatrix} \dfrac{\partial g}{\partial x} & \dfrac{\partial g}{\partial y} \end{pmatrix} \begin{pmatrix} \dfrac{\partial x}{\partial s} & \dfrac{\partial x}{\partial t} \\ \dfrac{\partial y}{\partial s} & \dfrac{\partial y}{\partial t} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial g}{\partial x}\dfrac{\partial x}{\partial s} + \dfrac{\partial g}{\partial y}\dfrac{\partial y}{\partial s} & \dfrac{\partial g}{\partial x}\dfrac{\partial x}{\partial t} + \dfrac{\partial g}{\partial y}\dfrac{\partial y}{\partial t} \end{pmatrix}.$$

This gives the change of variable formulae in functions of two variables.

EXAMPLE 2.5.4. Suppose that

$$f : \mathbb{R}^3 \to \mathbb{R}^3 : (x,y,z) \mapsto (x^2, x^2y, \mathrm{e}^z) \qquad \text{and} \qquad g : \mathbb{R}^3 \to \mathbb{R} : (u,v,w) \mapsto u^2 + v^2 - w^2.$$

We can write $f(x,y,z) = (f_1(x,y,z), f_2(x,y,z), f_3(x,y,z))$, where

$$f_1(x,y,z) = x^2 \qquad \text{and} \qquad f_2(x,y,z) = x^2y \qquad \text{and} \qquad f_3(x,y,z) = \mathrm{e}^z$$

for every $(x,y,z) \in \mathbb{R}^3$. Note that

$$\begin{aligned}(\mathbf{D}g)(u,v,w)(\mathbf{D}f)(x,y,z) &= \begin{pmatrix} \dfrac{\partial g}{\partial u} & \dfrac{\partial g}{\partial v} & \dfrac{\partial g}{\partial w} \end{pmatrix} \begin{pmatrix} \dfrac{\partial f_1}{\partial x} & \dfrac{\partial f_1}{\partial y} & \dfrac{\partial f_1}{\partial z} \\ \dfrac{\partial f_2}{\partial x} & \dfrac{\partial f_2}{\partial y} & \dfrac{\partial f_2}{\partial z} \\ \dfrac{\partial f_3}{\partial x} & \dfrac{\partial f_3}{\partial y} & \dfrac{\partial f_3}{\partial z} \end{pmatrix} = \begin{pmatrix} 2u & 2v & -2w \end{pmatrix} \begin{pmatrix} 2x & 0 & 0 \\ 2xy & x^2 & 0 \\ 0 & 0 & \mathrm{e}^z \end{pmatrix} \\ &= \begin{pmatrix} 2x^2 & 2x^2y & -2\mathrm{e}^z \end{pmatrix} \begin{pmatrix} 2x & 0 & 0 \\ 2xy & x^2 & 0 \\ 0 & 0 & \mathrm{e}^z \end{pmatrix} = \begin{pmatrix} 4x^3(1+y^2) & 2x^4y & -2\mathrm{e}^{2z} \end{pmatrix}.\end{aligned}$$

Now consider

$$h = g \circ f : \mathbb{R}^3 \to \mathbb{R} : (x,y,z) \mapsto x^4(1+y^2) - \mathrm{e}^{2z}.$$

It is easily seen that

$$(\mathbf{D}h)(x,y,z) = \begin{pmatrix} \dfrac{\partial h}{\partial x} & \dfrac{\partial h}{\partial y} & \dfrac{\partial h}{\partial z} \end{pmatrix} = \begin{pmatrix} 4x^3(1+y^2) & 2x^4y & -2\mathrm{e}^{2z} \end{pmatrix}.$$

This verifies the Chain rule.

EXAMPLE 2.5.5. Suppose that in Example 2.5.4, we would like to compute the derivative of $g \circ f$ at $(1,2,0)$. Note first of all that $f(1,2,0) = (1,2,1)$. Then by the Chain rule,

$$(\mathbf{D}(g \circ f))(1,2,0) = (\mathbf{D}g)(1,2,1)(\mathbf{D}f)(1,2,0) = \begin{pmatrix} 2 & 4 & -2 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 4 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 20 & 4 & -2 \end{pmatrix}.$$

† Proof of Theorem 2F. We need to show that

$$\lim_{\mathbf{x}\to\mathbf{x}_0} \frac{\|g(f(\mathbf{x})) - g(f(\mathbf{x}_0)) - (\mathbf{D}g)(\mathbf{y}_0)(\mathbf{D}f)(\mathbf{x}_0)(\mathbf{x}-\mathbf{x}_0)\|}{\|\mathbf{x}-\mathbf{x}_0\|} = 0. \tag{5}$$

Note first of all that writing $\mathbf{y} = f(\mathbf{x})$, we have

$$\begin{aligned}
&\|g(f(\mathbf{x})) - g(f(\mathbf{x}_0)) - (\mathbf{D}g)(\mathbf{y}_0)(\mathbf{D}f)(\mathbf{x}_0)(\mathbf{x}-\mathbf{x}_0)\| \\
&\quad = \|g(\mathbf{y}) - g(\mathbf{y}_0) - (\mathbf{D}g)(\mathbf{y}_0)(\mathbf{y}-\mathbf{y}_0) + (\mathbf{D}g)(\mathbf{y}_0)(f(\mathbf{x}) - f(\mathbf{x}_0)) - (\mathbf{D}g)(\mathbf{y}_0)(\mathbf{D}f)(\mathbf{x}_0)(\mathbf{x}-\mathbf{x}_0)\| \\
&\quad \le \|g(\mathbf{y}) - g(\mathbf{y}_0) - (\mathbf{D}g)(\mathbf{y}_0)(\mathbf{y}-\mathbf{y}_0)\| + \|(\mathbf{D}g)(\mathbf{y}_0)(f(\mathbf{x}) - f(\mathbf{x}_0) - (\mathbf{D}f)(\mathbf{x}_0)(\mathbf{x}-\mathbf{x}_0))\|
\end{aligned}$$

by the Triangle inequality. As in the proof of Theorem 2A, there exists a positive constant M', depending only on $(\mathbf{D}g)(\mathbf{y}_0)$, such that $\|(\mathbf{D}g)(\mathbf{y}_0)\mathbf{u}\| \le M'\|\mathbf{u}\|$ for every $\mathbf{u} \in \mathbb{R}^m$. It follows that

$$\|(\mathbf{D}g)(\mathbf{y}_0)(f(\mathbf{x}) - f(\mathbf{x}_0) - (\mathbf{D}f)(\mathbf{x}_0)(\mathbf{x}-\mathbf{x}_0))\| \le M'\|f(\mathbf{x}) - f(\mathbf{x}_0) - (\mathbf{D}f)(\mathbf{x}_0)(\mathbf{x}-\mathbf{x}_0)\|,$$

and so

$$\begin{aligned}
&\|g(f(\mathbf{x})) - g(f(\mathbf{x}_0)) - (\mathbf{D}g)(\mathbf{y}_0)(\mathbf{D}f)(\mathbf{x}_0)(\mathbf{x}-\mathbf{x}_0)\| \\
&\quad \le \|g(\mathbf{y}) - g(\mathbf{y}_0) - (\mathbf{D}g)(\mathbf{y}_0)(\mathbf{y}-\mathbf{y}_0)\| + M'\|f(\mathbf{x}) - f(\mathbf{x}_0) - (\mathbf{D}f)(\mathbf{x}_0)(\mathbf{x}-\mathbf{x}_0)\|
\end{aligned}$$

Since f is differentiable at $\mathbf{x}_0$, it follows that

$$\lim_{\mathbf{x}\to\mathbf{x}_0} \frac{\|f(\mathbf{x}) - f(\mathbf{x}_0) - (\mathbf{D}f)(\mathbf{x}_0)(\mathbf{x}-\mathbf{x}_0)\|}{\|\mathbf{x}-\mathbf{x}_0\|} = 0.$$

It follows that to prove (5), it suffices to show that

$$\lim_{\mathbf{x}\to\mathbf{x}_0} \frac{\|g(\mathbf{y}) - g(\mathbf{y}_0) - (\mathbf{D}g)(\mathbf{y}_0)(\mathbf{y}-\mathbf{y}_0)\|}{\|\mathbf{x}-\mathbf{x}_0\|} = 0. \tag{6}$$

To do this, we first of all recall Theorem 2B. Since f is differentiable at $\mathbf{x}_0$, there exist positive real numbers M and δ_1 such that

$$\|f(\mathbf{x}) - f(\mathbf{x}_0)\| \le M\|\mathbf{x}-\mathbf{x}_0\|$$

for every $\mathbf{x} \in A$ satisfying $\|\mathbf{x}-\mathbf{x}_0\| < \delta_1$. On the other hand, since g is differentiable at $\mathbf{y}_0$, it follows that

$$\lim_{\mathbf{y}\to\mathbf{y}_0} \frac{\|g(\mathbf{y}) - g(\mathbf{y}_0) - (\mathbf{D}g)(\mathbf{y}_0)(\mathbf{y}-\mathbf{y}_0)\|}{\|\mathbf{y}-\mathbf{y}_0\|} = 0.$$

Hence given any $\epsilon > 0$, there exists $\eta > 0$ such that

$$\|g(\mathbf{y}) - g(\mathbf{y}_0) - (\mathbf{D}g)(\mathbf{y}_0)(\mathbf{y}-\mathbf{y}_0)\| < \frac{\epsilon}{M}\|\mathbf{y}-\mathbf{y}_0\|$$

for every $\mathbf{y} \in B$ satisfying $\|\mathbf{y}-\mathbf{y}_0\| < \eta$. Also, f is continuous at $\mathbf{x}_0$, so there exists $\delta_2 > 0$ such that

$$\|\mathbf{y}-\mathbf{y}_0\| = \|f(\mathbf{x}) - f(\mathbf{x}_0)\| < \eta$$

for every $\mathbf{x} \in A$ satisfying $\|\mathbf{x}-\mathbf{x}_0\| < \delta_2$. Now let $\delta = \min\{\delta_1, \delta_2\}$. Then clearly

$$\|g(\mathbf{y}) - g(\mathbf{y}_0) - (\mathbf{D}g)(\mathbf{y}_0)(\mathbf{y}-\mathbf{y}_0)\| < \epsilon\|\mathbf{x}-\mathbf{x}_0\|$$

for every $\mathbf{x} \in A$ satisfying $\|\mathbf{x}-\mathbf{x}_0\| < \delta$. (6) follows immediately. ○

2.6. Gradients and Directional Derivatives

Recall the last remark in Section 2.2. Suppose that $f : A \to \mathbb{R}$, where $A \subseteq \mathbb{R}^3$ is an open set. Suppose further that f is differentiable at $\mathbf{x}_0 \in A$. Then the vector in $\mathbb{R}^3$ given by

$$(\nabla f)(\mathbf{x}_0) = \left(\frac{\partial f}{\partial x}(\mathbf{x}_0), \frac{\partial f}{\partial y}(\mathbf{x}_0), \frac{\partial f}{\partial z}(\mathbf{x}_0)\right)$$

is called the gradient of f at $\mathbf{x}_0$.

In this section, we shall use this to obtain a formula for tangent planes to surfaces in $\mathbb{R}^3$.

Example 2.6.1. Suppose that $f : \mathbb{R}^3 \to \mathbb{R}$, where

$$f(x, y, z) = \sqrt{x^2 + y^2 + z^2}$$

for every $(x, y, z) \in \mathbb{R}^3$. Then

$$\begin{aligned}(\nabla f)(x_0, y_0, z_0) &= \left(\frac{\partial f}{\partial x}(x_0, y_0, z_0), \frac{\partial f}{\partial y}(x_0, y_0, z_0), \frac{\partial f}{\partial z}(x_0, y_0, z_0)\right)\\ &= \left(\frac{x_0}{\sqrt{x_0^2 + y_0^2 + z_0^2}}, \frac{y_0}{\sqrt{x_0^2 + y_0^2 + z_0^2}}, \frac{z_0}{\sqrt{x_0^2 + y_0^2 + z_0^2}}\right)\\ &= \frac{(x_0, y_0, z_0)}{\|(x_0, y_0, z_0)\|}\end{aligned}$$

is the unit vector in the direction of (x_0, y_0, z_0).

Example 2.6.2. Suppose that $f : \mathbb{R}^3 \to \mathbb{R}$, where

$$f(x, y, z) = \mathrm{e}^{xy} + z$$

for every $(x, y, z) \in \mathbb{R}^3$. Then

$$(\nabla f)(x_0, y_0, z_0) = \left(\frac{\partial f}{\partial x}(x_0, y_0, z_0), \frac{\partial f}{\partial y}(x_0, y_0, z_0), \frac{\partial f}{\partial z}(x_0, y_0, z_0)\right) = (y_0\mathrm{e}^{x_0y_0}, x_0\mathrm{e}^{x_0y_0}, 1).$$

For every $\mathbf{x}_0 \in \mathbb{R}^3$ and every unit vector $\mathbf{n} \in \mathbb{R}^3$, the mapping

$$L : \mathbb{R} \to \mathbb{R}^3 : t \mapsto \mathbf{x}_0 + t\mathbf{n}$$

represents the line through $\mathbf{x}_0$ in the direction $\mathbf{n}$. It follows that for every function of the form $f : \mathbb{R}^3 \to \mathbb{R}$, the composite function

$$f \circ L : \mathbb{R} \to \mathbb{R} : t \mapsto f(\mathbf{x}_0 + t\mathbf{n})$$

represents the function f restricted to the line L.

Definition. Suppose that $f : \mathbb{R}^3 \to \mathbb{R}$. Then the limit

$$\lim_{t \to 0} \frac{f(\mathbf{x}_0 + t\mathbf{n}) - f(\mathbf{x}_0)}{t},$$

if it exists, is called the directional derivative of f at $\mathbf{x}_0$ in the direction $\mathbf{n}$.

REMARK. Note that

$$\lim_{t\to 0}\frac{f(\mathbf{x}_0+t\mathbf{n})-f(\mathbf{x}_0)}{t}=\frac{\mathrm{d}}{\mathrm{d}t}f(\mathbf{x}_0+t\mathbf{n})\bigg|_{t=0}.$$

THEOREM 2G. *Suppose that $f:\mathbb{R}^3\to\mathbb{R}$ is differentiable. Then all directional derivatives of f exist. Furthermore, for every $\mathbf{x}_0\in\mathbb{R}^3$ and every unit vector $\mathbf{n}\in\mathbb{R}^3$, the directional derivative of f at $\mathbf{x}_0$ in the direction $\mathbf{n}=(n_1,n_2,n_3)$ is given by the scalar product*

$$(\nabla f)(\mathbf{x}_0)\cdot\mathbf{n}=\left(\frac{\partial f}{\partial x}(\mathbf{x}_0)\right)n_1+\left(\frac{\partial f}{\partial y}(\mathbf{x}_0)\right)n_2+\left(\frac{\partial f}{\partial z}(\mathbf{x}_0)\right)n_3.$$

REMARK. Note that

$$(\mathbf{D}f)(\mathbf{x}_0)\mathbf{n}=(\nabla f)(\mathbf{x}_0)\cdot\mathbf{n},$$

where the left hand side denotes the matrix product of the total derivative $(\mathbf{D}f)(\mathbf{x}_0)$ and the column matrix $\mathbf{n}$.

PROOF OF THEOREM 2G. Consider the functions

$$L:\mathbb{R}\to\mathbb{R}^3:t\mapsto\mathbf{x}_0+t\mathbf{n}=(x_0+tn_1,y_0+tn_2,z_0+tn_3)=(L_1(t),L_2(t),L_3(t))$$

and

$$g=f\circ L:\mathbb{R}\to\mathbb{R}:t\mapsto f(\mathbf{x}_0+t\mathbf{n}).$$

Note that

$$\lim_{t\to 0}\frac{f(\mathbf{x}_0+t\mathbf{n})-f(\mathbf{x}_0)}{t}=\lim_{t\to 0}\frac{g(t)-g(0)}{t}=\frac{\mathrm{d}g}{\mathrm{d}t}(0)=(\mathbf{D}g)(0).$$

By the Chain rule, we have

$$(\mathbf{D}g)(0)=(\mathbf{D}f)(L(0))(\mathbf{D}L)(0).$$

Since

$$(\mathbf{D}f)(L(0))=(\mathbf{D}f)(\mathbf{x}_0)=\begin{pmatrix}\dfrac{\partial f}{\partial x}(\mathbf{x}_0) & \dfrac{\partial f}{\partial y}(\mathbf{x}_0) & \dfrac{\partial f}{\partial z}(\mathbf{x}_0)\end{pmatrix}$$

and

$$(\mathbf{D}L)(0)=\begin{pmatrix}\dfrac{\mathrm{d}L_1}{\mathrm{d}t}(0)\\ \dfrac{\mathrm{d}L_2}{\mathrm{d}t}(0)\\ \dfrac{\mathrm{d}L_3}{\mathrm{d}t}(0)\end{pmatrix}=\begin{pmatrix}n_1\\ n_2\\ n_3\end{pmatrix},$$

it follows that

$$(\mathbf{D}g)(0)=(\nabla f)(\mathbf{x}_0)\cdot\mathbf{n}.$$

The result follows immediately. ○

EXAMPLE 2.6.3. We continue our discussion in Example 2.6.2, where $f(x,y,z) = \mathrm{e}^{xy} + z$ for every $(x,y,z) \in \mathbb{R}^3$. The rate of change of the function f at $(1,0,1)$ in the direction of the unit vector $(1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})$ is given by

$$(\nabla f)(1,0,1) \cdot \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right) = (0,1,1) \cdot \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right) = \frac{2}{\sqrt{3}}.$$

We now investigate the directional derivative further. Suppose that $(\nabla f)(\mathbf{x}_0) \neq \mathbf{0}$. Then for every unit vector $\mathbf{n} \in \mathbb{R}^3$, we have

$$(\nabla f)(\mathbf{x}_0) \cdot \mathbf{n} = \|(\nabla f)(\mathbf{x}_0)\| \cos\theta,$$

where θ is the angle between $(\nabla f)(\mathbf{x}_0)$ and $\mathbf{n}$. This is maximum if $\theta = 0$ and minimum if $\theta = \pi$, corresponding respectively to $(\nabla f)(\mathbf{x}_0)$ and $\mathbf{n}$ being in the same direction and being in opposite directions. In other words, $(\nabla f)(\mathbf{x}_0)$ is in the direction in which f increases the fastest, while $-(\nabla f)(\mathbf{x}_0)$ is in the direction in which f decreases the fastest.

EXAMPLE 2.6.4. We continue our discussion in Examples 2.6.2 and 2.6.3, where $f(x,y,z) = \mathrm{e}^{xy} + z$ for every $(x,y,z) \in \mathbb{R}^3$. The maximum rate of change of the function f at $(1,0,1)$ is in the direction of the vector $(\nabla f)(1,0,1) = (0,1,1)$. Since $(0, 1/\sqrt{2}, 1/\sqrt{2})$ is the unit vector in this direction, it follows that the maximum rate of change of the function f at $(1,0,1)$ is equal to

$$(\nabla f)(1,0,1) \cdot \left(0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) = (0,1,1) \cdot \left(0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) = \sqrt{2}.$$

Suppose now that $f : \mathbb{R}^3 \to \mathbb{R}$ is differentiable. Consider a surface of the form

$$S = \{(x,y,z) \in \mathbb{R}^3 : f(x,y,z) = c\},$$

where $c \in \mathbb{R}$ is a constant. Suppose further that $(x_0,y_0,z_0) \in S$ and $(\nabla f)(x_0,y_0,z_0) \neq (0,0,0)$ (the reader is advised to draw a picture). Let $g : \mathbb{R} \to \mathbb{R}^3$ be a differentiable function such that $g(\mathbb{R}) \subseteq S$ and $g(0) = (x_0,y_0,z_0)$; in other words, g is a path on S that passes through the point (x_0,y_0,z_0) when $t = 0$. Then clearly

$$(f \circ g)(t) = c$$

for every $t \in \mathbb{R}$, so it follows from the Chain rule that

$$0 = (\mathbf{D}(f \circ g))(0) = (\mathbf{D}f)(x_0,y_0,z_0)(\mathbf{D}g)(0) = (\nabla f)(x_0,y_0,z_0) \cdot \mathbf{v},$$

where $\mathbf{v} = (\mathbf{D}g)(0)$ is a tangent vector to the path $g(t)$ at $t = 0$, and so is tangent to the surface S at (x_0,y_0,z_0). It follows that $(\nabla f)(x_0,y_0,z_0)$ must be normal to the surface S at (x_0,y_0,z_0). It also follows that

$$(\nabla f)(x_0,y_0,z_0) \cdot (x - x_0, y - y_0, z - z_0) = 0$$

is the equation of the tangent plane to the surface S at (x_0,y_0,z_0).

EXAMPLE 2.6.5. Consider the surface $2xy + z^3 = 3$ at the point $(1,1,1)$. Here $f(x,y,z) = 2xy + z^3$. It is easy to check that $(\nabla f)(1,1,1) = (2,2,3)$. Hence the equation of the tangent plane to the surface at $(1,1,1)$ is $(2,2,3) \cdot (x-1, y-1, z-1) = 0$; in other words, $2x + 2y + 3z = 7$.

Problems for Chapter 2

1. Compute the matrix of partial derivatives of each of the following functions:
   a) $f : \mathbb{R}^3 \to \mathbb{R}^2 : (x, y, z) \mapsto (\mathrm{e}^{x+y} + z, \sin(x + y + z) - \cos(x - y))$
   b) $f : \mathbb{R}^2 \to \mathbb{R}^4 : (x, y) \mapsto (x + y, x - y, 2x + y^2, y)$
   c) $f : \mathbb{R}^3 \to \mathbb{R}^3 : (x, y, z) \mapsto (x^2 + 2y + 3z^2, \sin(x^2 + y^2), \cos z)$
   d) $f : \mathbb{R}^2 \to \mathbb{R}^5 : (x, y) \mapsto (\sin x, \cos y, \sin y, \cos x, \mathrm{e}^{xy})$

2. For each of the following functions, determine precisely where the function is differentiable, and find the total derivatives at these points:
   a) $f : \mathbb{R}^2 \to \mathbb{R} : (x, y) \mapsto x^3 - y^3$
   b) $f : \mathbb{R}^2 \to \mathbb{R}^2 : (x, y) \mapsto (|x|, \mathrm{e}^{x+y})$

3. Suppose that $A \subseteq \mathbb{R}^n$ is an open set, and that $\mathbf{x}_0 \in A$. Suppose further that $f : A \to \mathbb{R}$ and $g : A \to \mathbb{R}$ are both differentiable at $\mathbf{x}_0$, and that $g(\mathbf{x}) > 0$ for every $\mathbf{x} \in A$. Explain why the function $h : A \to \mathbb{R}$, defined by

$$h(\mathbf{x}) = \frac{f^3(\mathbf{x}) + f(\mathbf{x})g^2(\mathbf{x})}{f^2(\mathbf{x}) + g(\mathbf{x})}$$

   for every $\mathbf{x} \in A$, is differentiable at $\mathbf{x}_0$, and find $(\mathbf{D}h)(\mathbf{x}_0)$.

4. Suppose that $g_1 : \mathbb{R}^4 \to \mathbb{R}$, $g_2 : \mathbb{R}^4 \to \mathbb{R}$ and $h : \mathbb{R}^2 \to \mathbb{R}$ are differentiable functions. Suppose further that $f : \mathbb{R}^4 \to \mathbb{R}$ is defined by

$$f(\mathbf{x}) = h(g_1(\mathbf{x}), g_2(\mathbf{x}))$$

   for every $\mathbf{x} \in \mathbb{R}^4$. Express $(\nabla f)(\mathbf{x})$ in terms of the partial derivatives of g_1, g_2 and h.
   [Remark: It is convenient to write $\mathbf{x} = (x_1, x_2, x_3, x_4)$.]

5. Suppose that $f : \mathbb{R}^2 \to \mathbb{R}$ and $g : \mathbb{R}^2 \to \mathbb{R}$ are both differentiable at $\mathbf{x}_0$. Prove that

$$(\nabla(fg))(\mathbf{x}_0) = f(\mathbf{x}_0)(\nabla g)(\mathbf{x}_0) + g(\mathbf{x}_0)(\nabla f)(\mathbf{x}_0).$$

6. Let n be a fixed positive integer. The function $f : \mathbb{R} \to \mathbb{R}$ is defined by

$$f(x) = \begin{cases} x^n \sin(1/x) & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

   a) For what positive integer values of n is f differentiable at 0?
   b) For what positive integer values of n is the derivative of f continuous at 0?

7. Consider the functions $f : \mathbb{R}^2 \to \mathbb{R}^2$ and $g : \mathbb{R}^2 \to \mathbb{R}^2$, given by

$$f(x, y) = (x^2y^3, \mathrm{e}^{x-y}) \qquad \text{and} \qquad g(u, v) = (v \sin u, \mathrm{e}^u v^2).$$

   a) Show that the function $h = g \circ f : \mathbb{R}^2 \to \mathbb{R}^2$ is differentiable at $(1, 1)$, and find $(\mathbf{D}h)(1, 1)$.
   b) Find $(\mathbf{D}f)(1, 1)$ and $(\mathbf{D}g)(1, 1)$.
   c) Explain why $(\mathbf{D}h)(1, 1) = (\mathbf{D}g)(1, 1)(\mathbf{D}f)(1, 1)$.

8. Consider the functions $f : \mathbb{R}^3 \to \mathbb{R}^2$ and $g : \mathbb{R}^2 \to \mathbb{R}^4$, defined by $f(x, y, z) = (x^2y, z \sin y)$ and $g(u, v) = (uv, u^2v^2, u + v, u - v^2)$.
   a) Find the total derivatives $\mathbf{D}f$ and $\mathbf{D}g$.
   b) Find the composition $h = g \circ f : \mathbb{R}^3 \to \mathbb{R}^4$.
   c) Find the total derivative $\mathbf{D}h$.
   d) Verify the Chain rule.

9. Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$, given by

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

   a) Show that the partial derivatives of f exist at every $(x_0, y_0) \in \mathbb{R}^2$, and evaluate them in terms of x_0 and y_0, taking extra care when $(x_0.y_0) = (0,0)$.
   b) Explain why f is not differentiable at $(0,0)$.

10. Consider the function $f : \mathbb{R} \to \mathbb{R}^2$, given by $f(u) = (au, bu)$, where $a, b \in \mathbb{R}$ are fixed. Consider also the function $g : \mathbb{R}^2 \to \mathbb{R}$, given by

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

   a) Find $(\mathbf{D}f)(0)$.
   b) Show that $\partial g/\partial x$ and $\partial g/\partial y$ exist at $(0,0)$, and find $(\mathbf{D}g)(0,0)$.
   c) Consider the function $h = g \circ f : \mathbb{R} \to \mathbb{R}$. Show that h is differentiable, and find $h'(0)$.
   d) Explain why $(\mathbf{D}g)(0,0)(\mathbf{D}f)(0) \neq h'(0)$.

11. Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$, given by

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

   a) Find the partial derivatives $\partial f/\partial x$ and $\partial f/\partial y$ at $(x_0, y_0) \in \mathbb{R}^2$, taking extra care in the case when $(x_0, y_0) = (0,0)$.
   b) Is f differentiable at $(0,0)$? Justify your assertion.

12. Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is an even function, so that $f(\mathbf{x}) = f(-\mathbf{x})$ for every $\mathbf{x} \in \mathbb{R}^n$. Suppose further that f is differentiable. Find $(\mathbf{D}f)(\mathbf{0})$.

13. Compute the directional derivative of the function $f(x,y,z) = xyz$ at the point $(1,0,1)$ and in the direction of the vector $(-1,0,1)$.

14. For each of the following, find an equation of the tangent plane of the graph of the function f at the point $(x_0, y_0, f(x_0, y_0))$:
   a) $f(x,y) = \sqrt{x^2+y^2}$ at $(x_0, y_0) = (1,1)$
   b) $f(x,y) = 4x^2 + y^2$ at $(x_0, y_0) = (1,2)$

15. Consider the surface $2x^2 + 3y^2 + 4z^2 = 9$. Suppose that a particle leaves the surface at the point $(1,1,1)$ along the normal directed towards the xy-plane, and with the constant speed of 1 unit per second. How long does it take for the particle to reach the xy-plane?