Chapter 2

# The Foundation of Calculus

In this chapter we shall take a look at some of the fundamental ideas of calculus that we shall build on throughout the book. How much new material you will find here depends on your calculus courses. If you have followed a fairly theoretical calculus sequence or taken a course in advanced calculus, almost everything may be familiar, but if your calculus courses were only geared towards calculations and applications, you should work through this chapter before you approach the more abstract theory in Chapter 3.

What we shall study in this chapter is a mixture of theory and technique. We begin by looking at the $\epsilon$-$\delta$-technique for making definitions and proving theorems. You may have found this an incomprehensible nuisance in your calculus courses, but when you get to real analysis, it becomes an indispensable tool that you have to master – the subject matter is now so abstract that you can no longer base your work on geometrical figures and intuition alone. We shall see how the $\epsilon$-$\delta$-technique can be used to treat such fundamental notions as convergence and continuity.

The next topic we shall look at is completeness of $\mathbb{R}$ and $\mathbb{R}^n$. Although it is often undercommunicated in calculus courses, this is the property that makes calculus work – it guarantees that there are enough real numbers to support our belief in a one-to-one correspondence between real numbers and points on a line. There are two ways to introduce the completeness of $\mathbb{R}$ – by least upper bounds and Cauchy sequences – and we shall look at them both. Least upper bounds will be an important tool throughout the book, and Cauchy sequences will show us how completeness can be extended to more general structures.

In the last section we shall take a look at four important theorems from calculus: the Intermediate Value Theorem, the Bolzano-Weierstrass Theorem, the Extreme Value Theorem, and the Mean Value Theorem. All of these theorems are based on the completeness of the real numbers, and they introduce themes that will be important later in the book.

## 2.1. Epsilon-delta and all that

One often hears that the fundamental concept of calculus is that of a *limit*, but the notion of limit is based on an even more fundamental concept, that of the *distance* between points. When something approaches a limit, the distance between this object and the limit point decreases to zero. To understand limits, we first of all have to understand the notion of distance.

### Norms and distances

As you know, the distance between two points $\mathbf{x} = (x_1, x_2, \dots, x_m)$ and $\mathbf{y} = (y_1, y_2, \dots, y_m)$ in $\mathbb{R}^m$ is

$$\|\mathbf{x} - \mathbf{y}\| = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_m - y_m)^2}.$$

If we have two numbers $x, y$ on the real line, this expression reduces to

$$|x - y|.$$

Note that the order of the points doesn't matter: $\|\mathbf{x} - \mathbf{y}\| = \|\mathbf{y} - \mathbf{x}\|$ and $|x - y| = |y - x|$. This simply means that the distance from $\mathbf{x}$ to $\mathbf{y}$ is the same as the distance from $\mathbf{y}$ to $\mathbf{x}$.

If you don't like absolute values and norms, these definitions may have made you slightly uncomfortable, but don't despair – there isn't really that much you need to know about absolute values and norms to begin with.

The first thing I would like to emphasize is:

*Whenever you see expressions of the form $\|\mathbf{x} - \mathbf{y}\|$,*
*think of the distance between $\mathbf{x}$ and $\mathbf{y}$.*

Don't think of norms or individual points; think of the distance between the points! The same goes for expressions of the form $|x - y|$ where $x, y \in \mathbb{R}$: Don't think of numbers and absolute values; think of the distance between two points on the real line!

The next thing you need to know is the *Triangle Inequality* which says that if $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$, then

$$\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|.$$

If we put $\mathbf{x} = \mathbf{u} - \mathbf{w}$ and $\mathbf{y} = \mathbf{w} - \mathbf{v}$, this inequality becomes

$$\|\mathbf{u} - \mathbf{v}\| \leq \|\mathbf{u} - \mathbf{w}\| + \|\mathbf{w} - \mathbf{v}\|.$$

Try to understand this inequality geometrically. It says that if you are given three points $\mathbf{u}, \mathbf{v}, \mathbf{w}$ in $\mathbb{R}^m$, the distance $\|\mathbf{u} - \mathbf{v}\|$ of going directly from $\mathbf{u}$ to $\mathbf{v}$ is always less than or equal to the combined distance $\|\mathbf{u} - \mathbf{w}\| + \|\mathbf{w} - \mathbf{v}\|$ of first going from $\mathbf{u}$ to $\mathbf{w}$ and then continuing from $\mathbf{w}$ to $\mathbf{v}$.

The Triangle Inequality is important because it allows us to control the size of the sum $\mathbf{x} + \mathbf{y}$ if we know the size of the individual parts $\mathbf{x}$ and $\mathbf{y}$.

**Remark:** It turns out that the notion of distance is so central that we can build a theory of convergence and continuity on it alone. This is what we are going to do in the next chapter where we introduce the concept of a metric space. Roughly

Chapter 3

# Metric Spaces

Many of the arguments you have seen in multivariable calculus are almost identical to the corresponding arguments in single-variable calculus, especially arguments concerning convergence and continuity. The reason is that the notions of convergence and continuity can be formulated in terms of distance, and the notion of distance between numbers that you need in single-variable theory is very similar to the notion of distance between points or vectors that you need in the theory of functions of severable variables. In more advanced mathematics, we need to find the distance between more complicated objects than numbers and vectors, e.g., between sequences, sets, and functions. These new notions of distance leads to new notions of convergence and continuity, and these again lead to new arguments surprisingly similar to those you have already seen in single- and multivariable calculus.

After a while it becomes quite boring to perform almost the same arguments over and over again in new settings, and one begins to wonder if there is a general theory that covers all these examples – is it possible to develop a general theory of distance where we can prove the results we need once and for all? The answer is yes, and the theory is called the theory of metric spaces.

A metric space is just a set $X$ equipped with a function $d$ of two variables which measures the distance between points: $d(x, y)$ is the distance between two points $x$ and $y$ in $X$. It turns out that if we put mild and natural conditions on the function $d$, we can develop a general notion of distance that covers distances between numbers, vectors, sequences, functions, sets, and much more. Within this theory we can formulate and prove results about convergence and continuity once and for all. The purpose of this chapter is to develop the basic theory of metric spaces. In later chapters we shall meet some of the applications of the theory.

## 3.1. Definitions and examples

As already mentioned, a metric space is just a set $X$ equipped with a function $d: X \times X \to \mathbb{R}$ that measures the distance $d(x, y)$ between points $x, y \in X$. For the theory to work, we need the function $d$ to have properties similar to the distance functions we are familiar with. So what properties do we expect from a measure of distance?

First of all, the distance $d(x, y)$ should be a nonnegative number, and it should only be equal to zero if $x = y$. Second, the distance $d(x, y)$ from $x$ to $y$ should equal the distance $d(y, x)$ from $y$ to $x$. Note that this is not always a reasonable assumption – if we, e.g., measure the distance from $x$ to $y$ by the time it takes to walk from $x$ to $y$, $d(x, y)$ and $d(y, x)$ may be different – but we shall restrict ourselves to situations where the condition is satisfied. The third condition we shall need says that the distance obtained by going directly from $x$ to $y$ should always be less than or equal to the distance we get when we go via a third point $z$, i.e.,

$$d(x, y) \leq d(x, z) + d(z, x).$$

It turns out that these conditions are the only ones we need, and we sum them up in a formal definition.

**Definition 3.1.1.** *A* metric space $(X, d)$ *consists of a nonempty set* $X$ *and a function* $d \colon X \times X \to [0, \infty)$ *such that:*

(i) *(Positivity) For all* $x, y \in X$*, we have* $d(x, y) \geq 0$ *with equality if and only if* $x = y$.

(ii) *(Symmetry) For all* $x, y \in X$*, we have* $d(x, y) = d(y, x)$.

(iii) *(Triangle Inequality) For all* $x, y, z \in X$*, we have*

$$d(x, y) \leq d(x, z) + d(z, y).$$

*A function* $d$ *satisfying conditions* (i)-(iii) *is called a* metric *on* $X$.

**Comment:** When it is clear – or irrelevant – which metric $d$ we have in mind, we shall often refer to "the metric space $X$" rather than "the metric space $(X, d)$".

Let us take a look at some examples of metric spaces.

**Example 1:** If we let $d(x, y) = |x - y|$, then $(\mathbb{R}, d)$ is a metric space. The first two conditions are obviously satisfied, and the third follows from the ordinary Triangle Inequality for real numbers:

$$d(x, y) = |x - y| = |(x - z) + (z - y)| \leq |x - z| + |z - y| = d(x, z) + d(z, y).$$

**Example 2:** If we let

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\| = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2},$$

then $(\mathbb{R}^n, d)$ is a metric space. The first two conditions are obviously satisfied, and the third follows from the triangle inequality for vectors the same way as above:

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\| = \|(\mathbf{x} - \mathbf{z}) + (\mathbf{z} - \mathbf{y})\| \leq \|\mathbf{x} - \mathbf{z}\| + \|\mathbf{z} - \mathbf{y}\| = d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y}).$$

**Example 3:** Assume that we want to move from one point $\mathbf{x} = (x_1, x_2)$ in the plane to another $\mathbf{y} = (y_1, y_2)$, but that we are only allowed to move horizontally and

# Differential Calculus in Normed Spaces

There are many ways to look at derivatives – we can think of them as rates of change, as slopes, as instantaneous speed, or as new functions derived from old ones according to certain rules. If we consider functions of several variables, there is even more variety – we have directional derivatives, partial derivatives, gradients, Jacobian matrices, total derivatives, and so on.

In this chapter we shall extend the notion even further, to normed spaces, and we need a unifying idea to hold on to. That idea will be *linear approximation*: Our derivatives will always be linear approximations to functional differences of the form $f(a+r) - f(a)$ for small $r$. Recall that if $f\colon \mathbb{R} \to \mathbb{R}$ is a function of one variable, $f(a+r) - f(a) \approx f'(a)r$ for small $r$; if $f\colon \mathbb{R}^n \to \mathbb{R}$ is s scalar function of several variables, $f(\mathbf{a}+\mathbf{r}) - f(\mathbf{a}) \approx \nabla f(\mathbf{a}) \cdot \mathbf{r}$ for small $\mathbf{r}$; and if $\mathbf{F}\colon \mathbb{R}^n \to \mathbb{R}^m$ is a vector valued function, $\mathbf{F}(\mathbf{a}+\mathbf{r}) - \mathbf{F}(\mathbf{a}) \approx J\mathbf{F}(\mathbf{a})\mathbf{r}$ for small $\mathbf{r}$, where $J\mathbf{F}(\mathbf{a})$ is the Jacobian matrix. The point of these approximations is that for a given $\mathbf{a}$, the right-hand side is always a *linear* function in $\mathbf{r}$, and hence easier to compute and control than the nonlinear function on the left-hand side.

At first glance, the idea of linear approximation may seem rather feeble, but, as you probably know from your calculus courses, it is actually extremely powerful. It is important to understand what it means. That $f'(a)r$ is a better and better approximation of $f(a+r) - f(a)$ for smaller and smaller values of $r$ doesn't just mean that the quantities get closer and closer – that is a triviality as they both approach 0. The real point is that they get closer and closer *even compared to the size of* $r$, i.e., the fraction

$$\frac{f(a+r) - f(a) - f'(a)r}{r}$$

goes to zero as $r$ goes to zero.

As you know from calculus, there is a geometric way of looking at this. If we put $x = a + r$, the expression $f(a + r) - f(a) \approx f'(a)r$ can be reformulated as $f(x) \approx f(a) + f'(a)(x - a)$ which just says that the tangent at $a$ is a very good approximation to the graph of $f$ in the area around $a$. This means that if you look at the graph and the tangent in a microscope, they will become indistinguishable as you zoom in on $a$. If you compare the graph of $f$ to any other line through $(a, f(a))$, they will cross at an angle and remain separate as you zoom in.

The same holds in higher dimensions. If we put $\mathbf{x} = \mathbf{a} + \mathbf{r}$, the expression $f(\mathbf{a} + \mathbf{r}) - f(\mathbf{a}) \approx \nabla f(\mathbf{a}) \cdot \mathbf{r}$ becomes $f(\mathbf{x}) \approx f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a})$ which says that the tangent plane at $\mathbf{a}$ is a good approximation to the graph of $f$ in the area around $\mathbf{a}$ – in fact, so good that if you zoom in on $\mathbf{a}$, they will after a while become impossible to tell apart. If you compare the graph of $f$ to any other plane through $(\mathbf{a}, f(\mathbf{a}))$, they will remain distinct as you zoom in.

**Notational Convention:** In the previous chapter, I was always very careful in specifying the norms – the norm in $U$ would be denoted by $\|\cdot\|_U$, while the norm in $V$ was denoted by $\|\cdot\|_V$. This has the advantage of always making it clear which norm I am referring to, and the disadvantage of making long formulas look rather cluttered. In the present chapter, I find that the disadvantages outweigh the advantages, and drop the subscripts. Hence:

*Unless otherwise specified, all norms in this chapter are denoted by $\|\cdot\|$. It should always be clear from the context which norm I am referring to, but to make things easier, I shall usually (but not always) operate with functions from $X$ to $Y$, and use $\mathbf{x}$ and $\mathbf{a}$ for elements in $X$, and $\mathbf{y}$ and $\mathbf{b}$ for elements in $Y$.*

## 6.1. The derivative

In this section, $X$ and $Y$ will be normed spaces over $\mathbb{K}$, where as usual $\mathbb{K}$ is either $\mathbb{R}$ or $\mathbb{C}$. Our first task will be to define derivatives of functions $\mathbf{F}\colon X \to Y$. After the discussion above, the following definition should not come as a surprise.

**Definition 6.1.1.** *Assume that $X$ and $Y$ are two normed spaces. Let $O$ be an open subset of $X$ and consider a function $\mathbf{F}\colon O \to Y$. If $\mathbf{a}$ is a point in $O$, a* derivative *of $\mathbf{F}$ at $\mathbf{a}$ is a bounded, linear map $A\colon X \to Y$ such that*

$$\sigma(\mathbf{r}) = \mathbf{F}(\mathbf{a} + \mathbf{r}) - \mathbf{F}(\mathbf{a}) - A(\mathbf{r})$$

*goes to $\mathbf{0}$ faster than $\mathbf{r}$, i.e., such that*

$$\lim_{\mathbf{r}\to\mathbf{0}} \frac{\|\sigma(\mathbf{r})\|}{\|\mathbf{r}\|} = 0.$$

The first thing to check is that a function cannot have more than one derivative.

**Lemma 6.1.2.** *Assume that the situation is as in the definition above. The function $\mathbf{F}$ cannot have more than one derivative at the point $\mathbf{a}$.*

**Proof.** If $A$ and $B$ are derivatives of $\mathbf{F}$ at $\mathbf{a}$, we have that both

$$\sigma_A(\mathbf{r}) = \mathbf{F}(\mathbf{a} + \mathbf{r}) - \mathbf{F}(\mathbf{a}) - A(\mathbf{r})$$

Chapter 7

# Measure and Integration

In calculus you have learned how to calculate the size of different kinds of sets: the length of a curve, the area of a region or a surface, the volume or mass of a solid. In probability theory and statistics you have learned how to compute the size of other kinds of sets: the probability that certain events happen or do not happen.

In this chapter we shall develop a general theory for the size of sets, a theory that covers all the examples above and many more. Just as the concept of a metric space gave us a general setting for discussing the notion of distance, the concept of a measure space will provide us with a general setting for discussing the notion of size.

In calculus we use integration to calculate the size of sets. In this chapter we turn the situation around: We first develop a theory of size and then use it to define integrals of a new and more general kind. As we shall sometimes wish to compare the two theories, we shall refer to integration as taught in calculus as *Riemann integration* in honor of the German mathematician Bernhard Riemann (1826-1866) and the new theory developed here as *Lebesgue integration* in honor of the French mathematician Henri Lebesgue (1875-1941).

Let us begin by taking a look at what we might wish for in a theory of size. Assume that we want to measure the size of subsets of a set $X$ (if you need something concrete to concentrate on, you may let $X = \mathbb{R}^2$ and think of the area of subsets of $\mathbb{R}^2$, or let $X = \mathbb{R}^3$ and think of the volume of subsets of $\mathbb{R}^3$). What properties do we want such a measure to have?

Well, if $\mu(A)$ denotes the size of a subset $A$ of $X$, we would expect

(i) $\mu(\emptyset) = 0$.

as nothing can be smaller than the empty set. In addition, it seems reasonable to expect:

(ii) If $A_1, A_2, A_3 \ldots$ is a disjoint sequence of sets, then

$$\mu(\bigcup_{n\in\mathbb{N}} A_n) = \sum_{n=1}^{\infty} \mu(A_n).$$

These two conditions are, in fact, all we need to develop a reasonable theory of size, except for one complication: It turns out that we cannot in general expect to measure the size of *all* subsets of $X$ – some subsets are just so irregular that we cannot assign a size to them in a meaningful way. This means that before we impose conditions (i) and (ii) above, we need to decide which properties the *measurable sets* (those we are able to assign a size to) should have. If we call the collection of all measurable sets $\mathcal{A}$, the statement $A \in \mathcal{A}$ is just a shorthand for "$A$ is measurable".

The first condition is simple; since we have already agreed that $\mu(\emptyset) = 0$, we must surely want to impose

(iii) $\emptyset \in \mathcal{A}$.

For the next condition, assume that $A \in \mathcal{A}$. Intuitively, this means that we should be able to assign a size $\mu(A)$ to $A$. If the size $\mu(X)$ of the entire space is finite, we ought to have $\mu(A^c) = \mu(X) - \mu(A)$, and hence $A^c$ should be measurable. We shall impose this condition even when $X$ has infinite size:

(iv) If $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$.

For the third and last condition, assume that $\{A_n\}$ is a sequence of disjoint sets in $\mathcal{A}$. In view of condition (ii), it is natural to assume that $\bigcup_{n\in\mathbb{N}} A_n$ is in $\mathcal{A}$. We shall impose this condition even when the sequence is not disjoint (there are arguments for this that I don't want to get involved in at the moment):

(v) If $\{A_n\}_{n\in\mathbb{N}}$ is a sequence of sets in $\mathcal{A}$, then $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$.

When we now begin to develop the theory systematically, we shall take the five conditions above as our starting point.

## 7.1. Measure spaces

Assume that $X$ is a nonempty set. A collection $\mathcal{A}$ of subsets of $X$ that satisfies conditions (iii)-(v) above is called a $\sigma$-algebra. More succinctly:

**Definition 7.1.1.** *Assume that $X$ is a nonempty set. A collection $\mathcal{A}$ of subsets of $X$ is called a* $\sigma$-algebra *if the following conditions are satisfied:*

(i) $\emptyset \in \mathcal{A}$.

(ii) *If $A \in \mathcal{A}$, then $A^c = X \setminus A \in \mathcal{A}$.*

(iii) *If $\{A_n\}_{n\in\mathbb{N}}$ is a sequence of sets in $\mathcal{A}$, then $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$.*

*The sets in $\mathcal{A}$ are called* measurable *if it is clear which $\sigma$-algebra we have in mind, and $\mathcal{A}$-measurable if the $\sigma$-algebra needs to be specified. If $\mathcal{A}$ is a $\sigma$-algebra of subsets of $X$, we call the pair $(X, \mathcal{A})$ a* measurable space.