# Chapter 9

# The Derivative

Remember that for a function of one real variable, we use the tangent line to begin the discussion about the derivative. The derivative measures a rate of change, the slope measures a rate of change for the line, and basically we worked to get these two concepts to mean the same thing. Can we do something similar for functions of more than one variable? We will need to restrict our attention to a small number of variables if we want to "visualize" the derivative, but the general idea should follow our intuition from the case of one variable and extend our understanding in the case of more variables.

The concept of *local linearity* played a large part in the justification of the derivative of a function of one variable. This local linearity is what gave us the idea of looking at the tangent line. Let's look at a function of two real variables, say $f(x, y) = x^2 + y^2$. It's graph is a surface in $\mathbb{R}^3$. If we "zoom" in on a point on the surface, what does the surrounding area look like? Think about standing on the surface of the earth. In a small area around you, the earth looks relatively "flat" — or looks like a two-dimensional linear object. In other words, we won't have a tangent line, but will have a tangent *plane*. Therefore, our "derivative" must take this into account! Of course, we don't expect the derivative of a function to be the equation of a plane!! We might ask, though, what information do we need in order to determine a plane? Think about it, if we know a normal vector, then we know the equation of the plane. How so?

A ***normal*** vector, $\mathbf{n}$, to the plane $P$ is a vector that is perpendicular to the plane. Thus, if $\mathbf{a} = \langle a, b, c\rangle$ is a point in the plane and $\mathbf{x} = \langle x, y, z\rangle$ is any other point in the plane, the vector $\mathbf{x} - \mathbf{a} = \langle x - a, y - b, z - c\rangle$ lies in the plane. If $\mathbf{n} = \langle A, B, C\rangle$ is a normal vector, we have that:

$$\mathbf{n} \cdot (\mathbf{x} - \mathbf{a}) = 0,$$

or

$$A(x - a) + B(y - b) + C(z - c) = 0,$$

which clearly gives us an equation of the plane $P$. Now, a normal vector would require three bits of information — an $x$-coordinate, a $y$-coordinate, and a $z$-coordinate. We are not likely to be able to get three distinct bits of information from a function of 2-variables, at least not all of the time. We may want to consider a different approach.

## 9.1 Partial Derivatives

We do know that we used the concept of the limit to arrive at the derivative for a function of one variable. We might be able to do the same for a function of more than one variable. We will need to recall and use the concepts of *open* sets.

A point $\mathbf{r} \in U \subseteq \mathbb{R}^m$ is an *interior point* of $U$ if there exists some $\delta > 0$ so that $B_\delta(\mathbf{r}) = \{\mathbf{x} \in \mathbb{R}^m \mid \|\mathbf{x} - \mathbf{r}\| < \delta\} \subset U$. A point $\mathbf{r} \in U \subseteq \mathbb{R}^m$ is an *exterior point* of $U$ if there exists some $\delta > 0$ so that $B_\delta(\mathbf{r}) \bigcap U = \emptyset$. A point $\mathbf{r} \in U \subseteq \mathbb{R}^m$ is a *boundary point* of $U$ if for every $\delta > 0$ $B_\delta(\mathbf{r})$ contains points that belong to $U$ and points that do not belong to $U$. A set $U \subseteq \mathbb{R}^m$ is *open* in $\mathbb{R}^m$ if all of its points are interior points.

Let $f(x, y)$ be a function and let $(a, b)$ lie in its domain $U$. Now, if we are to see how $f$ *changes* at $(a, b)$ we have to decide on which direction we are traveling. In other words, we have to specify a direction in which the variables change. We will use the directions that are given to us by the coordinate axes here.

If we fix $y$ at the value $b$ then the function $f(x, b)$ is a function of only one variable $x$ as $f$ moves on the line parallel to the $x$-axis going through $(a, b)$. We can measure its rate of change as

$$\lim_{h\to 0} \frac{f(a+h,b) - f(a,b)}{h},$$

if that limit exists. This limit is called the *partial derivative of $f$ with respect to $x$* at $(a, b)$ and is denoted by

$$\frac{\partial f}{\partial x}(a,b) = \lim_{h\to 0} \frac{f(a+h,b) - f(a,b)}{h}.$$

Likewise, the *partial derivative of $f$ with respect to $y$* at $(a, b)$ and is denoted by

$$\frac{\partial f}{\partial y}(a,b) = \lim_{k\to 0} \frac{f(a,b+k) - f(a,b)}{k}.$$

As we let $(a, b)$ vary over all points in the domain of $f$, we get the functions:

$$f_x(x,y) = \frac{\partial f}{\partial x}(x,y) = \lim_{h\to 0} \frac{f(x+h,y) - f(x,y)}{h} \tag{9.1}$$

$$f_y(x,y) = \frac{\partial f}{\partial y}(x,y) = \lim_{h\to 0} \frac{f(x,y+h) - f(x,y)}{h} \tag{9.2}$$

which represent the *partial derivatives of $f$ with respect to $x$ and $y$.*

This easily generalizes to functions of several variables.

**Definition 9.1** *Let $f\colon U \subseteq \mathbb{R}^m \to \mathbb{R}$ be a real-valued function of $m$ variables $x_1, \ldots, x_m$, defined on an open set $U \in \mathbb{R}^m$. The* partial derivative *of $f$ with respect to $x_i$ is a real-valued function $\partial f/\partial x_i$ of $m$ variables defined by*

$$\frac{\partial f}{\partial x_i}(x_1, \ldots, x_m) = \lim_{h\to 0} \frac{f(x_1, \ldots, x_i + h, \ldots, x_m) - f(x_1, ldots, x_m)}{h},$$

*if that limit exists.*

What this tells us that the when computing $\partial f/\partial x_i$, we may regard all of the other variables except $x_i$ as constants and applying the standard rules for differentiating functions of one variable — if possible. Otherwise, use the definition.

There are other commonly used notations for $\partial f/\partial x_i$. They are $f_{x_i}$, $D_{x_i}F$ and $D_i f$. Note that the partial derivative is the slope of the tangent line to the curve obtained by intersecting the graph of the function with the hyperplane $(a_1, \ldots, x_i, \ldots, a_m)$.
**Note**: In 2-space the partial derivative can be written in vector form as follows. Let $\mathbf{x} = (x, y)$ then

$$f_x(\mathbf{x}) = \lim_{h\to 0} \frac{f(\mathbf{x} + h\mathbf{i}) - f(\mathbf{x})}{h}$$
$$f_y(\mathbf{x}) = \lim_{h\to 0} \frac{f(\mathbf{x} + h\mathbf{j}) - f(\mathbf{x})}{h}$$

and these "look" more like our definition of derivative in one dimension.

## 9.2 Derivative of Vector-Valued Functions

Let $\mathbf{F}$ be a vector-valued function $\mathbf{F}\colon U \subseteq \mathbb{R}^m \to \mathbb{R}^n$. Recall that $\mathbf{F}$ can be written in terms of its component functions:

$$\mathbf{F}(x_1, \ldots, x_m) = (F_1(x_1, \ldots, x_m), \ldots, F_n(x_1, \ldots, x_n))$$

or as $\mathbf{F}(\mathbf{x}) = (F_1(\mathbf{x}), \ldots, F_n(\mathbf{x}))$. This means that each of the functions $F_1, \ldots, F_n$ will have $m$ partial derivatives. That means we have $n \times m$ partial derivatives to compute and track. The only way that we can reasonably do this is in a matrix.

By $D\mathbf{F}(\mathbf{x})$ we mean an $n \times m$ matrix of partial derivatives of the component functions evaluated at $\mathbf{x}$, provided each of the partial derivatives exists at $\mathbf{x}$. Thus

$$D\mathbf{F}(\mathbf{x}) = \begin{bmatrix} \frac{\partial F_1}{\partial x_1}(\mathbf{x}) & \frac{\partial F_1}{\partial x_2}(\mathbf{x}) & \ldots & \frac{\partial F_1}{\partial x_m}(\mathbf{x}) \\ \frac{\partial F_2}{\partial x_1}(\mathbf{x}) & \frac{\partial F_2}{\partial x_2}(\mathbf{x}) & \ldots & \frac{\partial F_2}{\partial x_m}(\mathbf{x}) \\ \vdots & \vdots & & \vdots \\ \frac{\partial F_n}{\partial x_1}(\mathbf{x}) & \frac{\partial F_n}{\partial x_2}(\mathbf{x}) & \ldots & \frac{\partial F_n}{\partial x_m}(\mathbf{x}) \end{bmatrix}$$

The $i$th row consists of all of the partial derivatives of the $i$th component function and the $j$th column is the matrix

$$\frac{\partial \mathbf{x}}{\partial x_j}(\mathbf{x}) = \mathbf{F}_{x_j}(\mathbf{x}) = \begin{bmatrix} \frac{\partial F_1}{\partial x_j}(\mathbf{x}) \\ \frac{\partial F_2}{\partial x_j}(\mathbf{x}) \\ \vdots \\ \frac{\partial F_n}{\partial x_j}(\mathbf{x}) \end{bmatrix}$$

Consider the case of $f\colon \mathbb{R} \to \mathbb{R}$. Then $Df(x)$ is a $1 \times 1$ matrix whose sole entry is the derivative of the sole component $f$ with the sole variable $x$. Hence, $Df(x) = f'(x)$ is the usual derivative of $f$ with respect to $x$.

If $f\colon \mathbb{R}^m \to \mathbb{R}$ is a real valued function of $m$ variables, then $Df(\mathbf{x})$ is a $1 \times m$ matrix

$$Df(\mathbf{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}) & \frac{\partial f}{\partial x_2}(\mathbf{x}) & \dots & \frac{\partial f}{\partial x_m}(\mathbf{x}) \end{bmatrix}.$$

We can think of this as a vector in $\mathbb{R}^m$ and in this case $Df(\mathbf{x})$ is called the *gradient* of $f$ at $\mathbf{x}$, and is denoted by *grad* $f(\mathbf{x})$ or $\nabla f(\mathbf{x})$. More on this later.

If $\mathbf{s}\colon [a, b] \to \mathbb{R}^n$ is a vector valued function of one variable — a path in $\mathbb{R}^n$ — then $D\mathbf{s}(t)$ is a $n \times 1$ matrix

$$D\mathbf{s}(t) = \begin{bmatrix} x_1'(t) \\ x_2'(t) \\ \vdots \\ x_n'(t) \end{bmatrix}.$$

When this is evaluated at a point $t_0$ and interpreted as a vector, $D\mathbf{s}(t_0)$ is the *tangent vector*, or *velocity vector*, of $\mathbf{s}$, and is denoted by $\mathbf{s}'(t_0)$.

## 9.3 The Derivative and Differentiability

**Definition 9.2** *A vector-valued function $\mathbf{F} = (F_1, \dots, F_n)\colon \mathbb{R}^m \to \mathbb{R}^n$, defined on an open set $U \subseteq \mathbb{R}^m$ is differentiable at $\mathbf{a} \in U$ if*

1. *all of the partial derivatives of all of the components $F_1, \dots, F_n$ of $\mathbf{F}$ exist at $\mathbf{a}$, and*
2. *the matrix of partial derivatives $D\mathbf{F}(\mathbf{a})$ of $\mathbf{F}$ at $\mathbf{a}$ satisfies*

$$\lim_{\mathbf{x} \to \mathbf{a}} \frac{\|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a}) - D\mathbf{F}(\mathbf{a})(\mathbf{x} - \mathbf{a})\|}{\|\mathbf{x} - \mathbf{a}\|} = 0. \tag{9.3}$$

If a vector-valued function $\mathbf{F}$ satisfies the conditions above, then the matrix $D\mathbf{F}(\mathbf{a})$ of partial derivatives is called the *derivative of* $\mathbf{F}$ *at* $\mathbf{a}$.

Let's study this second condition above. What does it mean? Let $m = n = 1$ and let's see what it tells us. In this case we are considering a function $f\colon \mathbb{R} \to \mathbb{R}$ and $Df(x) = f'(x)$. Thus the above limit becomes

$$\lim_{x \to a} \frac{|f(x) - [f(a) - f'(a)(x - a)]|}{|x - a|} = 0. \tag{9.4}$$

We have seen that the expression $L_a(x) = f(a) - f'(a)(x - a)$ before. It is called the *linear approximation* of $f$ at $a$. Of course, here it represents the equation of the tangent line to the graph of $f$ at $a$.

Now the limit in 9.4 is zero and the denominator goes to zero, so the numerator $|f(x) - L_a(x)|$ must also approach zero. This tells us that $L_a(x)$ approaches $f(x)$ as $x$ approaches $a$. Now, this is not especially shocking since any line through $(a, f(a))$ satisfies this condition. However, there is more to this formula than meets the eye. Rewrite Equation 9.4 as

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

This means that the slopes of $f(x)$ and $L_a(x)$ must approach each other. This definition thus says that the tangent line is the best linear approximation to the curve $y = f(x)$ near $x = a$. This is *local linearity*.

Next, consider the case of a function $f\colon \mathbb{R}^2 \to \mathbb{R}$ and let $\mathbf{a} = (a, b)$ and $\mathbf{x} = (x, y)$. Then

$$\begin{aligned} Df(\mathbf{a})(\mathbf{x} - \mathbf{a}) &= \begin{bmatrix} \frac{\partial f}{\partial x}(a, b) & \frac{\partial f}{\partial y}(a, b) \end{bmatrix} \cdot \begin{bmatrix} x - a \\ y - b \end{bmatrix} \\ &= \frac{\partial f}{\partial x}(a, b) \cdot (x - a) + \frac{\partial f}{\partial y}(a, b) \cdot (y - b). \end{aligned}$$

Then Equation 9.3 becomes

$$\lim_{(x,y) \to (a,b)} \frac{|f(x, y) - L_{(a,b)}(x, y)|}{\sqrt{(x - a)^2 + (y - b)^2}} = 0, \tag{9.5}$$

where

$$L_{(a,b)}(x, y) = f(a, b) + \frac{\partial f}{\partial x}(a, b) \cdot (x - a) + \frac{\partial f}{\partial y}(a, b) \cdot (y - b) \tag{9.6}$$

is the linearization of $f$ at $(a, b)$. Note that this is a linear equation in the two variables $x$ and $y$ and its graph is a plane. A plane that has the point $(a, b, f(a, b))$ in common with the graph of $f$ and is the unique plane that satisfies equation 9.5 is called the *tangent plane to* $f$ *at* $(a, b)$.

First, we need to theorems to tell us how to work with these *norms* that have replaced our *absolute values* that we used when looking at functions of one variable.

**Theorem 9.1 (Cauchy-Schwarz Inequality)** *Let* $\mathbf{v}$ *and* $\mathbf{w}$ *be vectors in* $\mathbb{R}^m$, $m \geq 1$. *Then* $|\mathbf{v} \cdot \mathbf{w}| \leq \|v\| \|w\|$.

The proof is left as an exercise.

Let $A$ be an $n \times m$ matrix and let $\mathbf{x} \in \mathbb{R}^m$. Then by $A \cdot \mathbf{x}$ we mean the matrix product of $A$ and $\mathbf{x}$ where $\mathbf{x}$ is thought of as an $m \times 1$ matrix. Then $A \cdot \mathbf{x}$ is an $n \times 1$ matrix, or vector in $\mathbb{R}^n$ and we can use $\|A \cdot \mathbf{x}\|$ to denote its norm.

The norm of an $n \times m$ matrix $A = [a_{ij}]$ is defined by

$$\|A\| = \left( \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij}^2 \right)^{1/2}.$$

**Theorem 9.2** *Let* $A$ *be and* $n \times m$ *matrix and let* $\mathbf{x} \in \mathbb{R}^m$ *be any vector. Then* $\|A \cdot \mathbf{x}\| \leq \|A\| \|\mathbf{x}\|$.

PROOF: Let $\mathbf{x} = (x_1, \ldots, x_m)$, then $A \cdot \mathbf{x}$ is

$$A \cdot \mathbf{x} = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1m} \\ a_{21} & a_{22} & \ldots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nm} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m \end{bmatrix}$$

The square of its magnitude is

$$\|A \cdot \mathbf{x}\|^2 = (a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m)^2 + (a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m)^2 + \cdots + (a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m)^2. \quad (9.7)$$

By the Cauchy-Schwarz Inequality with $\mathbf{a} = \langle a_{11}, a_{12}, \ldots, a_{1m} \rangle$ and $\mathbf{x}$ we have

$$(\mathbf{a} \cdot \mathbf{x})^2 \leq \|\mathbf{a}\|^2 \|\mathbf{x}\|^2$$

or

$$(a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m)^2 \leq (a_{11}^2 + a_{12}^2 + \cdots + a_{1m}^2)(x_1^2 + x_2^2 + \cdots + x_m^2).$$

This is the first expression, and we can estimate the remaining terms in the expression for $\|A \cdot \mathbf{x}\|^2$ and we get

$$\begin{aligned} \|A \cdot \mathbf{x}\|^2 &\leq (a_{11}^2 + a_{12}^2 + \cdots + a_{1m}^2)(x_1^2 + x_2^2 + \cdots + x_m^2) \\ &\quad (a_{21}^2 + a_{22}^2 + \cdots + a_{2m}^2)(x_1^2 + x_2^2 + \cdots + x_m^2) \\ &\quad + \cdots + (a_{n1}^2 + a_{n2}^2 + \cdots + a_{nm}^2)(x_1^2 + x_2^2 + \cdots + x_m^2) \\ &= (a_{11}^2 + a_{12}^2 + \cdots + a_{nm}^2)(x_1^2 + x_2^2 + \cdots x_m^2) \\ &= \left( \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij}^2 \right) \|\mathbf{x}\|^2 \end{aligned}$$

Therefore

$$\|A\cdot\mathbf{x}\| \leq \left(\sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}^2\right)^{1/2} \|\mathbf{x}\| = \|A\|\|\mathbf{x}\|.$$

∎

**Theorem 9.3** *Let* $\mathbf{F}\colon \mathbb{R}^m \to \mathbb{R}^n$ *be differentiable at* $\mathbf{a}\in\mathbb{R}^m$*. Then*

$$\|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a})\| \leq (1 + \|D\mathbf{F}(\mathbf{a})\|)\|\mathbf{x}-\mathbf{a}\|,$$

*for* $\mathbf{x}$ *near* $\mathbf{a}$*, such that* $\mathbf{x}\neq\mathbf{a}$*.*

Proof: By the definition of differentiability, since $\mathbf{F}$ is differentiable at $\mathbf{a}$

$$\lim_{\mathbf{x}\to\mathbf{a}} \frac{\|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a}) - DF(\mathbf{a})\|\mathbf{x}-\mathbf{a}\|}{\|\mathbf{x}-\mathbf{a}\|} = 0.$$

Now, we can interpret this in terms of $\epsilon$ and $\delta$. Since this limit exists, for any $\epsilon > 0$ there is a $\delta > 0$ so that

$$\left|\frac{\|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a}) - DF(\mathbf{a})\|\mathbf{x}-\mathbf{a}\|}{\|\mathbf{x}-\mathbf{a}\|} - 0\right| < \epsilon$$

whenever $0 < \|\mathbf{x}-\mathbf{a}\| < \delta$. Thus for $\epsilon = 1$, there is a $\delta > 0$ so that whenever $0 < \|\mathbf{x}-\mathbf{a}\| < \delta$ we have

$$\frac{\|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a}) - DF(\mathbf{a})\|\mathbf{x}-\mathbf{a}\|}{\|\mathbf{x}-\mathbf{a}\|} < 1,$$

which means that

$$\|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a}) - DF(\mathbf{a})\|\mathbf{x}-\mathbf{a}\| < \|\mathbf{x}-\mathbf{a}\|.$$

Now, by the Triangle Inequality and this above inequality we have

$$\begin{aligned}\|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a})\| &= \|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a}) - D\mathbf{F}(\mathbf{a})(\mathbf{x}-\mathbf{a}) + D\mathbf{F}(\mathbf{a})(\mathbf{x}-\mathbf{a})\| \\ &\leq \|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a}) - D\mathbf{F}(\mathbf{a})(\mathbf{x}-\mathbf{a})\| + \|D\mathbf{F}(\mathbf{a})(\mathbf{x}-\mathbf{a})\| \\ &\leq \|\mathbf{x}-\mathbf{a}\| + \|D\mathbf{F}(\mathbf{a})(\mathbf{x}-\mathbf{a})\| \\ &\leq \|\mathbf{x}-\mathbf{a}\| + \|D\mathbf{F}(\mathbf{a})\|\|\mathbf{x}-\mathbf{a}\| = (1+\|D\mathbf{F}(\mathbf{a})\|)\|\mathbf{x}-\mathbf{a}\|.\end{aligned}$$

for $0 \leq \|\mathbf{x}-\mathbf{a}\| < \delta$. ∎

**Theorem 9.4** *If* $\mathbf{F}\colon \mathbb{R}^m \to \mathbb{R}^n$ *is a vector valued function and if* $\mathbf{F}$ *is differentiable at* $\mathbf{a}$*, then it is continuous at* $\mathbf{a}$*.*

PROOF: We need to show that $\lim_{\mathbf{x}\to\mathbf{a}} \mathbf{F}(\mathbf{x}) = \mathbf{F}(\mathbf{a})$. This is equivalent to showing that $\lim_{\mathbf{x}\to\mathbb{Z}} \|F$
$bF(\mathbf{x}) - \mathbf{F}(\mathbf{a})\| = 0$. Since $\mathbf{F}$ is differentiable at $\mathbf{a}$, it follows from Theorem 9.3 that there is a $\delta > 0$ such that

$$\|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a})\| \leq (1 + \|D\mathbf{F}(\mathbf{a})\|)\|\mathbf{x} - \mathbf{a}\|, \tag{9.8}$$

whenever $0 < \|\mathbf{x} - \mathbf{a}\| < \delta$. Choose any $\epsilon > 0$ and let $\delta_1 < \min\{\delta, \epsilon(1 + \|D\mathbf{F}(\mathbf{a})\|)^{-1}\}$. Since $\delta_1 < \delta$ we have that if $0 < \|\mathbf{x} - \mathbf{a}\| < \delta_1$ then Equation 9.8 holds, so

$$\|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a})\| < (1 + \|D\mathbf{F}(\mathbf{a})\|)\delta_1 < (1 + \|D\mathbf{F}(\mathbf{a})\|)\frac{\epsilon}{1 + \|D\mathbf{F}(\mathbf{a})\|} = \epsilon.$$

Thus, $\mathbf{F}$ is continuous at $\mathbf{a}$. ∎

Note that the theorem says that $\mathbf{F}$ must be differentiable. This means more than just that the partial derivatives exist. Consider $f\colon \mathbb{R}^2 \to \mathbb{R}$ given by

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

Then, from the definition

$$\frac{\partial f}{\partial x}(0, 0) = \lim_{h\to 0} \frac{f(h, 0) - f(0)}{h} = \lim_{h\to 0} \frac{0 - 0}{h} = 0$$

and

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

Therefore, both partial derivatives exist and are 0 at $(0, 0)$. The limit of $f$ at 0, however, does not exist. We can show that by showing that it goes to two different values along different paths to $(0, 0)$. Along the $y$-axis we get

$$\lim_{x=0, y\to 0} \frac{xy}{x^2 + y^2} = \lim_{y\to 0} \frac{0}{y^2} = 0,$$

while along the line $y = x$ we get

$$\lim_{(x,y)\to(0,0)} \frac{xy}{x^2 + y^2} = \lim_{x\to 0} \frac{x^2}{x^2 + y^2} = \frac{1}{2}.$$

Thus this limit does not exist and $f$ cannot be continuous at $(0, 0)$.

This means that $f$ cannot be differentiable. How can we see this? We know that the partial derivatives exist at $(0, 0)$, so we now need to check the linear approximation condition, that is we need to see if

$$\lim_{(x,y)\to(0,0)} \frac{|f(x, y) - L_{(0,0)}(x, y)|}{\sqrt{x^2 + y^2}} = 0.$$

Since $\partial f/\partial x(0,0) = 0 = \partial f/\partial y(0,0)$ we have that

$$L_{(0,0)}(x,y) = f(0,0) + 0 \cdot (x-0) + 0 \cdot (y-0) = 0.$$

This makes the above limit

$$\lim_{(x,y)\to(0,0)} \frac{\left|\frac{xy}{x^2+y^2} - 0\right|}{\sqrt{x^2+y^2}} = \lim{(x,y) \to (0,0)} \frac{|xy|}{(x^2+y^2)^{3/2}}.$$

If we approach along the path $y = x$, we get

$$\lim{(x,y) \to (0,0)} \frac{|xy|}{(x^2+y^2)^{3/2}} = \lim_{x\to 0} \frac{x^2}{(2x^2)^{3/2}} = \lim_{x \to 0} \frac{1}{2^{3/2}x}$$

which does not exist. This means that the linearization condition cannot hold and $f$ cannot be differentiable at $(0,0)$.

All is not lost however. The existence of the partial derivatives should have some bearing on how the function behaves, and it certainly does.

**Theorem 9.5** *Let* $\mathbf{F}\colon \mathbb{R}^m \to \mathbb{R}^n$ *be a vector-values function with component functions* $F_1, \ldots, F_n \colon \mathbb{R}^m \to \mathbb{R}$*. If all partial derivatives* $\partial F_i/\partial x_j$ $(i = 1, \ldots, n,\ j = 1, \ldots, m)$ *are continuous at* $\mathbf{a}$*, then* $\mathbf{F}$ *is differentiable at* $\mathbf{a}$*.*

Proof: We will prove this for $m = 2$ and $n = 1$ first and then see how it extends to the general case. Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ and assume that $\partial f/\partial x$ and $\partial f/\partial y$ are continuous at $\mathbf{a} = (a,b)$. To show that $f$ is differentiable at $\mathbf{a}$ we must show that

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

Now write

$$f(\mathbf{x}) - f(\mathbf{a}) = f(x,y) - f(a,b) = f(x,y) - f(a,y) + f(a,y) - f(a,b).$$

By the Mean Value Theorem applied to $g(x) = f(x,y)$ by keeping $y$ fixed, there is a number $c_1$ between $x$ and $a$ so that

$$f(x,y) - f(a,y) = \frac{\partial f}{\partial x}(c_1, y)(x-a).$$

Similarly, there is a number $c_2$ between $y$ and $b$ so that

$$f(a,y) - f(a,b) = \frac{\partial f}{\partial y}(a, c_2)(y-b).$$

Thus,

$$f(x,y) - f(a,b) = \frac{\partial f}{\partial x}(c_1,y)(x-a) + \frac{\partial f}{\partial y}(a,c_2)(y-b)$$

and

$$\begin{aligned}
&|f(x,y) - f(a,b) - Df(a,b)(x-a, y-b)| \\
&\quad = \left| \frac{\partial f}{\partial x}(c_1, y)(x-a) + \frac{\partial f}{\partial y}(a, c_2)(y-b) - \frac{\partial f}{\partial x}(a,b)(x-a) - \frac{\partial f}{\partial y}(a,b)(y-b) \right| \\
&\quad = \left| \left( \frac{\partial f}{\partial x}(c_1, y) - \frac{\partial f}{\partial x}(a,b) \right)(x-a) + \left( \frac{\partial f}{\partial y}(a, c_2) - \frac{\partial f}{\partial y}(a,b) \right)(y-b) \right| \\
&\quad \le \left| \frac{\partial f}{\partial x}(c_1, y) - \frac{\partial f}{\partial x}(a,b) \right| |x-a| + \left| \frac{\partial f}{\partial y}(a, c_2) - \frac{\partial f}{\partial y}(a,b) \right| |y-b| \\
&\quad \le \left( \left| \frac{\partial f}{\partial x}(c_1, y) - \frac{\partial f}{\partial x}(a,b) \right| + \left| \frac{\partial f}{\partial y}(a, c_2) - \frac{\partial f}{\partial y}(a,b) \right| \right) \|(x-a, y-b)\|.
\end{aligned}$$

Finally,

$$\frac{|f(\mathbf{x}) - f(\mathbf{a}) - Df(\mathbf{a})(\mathbf{x}-\mathbf{a})|}{\|\mathbf{x}-\mathbf{a}\|} \le \left| \frac{\partial f}{\partial x}(c_1, y) - \frac{\partial f}{\partial x}(a,b) \right| + \left| \frac{\partial f}{\partial y}(a, c_2) - \frac{\partial f}{\partial y}(a,b) \right|.$$

Now, as $x \to a$ and $y \to b$, we must have that $c_1 \to a$ and

$$\frac{\partial f}{\partial x}(c_1, y) - \frac{\partial f}{\partial x}(a,b) \to 0,$$

because $\partial f / \partial x$ is continuous at $(a,b)$. Similarly, as $x \to a$ and $y \to b$

$$\frac{\partial f}{\partial y}(a, c_2) - \frac{\partial f}{\partial y}(a,b) \to 0,$$

and we are done.

Clearly, this technique of proof can be applied to any function $f\colon \mathbb{R}^m \to \mathbb{R}$ with $m \ge 2$, so that we can see the theorem is true for any such function. The general case of a vector valued function $\mathbf{F}\colon \mathbb{R}^m \to \mathbb{R}^n$ follows from the fact that everything is done componentwise. ∎

**Definition 9.3** *A function whose partial derivatives exist and are continuous is said to be* continuously differentiable*, or of* class $C^1$*. A function whose partials are $k$-times differentiable and continuous is said to be of class $C^k$. If all partials of all orders are differentiable and continuous, then the function is said to be of class $C^\infty$ or* smooth*. A function that is continuous but not differentiable is said to be of class $C^0$.*

Examples of smooth functions $f\colon \mathbb{R} \to \mathbb{R}$ are the sine function, cosine function and exponential function among many. The absolute value function is a function of class $C^0$. If $g(x) = \int_0^x |t|\,dt$, then $g$ is $C^1$ function.

## 9.4 Properties of Derivatives

**Theorem 9.6** *Assume that* $\mathbf{F}, \mathbf{G} \colon \mathbb{R}^m \to \mathbb{R}^n$ *and* $f, g \colon \mathbb{R}^m \to \mathbb{R}$ *are differentiable at* $\mathbf{a} \in \mathbb{R}^m$*, let* $c \in \mathbb{R}$*, and let* $\mathbf{v}, \mathbf{w} \colon \mathbb{R} \to \mathbb{R}^n$ *be differentiable at* $a \in \mathbb{R}$*.*

1. *The sum* $\mathbf{F} + \mathbf{G}$ *and difference* $\mathbf{F} - \mathbf{G}$ *are differentiable at* $\mathbf{a}$ *and*
$$D(\mathbf{F} \pm \mathbf{G})(\mathbf{a}) = D\mathbf{F}(\mathbf{a}) \pm D\mathbf{G}(\mathbf{a}).$$
2. *The product* $c\mathbf{F}$ *is differentiable at* $\mathbf{a}$ *and*
$$D(c\mathbf{F})(\mathbf{a}) = cD\mathbf{F}(\mathbf{a}).$$
3. *The product* $fg$ *is differentiable and*
$$D(fg)(\mathbf{a}) = g(\mathbf{a})Df(\mathbf{a}) + f(\mathbf{a})Dg(\mathbf{a}).$$
4. *If* $g(\mathbf{a}) \neq 0$*, then the quotient* $f/g$ *is differentiable at* $\mathbf{a}$ *and*
$$D\left(\frac{f}{g}\right)(\mathbf{a}) = \frac{g(\mathbf{a})Df(\mathbf{a}) - f(\mathbf{a})Dg(\mathbf{a})}{g(\mathbf{a})^2}.$$
5. *The dot product* $\mathbf{v} \cdot \mathbf{w}$ *is differentiable at* $a$ *and*
$$(\mathbf{v} \cdot \mathbf{w})'(a) = \mathbf{v}'(a) \cdot \mathbf{w}(a) + \mathbf{v}(a) \cdot \mathbf{w}'(a).$$
6. *If* $\mathbf{v}, \mathbf{w} \colon \mathbb{R} \to \mathbb{R}^3$*, then the cross product* $\mathbf{v} \times \mathbf{w}$ *is differentiable at* $a$ *and*
$$(\mathbf{v} \times \mathbf{w})'(a) = \mathbf{v}'(a) \times \mathbf{w}(a) + \mathbf{v}(a) \times \mathbf{w}'(a).$$

**Theorem 9.7 (The Chain Rule)** *Suppose* $\mathbf{F} \colon U \subseteq \mathbb{R}^m \to \mathbb{R}^n$ *is differentiable at* $\mathbf{a} \in U$*,* $U$ *is open in* $\mathbb{R}^m$*,* $\mathbf{G} \colon V \subseteq \mathbb{R}^n \to \mathbb{R}^p$ *is differentiable at* $\mathbf{F}(\mathbf{a}) \in V$*,* $V$ *is open in* $\mathbb{R}^n$*, and* $\mathbf{F}(U) \subseteq V$*. Then* $\mathbf{G}\circ$ *is differentiable at* $\mathbf{a}$ *and*
$$D(\mathbf{G} \circ \mathbf{F}(\mathbf{a}) = D\mathbf{G}(\mathbf{F}(\mathbf{a}))D\mathbf{F}(\mathbf{a}),$$
*where the right hand side is matrix multiplication.*

This is not hard to understand, but it does require careful bookkeeping in order to keep the different pieces straight. Let $\mathbf{F} \colon \mathbb{R}^2 \to \mathbb{R}^3$ be given by $\mathbf{F}(x, y) = (x^3 + y, x^2y^2, 2 + xy)$ and $\mathbf{G} \colon \mathbb{R}^3 \to \mathbb{R}^2$ be given by $\mathbf{G}(u, v, w) = (u^2 + v, uv^2 + uvw)$. Then $D(\mathbf{G} \circ \mathbf{F})(1, 1)$ is found by
$$D(\mathbf{G} \circ \mathbf{F})(1, 1) = D\mathbf{G}(\mathbf{F}(1, 1))D\mathbf{F}(1, 1) = D\mathbf{G}(2, 1, 3)D\mathbf{F}(1, 1).$$
$$D\mathbf{F}(1, 1) = \begin{bmatrix} 3x^2 & 1 \\ 2xy^2 & 2yx^2 \\ y & x \end{bmatrix}_{(x,y)=(1,1)} = \begin{bmatrix} 3 & 1 \\ 2 & 2 \\ 1 & 1 \end{bmatrix}$$

and

$$D\mathbf{G}(u,v,w) = \begin{bmatrix} 2u & 1 & 0 \\ v^2+vw & 2uv+uw & uv \end{bmatrix}_{(u,v,w)=(2,1,3)} = \begin{bmatrix} 4 & 1 & 0 \\ 4 & 10 & 2 \end{bmatrix},$$

so

$$D(\mathbf{G}\circ\mathbf{F})(1,1) = \begin{bmatrix} 4 & 1 & 0 \\ 4 & 10 & 2 \end{bmatrix}\begin{bmatrix} 3 & 1 \\ 2 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 14 & 6 \\ 34 & 26 \end{bmatrix}.$$

**Example 9.1** Let $f\colon \mathbb{R}^3 \to \mathbb{R}$ and let $\mathbf{G}\colon \mathbb{R}^3 \to \mathbb{R}^3$ be given by

$$\mathbf{G}(x,y,z) = (u(x,y,z), v(x,y,z), w(x,y,z)).$$

Assume that $f$ and $\mathbf{G}$ are differentiable and define $h\colon \mathbb{R}^3 \to \mathbb{R}$ by $h = f\circ\mathbf{G}$. Then

$$h(x,y,z) = f(\mathbf{G}(x,y,z)) = f(u(x,y,z), v(x,y,z), w(x,y,z)).$$

To find the partial derivatives of $h$ we need to compute the derivative

$$Dh = \begin{bmatrix} \frac{\partial h}{\partial x} & \frac{\partial h}{\partial y} & \frac{\partial h}{\partial z} \end{bmatrix}.$$

Now, use $D_1 f$, $D_2 f$ and $D_3 f$ to denote the partial derivatives of $f$ with respect to its variables, so $Df = [D_1 f \;\, D_2 f \;\, D_3 f]$ and

$$Dh = Df D\mathbf{G} = \begin{bmatrix} D_1 f & D_2 f & D_3 f \end{bmatrix} \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} & \frac{\partial u}{\partial z} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} & \frac{\partial v}{\partial z} \\ \frac{\partial w}{\partial x} & \frac{\partial w}{\partial y} & \frac{\partial w}{\partial z} \end{bmatrix},$$

and so

$$\begin{aligned} \frac{\partial h}{\partial x} &= D_1 f\frac{\partial u}{\partial x} + D_2 f\frac{\partial v}{\partial x} + D_3 f\frac{\partial w}{\partial x} \\ \frac{\partial h}{\partial y} &= D_1 f\frac{\partial u}{\partial y} + D_2 f\frac{\partial v}{\partial y} + D_3 f\frac{\partial w}{\partial y} \\ \frac{\partial h}{\partial z} &= D_1 f\frac{\partial u}{\partial z} + D_2 f\frac{\partial v}{\partial z} + D_3 f\frac{\partial w}{\partial z} \end{aligned}$$

Using $u$, $v$, and $w$ for the variable of $f$, we have that $D_1 f = \partial f/partialu$, $D_2 f = \partial f/partialv$, and $D_3 f = \partial f/\partial w$, and the above becomes

$$\begin{aligned} \frac{\partial h}{\partial x} &= \frac{\partial f}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial f}{\partial v}\frac{\partial v}{\partial x} + \frac{\partial f}{\partial w}\frac{\partial w}{\partial x} \\ \frac{\partial h}{\partial y} &= \frac{\partial f}{\partial u}\frac{\partial u}{\partial y} + \frac{\partial f}{\partial v}\frac{\partial v}{\partial y} + \frac{\partial f}{\partial w}\frac{\partial w}{\partial y} \\ \frac{\partial h}{\partial z} &= \frac{\partial f}{\partial u}\frac{\partial u}{\partial z} + \frac{\partial f}{\partial v}\frac{\partial v}{\partial z} + \frac{\partial f}{\partial w}\frac{\partial w}{\partial z} \end{aligned} \tag{9.9}$$

## 9.5 Gradient and Directional Derivative

Recall that the derivative $Df(\mathbf{x})$ of a differentiable real-valued function $f\colon \mathbb{R}^m \to \mathbb{R}$ is a $1 \times m$ matrix

$$Df(\mathbf{x}) = \left[\frac{\partial f}{\partial x_1}(\mathbf{x}) \quad \frac{\partial f}{\partial x_2}(\mathbf{x}) \quad \dots \quad \frac{\partial f}{\partial x_m}(\mathbf{x})\right].$$

This matrix, when interpreted as a vector in $\mathbb{R}^m$ is denoted by $\nabla f(\mathbf{x})$ or *grad* $f(\mathbf{x})$ and is called the *gradient* of $f$ at $\mathbf{x}$.

The gradient of $f\colon \mathbb{R}^2 \to \mathbb{R}$ is the vector field $\nabla f$ given by

$$\nabla f(x,y) = \left\langle \frac{\partial f}{\partial x}(x,y), \frac{\partial f}{\partial y}(x,y) \right\rangle.$$

The gradient of $f\colon \mathbb{R}^3 \to \mathbb{R}$ is defined similarly.

Now, the partial derivatives measure the change in the values of the function as we move parallel to the coordinate axes. That is somewhat restrictive and we may need to know how a function changes if we move in any arbitrary direction from a point.

Let $f(x,y)$ be a function differentiable on an open set $U \subseteq \mathbb{R}^2$, and let $\mathbf{p} = (a,b) \in U$. We want to compute the rate of change of $f(x,y)$ at $\mathbf{p}$ in the direction of a unit vector $\mathbf{u} = (u,v)$.

Now, recall that $\ell(t) = \mathbf{p} + t\mathbf{u} = (a+tu, b+tv)$ represents the line in $\mathbb{R}^2$ that passes through the point $\mathbf{p} = \ell(0)$ and whose direction is given by the the direction of $\mathbf{u} = (u,v)$. Now, we can find $f(\ell(t)) = f(\mathbf{p}+t\mathbf{u})$. The collection of points $\mathbf{s}(t) = (a+tu, b+tv, f(\mathbf{p}+t\mathbf{u}))$ is a curve on the surface of the graph of $f$. Note that $\mathbf{s}(t)$ lies in the plane that is perpendicular to the $xy$-plane along the line $\ell(t)$. We can then compute the tangent to the curve $\mathbf{s}(t)$ at $t=0$ and that will be the change in $f$ in the direction of $\mathbf{u}$.

**Definition 9.4** *Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be a real valued differentiable function. The* directional derivative *of $f$ at the point $\mathbf{p}$ in the direction of a unit vector $\mathbf{u}$ is given by*

$$D_{\mathbf{u}}f(a,b) = \frac{d}{dt}f(\mathbf{p}+t\mathbf{u})\bigg|_{t=0}.$$

This is simply

$$D_{\mathbf{u}}f(a,b) = \lim_{t\to 0}\frac{f(\mathbf{p}+t\mathbf{u}) - f(\mathbf{p}}{t} = \lim_{t\to 0}\frac{f(9a,b) + t(u,v)) - f(a,b)}{t}.$$

**Theorem 9.8** *Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be a differentiable function and let $\mathbf{p} = (a,b)$. Then*

$$D_{\mathbf{u}}f(a,b) = \nabla f(a,b) \cdot \mathbf{u},$$

*where $\mathbf{u}$ is a unit vector in $\mathbb{R}^2$.*

The directional derivative can be defined similarly for any function $f\colon \mathbb{R}^m \to \mathbb{R}$.

**Theorem 9.9 (Maximum Rate of Change)** *Let $f\colon u \subseteq \mathbb{R}^m \to \mathbb{R}$ and assume that $\nabla f(\mathbf{x}) \neq 0$ for $\mathbf{x} \in U$. The direction of the largest increase in $f$ at $\mathbf{x}$ is given by the vector $\nabla f$.*

PROOF: The rate of change in $f$ is given by its directional derivative, $D_{\mathbf{u}}f$. This is $D_{\mathbf{u}}f = \nabla f \cdot \mathbf{u}$. Now, $D_{\mathbf{u}}f = \|\nabla f\|\|\mathbf{u}\|\cos\theta$ where $\theta$ is the angle between the vector $\nabla f$ and $\mathbf{u}$. Since $\mathbf{u}$ is a unit vector, we have $D_{\mathbf{u}}f = \|\nabla f\|\cos\theta$. Between 0 and $2\pi$ the cosine function is a maximum at $\theta = 0$, so the largest value of $D_{\mathbf{u}}f$ must be when $\theta = 0$ or in the direction $D_{\mathbf{u}}f$. ∎

**Theorem 9.10** *Let $f\colon U \subseteq \mathbb{R}^2 \to \mathbb{R}$ be a differentiable function. Then the gradient of $f$ at $(a, b)$ is perpendicular to the level curve passing through $(a, b)$.*

This result generalizes to $m$-space as well.

**Theorem 9.11 (Properties of the Gradient)** *Let $f\colon \mathbb{R}^m \to \mathbb{R}$ be differentiable.*

1. *If $\nabla f(\mathbf{x}) = 0$ at some point $\mathbf{x}$, then $D_{\mathbf{u}}f(\mathbf{x}) = 0$ in all directions.*
2. *If $\nabla f(\mathbf{x}) \neq 0$, then there is a direction at $\mathbf{x}$ [given by $\nabla f(\mathbf{x})$] where $f$ increases most rapidly. The magnitude of the most rapid increase is : $\nabla f(\mathbf{x})\|$. The opposite direction, $-\nabla f(\mathbf{x})$, points in the direction of most rapid decrease in $f$. The rate of change of $f$ in the directions perpendicular to $\nabla f(\mathbf{x})$ is zero.*
3. *The gradient at $\mathbf{a}$ is perpendicular to the level hypersurface that passes through $\mathbf{a}$.*