Part I

Differentiation & Integration

Chapter 1

Fundamental concepts

1.1 Motivation

Differential equations are essential to most branches of mathematical physics and applied mathematics, as we live in a changing world. For example, it is an empirically observed fact, beloved of Physics interviewers asking questions about tea, milk, and mysterious absences, that the rate of change of the temperature of a body is proportional to the difference in temperature between the body and its surroundings. So, an obvious question is how does the body's temperature vary with time?

Let's set the problem up mathematically.

- Let us call the temperature of the body $\theta(t)$: "theta" for those without the benefits of a classical education. We say that θ is a **dependent** variable because it **depends** on (i.e. varies with) time t.
- We also say that θ is a **function** of its **argument** t, and we wish to determine the functional relationship between θ and t.
- We say that t is the **independent** variable.
- We also need to say what the temperature of the surroundings is, and for simplicity let us suppose that it is a constant θ_0.

Then we need a notation for "the rate of change", which we write as an (ordinary) derivative:

$$\frac{d}{dt}\theta \propto \theta - \theta_0 \to \frac{d}{dt}\theta = -k(\theta - \theta_0),$$

where $k > 0$ is a constant. The sign follows from the reasonable assumption that if $\theta > \theta_0$ we expect the body to cool down towards θ_0, while if $\theta < \theta_0$ we expect it to heat up.

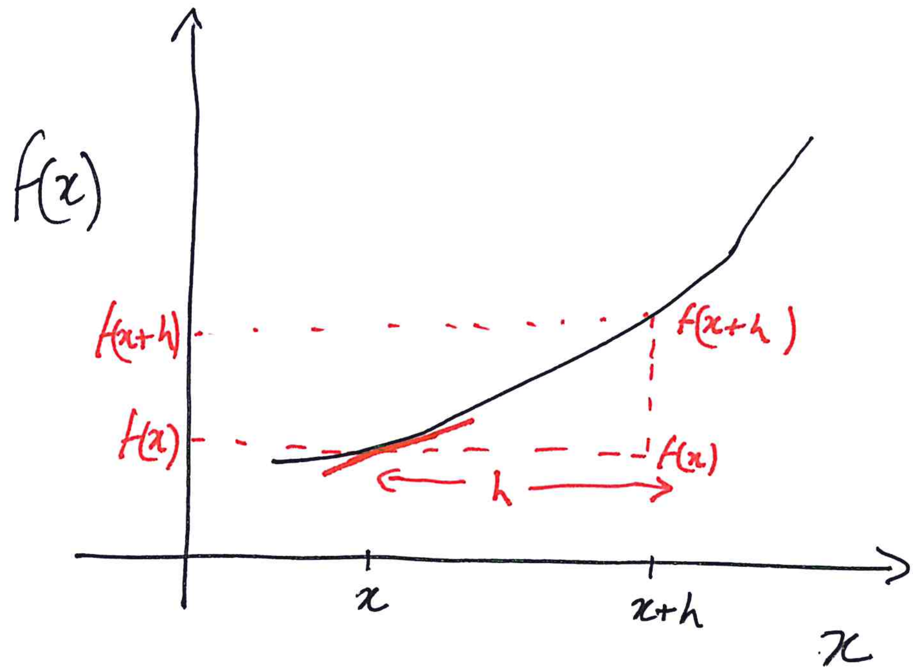

Figure 1.1: Schematic representation of the derivative of a function $f(x)$ at a point as the slope of the graph of the function at that point.

Now, how might we solve this **differential equation**? Well, from the name of the course, you hopefully have come to the right place... but there are a lot of open questions:

- how is a derivative defined?

- have we got enough information to solve the equation?

- how would we solve the equation?

These are the sort of questions which will be addressed in this course. More formal issues of existence/uniqueness and validity will be addressed in the Analysis courses... but there's still plenty in this course to keep us all busy!

1.2 Preliminary definitions

1.2.1 Definition of a derivative

We define the **derivative** of a **function** $f(x)$ with respect to x as the function defined by the **limit**:

$$\frac{df}{dx} = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}. \tag{1.1}$$

We are not going to define functions formally in this course, but loosely they are mathematical expressions which take inputs and give well-defined outputs for those inputs. In particular, for a given input, the output is uniquely defined. So, if we know x, we can work out $f(x)$. Similarly, if we know x and $f(x)$ we can use the expression (1.1) to work out df/dx at x, assuming of course the limit exists.

Notationally, it is sometimes important to distinguish between the function, defined over all its possible input points x (i.e. over its **domain**) and the specific value of the function for a particular input value of $x = x_0$ say. Applying (1.1), we write

$$\left.\frac{df}{dx}\right|_{x=x_0} = \frac{df}{dx}(x_0) = \lim_{h\to 0} \frac{f(x_0+h) - f(x_0)}{h}. \tag{1.2}$$

As shown in the schematic figure 1.1, the value of df/dx at $x = x_0$ is the slope of the graph of $f(x)$ at the point $x = x_0$. It should always be clear by context or notation whether one is discussing the function, or its specific value for a given input argument.

1.2.2 Differentiability

For the function $f(x)$ to be **differentiable**, and so for the function df/dx to be well-defined at the point x_0, the left-hand limit (i.e. h is negative and approaches zero from below) and the right-hand limit (h is positive and approaches zero from above) must be defined and equal, i.e.

$$\lim_{h\to 0^-} \frac{f(x_0+h) - f(x_0)}{h} = \lim_{h\to 0^+} \frac{f(x_0+h) - f(x_0)}{h}.$$

This is actually quite a strong restriction on the "smoothness" of the function $f(x)$. As an example, $f(x) = |x|$ is not differentiable at $x = 0$. Deliberately in this course we do not define what is meant by a "function" or a "limit", but don't worry, these concepts will come up again...

Notation

Writing the derivative of a function $f(x)$ as df/dx is known as Leibniz notation. Notice how the denominator shows what the argument of the function is. Of course, other mathematicians had other notations, and two important ones are Lagrange's notation $f'(x)$ (as we shall see this notation often has the argument specifically quoted to avoid ambiguity) and Newton's notation of a superscript dot, $\dot{f}$, which is typically only used when the independent variable is time.

Furthermore, the definition of derivative can be applied recursively, defining derivatives of derivatives, i.e. second derivatives etc:

$$\frac{d}{dt}\left(\frac{df}{dt}\right) = \frac{d^2 f}{dt^2} = f''(t) = \ddot{f}.$$

1.3 Big O and little o notation

Two very useful concepts in applied mathematics are big O (pronounced "Oh") and little o, which is sometimes written $\underline{o}$ to distinguish clearly between the two symbols. These two concepts, sometimes called **order parameters**, are used to give comparative scalings between functions sufficiently close to some limiting point x_0. Conventionally, these concepts are used with an abuse of the equals notation, but let's all keep calm.

1. Definition "little oh": $f(x)$ is $o[g(x)]$ as $x \to x_0$ if

$$\lim_{x \to x_0} \frac{f(x)}{g(x)} = 0,$$

   commonly written as $f(x) = o[g(x)]$.

2. Definition "big Oh": $F(x)$ is $O[G(x)]$ as $x \to x_0$ if:

   (a) for the case where x_0 is finite, there exist two positive finite constants M and δ such that for all x with $|x - x_0| < \delta$,

$$|F(x)| \leq M|G(x)|;$$

   (b) for the case where x_0 is plus (or minus) infinity, there exist two positive finite constants M and x_1 such that for all $x > x_1$ for plus infinity, (or equivalently for all $x < -x_1$ for minus infinity):

$$|F(x)| \leq M|G(x)|.$$

Similarly, this is also commonly written as $F(x) = O[G(x)]$.

The equals notation is most commonly used, though really these concepts can be thought of as conditions for the functions f and F to belong to sets of functions with the required property of varying in a particular way as they approach the special limiting point $x = x_0$.

Note from the definitions, $f = o(g) \rightarrow f = O(g)$ but not vice versa. Very commonly, the functions $g(x)$ and $G(x)$ are powers of x but that is not necessary. Loosely:

1. If, in the definition "little oh", $g(x)$ tends towards infinity as $x \rightarrow x_0$, $f(x)$ is definitely growing more slowly.

2. If, in the definition "little oh", $g(x)$ tends towards zero as $x \rightarrow x_0$, $f(x)$ is definitely decaying to zero even more rapidly.

3. If, in the definition "big Oh", $G(x)$ tends towards infinity at some rate as $x \rightarrow x_0$, $F(x)$ can at most be going to infinity at a fixed multiple of that rate.

4. If, in the definition "big Oh", $G(x)$ tends towards zero at some rate as $x \rightarrow x_0$, $F(x)$ is definitely eventually decaying to zero at least as quickly as a fixed multiple of that rate.

Some examples:

- $x = o(\sqrt{x})$ as $x \rightarrow 0$.

- $x = O(\sqrt{x})$ as $x \rightarrow 0$.

- $\sin 2x = O(x)$ as $x \rightarrow 0$.

- $\sqrt{x} = o(x)$ as $x \rightarrow \infty$.

- $\cos(x) = O(1)$ for all x.

1.3.1 Tangent line at x_0

Armed with these parameters, we can now rewrite the (specific) derivative definition (1.2):

$$\left.\frac{df}{dx}\right|_{x_0} = \frac{f(x_0 + h) - f(x_0)}{h} + \frac{o(h)}{h}. \tag{1.3}$$

The equivalence of this definition and (1.2) can be proved by contradiction. Assume that

$$\left.\frac{df}{dx}\right|_{x_0} = \frac{f(x_0+h)-f(x_0)}{h} + \frac{R(h)}{h},$$

for some remainder function $R(h)$ which is categorically **not** $o(h)$. Take limits as $h \to 0$ of both sides of the equation. The left-hand side is a constant which doesn't change in the limit. The first term on the right-hand side is by definition (1.2), and so is the same as the left-hand side. But by construction,

$$\lim_{h\to 0} \frac{R(h)}{h} \neq 0,$$

and we have a contradiction!

Multiplying (1.3) across by h, we obtain

$$f(x_0+h) \;=\; f(x_0) + \left(\left.\frac{df}{dx}\right|_{x_0}\right) h + o(h). \tag{1.4}$$

If we now identify $x = x_0 + h$, $y(x) = f(x)$, $y_0 = f(x_0)$, (1.4) can be related to the equation of the tangent line at x_0 to the curve $y = f(x)$, i.e.

$$y = y_0 + m(x - x_0),$$

where the slope $m = df/dx$ at the point $x = x_0$.

Chapter 2

Rules of differentiation

There are many useful **rules** of differentiation. Let us investigate how the fundamental definitions presented above can lead to these rules.

2.1 Chain rule

It is entirely possible that the argument of the function we are considering is itself a function of the dependent variable, i.e. $f(x) = F[g(x)]$ for some functions F and g. For example, what is the derivative with respect to x of $f(x) = \sin(x^2 - x + 2)$? Here $F(X) = \sin(X)$, and $g(x) = x^2 - x + 2$. Let us apply what we have learnt.

Note that we want to work out the derivative of f with respect to x. From (1.1),

$$\frac{df}{dx} = \lim_{h \to 0} \frac{F[g(x+h)] - F[g(x)]}{h},$$

and then using (1.4), the right-hand side can be written as

$$\frac{df}{dx} = \lim_{h \to 0} \frac{F\left[g(x) + h\frac{dg}{dx} + o(h)\right] - F[g(x)]}{h}, \tag{2.1}$$

where we understand that dg/dx is evaluated at x, and we assume that g is differentiable. We also remember that $o(h)$ represents a term $T_h(h)$ such that, by definition

$$\lim_{h \to 0} \frac{T_h(h)}{h} = 0.$$

Now, let us relabel

$$g(x) + h\frac{dg}{dx} + o(h) = X + H; \ X = g(x), \ H = h\frac{dg}{dx} + o(h).$$

Then, using (1.4),

$$F(X+H) = F(X) + H\frac{dF}{dX} + o(H), \tag{2.2}$$

where here now $o(H)$ represents a term $T_H(H)$ such that, by definition

$$\lim_{H\to 0} \frac{T_H(H)}{H} = 0.$$

Of course, by the definition of H, T_H may also be thought of as a function of h. Now let us consider this term $T_H(H)$ which is $o(H)$. Importantly, in this particular context, since it appears in (2.1), we are only interested in how this term behaves as $h \to 0$. There are two different cases to consider.

1. If $dg/dx = 0$, then H is also a term which is $o(h)$ as $h \to 0$, and so clearly a term that is $o(H)$ as $H \to 0$ (i.e. a term that goes to zero "faster" than H as H goes to zero) is also $o(h)$ as $h \to 0$, (i.e. goes to zero "faster" than h as $h \to 0$) since H itself goes to zero "faster" than h as $h \to 0$.

2. If dg/dx is not equal to zero, we can make two further observations.

   - H is a term which is $O(h)$, but is most definitely not a term which is $o(h)$ as $h \to 0$.
   - Therefore, in this particular context, H is effectively linearly proportional to h.

   Therefore, here, if a term is $o(H)$ as $H \to 0$, it must also be $o(h)$ as $h \to 0$, which is what we require.

Armed with this result, (2.2) becomes

$$\begin{aligned} F(X+H) &= F(X) + \left[h\frac{dg}{dx} + o(h)\right]\frac{dF}{dX} + o(H), \\ &= F\left[g(x)\right] + \left[h\frac{dg}{dx}\right]\frac{dF}{dg} + o(h), \end{aligned} \tag{2.3}$$

as $h \to 0$. Here, we write the derivative of F with respect to its argument $X = g(x)$ as dF/dg. We could have written it as $F'[g(x)]$ if we'd rather follow Lagrange of course. Also, we've exploited the property of the order parameter o that any constant multiple of a term which is $o(h)$ is also $o(h)$ by definition.

If we now substitute (2.3) for the first term in the numerator back into (2.1), we obtain

$$\begin{aligned}\frac{df}{dx} &= \lim_{h\to 0}\frac{F\left[g(x)\right] + h\frac{dg}{dx}\frac{dF}{dg} + o(h) - F[g(x)]}{h},\\ &= \lim_{h\to 0}\frac{dF}{dg}\frac{dg}{dx} + \frac{o(h)}{h}.\end{aligned}$$

Finally, taking the limit, we have the **chain rule**

$$\frac{df}{dx} = \frac{dF}{dg}\frac{dg}{dx} = \frac{df}{dg}\frac{dg}{dx}, \tag{2.4}$$

where we have made a final notational observation that $F(g)$ and $f(g)$ are the same function by definition. For our specific example (of course):

$$\frac{d}{dx}\left[\sin(x^2 - x + 2)\right] = \left[\cos(x^2 - x + 2)\right](2x - 1).$$

2.2 Product rule

Consider the situation where $f(x) = u(x)v(x)$, i.e. f can be written as the product of two other functions u and v. Then, as you are asked to establish on the example sheet from the definition of a derivative:

$$\frac{df}{dx} = v\frac{du}{dx} + u\frac{dv}{dx}. \tag{2.5}$$

Indeed, the so-called quotient rule is just a special case of the product rule:

$$f = \frac{u}{v} \to f' = \frac{u'v - v'u}{v^2}.$$

2.3 (General) Leibniz rule

The product rule can be generalized to higher order derivatives very straightforwardly by recursive application.

$$\begin{aligned} f &= uv,\\ f' &= u'v + uv',\\ f'' &= u''v + u'v' + u'v' + uv'',\\ &= u''v + 2u'v' + uv'',\\ f''' &= u'''v + u''v' + 2u''v' + 2u'v'' + u'v'' + uv''',\\ &= u'''v + 3u''v' + 3u'v'' + uv'''.\end{aligned}$$

Hopefully, this is reminiscent of Pascal's triangle and the binomial theorem. The general Leibniz rule is:

$$\begin{aligned} f^{(n)}(x) &= u^{(n)}v + nu^{(n-1)}v' + \frac{n(n-1)}{2!}u^{(n-2)}v'' + \dots \\ &\quad \dots + \left(\begin{array}{c} n \\ r \end{array} \right) u^{(n-r)}v^{(r)} + \dots uv^{(n)}, \end{aligned}$$

where a superscript (n) denotes the n^{th} derivative, and the binomial coefficient

$$\left(\begin{array}{c} n \\ r \end{array} \right)$$

denotes the number of combinations of r elements that can be taken from n elements without replacement. Of course, this "rule" relies on the function f and its first $n-1$ derivatives all being differentiable.

2.4 Taylor Series

2.4.1 Taylor's Theorem

Remember (1.4), which can be rewritten as

$$f(x_0 + h) = f(x_0) + h \left. \frac{df}{dx} \right|_{x=x_0} + o(h).$$

Provided the first n derivatives of f exist, this can be generalized to

$$\begin{aligned} f(x_0 + h) &= f(x_0) + h \left. \frac{df}{dx} \right|_{x=x_0} + \frac{h^2}{2!} \left. \frac{d^2 f}{dx^2} \right|_{x=x_0} + \dots \\ &\quad \dots + \frac{h^n}{n!} \left. \frac{d^n f}{dx^n} \right|_{x=x_0} + E_n, \end{aligned} \tag{2.6}$$

where $E_n = o(h^n)$ as $h \to 0$. In fact, if $f^{(n+1)}$ exists $\forall x \in (x_0, x_0 + h)$ and f^n is continuous on $[x_0, x_0 + h]$, then for some $x_0 < x_n < x_0 + h$,

$$E_n = \frac{f^{(n+1)}(x_n)}{(n+1)!} h^{n+1},$$

and so $E_n = O(h^{n+1})$ as $h \to 0$. This is a statement of **Taylor's Theorem**. Note that it is an exact statement which expresses the value of a function f at a point $x_0 + h$ in terms of the value of the function at x_0, its derivatives at x_0, and an error term E_n whose behaviour we know as h gets smaller.

2.4.2 Taylor Polynomials

Assume that $x = x_0 + h$. Then, (2.6) can be rewritten as

$$f(x) = f(x_0) + (x - x_0)f'(x_0) + \dots + \frac{(x - x_0)^n}{n!} f^{(n)}(x_0) + E_n.$$

The first n terms on the right hand side are the first n terms of the **Taylor series**, or alternatively the nth order Taylor polynomial of $f(x)$ about the point $x = x_0$. It gives a **local** approximation to the function, and interestingly, in a sense, it is the best possible local approximation to the function.

2.4.3 Determination of coefficients

The coefficients of the Taylor series are determined from evaluations of higher and higher order derivatives of the underlying function. Without loss of generality (wlog) consider an expansion of $f(x)$ about $x = 0$. Then the "obvious" series representation of a sufficiently smooth function (i.e. one which has sufficiently many higher derivatives) $f(x)$ is

$$f(x) = a_0 + a_1 x + a_2 x^2 + \dots + a_n x^n + \dots$$

It is hopefully straightforward to establish that:

$$f(0) = a_0,\ f'(0) = a_1,\ f''(0) = 2a_2,\ \dots\ f^{(n)}(0) = n! a_n.$$

2.5 L'Hôpital's Rule

Taylor series representations are very useful to understand L'Hôpital's rule, which can be used to determine the value of **indeterminate forms** such as ∞/∞ and $0/0$. The simplest version of L'Hôpital's rule is for the case where $f(x)$ and $g(x)$ are both differentiable at $x = x_0$ and

$$\lim_{x \to x_0} f(x) = \lim_{x \to x_0} g(x) = 0.$$

Then if $g'(x_0) \neq 0$,

$$\lim_{x \to x_0} \frac{f(x)}{g(x)} = \lim_{x \to x_0} \frac{f'(x)}{g'(x)}, \tag{2.7}$$

provided the limit exists (as is often the case, there can be quite scary pathological cases beloved of clear-thinking analysts. . .)

A proof for this special case (the rule applies in much more general circumstances) is quite short. From the Taylor series representations:

$$\begin{aligned}
f(x) &= f(x_0) + (x - x_0)f'(x_0) + o(x - x_0); \\
&= 0 + (x - x_0)f'(x_0) + o(x - x_0); \\
g(x) &= g(x_0) + (x - x_0)g'(x_0) + o(x - x_0); \\
&= 0 + (x - x_0)g'(x_0) + o(x - x_0); \\
\to \frac{f}{g} &= \frac{f' + \frac{o(x-x_0)}{(x-x_0)}}{g' + \frac{o(x-x_0)}{(x-x_0)}}. \qquad (2.8)
\end{aligned}$$

Therefore, taking the limit:

$$\lim_{x \to x_0} \frac{f(x)}{g(x)} = \lim_{x \to x_0} \frac{f' + \frac{o(x-x_0)}{(x-x_0)}}{g' + \frac{o(x-x_0)}{(x-x_0)}} = \frac{f'(x_0)}{g'(x_0)}. \qquad (2.9)$$

Interestingly, the rule can be generalised to higher orders. For example, consider

$$f(x) = 3\sin(x) - \sin(3x); \;\; g(x) = 2x - \sin(2x).$$

For these functions, $f(0) = g(0) = f'(0) = g'(0) = f''(0) = g''(0) = 0$. As an exercise, show that

$$\lim_{x \to 0} \frac{f(x)}{g(x)} = 3 = \lim_{x \to 0} \frac{f'''(x)}{g'''(x)}.$$

Chapter 3

Integration

It might be surprising, but integration can be thought of as the solution of a particular kind of differential equation.

3.1 Integrals as sums

Here, we are going to take the applied mathematician's cop-out of considering "appropriately well-behaved functions". Rigorously established calculus a.k.a. analysis is one of the intellectual triumphs of humanity, which is something to look forward to in other courses. Here, we have to remember that there are many things which are true that are not proven.

3.1.1 Integrals as Riemann sums

Let us **define** a definite integral as an infinite sum, and so

$$\int_a^b f(x)dx \equiv \lim_{\Delta x \to 0} \sum_{n=0}^{N-1} f(x_n)\Delta x = \lim_{N\to\infty} \sum_{n=0}^{N-1} f(x_n)\Delta x, \tag{3.1}$$

where $\Delta x = (b-a)/N$ and $x_n = a + n\Delta x$, as shown in figure 3.1.

An interesting question to address is how the difference between the "area under the curve" concept for integration and the above definition in terms of a sum of rectangles changes as $N \to \infty$. So let us consider a particular finite value of N, and consider the magnitude of the difference between the area under the curve, and the area of the sum of rectangles. Clearly, this difference is made up of a sum of such differences associated with each rectangle, so let us consider one such rectangle, wlog between x_n and x_{n+1}, as shown schematically in figure 3.1. How can we estimate the error $|\epsilon|$ for this particular rectangle?

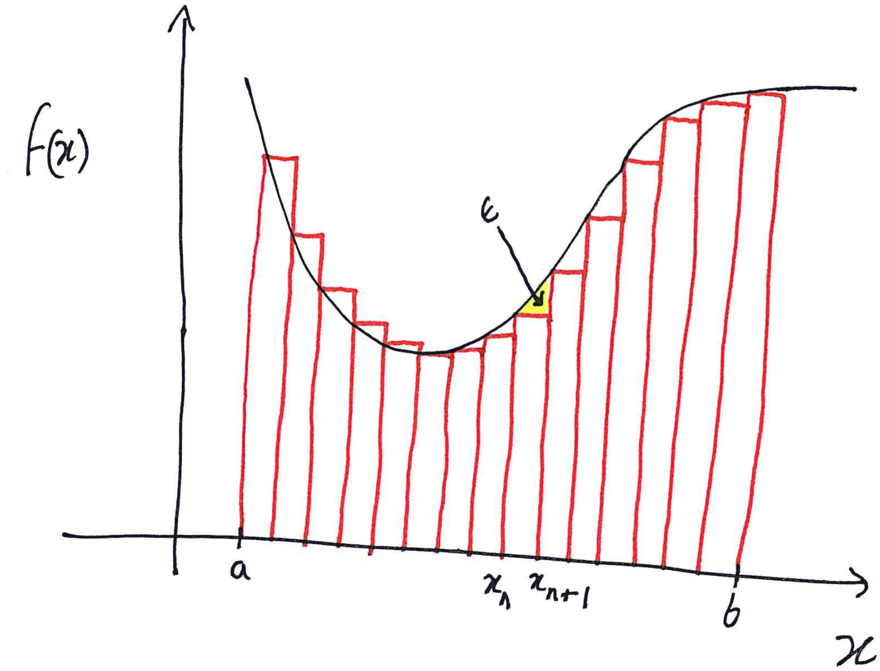

Figure 3.1: Pictorial representation of the Riemann sum approximation to an integral, showing the error ϵ between the area under the curve and the rectangle with base $x_n \leq x \leq x_{n+1}$.

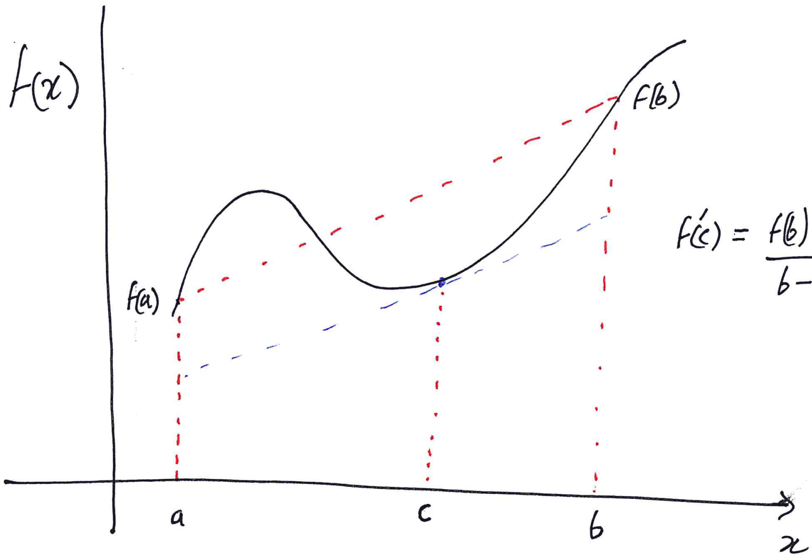

Figure 3.2: Schematic representation of the plausibility of the Mean Value Theorem. A point c can always be found such that the (dashed) tangent line to the curve at c is parallel to the (dashed) line connecting $f(a)$ and $f(b)$.

Well, the first thing we need to do is get an estimate for $|f(x)-f(x_n)|$ for $x_n \leq x \leq x_{n+1}$. Fortunately, we've got a theorem for that. We can appeal to the **Mean Value Theorem**, (MVT) which states for appropriately well-behaved functions, that there exists a point $a < c < b$ such that

$$f(b) = f(a) + (b-a)f'(c).$$

This is eminently plausible whenever you draw a curve representing a function, as shown in figure 3.2 but the rigorous proof is beyond the scope of this course. Therefore, for every x, there is going to be point $x_n < x_c < x$ such that

$$f(x) = f(x_n) + (x - x_n)f'(x_c).$$

Therefore,

$$|f(x) - f(x_n)| \leq \max_{x\in[a,b]} |f'(x)||x - x_n|.$$

So, provided $|f'(x)|$ remains bounded in this interval (related to the function being appropriately well-behaved of course) the absolute magnitude of the difference between the function and the height of the rectangle throughout the interval is

$$|f(x) - f(x_n)| \leq M|x - x_n| \leq M(x_{n+1} - x_n) = M\Delta x,$$

where M is a finite positive number.

M could be quite large if the function oscillates wildly like Morrissey, but all we care about is that it remains finite, and so we know that the difference in height between the Riemann sum rectangle, and a rectangle definitely big enough to contain the area under the curve is proportional to Δx, and so is $O(\Delta x)$. Since the **width** of this interval is $x_{n+1} - x_n = \Delta x$, and remembering the definiton of big Oh, we see that

$$|\epsilon| = O(\Delta x^2) = O([b-a]^2/N^2). \tag{3.2}$$

So, the area under the graph from x_n to x_{n+1} is $f(x_n)\Delta x + O([b-a]^2/N^2)$. Now if we add up N of such intervals, we have to add up N of such differences, and so the area under the curve from a to b is

$$\begin{aligned} A &= \lim_{N\to\infty} \left[\sum_{n=0}^{N-1} f(x_n)\Delta x + N \times O([b-a]^2/N^2) \right], \\ &= \lim_{N\to\infty} \left[\sum_{n=0}^{N-1} f(x_n)\Delta x + O([b-a]^2/N) \right], \\ &= \int_a^b f(x)dx, \end{aligned}$$

by definition, as the total error goes to zero as $N \to \infty$. Therefore, the definition of an integral as the limit of a sum of rectangles is consistent with the integral being the area under the curve (for functions with bounded derivatives which satisfy the conditions of the MVT...blah blah blah)

3.2 Fundamental Theorem of Calculus

The MVT is also central to the **Fundamental Theorem of Calculus**, (FTC . . . I love TLAs . . .) which essentially establishes that integration is "anti-differentiation". The (first part of the) FTC states that if $F(x)$ is **defined** as the function

$$F(x) = \int_a^x f(t)dt + F(a), \tag{3.3}$$

then

$$\frac{dF}{dx} = f(x). \tag{3.4}$$

Another way to interpret the FTC is that an integral is the solution of a particularly simple differential equation, since $F(x)$ is the solution of the differential equation (3.4) with the initial condition $F(a)$ being given.

A sketch of the proof is as follows. By definition

$$\frac{dF}{dx} = \lim_{h \to 0} \frac{1}{h} \left[\int_a^{x+h} f(t)dt - \int_a^x f(t)dt \right].$$

The two overlapping parts cancel, and so

$$\frac{dF}{dx} = \lim_{h \to 0} \frac{1}{h} \int_x^{x+h} f(t)dt.$$

By a similar argument to the one presented above using the MVT leading to (3.2), the integral is approximated by h multiplied by the value of the function at the lower limit, with an error of order $O(h^2)$:

$$\int_x^{x+h} f(t)dt = f(x)h + O(h^2), \tag{3.5}$$

and so

$$\frac{dF}{dx} = \lim_{h \to 0} \frac{1}{h} \left[f(x)h + O(h^2) \right] = f(x),$$

as required, since a term which is $(1/h) \times O(h^2)$ is $O(h)$ as $h \to 0$ and hence $\to 0$ as $h \to 0$.

Corollaries to the FTC

- It is (relatively) straightforward to establish that

$$\frac{d}{dx} \int_x^b f(t)dt = -f(x).$$

- Using the chain rule (2.4),

$$\frac{d}{dx} \int_a^{g(x)} f(t)dt = f[g(x)]g'(x).$$

Notation

Two alternative notations for an **indefinite** integral are

$$\int f(x)dx = \int^x f(t)dt.$$

3.3 Integration by substitution

Calculation of integrals can often (particularly in STEP ...) seem like an art involving recognition. If the integrand contains a function of a function, it can sometimes help with recognition to substitute for the inner function and remember the chain rule. For example:

$$\begin{aligned} I &= \int \frac{(1-2x)}{\sqrt{x-x^2}}dx; \quad u = x - x^2,\ du = (1-2x)dx; \\ &= \int \frac{du}{u^{1/2}} = 2u^{1/2} + C = 2\sqrt{x-x^2} + C, \end{aligned}$$

where C is a constant.

3.3.1 Trigonometric substitution

There are of course several identities, which are associated with particular expressions in the integrand leading to particular substitutions. Here are lists of identities I, expressions E and substitutions S:

- I: $\cos^2\theta + \sin^2\theta = 1$; E: $\sqrt{1-x^2}$; S: $x = \sin\theta$.
- I: $1 + \tan^2\theta = \sec^2\theta$; E: $1 + x^2$; S: $x = \tan\theta$.
- I: $\cosh^2 u - \sinh^2 u = 1$; E: $\sqrt{x^2-1}$; S: $x = \cosh u$.
- I: $\cosh^2 u - \sinh^2 u = 1$; E: $\sqrt{1+x^2}$; S: $x = \sinh u$.
- I: $1 - \tanh^2 u = \mathrm{sech}^2 u$; E: $1 - x^2$; S: $x = \tanh u$.

As an example/exercise, show that

$$I = \int \sqrt{2x - x^2}dx = \frac{1}{2}\arcsin(x-1) + \frac{1}{2}(x-1)\sqrt{1-(x-1)^2} + C,$$

for C a constant and arcsin being the inverse sine function.

3.4 Integration by parts

"Integration by parts" exploits the product rule (2.5). Applying this rule:

$$\int uv'dx = uv - \int vu'dx. \tag{3.6}$$

- Example 1:

$$\begin{aligned} I &= \int_0^\infty xe^{-x}dx; \quad u = x,\ v' = e^{-x}; \\ &= [-xe^{-x}]_0^\infty - \int_0^\infty (-e^{-x})dx; \quad u' = 1,\ v = -e^{-x}; \\ &= [-e^{-x}]_0^\infty = 1. \end{aligned}$$

- Example 2:

$$\begin{aligned} I &= \int \ln x dx; \quad u = \ln x,\ v' = 1; \\ &= x\ln x - \int x\left(\frac{1}{x}\right)dx; \quad u' = \frac{1}{x},\ v = x, \\ &= x\ln x - x + C. \end{aligned}$$

Chapter 4

Partial differentiation

Here we generalize differentiation to multivariate functions.

4.1 Functions of several variables

It is of course possible to imagine **multivariate functions**, i.e. functions which have more than one argument, and so depend on more than one independent variable. Examples include:

- the height of terrain, which depends on both latitude and longitude;
- the density of gas, which depends on both temperature and pressure;
- the deflection of a guitar string, which depends on both its $x-$location along the fretboard and time (assuming it is being played).

So, consider a function $f(x, y)$, which is best represented by a **contour plot**, as shown schematically in figure 4.1 where each contour line is a curve on which the function takes a constant value. Now an obvious question is what is the slope of a hill? The answer, as any skier knows, depends on the direction that the skier is facing. To answer this question properly the first thing to do is find the slope in directions parallel to the coordinate axes. To do that, we need to define a new concept: a **partial derivative**.

4.2 Partial derivatives

The **partial derivative** of $f(x, y)$ with respect to x is defined as the function

$$\left.\frac{\partial f}{\partial x}\right|_y = \lim_{\delta x \to 0} \frac{f(x + \delta x, y) - f(x, y)}{\delta x}. \tag{4.1}$$

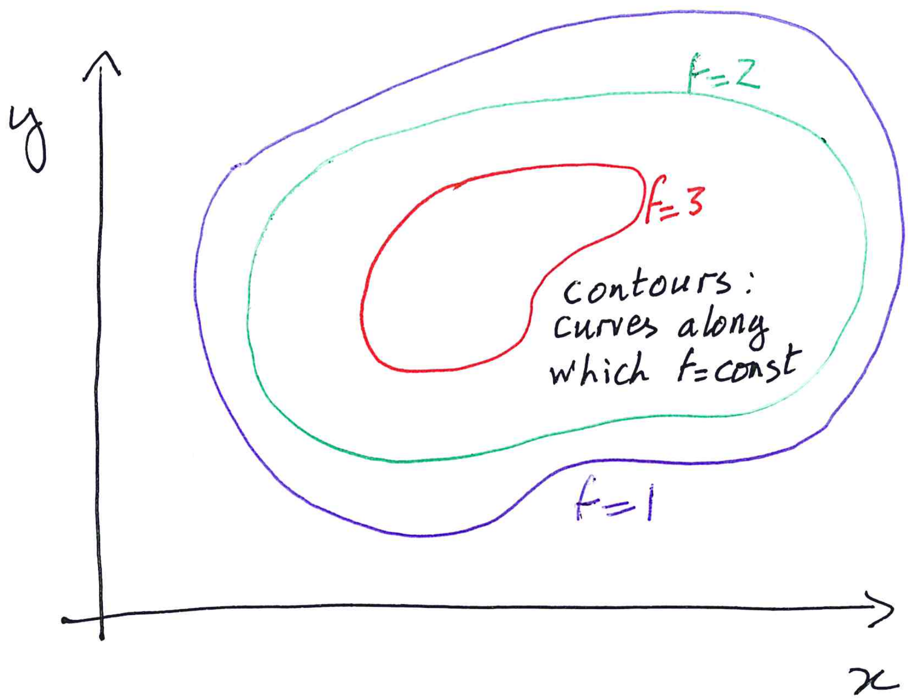

Figure 4.1: Schematic representation of a function $f(x, y)$ as a contour plot. Each line corresponds to a constant value of f.

Note that y is held constant, as the partial derivative of $f(x, y)$ with respect to x is the rate of change with respect to x as y is held constant. Essentially this corresponds to the slope of the hill experienced when moving purely left to right (in the positive $x-$direction) in the figure.

Similarly, the **partial derivative** of $f(x, y)$ with respect to y is defined as the function

$$\left.\frac{\partial f}{\partial y}\right|_x = \lim_{\delta y \to 0} \frac{f(x, y + \delta y) - f(x, y)}{\delta y}, \tag{4.2}$$

i.e. the slope of the hill experienced when moving from the base of the figure upwards (in the positive $y-$direction). Note the curly "partial" derivative symbol.

4.2.1 Partial derivative computation example

Computing partial derivatives is straightforward. All you need to remember is what to keep fixed. Consider the example

$$f(x, y) = x^2 + y^3 + e^{xy^2}.$$

Then

$$\begin{aligned} \left.\frac{\partial f}{\partial x}\right|_y &= 2x + y^2 e^{xy^2}. \\ \left.\frac{\partial f}{\partial y}\right|_x &= 3y^2 + 2xy e^{xy^2}. \end{aligned}$$

We can also compute second derivatives of course:

$$\begin{aligned} \frac{\partial^2 f}{\partial x^2} &= 2 + y^4 e^{xy^2}, \\ \frac{\partial^2 f}{\partial y^2} &= 6y + 2x e^{xy^2} + 4x^2 y^2 e^{xy^2}, \\ \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial^2 f}{\partial x \partial y} &= 2y e^{xy^2} + 2xy^3 e^{xy^2}, \\ \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial y \partial x} &= 2y e^{xy^2} + 2xy^3 e^{xy^2}. \end{aligned}$$

Notice that for this function

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}, \tag{4.3}$$

and so it doesn't matter in which order the partial derivatives are taken. Intuitively, this makes sense, but as ever in mathematics you have to be careful. **Sufficient** conditions for (4.3) to hold (which is a special case of **Schwarz's theorem**, or **Clairaut's theorem** if you are more saucisson than bratwurst) are that the second derivatives are continuous at the point of interest (loosely, that the function tends to the same limit whichever direction you approach the point).

Partial derivative notation

It is necessary to be careful to avoid ambiguity in which arguments specifically are being held fixed when the partial derivatives are being taken. If there is no indication which arguments are being held fixed, that implies that all arguments are being held fixed, except of course the argument with respect to which the partial derivative is being taken! Clear? As a clarifying example, for the function $f(x, y, z)$,

$$\frac{\partial f}{\partial x} = \left.\frac{\partial f}{\partial x}\right|_{y,z} \neq \left.\frac{\partial f}{\partial x}\right|_{y},$$

since in the last expression z may vary.

There is also an alternative subscript notation:

$$f_x \equiv \frac{\partial f}{\partial x}; f_{xy} \equiv \frac{\partial^2 f}{\partial y \partial x}.$$

Note the opposite ordering: in the subscript the derivative on the right is taken after the derivative on the left, while in the fractional notation the derivative on the left is taken after the derivative on the right. Cool eh?

4.3 Chain rule revisited

The chain rule (2.4) can be generalized to higher dimensions. Consider the increment by which $f(x, y)$ (a function with continuous second derivatives everywhere in the domain of interest) varies between two nearby points $(x + \delta x, y + \delta y)$ and (x, y) i.e.

$$\delta f = f(x + \delta x, y + \delta y) - f(x, y).$$

The quantity δf is also called a **differential** of the function f. As shown in figure 4.2, this can be made up of two changes:

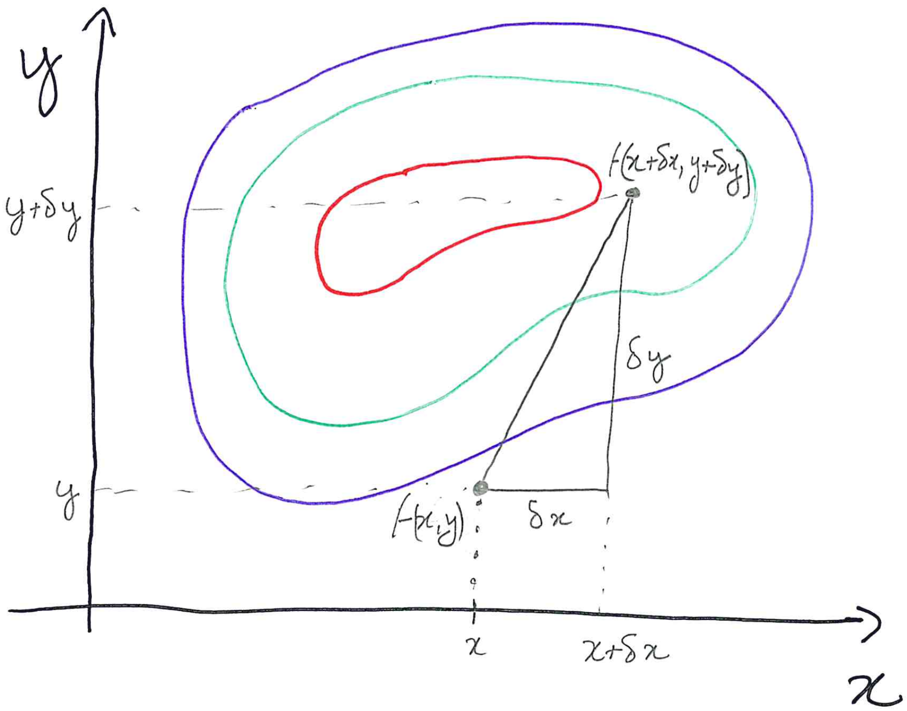

Figure 4.2: Schematic representation of a function $f(x, y)$ as a contour plot. Each line corresponds to a constant value of f, showing two representative points $(x + \delta x, y + \delta y)$ and (x, y).

1. as the vertical coordinate changes from $y + \delta y$ to y, keeping the horizontal coordinate fixed at $x + \delta x$;

2. as the horizontal coordinate changes from $x + \delta x$ to x, keeping the vertical coordinate fixed at y;

and so

$$\begin{aligned}\delta f &= f(x + \delta x, y + \delta y) - f(x + \delta x, y) \\ &\quad + f(x + \delta x, y) - f(x, y).\end{aligned}$$

Applying (4.1), (4.2) and (1.4), we obtain:

$$\delta f = \left[\frac{\partial f}{\partial y}(x + \delta x, y)\right] \delta y + o(\delta y) + \left[\frac{\partial f}{\partial x}(x, y)\right] \delta x + o(\delta x).$$

Note that the two partial derivatives on the right hand side are evaluated at different points, so let's bring the first term to be evaluated at (x, y) as well. Applying (4.2) and (1.4) to the first term on the right hand side, note that the derivative at $(x+\delta x, y)$ can be related to a derivative at (x, y), obtaining

$$\frac{\partial f}{\partial y}(x + \delta x, y) = \frac{\partial f}{\partial y}(x, y) + \left[\frac{\partial^2 f}{\partial x \partial y}(x, y)\right] \delta x + o(\delta x).$$

Therefore,

$$\begin{aligned}\delta f &= \left[\frac{\partial f}{\partial y}(x, y)\right] \delta y + \left[\frac{\partial f}{\partial x}(x, y)\right] \delta x \\ &\quad + \left[\frac{\partial^2 f}{\partial x \partial y}(x, y)\right] \delta x \delta y + o(\delta x) + o(\delta y). \qquad (4.4)\end{aligned}$$

Now the two partial derivatives are being evaluated at the same point (x, y), and so taking the limit as $\delta x = \delta y \to 0$ we obtain the **differential form** of the chain rule:

$$df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy, \qquad (4.5)$$

relating an **infinitesimal** or **differential** df of $f(x, y)$ to the differentials dx and dy.

We take this particular direction $\delta x = \delta y \to 0$ for the limit for simplicity, though if the function f is sufficiently smooth, the value of the limit will be the same whatever direction we take. Fundamentally, we want to ensure that the term involving the second derivative is either $o(\delta x)$ or $o(\delta y)$ as $\delta x \to 0$ and $\delta y \to 0$, and so (4.4) is equivalent to

$$\delta f = \left[\frac{\partial f}{\partial x}(x,y)\right]\delta x + \left[\frac{\partial f}{\partial y}(x,y)\right]\delta y + o(\delta x) + o(\delta y). \tag{4.6}$$

The differential form (4.5) should be understood as a shorthand for (4.6) in two canonical circumstances.

4.3.1 Integral form of chain rule

First, the differential form can be understood to be summed over the differentials to define integrals, i.e.

$$\int df = \int \frac{\partial f}{\partial x}dx + \int \frac{\partial f}{\partial y}dy. \tag{4.7}$$

4.3.2 Multivariate chain rule

Alternatively, we can understand the differential form to be divided by another **infinitesimal** before taking the limit. The simplest case of this second example is when $[x(t), y(t)]$ defines a path in space, where t is a parameter along the path, and may be thought of as a label of the "time". Therefore the function $f[x(t), y(t)]$ is clearly a function of time, and so

$$\frac{df}{dt} = \lim_{\delta t \to 0}\frac{\delta f}{\delta t} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}. \tag{4.8}$$

Notice that, since x and y are functions of t alone, there are **ordinary** derivatives of x and y with respect to t on the right hand side of this equation.

Another commonly occuring case is when $y(x)$, and so $f[x, y(x)]$. Therefore, the rate of change of f with respect to x is:

$$\frac{df}{dx} = \frac{\partial f}{\partial x}\frac{dx}{dx} + \frac{\partial f}{\partial y}\frac{dy}{dx} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\frac{dy}{dx}. \tag{4.9}$$

Thus the rate of change of f with respect to x is the sum of the rate of change of f with respect to x **keeping** y **fixed** plus the rate of change of f with respect to y **keeping** x **fixed** multiplied by the rate of change of y with respect to x.

Chapter 5

Applications of the chain rule

The chain rule is of course enormously important, and has many applications.

5.1 Change of variables

The chain rule naturally plays a central role when we change the independent variables, for example through a coordinate transformation. For example the function f may be a function of x and y or equivalently a function of r and θ, i.e. $f[x(r,\theta), y(r,\theta)]$. Therefore

$$\begin{aligned}
\left.\frac{\partial f}{\partial r}\right|_\theta &= \left.\frac{\partial f}{\partial x}\right|_y \left.\frac{\partial x}{\partial r}\right|_\theta + \left.\frac{\partial f}{\partial y}\right|_x \left.\frac{\partial y}{\partial r}\right|_\theta, \\
\left.\frac{\partial f}{\partial \theta}\right|_r &= \left.\frac{\partial f}{\partial x}\right|_y \left.\frac{\partial x}{\partial \theta}\right|_r + \left.\frac{\partial f}{\partial y}\right|_x \left.\frac{\partial y}{\partial \theta}\right|_r.
\end{aligned}$$

5.2 Implicit differentiation

Consider the expression $F(x, y, z) = C$ for some constant C. This defines a surface in 3D space, and so it **implicitly** defines a functional relationship between one of the coordinates x, y and z and the other two, i.e.

$$z = z(x, y) \text{ or } \; x = x(y, z) \text{ or } \; y = y(x, z).$$

Let us consider a concrete example:

$$xy^2 + yz^2 + z^5 x = 5. \tag{5.1}$$

- It is easy to express x explicitly as a function of the other two variables:

$$x = \frac{5 - yz^2}{y^2 + z^5}.$$

- It is possible to express $y(x,z)$ by considering the roots of a quadratic.
- However, we cannot find $z(x,y)$ explicitly, since that would require the (impossible) calculation of the roots of a quintic $z = z(x,y)$.

Nevertheless it is still straightforward to calculate how z varies with x while y remains fixed, by taking the partial derivative with respect to z (keeping y fixed) of (5.1):

$$\to y^2 + 2yz \left.\frac{\partial z}{\partial x}\right|_y + z^5 + 5xz^4 \left.\frac{\partial z}{\partial x}\right|_y = 0.$$

Rearranging,

$$\left.\frac{\partial z}{\partial x}\right|_y = -\left(\frac{y^2+z^5}{2yz+5xz^4}\right).$$

In general, wlog one can think of $F[x,y,z(x,y)] = C$. Then, generalizing the chain rule in differential form (4.5) to three variables:

$$dF = \left.\frac{\partial F}{\partial x}\right|_{y,z} dx + \left.\frac{\partial F}{\partial y}\right|_{x,z} dy + \left.\frac{\partial F}{\partial z}\right|_{x,y} dz,$$

where for clarity, the variables which are being kept fixed are listed. Now consider the partial derivative of F with respect to x keeping only y fixed, which must be:

$$\begin{aligned}
\left.\frac{\partial F}{\partial x}\right|_y &= \left.\frac{\partial F}{\partial x}\right|_{y,z} \left.\frac{\partial x}{\partial x}\right|_y + \left.\frac{\partial F}{\partial y}\right|_{x,z} \left.\frac{\partial y}{\partial x}\right|_y + \left.\frac{\partial F}{\partial z}\right|_{x,y} \left.\frac{\partial z}{\partial x}\right|_y, \\
&= \left.\frac{\partial F}{\partial x}\right|_{y,z} + \left.\frac{\partial F}{\partial z}\right|_{x,y} \left.\frac{\partial z}{\partial x}\right|_y = 0,
\end{aligned}$$

since F is of course a constant! The critical issue is that z is still allowed to vary for the term on the left hand side, so it is possible for the point (x,y,z) to remain on the surface, so F remains constant, and so $\partial F/\partial x|_y = 0$. On the other side for example $\partial F/\partial x|_{y,z}$ can be non-zero, as it quantifies how F is varying with x at **fixed** y and z, and so effectively is **tangent** to the surface, rather than on it. Therefore

$$\left.\frac{\partial z}{\partial x}\right|_y = -\frac{\left.\frac{\partial F}{\partial x}\right|_{y,z}}{\left.\frac{\partial F}{\partial z}\right|_{x,y}}.$$

Similarly

$$\begin{aligned}
\left.\frac{\partial x}{\partial y}\right|_z &= -\frac{\left.\frac{\partial F}{\partial y}\right|_{x,z}}{\left.\frac{\partial F}{\partial x}\right|_{y,z}}, \\
\left.\frac{\partial y}{\partial z}\right|_x &= -\frac{\left.\frac{\partial F}{\partial z}\right|_{x,y}}{\left.\frac{\partial F}{\partial y}\right|_{x,z}},
\end{aligned}$$

and so

$$\left.\frac{\partial x}{\partial y}\right|_z \left.\frac{\partial y}{\partial z}\right|_x \left.\frac{\partial z}{\partial x}\right|_y = -1. \tag{5.2}$$

5.2.1 Reciprocal rules

Normal reciprocal rules apply for partial derivatives, **provided** the same variables are being held constant. For example, for the coordinate change from (x, y) to (r, θ) and vice versa,

$$\frac{\partial r}{\partial x} \neq \frac{1}{\frac{\partial x}{\partial r}},$$

since the left hand side involves holding y fixed since $r(x, y)$, while the right hand side involves holding θ fixed since $x(r, \theta)$. But

$$\left.\frac{\partial r}{\partial x}\right|_y = \frac{1}{\left.\frac{\partial x}{\partial r}\right|_y}.$$

5.3 Families of integrals with parameters

Consider a family of functions $f(x, c)$ where the parameter c labels the different members of the family, as shown schematically in figure 5.1. It is then possible to define a class of integrals $I(b, c)$ which depend on two parameters:

$$I(b, c) = \int_0^b f(x, c)dx.$$

By the fundamental theorem of calculus, and the definition of a partial derivative:

$$\left.\frac{\partial I}{\partial b}\right|_c = f(b, c).$$

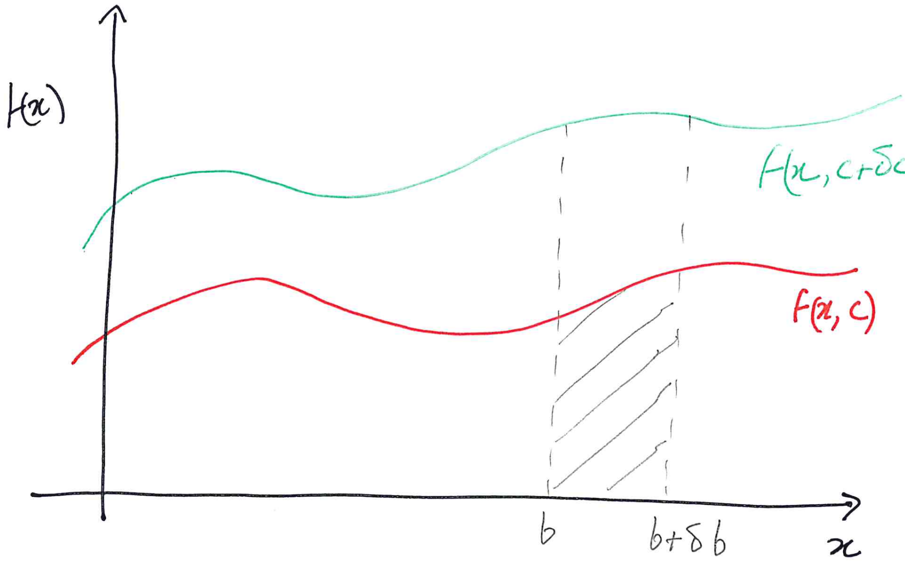

Figure 5.1: Schematic representation of integrals associated with a family of functions $f(x, c)$ as the parameter c is varied.

The partial derivative with respect to c is a little more complicated, but don't worry just plug (the various definitions) and chug (the manipulations):

$$\begin{aligned}
\left.\frac{\partial I}{\partial c}\right|_b &= \lim_{\delta c \to 0} \frac{1}{\delta c}\left[\int_0^b f(x, c+\delta c)dx - \int_0^b f(x, c)dx\right], \\
&= \lim_{\delta c \to 0}\left[\int_0^b \frac{f(x, c+\delta c) - f(x, c)}{\delta c}dx\right], \\
&= \int_0^b \left.\frac{\partial f}{\partial c}\right|_x dx.
\end{aligned}$$

These two results can be combined with the chain rule (4.8) if the parameters b and c are themselves functions of a third parameter, t say. Then

$$\begin{aligned}
I[b(t), c(t)] &= \int^{b(t)} f[x, c(t)]dx, \\
\frac{dI}{dt} &= \frac{\partial I}{\partial b}\frac{db}{dt} + \frac{\partial I}{\partial c}\frac{dc}{dt}, \\
&= f(b, c)\frac{db}{dt} + \frac{dc}{dt}\int_0^b \left.\frac{\partial f}{\partial c}\right|_x dx.
\end{aligned}$$

As an example:

$$\begin{aligned} I(t) &= \int_0^t f(x,t)dx, \\ \frac{dI}{dt} &= f(t,t) + \int_0^t \left.\frac{\partial f}{\partial t}\right|_x dx. \end{aligned}$$