1. Derivatives of Functions of a Single Variable

1a) Our ultimate interest is derivatives of multivariable functions, but we begin with a quick review of derivatives of single variable functions.

Consider a function \( f : X \rightarrow Y \), where \( X \) and \( Y \) are subsets of \( \mathbb{R} \). Let \( x_0 \) be any point in the interior of \( X \). Then, for given \( x_0 \), we may define a new function \( q = q(h) \), called a difference quotient, with values given by

\[
q(h) \equiv \frac{f(x_0 + h) - f(x_0)}{h} = \frac{\Delta f(x)}{\Delta x}.
\]

The domain of the function is all non-zero \( h \) values such that \( x_0 + h \) belongs to \( X \). \( x_0 \) is a parameter rather than a variable of the function; there is a different \( q \) function for each different value of \( x_0 \).

1b) Geometrically, the difference quotient \( q = q(h) \) has the interpretation of the slope of a straight line joining the points \( (x_0, f(x_0)) \) and \( (x_0 + h, f(x_0 + h)) \), as shown in Figure 1. Plainly, different values of \( h \) will give different lines and hence, in most cases, different slopes, i.e., different values of \( q \). Note that, while a positive \( h \) is illustrated in Figure 1, \( h \) can be either positive or negative.

1c) The derivative of the function \( f \) at \( x_0 \), denoted \( f'(x_0) \), is defined by:

\[
f'(x_0) \equiv \lim_{h \to 0} q(h) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h},
\]

provided the limit exists and is finite.

The limit may or may not exist. If it does, then the function is said to be differentiable at \( x_0 \). A function may be differentiable at some points and not others; if it is differentiable at every point in its domain, the function is said to be differentiable.

1d) Geometrically, the derivative has the interpretation of the slope of a tangent line to the function at \( x_0 \), as illustrated in Figure 2. Intuitively, a function is differentiable at a point if its graph is “smooth” at that point so that a well-defined tangent exists.

1e) If \( f \) is differentiable at \( x_0 \), then \( f \) is continuous at \( x_0 \).

1f) Sometimes we generalise the definition of a derivative to allow derivatives to be infinite (geometrically, infinite derivatives occur where a curve becomes vertical at a single point). It is important to note, however, that most theorems concerning derivatives are proved on the assumption that the derivative is finite and these theorems may not be valid if infinite derivatives are allowed. In particular, the existence of an infinite limit in (2) does not guarantee that the function is continuous at \( x_0 \).
2. Derivative Rules and Examples

2a) By applying the definition in (2), the following derivatives are easily calculated (in the first two derivatives, \( c \) is any constant):

\[
\begin{align*}
    f(x) &= c & \Rightarrow & & f'(x_0) &= 0 \\
    f(x) &= c \cdot x & \Rightarrow & & f'(x_0) &= c \\
    f(x) &= x^2 & \Rightarrow & & f'(x_0) &= 2x_0
\end{align*}
\]

Somewhat harder to prove are:

\[
\begin{align*}
    f(x) &= e^x & \Rightarrow & & f'(x_0) &= e^{x_0} \\
    f(x) &= \ln x \ (x > 0) & \Rightarrow & & f'(x_0) &= \frac{1}{x_0} \\
    f(x) &= x^r \ (x_0 > 0, r \in \mathbb{R}) & \Rightarrow & & f'(x_0) &= rx_0^{r-1}
\end{align*}
\]

Note. Especially when a differentiation formula holds for all \( x_0 \) in the function’s domain, we usually drop the “0” subscript from \( x_0 \) when writing the differentiation formula, e.g., if \( f(x) = x^2 \), then we write \( f'(x) = 2x \).

2b) The derivative of a function \( f \), denoted \( f' \), is itself a function, with a domain consisting of all the points in the domain of \( f \) at which \( f \) is differentiable. Since \( f' \) is a
function, it may itself potentially be differentiated. The derivative of \( f' \) is called the **second derivative** of the original function \( f \) and is denoted \( f'' \). Similarly, the third derivative of \( f \) is the (first) derivative of \( f'' \) and is denoted \( f''' \). e.g., if \( f(x) = x^2 \), then \( f'(x) = 2x, \ f''(x) = 2 \) and \( f'''(x) = 0 \).

**2c)** We can often calculate derivatives of complicated functions from derivatives of simpler functions. Specifically, if \( f'(x) \) and \( g'(x) \) both exist at some point \( x \), then

\[
\begin{align*}
    h(x) = f(x) + g(x) & \implies h'(x) = f'(x) + g'(x) \\
    h(x) = f(x) - g(x) & \implies h'(x) = f'(x) - g'(x) \\
    h(x) = f(x) \cdot g(x) & \implies h'(x) = f'(x) \cdot g(x) + f(x) \cdot g'(x) \\
    h(x) = f(x)/g(x) & \implies h'(x) = \frac{f'(x) \cdot g(x) - f(x) \cdot g'(x)}{[g(x)]^2} \quad (g(x) \neq 0)
\end{align*}
\]

Combining these rules with the derivatives in **2a)**, it is easy to see, for example, that:

\[
\begin{align*}
    h(x) = 3x + 2x^2 - 4x^3 & \implies h'(x) = 3 + 4x - 12x^2 \\
    h(x) = \frac{2 - x}{\ln x + x^2} & \implies h'(x) = \frac{-1 \cdot (\ln x + x^2) - (2 - x) \cdot (1/x + 2x)}{[\ln x + x^2]^2} \\
    h(x) = e^x \cdot \ln x & \implies h'(x) = e^x \cdot \ln x + e^x \cdot \frac{1}{x}.
\end{align*}
\]
2d) For some purposes (especially the calculation of derivatives at the endpoints of a closed interval), we need the concept of a one-sided derivative. If \( f \) is defined on the semi-open interval \((c, x_0]\), we may define the left-hand derivative
\[
f'_{-}(x_0) \equiv \lim_{h \to 0^-} \frac{f(x_0 + h) - f(x_0)}{h}.
\]
The symbolism \( h \to 0^- \) denotes that \( h \) approaches zero from below. It is the left-hand limit (the definition of a left-hand limit is the same as that of a regular limit except that we only consider \( x \) values below \( x_0 \)).

Similarly, if \( f \) is defined on the semi-open interval \([x_0, c)\), we may define the right-hand derivative
\[
f'_{+}(x_0) \equiv \lim_{h \to 0^+} \frac{f(x_0 + h) - f(x_0)}{h},
\]
where the symbolism \( h \to 0^+ \) denotes that \( h \) approaches zero from above.

Note that if we are considering the derivative of \( f \) on a closed interval \([a, b]\), we normally write \( f'(a) \) and \( f'(b) \) for the one-sided derivatives at the endpoints, the one-sided nature of the derivatives being understood.

2e) **Mean Value Theorem for Derivatives.** Assume that \( f \) is continuous on a closed interval \([a, b]\) and that \( f \) is differentiable on the open interval \((a, b)\). Then there is at least one point \( c \in (a, b) \) such that
\[
f(b) - f(a) = f'(c) \cdot (b - a).
\]

2f) **Intermediate Value Theorem for Derivatives.** Assume that \( f \) is differentiable on a closed interval \([a, b]\). For any two points \( x_1 < x_2 \in [a, b] \) such that \( f'(x_1) \neq f'(x_2) \), \( f' \) takes on every value between \( f'(x_1) \) and \( f'(x_2) \) somewhere on the open interval \((x_1, x_2)\).

2g) **(Special Case of) Taylor’s Theorem.** Consider \( f: I \to \mathbb{R} \), where \( I \) is an interval in \( \mathbb{R} \) and \( f \) has a continuous second derivative. For any points \( x_0 \) and \( x_0 + h \) in \( I \), there exists \( c \) between \( x_0 \) and \( x_0 + h \) such that
\[
f(x_0 + h) = f(x_0) + f'(x_0)h + \frac{1}{2}f''(c)h^2.
\]

3. **Differentials**

3a) For some purposes, the equation of the tangent line to a function at a point \( x_0 \) is of interest. It provides a linear approximation to the function in the vicinity of \( x_0 \). Its equation is given by:
\[
y_t = f(x_0) + f'(x_0)(x - x_0)
\]
(the \( t \) subscript is to indicate that it is the \( y \) value along the tangent line).

Another way of looking at the tangent line is in terms of the changes in \( x \) and \( y \) that occur along it. Re-arranging (5) gives:
\[
y_t - f(x_0) = f'(x_0)(x - x_0),
\]
which may be written as
\[ dy \equiv f'(x_0)dx, \tag{6} \]
where
\[ dy \equiv y_t - f(x_0) \quad \text{and} \quad dx \equiv x - x_0, \]
i.e., \(dy\) is the change in \(y\) along the tangent line which results from the change in \(x\) given by \(dx\). The expressions \(dy\) and \(dx\) satisfying (6) are known as differentials. Note that there is no implication that the changes \(dx\) and \(dy\) are small or infinitesimal.

Note. Sometimes \(dy\) is written as \(d f\).

If now we divide both sides of (6) by \(dx\), we get:
\[ \frac{dy}{dx} = f'(x_0), \]
i.e., the derivative \(f'(x_0)\) equals the ratio of the differentials, \(dy\) and \(dx\) — a conclusion that the notation has been designed to suggest. There is nothing very profound about this: it is essentially a matter of definition. In particular, note that this approach does not supply an alternative way to define a derivative, since the definition of a differential in (6) presupposes that the derivative \(f'(x_0)\) is already defined.

3b) Suppose that \(u = f(x)\) and \(v = g(x)\). Then it is easy to show that
\[ d(u \pm v) = du \pm dv \]
\[ d(u \cdot v) = du \cdot v + u \cdot dv \]
\[ d\left( \frac{u}{v} \right) = \frac{du \cdot v - u \cdot dv}{v^2}. \]

4. Partial Derivatives of Multivariable Functions

4a) Consider a function \(f: X \to Y\) where \(X \subset \mathbb{R}^n\) and \(Y \subset \mathbb{R}\), i.e., \(f\) is a function of \(n\) variables. Let \(x^0\) be any point belonging to the interior of \(X\). As with functions of a single variable, we define a difference quotient for given \(x^0 \in X\):
\[ q(h) \equiv \frac{f(x_1^0 + h, x_2^0, \ldots, x_n^0) - f(x_1^0, x_2^0, \ldots, x_n^0)}{h}. \tag{7} \]
The domain of the function is all non-zero \(h\) values such that \((x_1^0 + h, x_2^0, \ldots, x_n^0)\) belongs to \(X\). \(x^0\) is a parameter rather than a variable of the function; there is a different \(q\) function for each different value of \(x^0\).

The partial derivative of \(f\) with respect to \(x_1\) at the point \(x^0 = (x_1^0, x_2^0, \ldots, x_n^0)\) is given by
\[ f_1(x_1^0, x_2^0, \ldots, x_n^0) = \lim_{h \to 0} \frac{f(x_1^0 + h, x_2^0, \ldots, x_n^0) - f(x_1^0, x_2^0, \ldots, x_n^0)}{h}. \]
Partial derivatives with respect to \(x_2, \ldots, x_n\) are similarly defined (i.e., to get the derivative with respect to \(x_i\), one adds \(h\) to \(x_i^0\) rather than to \(x_1^0\) in the above formula).
4b) Notation. We denote the partial derivative of \( f(x_1, \ldots, x_n) \) with respect to \( x_i \) by
\[
 f_i(x_1, x_2, \ldots, x_n) \quad \text{or} \quad \frac{\partial}{\partial x_i} f(x_1, \ldots, x_n) \quad \text{or} \quad D_i f(x_1, x_2, \ldots, x_n).
\]
If \( y = f(x_1, \ldots, x_n) \), then the partial derivative with respect to \( x_i \) may also be written as \( \partial y/\partial x_i \).

If function values take a form like \( f(x, z) \), then we may write partial derivatives as \( f_x(x, z) \) and \( f_z(x, z) \) or as \( \frac{\partial}{\partial x} f(x, z) \) and \( \frac{\partial}{\partial z} f(x, z) \) or as \( D_x f(x, z) \) and \( D_z f(x, z) \), respectively.

4c) Another useful way of looking at partial derivatives is as follows. Given a function \( f \) of \( n \) variables, we can define a function of \( x_1 \) alone by holding \( x_i \) constant and equal to \( x_0^i \) for all \( i \neq 1 \) and then letting
\[
g(x_1) \equiv f(x_1, x_0^2, x_0^3, \ldots, x_0^n).
\]
Clearly, different values of \((x_0^2, x_0^3, \ldots, x_0^n)\) will typically yield different \( g \) functions. For a given value of \((x_0^2, x_0^3, \ldots, x_0^n)\) and hence a given \( g \) function, the domain of \( g \) consists of all \( x_1 \) values with the property that \((x_1, x_0^2, x_0^3, \ldots, x_0^n)\) belongs to the domain of \( f \).

Another way of defining the partial derivative of \( f \) with respect to \( x_1 \) at the point \( x^0 \) is to define it as the (ordinary) derivative of \( g \) with respect to \( x_1 \) at \( x_1^0 \), i.e.,
\[
f_1(x_1^0, x_2^0, \ldots, x_n^0) = g'(x_1^0).
\]
For example, if
\[
f(x_1, x_2) = 2x_1^2 + x_1 x_2,
\]
then we may define
\[
g(x_1) = 2x_1^2 + x_1^0 x_2.
\]
Noting that \( x_2^0 \) is a constant, it follows that
\[
g'(x_1) = 4x_1 + x_2^0
\]
and hence that
\[
f_1(x_1^0, x_2^0) = g'(x_1^0) = 4x_1^0 + x_2^0.
\]

Note. As with ordinary derivatives, we usually drop the zero if a differentiation formula holds over a function’s whole domain. Thus we would write:
\[
f_1(x_1, x_2) = g'(x_1) = 4x_1 + x_2.
\]

4d) Partial derivatives with respect to all the other variables \( x_i \) \((i \neq 1)\) may be defined similarly. Thus if
\[
f(x_1, x_2) = 2x_1^2 + x_1 x_2,
\]
as above, then the partial derivative with respect to \( x_2 \) is found by forming a function of \( x_2 \) alone by holding \( x_1 \) constant, and then differentiating that function of \( x_2 \). The result is

\[
f_2(x_1, x_2) = x_1.
\]

4e) Since partial derivatives are themselves functions (with a domain consisting of all points at which the original function is partially differentiable), they may themselves potentially be differentiated. We write \( f_{ij}(x_1, \ldots, x_n) \) for the partial derivative with respect to \( x_i \) of the partial derivative with respect to \( x_j \) (both \( i = j \) and \( i \neq j \) are possible). Alternative notations are

\[
\frac{\partial^2}{\partial x_i \partial x_j} f(x_1, \ldots, x_n) \quad \text{and} \quad D_{ij} f(x_1, \ldots, x_n) \quad \text{and} \quad \frac{\partial^2 y}{\partial x_i \partial x_j},
\]

assuming in the last case that \( y = f(x_1, \ldots, x_n) \). Second order partial derivatives may themselves potentially be differentiated to form third order partial derivatives and so on.

4f) If a function \( f : X \to Y \) has first order partial derivatives on an open set \( U \subset X \) that are themselves continuous functions on \( U \), we say that the function is \( C^1 \) on \( U \) or that \( f \in C^1 \). \( C^1 \) functions are also termed continuously differentiable.

More generally, if a function has \( n \)th order partial derivatives that are themselves continuous functions, we say that the function is \( C^n \) or that \( f \in C^n \).

4g) Young’s Theorem. If \( f \in C^2 \) on an open region \( U \) (i.e., \( f \) has continuous second order partial derivatives on \( U \)), the order in which the partial derivatives are taken does not matter, i.e.,

\[
f_{ij}(x) = f_{ji}(x) \quad \text{for all } x \in U.
\]

4h) The vector of partial derivatives of \( f \) at \( x^0 \) is known as its gradient vector and is denoted by \( \nabla f(x^0) \), i.e.,

\[
\nabla f(x^0) = [f_1(x^0), f_2(x^0), \ldots, f_n(x^0)]
\]

4i) Now consider a plane that is tangent to the surface of a multivariable function. The equation of this plane is:

\[
y_t = f(x^0) + \nabla f(x^0) \cdot (x - x^0) = f(x^0) + \sum_{i=1}^{n} f_i(x^0)(x_i - x_i^0).
\]

To see that this is indeed the equation of the plane tangent to \( f \) at \( x^0 \), observe that, treating \( x^0 \) as a vector of constants, \( \partial y_t / \partial x_i = f_i(x^0) \), so that the rate of change of \( y_t \) with respect to changes in any \( x_i \) is the same as that of the function \( f \), as one would expect from a tangent plane.

As with a function of a single variable, we may alternatively express the equation of the tangent plane in terms of differences.

\[
y_t - f(x^0) = \nabla f(x^0) \cdot (x - x^0).
\]
This may be written as:

\[ dy = \nabla f(x^0) \cdot dx = \sum_{i=1}^{n} f_i(x^0) \cdot dx_i, \]

where \( dy \) and \( dx \) are differentials. If we set \( dx_i = 0 \) for all \( i \neq j \), then this reduces to

\[ dy = f_j(x^0) \cdot dx_j. \]

Dividing both sides by \( dx_j \) gives

\[ \frac{dy}{dx_j} \bigg|_{dx_i=0(i\neq j)} = f_j(x^0). \]

Thus partial derivatives are given as the ratio of differentials. Again, this just a matter of re-arranging a definition; it doesn’t provide an alternative way to define a partial derivative since the definition depends on that definition already existing. Rather unhappily in this case, the ratio of differentials doesn’t involve the partial derivative symbol \( \partial \) that we would prefer.

4j) Suppose that \( u = f(x_1, \ldots, x_m) \) and \( v = g(x_1, \ldots, x_n) \) (with \( m \) and \( n \) not necessarily equal, so that \( u \) and \( v \) need not depend on the same variables). Then it is easy to show that

\[
\begin{align*}
    d(u \pm v) &= du \pm dv \\
    d(u \cdot v) &= du \cdot v + u \cdot dv \\
    d \left( \frac{u}{v} \right) &= \frac{du \cdot v - u \cdot dv}{v^2}
\end{align*}
\]

5. **The Total Derivative**

5a) As noted above, the vector of partial derivatives of \( f \) at \( x^0 \) is known as its gradient vector and is denoted by \( \nabla f(x^0) \), i.e.,

\[ \nabla f(x^0) = [f_1(x^0), f_2(x^0), \ldots, f_n(x^0)]. \]

Simon and Blume term this the total derivative of \( f \). The idea is as follows. Consider the tangent line in Figure 2. The equation of this line is:

\[ y_t = f(x_0) + f'(x_0)(x - x_0) \]

Thus the derivative of \( f \) is the coefficient of \( x - x_0 \) in the equation of the tangent line. Now consider a plane that is tangent to the surface of a multivariable function. The equation of this plane is:

\[ y_t = f(x^0) + \nabla f(x^0) \cdot (x - x^0). \]

By analogy with the one-dimensional case, one identifies the vector of coefficients in the equation of the tangent plane, \( \nabla f(x^0) \), as the total derivative of the function.
In this, Simon and Blume are departing somewhat from conventional terminology. Conventionally, one says that, if the total derivative exists, then it equals \( \nabla f(x^0) \). However, the total derivative may not exist, even if \( \nabla f(x^0) \) exists. The reason for the distinction is as follows. As noted earlier, if a function of one variable is differentiable, then it is continuous. By analogy, one wishes to say that, if the total derivative of a multivariable function exists, then the function is continuous. However, the existence of \( \nabla f(x^0) \) does not ensure that \( f \) is continuous at \( x^0 \). Accordingly, for the total derivative to be said to exist, we require an additional condition, namely that

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

This ensures that \( f \) is continuous at \( x^0 \).

It is a simple exercise to confirm that this condition is satisfied by the derivative of a function of a single variable. The total derivative may be denoted \( Df(x^0) \).

5b) A sufficient condition for \( f \) to have a total derivative at \( x^0 \in U \), where \( U \subset \mathbb{R}^n \) is an open set, is that \( f \) be a \( C^1 \) function on \( U \).

6. The Chain Rule for Functions of a Single Variable

6a) Given a function \( f: X \to Y \) with values given by \( y = f(x) \) and another function \( g: Y \to Z \) with values given by \( z = g(y) \), where \( Y \) includes \( f(X) \) (i.e., the domain of \( g \) includes the range of \( f \)), we may define the composite function, \( h: X \to Z \), where

\[
h(x) \equiv g(f(x)) \quad \text{for all } x \in X.
\]

We are interested in the derivative of \( h(x) \) with respect to \( x \). Clearly the way that \( x \) affects \( h(x) \) is via a two stage procedure. A change in \( x \) causes a change in \( f(x) \) and, in turn, the change in \( f(x) \) causes a change in \( g(f(x)) \), which is \( h(x) \). It turns out that the measure of the overall rate of change, \( h'(x) \), is the product of the rates of change at each stage of the procedure, \( f'(x) \) and \( g'(f(x)) \). (Note that \( g'(f(x)) \) is the derivative of \( g \) with respect to \( y \), \( g'(y) \), evaluated at \( y = f(x) \).) Formally, we have the following:

(Chain Rule.) If \( g \) is differentiable at \( y_0 = f(x_0) \) and \( f \) is differentiable at \( x_0 \), then

\[
h'(x_0) = g'(f(x_0)) \cdot f'(x_0).
\]

Note. Rather than write \( g'(f(x_0)) \), the derivative is sometimes written as \( g'(y_0) \) where \( y_0 = f(x_0) \).

6b) To understand this rule, we may approach it in various ways. Consider first the case where both \( f \) and \( g \) are linear. Specifically, suppose that

\[
g(y) = a + by \quad \text{and} \quad f(x) = c + dx,
\]
where \(a, b, c\) and \(d\) are constants. Then clearly

\[
h(x) = g(f(x)) = a + b \cdot f(x) \\
= a + b(c + dx) \\
= a + bc + bdx
\]

so that

\[
h'(x) = bd,
\]
i.e.,

\[
h'(x) = g'(y) \cdot f'(x)
\]
as per the chain rule. The chain rule given above simply asserts that the same relationship also holds for non-linear functions.

6c) A perhaps more intuitive grasp of the chain rule can be reached as follows. Using the notation \(z = g(y)\) and \(y = f(x)\), we may consider the effect of a discrete change in \(x\), which leads to a change in \(y\), which in turn leads to a change in \(z\). Plainly, we have the following identity:

\[
\frac{\Delta z}{\Delta x} = \frac{\Delta z}{\Delta y} \cdot \frac{\Delta y}{\Delta x},
\]

since the two \(\Delta y\) terms on the RHS cancel out. Assuming the relevant derivatives exist, if now we let \(\Delta x\) go to zero (implying that \(\Delta y \to 0\)), then we get the derivative expressions:

\[
\frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx}.
\]  

(9)

The expression on the left is \(h'(x)\) and the expressions on the right are \(g'(y)\) and \(f'(x)\). Thus (9) is an alternative (and perhaps more easily remembered) form of the chain rule.1 The advantage of the first version in equation (8), however, is that it makes explicit the values at which the various derivatives are evaluated, i.e., to get the derivative of \(h\) at \(x_0\), \(g'\) is evaluated at \(f(x_0)\) and \(f'\) is evaluated at \(x_0\).

6d) Some examples of the use of the chain rule follow.

Example 1. If \(g(y) = 2 + y^2\) and \(f(x) = 3x\), then we may define the composite function:

\[
h(x) = g(f(x)) \\
= g(3x) \\
= 2 + (3x)^2 = 2 + 9x^2
\]

---

1 The argument in the text is not fully rigorous since it is possible that for some non-zero \(\Delta x\), the corresponding \(\Delta y\) will be zero and hence \(\Delta z/\Delta y\) is undefined. A more elaborate argument gets around this difficulty.
Applying the chain rule to this expression gives:

\[ h'(x) = g'(f(x)) \cdot f'(x) \]
\[ = g'(3x) \cdot f'(x) \]
\[ = 2(3x) \cdot 3 \]
\[ = 18x. \]

Note that this is the same as the answer we get from differentiating the expanded version of \( h(x) \), namely \( h(x) = 2 + 9x^2 \).

**Example 2.** Let \( z = g(y) = y^2 + 4y \), let \( y = f(x) = 9x - 7 \) and define \( h(x) = g(f(x)) \).

Then

\[ h'(x) = \frac{g'(y) \cdot f'(x)}{dz/dy \; dy/dx} \]
\[ = (2y + 4) \cdot 9 \]
\[ = [2(9x - 7) + 4] \cdot 9 \] [substituting \( f(x) \) for \( y \)]
\[ = [18x - 10] \cdot 9 \]
\[ = 162x - 90 \]

To confirm this result, expand out \( h(x) = g(f(x)) \):

\[ h(x) = (9x - 7)^2 + 4(9x - 7) \]
\[ = 81x^2 - 126x + 49 + 36x - 28 \]
\[ = 81x^2 - 90x + 21 \]

so that

\[ h'(x) = 162x - 90, \]

which is the same answer as was derived from the chain rule.

6e) Observe finally, that the chain rule generalises to composite functions of arbitrarily many levels. Thus if \( z = a(y), y = b(x), x = c(w), w = d(v) \) and \( v = e(u) \) and we form the composite function

\[ h(u) = a(b(c(d(e(u))))), \]

then

\[ h'(u_0) = a'(b(c(d(e(u_0))))) \cdot b'(c(d(e(u_0)))) \cdot c'(d(e(u_0))) \cdot d'(e(u_0)) \cdot e'(u_0) \]
\[ = a'(y_0) \cdot b'(x_0) \cdot c'(w_0) \cdot d'(v_0) \cdot e'(u_0), \]

where \( y_0 = b(c(d(e(u_0)))), x_0 = c(d(e(u_0))), w_0 = d(e(u_0)) \) and \( v_0 = e(u_0) \). Equivalently,

\[ \frac{dz}{du} = \frac{dz}{dy} \cdot \frac{dy}{dx} \cdot \frac{dx}{dw} \cdot \frac{dw}{dv} \cdot \frac{dv}{du} \]
7. Chain Rule for Multivariable Functions

7a) Suppose that we have \( n \) functions \( x_1 = x_1(t), x_2 = x_2(t), \ldots, x_n = x_n(t) \) defined on a common domain \( T \subset \mathbb{R} \) and that we have a function \( f(x_1, x_2, \ldots, x_n) \) with a domain that includes \( (x_1(t), x_2(t), \ldots, x_n(t)) \) for all \( t \in T \). Then we may define a composite function \( F \) with values given by

\[
F(t) = f(x_1(t), x_2(t), \ldots, x_n(t))
\]

and a domain of \( T \subset \mathbb{R} \).

We are interested in the derivative \( F'(t) \). Plainly a change in \( t \) will potentially affect the value of \( F \) via each of the functions \( x_1(t), x_2(t), \ldots, x_n(t) \), i.e., there are potentially \( n \) different channels whereby a change in \( t \) can affect \( F(t) \). It turns out that the overall effect on \( F(t) \) is the sum of the effects operating through each channel. Formally, we have the following.

7b) (Chain Rule) Let \( x_i \) \((i=1, \ldots, n)\), \( f \) and \( F \) be defined as above. Assume that each \( x_i \) function is \( C^1 \) on an interval containing \( t = t^0 \) and that \( f \) is \( C^1 \) on an \( \epsilon \) ball about \( x^0 = x(t_0) \). Then we have the following:

\[
F'(t^0) = \sum_{i=1}^{n} f_i(x_1^0, x_2^0, \ldots, x_n^0) \cdot x'_i(t^0).
\]

Equivalently,

\[
F'(t^0) = \sum_{i=1}^{n} \frac{\partial f(x_1, x_2, \ldots, x_n)}{\partial x_i} \cdot \frac{dx_i(t)}{dt},
\]

where \( \partial f(x_1, x_2, \ldots, x_n)/\partial x_i \) is evaluated at \( x = x^0 \) and \( dx_i(t)/dt \) is evaluated at \( t = t^0 \). Further, \( F \) is \( C^1 \) at \( t_0 \).

7c) This rule is most easily understood in the linear case. Suppose

\[
f(x_1, \ldots, x_n) = a + b_1 x_1 + b_2 x_2 + \cdots + b_n x_n \quad (10)
\]

and

\[
x_i = c_i + d_i t \quad \text{for } i = 1, \ldots, n. \quad (11)
\]

If we form the composite function by substituting (11) into (10), we get

\[
F(t) = a + b_1 \underbrace{(c_1 + d_1 t)}_{x_1} + b_2 \underbrace{(c_2 + d_2 t)}_{x_2} + \cdots + b_n \underbrace{(c_n + d_n t)}_{x_n}.
\]

---

2 A more formal statement is as follows. We have a vector-valued function \( x: T \rightarrow U \), where \( T \subset \mathbb{R} \) and \( U \subset \mathbb{R}^n \), with values given by \( x = (x_1(t), \ldots, x_n(t)) \). We also have a function \( f: X \rightarrow Y \), where \( X \subset \mathbb{R}^n \) and \( Y \subset \mathbb{R} \), with values given by \( y = f(x_1, \ldots, x_n) \). Further, the range of \( x \) belongs to the domain of \( f \), i.e., \( x(T) \subset X \). We form the composite function \( F: T \rightarrow Y \), with values given by \( F(t) = f(x_1(t), x_2(t), \ldots, x_n(t)) \).
Thus
\[
F'(t) = b_1d_1 + b_2d_2 + \cdots + b_nd_n
\]
\[
= \sum_{i=1}^{n} b_id_i
\]
\[
= \sum_{i=1}^{n} f_i(x_1, \ldots, x_n) \cdot x'_i(t)
\]

7d) Example. If \( f(x_1, x_2) = 3x_1 - 4x_2^2 \) and \( x_1 = t^2, \ x_2 = 6t \), then we may form:
\[
F(t) = f(x_1(t), x_2(t)).
\] (12)

Applying the chain rule, we have:
\[
F'(t) = f_1(x_1, x_2) \cdot x'_1(t) + f_2(x_1, x_2) \cdot x'_2(t)
\]
\[
= 3 \cdot 2t - 8x_2 \cdot 6
\]
\[
= 6t - 8(6t) \cdot 6
\]
\[
= -282t.
\]

To verify this result, write out (12) explicitly by substituting \( x_1 = t^2 \) and \( x_2 = 6t \) into \( f(x_1, x_2) = 3x_1 - 4x_2^2 \). This gives:
\[
F(t) = 3t^2 - 4(6t)^2
\]
\[
= 3t^2 - 4 \cdot 36t^2
\]
\[
= -141t^2
\]

Thus \( F'(t) = -282t \), the same answer as was derived above using the chain rule.

8. Directional Derivatives

8a) A generalisation of the partial derivative is the directional derivative. Given a function \( f: X \rightarrow Y \) where \( X \subseteq \mathbb{R}^n \) and \( Y \subseteq \mathbb{R} \), we may choose any \( v \in \mathbb{R}^n \) and define \( g: \mathbb{R} \rightarrow \mathbb{R} \), where:
\[
g(t) = f(x^0 + tv).
\] (13)

Plainly, there is a different \( g \) function for each \( x^0 \) and each \( v \).

At \( t = 0 \), we get \( f(x^0) \). As \( t \) is varied, the argument of the function is changed in the (positive or negative) direction of \( v \). The directional derivative measures the rate of change of the value of the function with respect to these movements. Formally, we may define the directional derivative as follows:
\[
f'(x^0; v) = \lim_{t \to 0} \frac{f(x^0_1 + tv_1, x^0_2 + tv_2, \ldots, x^0_n + tv_n) - f(x^0_1, x^0_2, \ldots, x^0_n)}{t}
\]
\[
= \lim_{t \to 0} \frac{f(x^0 + tv) - f(x^0)}{t}.
\] (14)
Observe that
\[ g'(0) = \lim_{t \to 0} \frac{g(t) - g(0)}{t} \quad \text{by definition of derivative} \]
\[ = \lim_{t \to 0} \frac{f(x^0 + tv) - f(x^0)}{t} \quad \text{from (13).} \tag{15} \]
Comparing (14) and (15), we see that
\[ f'(x^0; v) = g'(0). \]
If the conditions of the chain rule are satisfied, we may compute \( g'(0) \) as follows from (13):
\[ g'(t) = \sum_{i=1}^{n} \frac{\partial f(x^0 + tv_i)}{\partial (x^0 + tv_i)} \cdot \frac{\partial (x^0_i + tv_i)}{\partial t} \]
\[ = \sum_{i=1}^{n} f_i(x^0 + tv_i)v_i. \tag{16} \]
Evaluating (16) at \( t = 0 \) gives:
\[ g'(0) = \sum_{i=1}^{n} f_i(x^0)v_i. \tag{17} \]
In matrix notation, this is
\[ g'(0) = \nabla f(x^0) \cdot v, \]
i.e.,
\[ f'(x^0; v) = \nabla f(x^0) \cdot v. \]
Plainly, partial derivatives correspond to the special case where \( v = e_i \), where \( e_i \) is the \( i \)th unit vector. Observe too that multiplying the vector \( v \) by \( \lambda \) will multiply the value of the directional derivative by \( \lambda \).

8b) Recall that the cosine of the angle \( \theta \) between two vectors \( \mathbf{a} \) and \( \mathbf{b} \) is given by
\[ \cos \theta = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|}. \]
Applying this to \( \nabla f(x^0) \) and \( v \) gives:
\[ \cos \theta = \frac{\nabla f(x^0) \cdot v}{\|\nabla f(x^0)\| \|v\|}. \]
Re-arranging:
\[ \nabla f(x^0) \cdot v = \|\nabla f(x^0)\| \|v\| \cos \theta. \]
If we specify that \( \|v\| = 1 \), then this becomes:
\[ \nabla f(x^0) \cdot v = \|\nabla f(x^0)\| \cos \theta. \]
Now the LHS is of course \( f'(x^0; v) \), the directional derivative of \( f \) in the direction \( v \), evaluated at \( x^0 \). The value of this expression will be maximised when \( \cos \theta = 1 \), which occurs when the angle between the vectors is zero, i.e., when \( v \) is a positive multiple of \( \nabla f(x^0) \). This means that, normalising direction vectors to have a length of 1, the directional derivative at \( x^0 \) is maximised when \( v \) is a positive multiple of \( \nabla f(x^0) \). We therefore say that \( \nabla f(x^0) \) points in the direction from \( x^0 \) in which \( f \) increases most rapidly.
9. Generalised Chain Rule

9a) A more general version of the Chain Rule may be defined as follows: Suppose that we have \( n \) functions \( x_1 = x_1(t_1, \ldots, t_m), x_2 = x_2(t_1, \ldots, t_m), \ldots, x_n = x_n(t_1, \ldots, t_m) \) defined on a common domain \( T \subset \mathbb{R}^m \) and that we have a function \( f(x_1, x_2, \ldots, x_n) \) with a domain that includes
\[
(x_1(t_1, \ldots, t_m), x_2(t_1, \ldots, t_m), \ldots, x_n(t_1, \ldots, t_m))
\]
for all \((t_1, \ldots, t_m) \in T\). Letting \( t = (t_1, \ldots, t_m) \), we may define a composite function
\[
F(t) = f(x_1(t), x_2(t), \ldots, x_n(t))
\]
with domain \( T \subset \mathbb{R}^m \).

9b) (Generalised Chain Rule) Let \( x_i (i = 1, \ldots, n), f \) and \( F \) be defined as above. Assume that each \( x_i \) function is \( C^1 \) on an \( \epsilon \) ball about \( t^0 \) and that \( f \) is \( C^1 \) on an \( \epsilon \) ball about \( x(t^0) \). Then we have the following:
\[
F_j(t^0) = \sum_{i=1}^{n} D_i f(x^{0}_1, x^{0}_2, \ldots, x^{0}_n) \cdot D_j x_i(t^0_1, \ldots, t^0_m),
\]
where \( D_i f \) denotes the partial derivative of \( f \) with respect to \( x_i \) and \( D_j x_i \) denotes the partial derivative of \( x_i \) with respect to \( t_j \). Equivalently:
\[
F_j(t^0) = \sum_{i=1}^{n} \frac{\partial f(x_1, x_2, \ldots, x_n)}{\partial x_i} \cdot \frac{\partial x_i(t_1, \ldots, t_m)}{\partial t_j},
\]
where \( \partial f(x_1, x_2, \ldots, x_n)/\partial x_i \) is evaluated at \( x = x^0 \) and \( \partial x_i(t_1, \ldots, t_m)/\partial t_j \) is evaluated at \( t = t^0 \).

9c) Once again, the rule is most easily understood in the linear case. Suppose
\[
f(x_1, \ldots, x_n) = a + b_1 x_1 + b_2 x_2 + \cdots + b_n x_n
\]
as before, and
\[
x_i = c_i + d_{i1} t_1 + d_{i2} t_2 + \cdots + d_{im} t_m \quad \text{for } i = 1, \ldots, n.
\]
If we form the composite function by substituting (19) into (18), we get
\[
F(t) = a + \sum_{i=1}^{m} b_i (c_i + d_{i1} t_1 + d_{i2} t_2 + \cdots + d_{im} t_m).
\]
Thus, differentiating with respect to \( t_j \):
\[
F_j(t) = \sum_{i=1}^{n} b_i d_{ij}
\]
\[
= \sum_{i=1}^{n} D_i f(x_1, \ldots, x_n) \cdot D_j x_i(t_1, \ldots, t_m),
\]
as required.
9d) Example. If
\[ f(x_1, x_2) = x_1 + x_2^3, \quad x_1 = t_1 - t_2^2 \quad \text{and} \quad x_2 = 2t_1 - 3t_2, \]
then we may form
\[ F(t_1, t_2) = f(x_1(t_1, t_2), x_2(t_1, t_2)). \] (20)

Applying the chain rule, we have:
\[ F_1(t_1, t_2) = \frac{\partial f(x_1, x_2)}{\partial x_1} \cdot \frac{\partial x_1(t_1, t_2)}{\partial t_1} + \frac{\partial f(x_1, x_2)}{\partial x_2} \cdot \frac{\partial x_2(t_1, t_2)}{\partial t_1} \]
\[ = 1 \cdot 1 + 3(x_2)^2 \cdot 2 \]
\[ = 1 + 3(2t_1 - 3t_2)^2 \cdot 2 \]
\[ = 1 + 3(4t_1^2 - 12t_1t_2 + 9t_2^2) \cdot 2 \]
\[ = 1 + 24t_1^2 - 72t_1t_2 + 54t_2^2 \]
and
\[ F_2(t_1, t_2) = \frac{\partial f(x_1, x_2)}{\partial x_1} \cdot \frac{\partial x_1(t_1, t_2)}{\partial t_2} + \frac{\partial f(x_1, x_2)}{\partial x_2} \cdot \frac{\partial x_2(t_1, t_2)}{\partial t_2} \]
\[ = 1 \cdot (-2t_2) + 3(x_2)^2 \cdot (-3) \]
\[ = -2t_2 + 3(2t_1 - 3t_2)^2 \cdot (-3) \]
\[ = -2t_2 + 3(4t_1^2 - 12t_1t_2 + 9t_2^2) \cdot (-3) \]
\[ = -2t_2 - 36t_1^2 + 108t_1t_2 - 81t_2^2 \]

To verify these results, write out (20) explicitly by substituting \( x_1 = t_1 - t_2^2 \) and \( x_2 = 2t_1 - 3t_2 \) into \( f(x_1, x_2) = x_1 + x_2^3 \). This gives:
\[ F(t_1, t_2) = t_1 - t_2^2 + [2t_1 - 3t_2]^3 \]
\[ = t_1 - t_2^2 + [2t_1 - 3t_2]^2(2t_1 - 3t_2) \]
\[ = t_1 - t_2^2 + [4t_1^2 - 12t_1t_2 + 9t_2^2](2t_1 - 3t_2) \]
\[ = t_1 - t_2^2 + 8t_1^3 - 36t_1^2t_2 + 54t_1t_2^2 - 27t_2^3 \]

Thus
\[ F_1(t_1, t_2) = 1 + 24t_1^2 - 72t_1t_2 + 54t_2^2 \]
and
\[ F_2(t_1, t_2) = -2t_2 - 36t_1^2 + 108t_1t_2 - 81t_2^2 , \]
which are the same answers as derived above using the chain rule.

Note. The above example was specially chosen so that it is possible to differentiate it both with and without the chain rule. In many cases the chain rule will be the only way to differentiate the function.
10. Functions from $\mathbb{R}^n$ to $\mathbb{R}^m$

10a) The preceding analysis can be extended to functions $f$ from $\mathbb{R}^n$ to $\mathbb{R}^m$ simply by applying the preceding analysis to each component function $f^i$ of $f$. In particular, if each component function has a total derivative, $Df^i(x)$, then the total derivative of $f$ is simply the vector of these derivatives, i.e.,

$$Df(x) = \begin{bmatrix} Df^1(x) & \cdots & Df^m(x) \end{bmatrix} = \begin{bmatrix} \frac{\partial f^1(x)}{\partial x_1} & \ldots & \frac{\partial f^1(x)}{\partial x_n} \\ \frac{\partial f^2(x)}{\partial x_1} & \ldots & \frac{\partial f^2(x)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f^m(x)}{\partial x_1} & \ldots & \frac{\partial f^m(x)}{\partial x_n} \end{bmatrix}$$

9b) Similarly, the chain rule can be given a matrix form. Specifically, given a function $x: T \to X$, where $T \subset \mathbb{R}^s$ and $X \subset \mathbb{R}^n$, and a function $f: X \to Y$, where $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^m$, we form the composite function $F: T \to Y$, where

$$F(t) = f(x(t)).$$

Then:

$$DF(t) = \begin{bmatrix} \frac{\partial F^1(t)}{\partial t_1} & \cdots & \frac{\partial F^1(t)}{\partial t_s} \\ \vdots & \ddots & \vdots \\ \frac{\partial F^m(t)}{\partial t_1} & \cdots & \frac{\partial F^m(t)}{\partial t_s} \end{bmatrix} \begin{bmatrix} \frac{\partial f^1(x)}{\partial x_1} & \ldots & \frac{\partial f^1(x)}{\partial x_n} \\ \frac{\partial f^2(x)}{\partial x_1} & \ldots & \frac{\partial f^2(x)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f^m(x)}{\partial x_1} & \ldots & \frac{\partial f^m(x)}{\partial x_n} \end{bmatrix} \begin{bmatrix} \frac{\partial x_1(t)}{\partial t_1} & \ldots & \frac{\partial x_1(t)}{\partial t_s} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_n(t)}{\partial t_1} & \cdots & \frac{\partial x_n(t)}{\partial t_s} \end{bmatrix}$$

which is analogous in form to the chain rule for functions of a single variable.

Note that the equality of the $(i, j)$th element for the two sides constitutes an instance of the previous chain rule for each $i, j$. 