SOME PRELIMINARY NOTIONS FOR MATH 8110 – THEORY OF PARTIAL DIFFERENTIAL EQUATIONS

MARCELO M. DISCONZI*

Abstract. These notes review some mathematical notions, notation, and conventions that will be used in MATH 8110. The target audience are non-math majors taking the course.

Contents

1. Introduction

The purpose of these notes is to review some basic notions, set up the notation, and give students an idea of some of the more basic material that will be required for the course. While not all of this material is necessarily a pre-requisite, it is something that students are expected to quickly acquaint themselves with. Some students may find most of what follows very basic. If you feel that way, you are still urged to go over these notes carefully and make sure you understand them, as some of its aspects contain subtleties that are many times overlooked when one first learns the material.

While we tried to give a sufficiently precise treatment of the concepts involved, our approach is primarily *pragmatic*, invoking certain mathematical ideas only as long as they are necessary to solving partial differential equations in $\mathbb{R}^n$. As a result, we have avoided the level of rigor usually employed when many of these ideas are first introduced to a mathematical audience.

The conventions and notation described below will be adopted throughout the course, unless stated otherwise. While for the most part, they are compatible with those of the course textbook [3], there are some differences, so please be alert.

2. Vectors and coordinates

Recall that $\mathbb{R}^n$ consists of the set of n-tuples $(x_1, x_2, \dots, x_n)$, where each x_i, $i = 1, 2, \dots, n$, is a real number. This last statement is written briefly as $x_i \in \mathbb{R}$, where $\mathbb{R}$ denotes the set of real numbers and $\in$ means "belongs to" (the symbol $\notin$ will be used later on, and it means "does not belong to"). We write $x \in \mathbb{R}^n$ to indicate that x is an element of $\mathbb{R}^n$, i.e., $x = (x_1, x_2, \dots, x_n)$. n is called the **dimension** of the space $\mathbb{R}^n$. Each x_i is called a **component** or **coordinate** of x

*Vanderbilt University, Nashville, TN, USA. marcelo.disconzi@vanderbilt.edu.

(x_1 is the first component or first coordinate, x_2 the second component or coordinate, and so on). We shall use the terminology **point** and **vector** interchangeably for elements of $\mathbb{R}^n$. "Component" and "coordinate" will also be used interchangeably. Notice that we *do not* use an "arrow," i.e., the notation $\vec{x}$ (that you might have seen in Physics, Linear Algebra, or Multi-variable Calculus) for elements of $\mathbb{R}^n$. In particular, the **zero element** of $\mathbb{R}^n$, also called the **origin**, will be denoted simply by 0, i.e. $0 \in \mathbb{R}^n$ corresponds to the vector whose components are all zero: $0 = (0, 0, \ldots, 0)$.

While we typically use letters such as x, y, and z to denote elements of $\mathbb{R}^n$, sometimes, when dealing with $\mathbb{R}^2$ and $\mathbb{R}^3$, we reserve them for the components of a vector. For example, $v = (x, y, z) \in \mathbb{R}^3$. On the other hand, in many situations we have to label a *set of points* in $\mathbb{R}^n$, in which case we use a subscript. For example, suppose we are given N points[1] in $\mathbb{R}^n$. We can denote them by $x_1, x_2, \ldots, x_N$. Here, we have $x_i \in \mathbb{R}^n$ for each $i = 1, 2, \ldots, N$. In this case x_i *should not be confused* with the i^{th} component of a vector, being rather the i^{th} vector in the collection of N vectors $x_1, x_2, \ldots, x_n$. Naturally, in such a situation a different notation is needed for the components of the vectors $x_1, x_2, \ldots, x_N$. For example, if we want to write the components of, say, the vector x_3, we can use two indices, as

$$x_3 = (x_{31}, x_{32}, \ldots, x_{3n}).$$

Above, the first index labels which vector we are talking about, i.e., the third vector in this case, whereas the second index denotes the corresponding component of the third vector. More generally, we write,

$$x_i = (x_{i1}, x_{i2}, \ldots, x_{in}),$$

where i indicates the i^{th} vector in the collection $x_1, x_2, \ldots, x_N$, and the second index indicates the corresponding component of the i^{th} vector. Therefore, we can denote all components of all the vectors in our collection by x_{ij}, where i varies from 1 to N and labels which vector we are talking about, and j varies from 1 to n and denotes the j^{th} component of the i^{th} vector.

Yet sometimes, it will be more convenient to denote the coordinates of a point with *upper indices* or *super-scripts*, i.e.,

$$x = (x^1, x^2, \ldots, x^n),$$

in which case the components of a collection of N vectors can be written as x_i^j, with i and j holding the same interpretation as in the last paragraph, i.e., i varies from 1 to N and labels which vector we are talking about, and j varies from 1 to n and denotes the j^{th} component of the i^{th} vector[2].

All of the above may look very confusing at first sight. Sometimes x_i is a component, sometimes it denotes a point (=vector!) in $\mathbb{R}^n$; but sometimes components are denoted by x and y. The point, of course, is that it may be more convenient to use one notation over the other. It all depends on the particular problem we are addressing, and we need to have enough flexibility to use the most convenient notation in each case.

The important thing to have in mind is that what determines which type of object one has is not its notation, but rather how it is stated in the relevant context. For instance, suppose you are given a homework problem that begins with "Let $x \in \mathbb{R}^n$...". In this case, you are told that x is a vector with n components. Similarly, you could find a statement in the textbook that reads "Consider three points, x_1, x_2, x_3, in $\mathbb{R}^3$...". Here, you are told that each x_i, $i = 1, 2, 3$, is a three-component

[1]Note that N and n have nothing to do with each other. N denotes the number of points we are given or have chosen, so if we pick, say, ten points, we have $N = 10$. n denotes the dimension of the space. In a given problem, n is usually fixed (say, we are working on the three-dimensional space $\mathbb{R}^3$), while N can vary (first we are given ten points, $N = 10$; later on we pick six points, $N = 6$, etc.).

[2]Although we shall not explore it in this course, there is in fact a deeper meaning in the distinction between denoting coordinates as x^j and x_j. The mathematically inclined reader can search for the concept of the *dual of a vector space*, whereas Physics students may want to look at the distinction between *vectors* and *co-vectors*, or between *covariant* and *contravariant* coordinates.

vector[3]. Similarly, whenever you write your solutions, you should make clear what you mean by each object you introduce. Except in very special cases, you should not, and cannot, assume that it is obvious what is meant by the notation you are using. For example, if you are given a problem that says "Let $x \in \mathbb{R}^n \dots$", then it will be clear that by x_i you mean the components of x. On the other hand, if the problem makes no reference to x, then you cannot write things like x_i assuming that the reader will know what you mean, as it could be interpreted in several different ways. In this case, you have to write in your solutions something like "Let $x \in \mathbb{R}^n$...", or whatever else is meant in the context at hand.

While notation can, in fact, be a source of a great deal of confusion, it is important to know how make the most of the convenience that different notations offer. Naturally, the importance resides in the concepts themselves rather than in how one expresses them, although having a clear way of expressing mathematical ideas is definitely preferable. This is not, of course, much different than when you first learned about variables and functions, probably sticking to writing x for the variable and y for the function, learning, later on, that you could use different letters without changing the mathematical content of the problems.

The **canonical** or **standard** vectors in $\mathbb{R}^n$, denoted by e_i, $i = 1, \dots, n$, are the vectors with 1 in the i^{ih} component and zero in the remaining ones, i.e.,

$$\begin{aligned} e_1 &= (1, 0, 0, \dots, 0) \\ e_2 &= (0, 1, 0, \dots, 0) \\ e_3 &= (0, 0, 1, \dots, 0) \\ &\vdots \qquad\qquad \vdots \\ e_n &= (0, 0, 0, \dots, 1). \end{aligned}$$

We can write the above also as

$$\begin{gathered} e_i = (0, 0, \dots, 1, 0, 0, \dots, 0) \\ \uparrow \\ i^{\text{th}} \text{ component.} \end{gathered}$$

According to the conventions previously discussed, the components of e_i can be described as

$$e_{ij} = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$$

The **norm** of a vector $x \in \mathbb{R}^n$, denoted by $|x|$, is defined by

$$|x| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

Notice that $|x|$ measures the distance of x to the origin, or, equivalently, the "length" of the vector x. If $x, y \in \mathbb{R}^n$, then $|x - y|$ is simply the distance between x and y.

The **inner product** or **dot product** between two vectors x and y of $\mathbb{R}^n$ is denoted by both $\langle x, y \rangle$ and $x \cdot y$, and it is defined as

$$\langle x, y \rangle = x_1 y_1 + x_2 y_2 + \cdots x_n y_n,$$

or, more concisely,

$$\langle x, y \rangle = \sum_{i=1}^{n} x_i y_i.$$

Notice that $|x| = \sqrt{\langle x, x \rangle}$. Two vectors are said to be **orthogonal** if their inner product is zero.

[3]Notice that it is just a coincidence that the number 3 appears twice here: we could have more, or less, than three points in the three-dimensional space $\mathbb{R}^3$; see footnote 1.

3. Functions

We recall that a **function** is a rule that assigns for each element in a set A, one, and only one, element in a set B. The set A is called the **domain** of the function and B its **co-domain**. The notation

$$f : A \to B$$

is used to indicate that f is a function with domain A and co-domain B. We sometimes say that f **takes values in** B to refer to the co-domain of a function.

For the most part, we shall be dealing with functions defined in $\mathbb{R}^n$ (or in a subset of $\mathbb{R}^n$, see section 4) and taking values in $\mathbb{R}$, i.e., $f : \mathbb{R}^n \to \mathbb{R}$. In this case, if we write $f(x)$, it is to be understood that $x \in \mathbb{R}^n$, i.e., $x = (x_1, x_2, \ldots, x_n)$ and $f(x) = f(x_1, x_2, \ldots, x_n)$. A function that takes values in $\mathbb{R}$ is called a **real valued function**. The following are examples of real valued functions:

$$\begin{array}{ll} (i) & f : \mathbb{R} \to \mathbb{R}, \\ & f(x) = 2x + 1. \\ (ii) & g : \mathbb{R}^3 \to \mathbb{R}, \\ & g(x, y, z) = xy + xz + yz. \\ (iii) & h : \mathbb{R}^n \to \mathbb{R}, \\ & h(x) = \langle x, x \rangle. \end{array}$$

The above are typical examples of how a function is usually presented. Consider example (ii). First, we are given something like $g : A \to B$, which indicates how we are naming the function (g in this case), and the domain and co-domain of the function ($\mathbb{R}^3$ and $\mathbb{R}$, respectively); after all, if a function is a rule between two sets, upon defining a function we better say what these two sets are. Next, we define the rule that associates to each element in A one, and only one, element in B. In this example, the rule is the following: given an element $(x, y, z) \in \mathbb{R}^3$, the corresponding element in $\mathbb{R}$ is obtained by computing $xy + xz + yz$. For example, $g(1, -2, 3) = 1 \times (-2) + 1 \times 3 + (-2) \times 3 = -5$. Although this all sounds very trivial, you should make sure that you understand the notation involved.

Less often, we shall need functions with domain $\mathbb{R}^n$ (or a subset of $\mathbb{R}^n$, see section 4) and $\mathbb{R}^m$, where n and m may or may not be equal, depending on the particular problem. A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is called a **vector valued function**. The particular case when $n = m$ is called a **vector field**. The following are examples:

$$\begin{array}{ll} (iv) & u : \mathbb{R}^2 \to \mathbb{R}^3, \\ & u(x) = (x_1 x_2, x_1 - x_2, x_1 + x_2). \\ (v) & v : \mathbb{R}^2 \to \mathbb{R}^2, \\ & v(x, y) = (\frac{\sqrt{2}}{2}x - \frac{\sqrt{2}}{2}y, \frac{\sqrt{2}}{2}x + \frac{\sqrt{2}}{2}y) \\ (vi) & w : \mathbb{R}^n \to \mathbb{R}^n, \\ & w(x) = -x. \end{array}$$

Notice the difference between examples $(i) - (iii)$ and $(iv) - (vi)$. In $(i) - (iii)$ the answer is a *real number*, while in $(iv) - (vi)$ the answer is always a *vector* with more than one component. For instance, given $(1, 1) \in \mathbb{R}^2$, $u(1, 1)$ is the vector $(1, 0, 2) \in \mathbb{R}^3$.

We remark that sometimes functions have a nice geometric interpretation. You are invited to explore the geometric meaning of the function in example (v). By assigning values to x and y and drawing both (x, y) and $v(x, y)$, you should be able to verify that v rotates vectors in $\mathbb{R}^2$ by $45°$ counter-clockwise. The function w in example (vi) is a "reflection": it sends each point $x \in \mathbb{R}^n$ to its antipodal point.

Looking at $(iv) - (vi)$, we see that a vector valued function can be thought of as a vector where *each component is a real-valued function.* Thus, we can write,

$$f : \mathbb{R}^n \to \mathbb{R}^m,$$
$$f = (f_1, f_2, \ldots, f_m),$$

where

$$f_i : \mathbb{R}^n \to \mathbb{R},\ i = 1, \ldots, m.$$

Even more explicitly, since each f_i has domain $\mathbb{R}^n$, we can write,

$$\begin{aligned} f &: \mathbb{R}^n \to \mathbb{R}^m, \\ f(x) &= (f_1(x), f_2(x), \ldots, f_m(x)), \\ &= (f_1(x_1, x_2, \ldots, x_n), f_2(x_1, x_2, \ldots, x_n), \ldots, f_m(x_1, x_2, \ldots, x_n)). \end{aligned}$$

Notice that it is not always true that a "rule between sets" gives a well-defined function. Consider

$$f : \mathbb{R}^4 \to \mathbb{R},$$
$$f(x) = x.$$

While $f(x) = x$ seems a perfectly well-defined rule, it is inconsistent with $f : \mathbb{R}^4 \to \mathbb{R}$, in that this last statement says that the co-domain is $\mathbb{R}$, but $x \in \mathbb{R}^4$. We can correct this by either changing the co-domain to $\mathbb{R}^4$, or by changing the form of f to make it real valued, for example, $f(x) = |x|$.

Consider a function $f : A \to B$. Notice that not all elements of B have to be "hit" by the function. For instance, we can define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = x^2$. Then the co-domain is the set of all real numbers, but if we pick $-1 \in \mathbb{R}$, there is no x in the domain such that $f(x) = -1$. Given an element $y \in B$, we say that it is the **image** (through f) of the element $x \in A$ if $f(x) = y$. The **range** (sometimes also called the **image set** or just image for short) of a function $f : A \to B$ is the set of elements in B that are the image through f of at least one element from A. In the previous example, the range of $f(x) = x^2$ is the set of non-negative real numbers.

4. Sets and subsets, and more about functions

In many situations it will be necessary to work with functions defined on subsets of $\mathbb{R}^n$. We shall also need to look at subsets of points and functions with certain properties, and therefore, we introduce here some general notions of how to carry out such definitions.

A common way to define a set of elements with a certain property is to use **curly brackets to mean "set"** and **the symbol $\Big|$ to mean "such that"**. For example, consider the set Q defined by

$$Q = \Big\{ x \in \mathbb{R}^2 \,\Big|\, x_2 \geq 0 \Big\}.$$

One reads this as "Q is the set of all points in $\mathbb{R}^2$ such that the second coordinate is non-negative." "All points in $\mathbb{R}^2$" is indicated by $x \in \mathbb{R}^2$; "such that" is indicated by $\Big|$; and "the second coordinate is non-negative" is indicated by $x_2 \geq 0$. In other words, Q consists of the first and second quadrants together.

We can have more than one property defining a set. For example,

$$I = \Big\{ x \in \mathbb{R}^2 \,\Big|\, x_1 \geq 0, \text{ and } x_2 \geq 0 \Big\}$$

consists of all points in $\mathbb{R}^2$ such that the first coordinate is non-negative *and* the second coordinate is non-negative as well. In other words, I is the first quadrant on the plane. Most of the time we

omit the conjunction "and," listing the defining properties of a set separated simply by a comma. Thus, the set I above could also be defined as

$$I = \Big\{ x \in \mathbb{R}^2 \,\Big|\, x_1 \geq 0, x_2 \geq 0 \Big\}.$$

Notice that above a certain convention for how to denote coordinates in $\mathbb{R}^2$ is implicitly understood (see the discussion on section 2). In most cases we assume that the notation for the elements defining a set speaks for itself. For instance, the reader should have no trouble in understanding that

$$S^1 = \Big\{ (x, y) \in \mathbb{R}^2 \,\Big|\, x^2 + y^2 = 1 \Big\}$$

defines a circle of radius one centered at the origin, with x and y denoting the first and second coordinates in $\mathbb{R}^2$, respectively. Another example is

$$B^1 = \Big\{ (x, y) \in \mathbb{R}^2 \,\Big|\, x^2 + y^2 \leq 1 \Big\}.$$

The set B^1 is a "ball" of radius one centered at the origin, i.e., not only the circle of radius one but also its interior. Yet another simple example is

$$A = \Big\{ (x, y) \in \mathbb{R}^2 \,\Big|\, x^2 + y^2 \leq 1,\, x^2 + y^2 \geq \frac{1}{4} \Big\}.$$

How does A look like? Recall that $x^2 + y^2$ is the *square* of the distance from the origin to (x, y). Thus, it reads, "A is the set of all points in $\mathbb{R}^2$ such that their distance to the origin is less than or equal to one, *and* their distance to the origin is greater than or equal to $\frac{1}{2}$." We conclude that A is the annular region between the circle of radius one and the circle of radius $\frac{1}{2}$. Notice that A could also have been written as

$$A = \Big\{ (x, y) \in \mathbb{R}^2 \,\Big|\, \frac{1}{4} \leq x^2 + y^2 \leq 1 \Big\}.$$

The definition of a set can also carry a *parameter*. Consider

$$S_r = \Big\{ (x, y) \in \mathbb{R}^2 \,\Big|\, x^2 + y^2 = r^2 \Big\}.$$

Clearly, S_r consists of the circle of radius r centered at the origin. Notice that r does not appear together with $(x, y) \in \mathbb{R}^2$, i.e., we are not saying that S_r is "the set of r's with a certain property." Rather, *given* r, we define S_r so that for each r we pick, there is a different set S_r. In other words, r is something we typically fix first, depending on the nature of the problem we want to deal with, and then we define the set S_r.

Consider this other example:

$$S_r(x_0, y_0) = \Big\{ (x, y) \in \mathbb{R}^2 \,\Big|\, (x - x_0)^2 + (y - y_0)^2 = r^2 \Big\}.$$

$S_r(x_0, y_0)$ is the circle of radius r centered at (x_0, y_0). Again, notice that (x_0, y_0) is something *fixed* for the definition of the set, i.e., we are not considering "all (x_0, y_0) in $\mathbb{R}^2$ such that...". A slightly more elaborate example is

$$\Pi_y = \Big\{ x \in \mathbb{R}^n \,\Big|\, \langle x, y \rangle = 0 \Big\}.$$

Π_y is the "set of all vectors in $\mathbb{R}^n$ such that their inner product with the (fixed) element y is equal to zero." In other words, Π_y is the set of all vectors in $\mathbb{R}^n$ that are orthogonal to y (the reader familiar with Linear Algebra will recognize Π_y as a $n-1$-dimensional place through the origin in $\mathbb{R}^n$). Notice that, similarly to r above, the element y is fixed beforehand.

Now that we know how to define sets in $\mathbb{R}^n$, we can look at functions whose domain is not the whole of $\mathbb{R}^n$ but a subset of it. Consider

$$B_1^n = \Big\{ x \in \mathbb{R}^n \,\Big|\, |x| < 1 \Big\}.$$

B_1^n is "the ball of radius one inside $\mathbb{R}^n$," i.e., the set of all vectors in $\mathbb{R}^n$ whose norm is less than one. Define

$$f : B_1^n \to \mathbb{R},$$
$$f(x) = |x|.$$

By definition, this function has domain B_1^n, and thus it does not make sense to ask what $f(x)$ is if $x \notin B_1^n$ (remember that $\notin$ means "does not belong to"). Some readers may find this example a little silly, as we can perfectly compute $|x|$ for any $x \in \mathbb{R}^n$, and not only for those in B_1^n. However, such readers should remember that a function is a rule between two sets, A and B, and as such it does not make sense to ask what should be assigned to elements outside[4] A. If this puzzles you, you can consider that this is a "minimalist" approach: if in a particular problem or application we only care about points in B_1^n, we restrict ourselves to defining functions on in B_1^n, not caring about what happens outside B_1^n.

While for some readers the discussion of the last paragraph may sound conceptually abstract, there are of course obvious cases where the defining rule of a function only holds in a subset of $\mathbb{R}^n$. E.g., let

$$\mathbb{R}_0^n = \left\{ x \in \mathbb{R}^n \,\middle|\, x \neq 0 \right\},$$

i.e., $\mathbb{R}_0^n$ is $\mathbb{R}^n$ except for the origin. Then

$$f : \mathbb{R}_0^n \to \mathbb{R},$$
$$f(x) = \frac{1}{|x|}$$

is a well defined function, but the same expression cannot be used on the whole of $\mathbb{R}^n$ as we would otherwise divide by zero. The set $\mathbb{R}_0^n$ is more commonly written as

$$\mathbb{R}_0^n = \mathbb{R}^n \backslash \{0\},$$

where $\backslash$ means **minus**, i.e, $\mathbb{R}_0^n$ is the set of all elements in $\mathbb{R}^n$ minus (i.e., except) the set consisting only of the element zero, $\{0\}$.

Another example is

$$f : \left\{ x \in \mathbb{R}^n \,\middle|\, |x| \leq 1 \right\} \to \mathbb{R},$$
$$f(x) = \sqrt{1 - |x|}.$$

f will not be real valued unless we restrict it to $|x| \leq 1$. In this last example, we took a shortcut: instead of first defining the set $\left\{ x \in \mathbb{R}^n \,\middle|\, |x| \leq 1 \right\}$, and then defining f as having that set as domain, we preferred to define f and its domain simultaneously. This is useful to avoid "giving names" to several different sets every time we define a new function.

We finish, noticing that there is a natural way to define $f(U)$, where U is some subset of the domain of f. If, say, $f : \mathbb{R}^n \to \mathbb{R}$ and $U \subset \mathbb{R}^n$, then

$$f(U) = \left\{ f(x) \,\middle|\, x \in U \right\}.$$

In particular, if $f : A \to B$, then $f(A)$ is simply the range of f.

[4]The fact that the same formula, $f(x) = |x|$, holds for elements outside B_1^n means that our function f, initially defined on B_1^n, can be *extended* to the whole of $\mathbb{R}^n$ and, moreover, this extension is in a sense "the obvious" and "best" one. We shall not get into this type of conceptual subtlety in our course, but the mathematically inclined students is welcome to discuss this with me during office hours.

5. Some topological notions in $\mathbb{R}^n$

In this section we try to introduce some ideas from Topology without assuming any prior knowledge of the topic. Our discussion will be applicable only to $\mathbb{R}^n$ and will be, as said in the introduction, pragmatic and lacking rigor. This will make the way some ideas are introduced somewhat awkward from a fully mathematical point of view. The interested reader is referred, for example, to [2] for a more consistent treatment.

The symbol $\subseteq$ means **subset**, and $A \subseteq B$ reads "A is a subset of B. The case $A = B$ is *not* excluded (every set is a subset of itself). If we want to say that A is a subset of B but cannot equal B itself, we use the symbol $\subset$, so $A \subset B$. Sometimes we also say that "A is contained in B." The symbol $\not\subset$ means "not a subset," e.g. $\mathbb{R}^3 \not\subset \mathbb{R}^2$.

Recall that $(a, b) \subset \mathbb{R}$ is an open interval, whereas $[a, b] \subset \mathbb{R}$ is a closed interval. $(a, b]$ and $[b, a)$ are neither open nor closed. Notice that an open interval has the following property. Pick any $x \in (a, b)$. Then, we can always find another interval I_x containing x such that $I_x \subset (a, b)$. Said in a slightly different way, given any $x \in (a, b)$, we can always find a number $\varepsilon > 0$ such that the interval $(x - \varepsilon, x + \varepsilon)$ is contained in (a, b), i.e., $(x - \varepsilon, x + \varepsilon) \subset (a, b)$. A very similar idea is used to talk about open sets in $\mathbb{R}^n$. For this, we need the following notation and terminology, which will be adopted from now on. The **ball of radius r and center** x_0 in $\mathbb{R}^n$ is defined as

$$B_r(x_0) = \Big\{ x \in \mathbb{R}^n \;\Big|\; |x - x_0| < r \Big\}. \tag{5.1}$$

Sometimes we write $B^n_r(x_0)$ to emphasize that this is a subset of $\mathbb{R}^n$, and the case $x_0 = 0$ is sometimes abbreviated B_r — in which case we refer simply to the ball of radius r.

A set $U \subset \mathbb{R}^n$ is called **open** if for any $x \in U$ there exists a $r > 0$ such that $B_r(x) \subset U$. Notice that $B_r(x)$ is always open for any $x \in \mathbb{R}^n$. The **complement of a set** $U \subset \mathbb{R}^n$, denoted U^c, is the set of elements in $\mathbb{R}^n$ that do *not* belong to U. More precisely,

$$U^c = \Big\{ x \in \mathbb{R}^n \;\Big|\; x \notin U \Big\}.$$

Notice that this is the same as

$$U^c = \mathbb{R}^n \backslash U.$$

A set is called **closed** if its complement is open. As an instructive exercise, the reader is invited to show that the set

$$S^1 = \Big\{ (x, y) \in \mathbb{R}^2 \;\Big|\; x^2 + y^2 = 1 \Big\}$$

is closed, by first identifying its complement, and then showing that it is open.

The **intersection** of two sets A and B contained in $\mathbb{R}^n$, denoted $A \cap B$, is the set of all elements that belong simultaneously to A and B. We write,

$$A \cap B = \Big\{ x \in \mathbb{R}^n \;\Big|\; x \in A, x \in B \Big\}.$$

The reader can verify the following examples:

$$B_{\frac{1}{2}} \cap B_1 = B_{\frac{1}{2}},$$

$$(-2, 2) \cap (0, 4) = (0, 2),$$

and

$$B_1 \cap B^c_{\frac{1}{2}} = \Big\{ (x, y) \in \mathbb{R}^2 \;\Big|\; \frac{1}{4} \leq x^2 + y^2 < 1 \Big\}.$$

When A and B have no element in common, their intersection results in the **empty set**, denoted $\varnothing$. For example, let $x = (2, 2, 2) \in \mathbb{R}^3$ and $y = (-2, -2, -2) \in \mathbb{R}^3$, then

$$B_1(x) \cap B_1(y) = \varnothing.$$

Inductively, we can define the intersection of more than two sets, in which case we write $A \cap B \cap C$, etc. In fact, we can take the intersection of infinitely many sets. For example, consider the sets $I_i \subset \mathbb{R}$ defined by

$$I_i = [-\frac{1}{i}, \frac{1}{i}],\ i = 1, 2, 3, \ldots$$

Then

$$\bigcap_{i=1}^{\infty} I_i = \{0\}.$$

To convince yourself of the above, draw the first few I_i's and see what results from their intersection.

Given a set $U \subset \mathbb{R}^n$, we can consider the "smallest" closed set containing U. To make this notion more precise, let $C(U)$ be the collection of all closed sets containing U. I.e., $F \in C(U)$ if, and only if, F is a closed set and $U \subset F$. If we consider the intersection of all F's with these properties, then we obtain a set called the **closure** of U, denoted $\overline{U}$. Thus,

$$\overline{U} = \bigcap_{F \in C(U)} F.$$

To understand this notion, consider that, if U is a closed set, then its closure is U itself, i.e., $\overline{U} = U$ whenever U is closed. If U is not closed, then $\overline{U}$ consists of U plus the points that are "missing" to make U closed. For instance, $[0, 1)$ is not closed, because the endpoint 1 is not included, thus

$$\overline{[0, 1)} = [0, 1].$$

Different sets can have the same closure, e.g., the closure of $(0, 1)$, $(0, 1]$, and $[0, 1)$ are all equal to $[0, 1]$. The reader should notice that $\overline{U}$ is *always* a closed set[5].

As another example, consider B_1, whose definition was given in (5.1) and in the text that immediately followed. B_1 is not closed because, by its definition, it includes only the points whose distance to the origin is *less* than one, i.e., it is "missing" the points whose distance is exactly *equal* to one. Thus

$$\overline{B_1} = \Big\{ x \in \mathbb{R}^n \,\Big|\, |x| \leq 1 \Big\}.$$

Analogously,

$$\overline{B_r(x_0)} = \Big\{ x \in \mathbb{R}^n \,\Big|\, |x - x_0| \leq r \Big\}.$$

Notice that the crucial difference of the above to (5.1) is that $<$ has been replaced by $\leq$.

When forming $\overline{B_r(x_0)}$, the points that were "missing" consists of a $n-1$-dimensional sphere of radius r and center x_0 (if that is not clear, draw a picture of $\overline{B_r(x_0)}$ in two and three dimensions). Therefore we define the $n-1$ **sphere of radius** r **and center** x_0 inside $\mathbb{R}^n$ as

$$S_r(x_0) = \Big\{ x \in \mathbb{R}^n \,\Big|\, |x - x_0| = r \Big\}.$$

We write $S_r^{n-1}(x_0)$ if we want to emphasize that $S_r(x_0)$ is a subset[6] of $\mathbb{R}^n$, and S_r for $S_r(0)$ — in this last case we refer simply to the sphere of radius r. The reader should check that $S_r(x_0)$ is a closed set.

[5] It is an instructive exercise to show this, as follows. By definition, one has to show that the complement of $\overline{U}$ is open. This is done using the so-called de Morgan's law (see, for instance, [1]): $\overline{U}^c = (\cap_F F)^c = \cup_F F^c$. Since each F^c is open because each F is closed, and since (as the reader can check) the union of open sets is open, we conclude that $\overline{U}^c$ is open, and thus $\overline{U}$ is closed.

[6] Notice that we write $S_r^{n-1}(x_0)$, and not $S_r^n(x_0)$, because $S_r(x_0)$ always has "one less dimension" than $\mathbb{R}^n$. For example, in two-dimensions, the ball of radius one is $\{(x, y) \in \mathbb{R}^2 \,|\, x^2 + y^2 \leq 1\}$, but the sphere is only the circle $\{(x, y) \in \mathbb{R}^2 \,|\, x^2 + y^2 = 1\}$. Although we shall not define the concept of dimension for general sets, this should be intuitive. Also, see the comments at the end of this section.

$S_r(x_0)$ can be thought of as the outermost points of $\overline{B_r(x_0)}$, or the "boundary" of $B_r(x_0)$. Furthermore, $\overline{B_r(x_0)}$ can be decomposed into two pieces, one open and one closed, namely, $B_r(x_0)$ and $S_r(x_0)$, and such that $B_r(x_0) \cap S_r(x_0) = \varnothing$. Next, we define some similar notions for sets that are not necessarily a ball. To do that, we need to first recall what the union of sets is.

The **union** of of two sets A and B contained in $\mathbb{R}^n$, denoted $A \cup B$, is defined as the set of points that belong to A or B, i.e.

$$A \cup B = \Big\{ x \in \mathbb{R}^n \;\Big|\; x \in A, \text{ or } x \in B \Big\}.$$

For example, $(0,2) \cup (1,3) = (0,3)$, and $B_1 \cup S_1 = \overline{B_1}$. As with intersections, we can consider the union of more than two sets, and even of infinitely many sets.

Given a set $U \subset \mathbb{R}^n$, we can consider the "largest" open set contained in U. More precisely, let $O(U)$ be the collection of all open sets contained in U. I.e., $E \in O(U)$ if, and only if, E is an open set and $E \subset U$. The **interior** of U, denoted $\mathring{U}$, is defined as

$$\mathring{U} = \bigcup_{E \in O(U)} E.$$

We point out that $\mathring{U}$ is *always* an open set.

The **boundary** of a set $U \subset \mathbb{R}^n$, denoted ∂U, is defined by

$$\partial U = \overline{U} \backslash \mathring{U}.$$

From the definition, it follows that $\partial U \cap \mathring{U} = \varnothing$.

Another important concept is that of connectedness. A set $U \subset \mathbb{R}^n$ is said to be **connected** if any two points in U can be joined by a *continuous* curve[7]. The basic intuition is that a connected set cannot be split into two parts that "do not communicate with each other." For instance, B_r is a connected set, as are $(0,1) \subset \mathbb{R}$, $[1,3)$, and S_r. The sets $\mathbb{R}\backslash\{0\}$, and $(-2,-1) \cup (1,2)$, on the other hand, are not connected.

We can now define one of the main types of sets of interest in this course. A **domain** in $\mathbb{R}^n$ is a *connected* and *open* set[8]. The notation Ω will always denote a domain, unless stated otherwise. The balls $B_r(x_0)$ are examples of domains. The spheres $S_r(x_0)$ are not domains because they fail to be open.

Given a domain Ω, we can consider its boundary $\partial\Omega$ as defined earlier. A domain is said to have **smooth boundary** if the following property holds: *locally*, $\partial\Omega$ can be written as the graph of an infinitely differentiable function.

Probably the best way to understand this definition is via the following example. Consider the ball of radius one in $\mathbb{R}^3$. Recall that it is the set

$$B_1 = \Big\{ (x,y,z) \in \mathbb{R}^3 \;\Big|\; \sqrt{x^2+y^2+z^2} < 1 \Big\}.$$

Its boundary is the two-dimensional sphere

$$S_1 = \Big\{ ((x,y,z) \in \mathbb{R}^3 \;\Big|\; \sqrt{x^2+y^2+z^2} = 1 \Big\},$$

i.e., $\partial B_1 = S_1$. Notice that S_1 cannot be the graph of a function, as it fails the analogue of the vertical line test in three-dimensions (any straight line through S_1 will cross it in two points). However, small pieces (this is roughly the meaning of "locally") of S_1 can always be written as a graph, as we now show.

[7]To be precise, this is actually the definition of what we call a *path connected* set. Giving the precise definition of connectedness would require a more extensive discussion. We point out, however, that in the main case of interest, i.e., when the set U is open, the concepts of connectedness and path connectedness agree. The interested reader is referred to [2] for details.

[8]The concept of domain here has nothing to do with the domain of a function.

Consider first the upper hemisphere of S_1, which we denote by S_1^+. It corresponds to the points that satisfy not only $\sqrt{x^2+y^2+z^2} = 1$ but also that are "above" the xy-plane, i.e., $z \geq 0$. Therefore, S_1^+ is given by the points that satisfy

$$\sqrt{x^2+y^2+z^2} = 1 \text{ and } z \geq 0.$$

Solving for z and using $z \geq 0$ to pick the positive square root gives

$$z = \sqrt{1-x^2-y^2}.$$

But this defines z as a function of x and y. Therefore, the upper hemisphere S_1^+ is the graph of the function $h(x,y) = \sqrt{1-x^2-y^2}$, whose domain[9] is the set $\{(x,y) \in \mathbb{R}^2 \,\big|\, x^2+y^2 \leq 1\}$.

Similarly, the lower hemisphere, denoted by S_1^-, is given by

$$\sqrt{x^2+y^2+z^2} = 1 \text{ and } z \leq 0.$$

Solving for z and using $z \leq 0$ to pick now the negative square root gives

$$z = -\sqrt{1-x^2-y^2}.$$

Again, z is a function of x and y, whose domain is the set $\{(x,y) \in \mathbb{R}^2 \,\big|\, x^2+y^2 \leq 1\}$. Since S_1^+ and S_1^- completely cover the sphere, i.e. $S_1^+ \cup S_-^1 = S_1$, we have shown that locally S^1 can always be written as the graph of a function.

The above does not quite show yet that B_1 has smooth boundary. To do so, we need to show that the functions whose graphs give S_1 are infinitely differentiable. Consider the case S_1^+. Taking derivatives of $\sqrt{1-x^2-y^2}$, we will find expressions that involve

$$\frac{1}{\sqrt{1-x^2-y^2}}.$$

The above, and its derivatives, will be well-defined as long as $x^2+y^2 < 1$, but for points such that $x^2+y^2 = 1$, we would be dividing by zero. The same statement holds for S_1^-.

Noticing that the points satisfying $x^2+y^2 = 1$ correspond exactly to the equator of the sphere, we conclude that the above shows that S_1 can always be written locally as the graph of an infinitely differentiable function, *except* for the points on the equator. To remedy this problem, we simply notice that there is nothing in the definition of a smooth boundary that requires us to limit ourselves to only two pieces of $\partial\Omega$. The points on the equator can also be shown to be part of graphs of certain infinitely differentiable functions, except that now z will be one of the variables and x or y will be the function. For instance, a similar reasoning as above shows that if we consider the right hemisphere, given by

$$\sqrt{x^2+y^2+z^2} = 1 \text{ and } x \geq 0,$$

and the left hemisphere, given by

$$\sqrt{x^2+y^2+z^2} = 1 \text{ and } x \leq 0,$$

then we can solve for x in terms of y and z, obtaining the functions

$$x = \sqrt{1-y^2-z^2}$$

and

$$x = -\sqrt{1-y^2-z^2}.$$

[9]For technical reasons, when one says that $\partial\Omega$ is locally the graph of a smooth function, it is convenient to define the domain of the function giving the graph to be always an open set. In this case we would take $\{(x,y) \in \mathbb{R}^2 \,\big|\, x^2+y^2 < 1\}$. We shall, however, avoid this kind of subtlety here.

These functions both have domain $y^2 + z^2 \leq 1$, and they are infinitely differentiable, except for the points satisfying $y^2 + z^2 = 1$. These points constitute the "principal meridian" of the sphere. But notice that these points are either on the upper hemisphere or on the lower hemisphere, which we have already shown to be the graph of an infinitely differentiable function except for points on the equator. However, the only points that are simultaneously on the equator and on the principal meridian are the points $(1, 0, 0)$ and $(-1, 0, 0)$. Therefore, we have shown that all points on the sphere, except those two, belong to the graph of an infinitely differentiable function. The reader can now probably imagine how those last two points are shown to also satisfy this property: one writes y as a (two) function(s) of x and z, and argue analogously.

The conclusion is that by using enough "caps" we can cover the sphere with pieces that are always the graph of an infinitely differentiable function, as shown in figure 1.

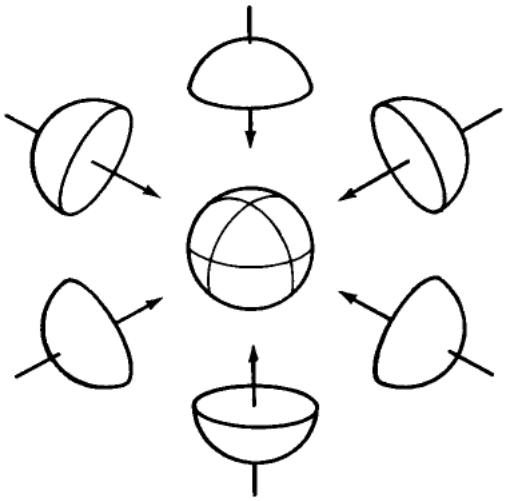

FIGURE 1. Covering the sphere with several caps (credit: M. P. do Carmo, *Riemannian Geometry.* Birkhäuser (1992)).

Hence, B_1 is a domain with smooth boundary. Of course, there is nothing special about the radius equal to one or the origin as the center. Nor is there anything special about the dimension $n = 3$. Similar arguments can be used to show that $B_r(x_0) \subset \mathbb{R}^n$ is always a domain with smooth boundary.

Some readers may find this last discussion too lengthy or complicated, but they should be able to grasp it quickly after working out some of the details by themselves. We insist that this last example be understood thoroughly, as B_1 is the prototypical example of domains with smooth boundary — which will be one of the most important sets used in the course.

A domain is called **bounded** if it can be enclosed inside a ball of radius r for some $r > 0$. For example, the region

$$A = \left\{ (x, y, z) \in \mathbb{R}^3 \,\middle|\, \frac{1}{16} < x^2 + y^2 + z^2 < 1 \right\}$$

is bounded because it lies inside[10] the ball of radius 2; while

$$\widetilde{A} = \left\{ (x, y, z) \in \mathbb{R}^3 \,\middle|\, x > 0, y > 0, z > 0 \right\}$$

is not bounded.

[10]Of course, A also lies inside the ball of radius 3, or 1.5... The definition only requires that it belongs to a ball of radius r for *some* $r > 0$.

Lastly, we make a comment about dimension. It should be intuitive that $B_r \subset \mathbb{R}^n$ has n dimensions, whereas S_r has $n-1$ dimensions (the reader can consider the case $n=3$ for simplicity if necessary). This is because in B_r we have n variables $x_1, x_2, \dots x_n$, but in S_r one variable can always be written as a function of the remaining $n-1$ variables, as we have shown above. Hence, there are truly only $n-1$ independent coordinates in S_r. Although we shall not define the concept of dimension here, the reader should keep this intuitive notion in mind: that a domain Ω in $\mathbb{R}^n$ is a n-dimensional spaces, whereas its boundary $\partial\Omega$ is a $n-1$-dimensional space[11]. This will be important later on when we discuss functions defined on Ω and $\partial\Omega$, i.e., $f : \Omega \to \mathbb{R}$ and $g : \partial\Omega \to \mathbb{R}$, respectively. f is then a function of n variables, while g is a function of $n-1$ varaibles.

6. Partial derivatives

In this section we recall some basic notions about partial derivatives. The reader can check the standard literature for a more detailed review of the topics here presented. The **partial derivative** with respect to the i^{ih} coordinate will be denoted by

$$\frac{\partial}{\partial x_i}, \text{ or } \partial_i,$$

with higher order derivatives denoted accordingly, e.g.,

$$\frac{\partial^2}{\partial x_i \partial x_j}, \partial^2_{ij}, \text{ or simply } \partial_{ij}.$$

We sometimes speak simply of "derivative" to mean "partial derivative." If f is a function of n variables (for instance, a function defined in $\mathbb{R}^n$), i.e., $f(x_1, x_2, \dots, x_n)$, its derivative with respect to the i^{ih} coordinate is also denoted by a subscript:

$$f_i = \partial_i f,$$

or yet

$$f_{x_i} = \partial_i f.$$

Notice that if $x \in \mathbb{R}^n$, then $f(x)$ means that f is a function of n variables, i.e., $f(x) = f(x_1, x_2, \dots, x_n)$, in which case $f_i(x)$ means the derivative of f with respect to the i^{ih} coordinate evaluated at x. You should be careful not to confuse the notation f_i for partial derivatives with the notation for the i^{th} component of a vector valued function.

A function is said to be **k-times differentiable** if all its partial derivatives up to order k exist, and **k-times continuously differentiable** if all its partial derivatives up to order k exist and are continuous. It is important to notice the difference between these two concepts. The reader is encouraged to try to find an example of a function whose derivative exists but is not continuous; i.e., find $f : \mathbb{R} \to \mathbb{R}$, such that f' is well-defined, but f' is not a continuous function. We denote by $C^k(\Omega)$ the set of real-valued k-times continuously differentiable functions defined on Ω (recall that Ω is always a domain in $\mathbb{R}^n$). More precisely,

$$C^k(\Omega) = \left\{ f : \Omega \to \mathbb{R} \;\middle|\; \text{ all partial derivatives of } f \text{ up to order } k \text{ exist and are continuous } \right\}.$$

The sets $C^k(\overline{\Omega})$ and $C^k(\partial\Omega)$ are defined similarly[12].

From now on, it will be assumed that the functions involved are sufficiently differentiable, so that all the formulas involving derivatives will make sense.

[11] To be more precise, the boundary is a space of *at most* $n-1$ dimensions. However, in all cases relevant for this course, it will in fact consist of a $n-1$ dimensional space.

[12] The reader should make sure that he or she understands the definition of derivative of a function defined on a *closed* set, such as $\overline{\Omega}$.

A very important tool to compute derivatives is provided by the **chain rule**. Recall that if $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ are differentiable, then

$$(f \circ g)'(x) = f'(g(x))g'(x),$$

where $f \circ g$ is the **composition** of f and g. We can omit x and write the above as

$$(f \circ g)' = (f' \circ g)g'.$$

Next, we recall how this rule generalizes to functions of several variables.

Consider $g : (a, b) \to \Omega$ and $f : \Omega \to \mathbb{R}$, so that $f \circ g : \mathbb{R} \to \mathbb{R}$ is well-defined. We can write $g = (g_1, g_2, \ldots, g_n)$. Thus,

$$(f \circ g)'(x) = \sum_{i=1}^{n} \partial_i f(g(x))g_i'(x),$$

or simply

$$(f \circ g)' = \sum_{i=1}^{n} (\partial_i f \circ g)g_i'.$$

Sometimes one sees the above written as

$$(f \circ g)' = \sum_{i=1}^{n} \partial_i f g_i'.$$

In this last expression, it is implicitly understood that for each $x \in (a, b)$, $\partial_i f$ is to be evaluated at $g(x)$, i.e., $\partial_i f(g(x))$, even though the composition $\circ g$ in $\partial_i f \circ g$ has been omitted. Although this may be a bit confusing at first sight, it is the only thing that makes sense, since we are computing the derivative of the composition $f \circ g$.

There is a simple mnemonics for the chain rule. Recall that for single variable functions, we can write $x = g(t)$ and $f = f(x)$, so that

$$\frac{d}{dt}(f \circ g) = \frac{df}{dx}\frac{dx}{dt}, \tag{6.1}$$

where of course $\frac{dx}{dt} = g'$.

In the case $g : (a, b) \to \Omega$ and $f : \Omega \to \mathbb{R}$, let us write $x = g(t)$, so that $x = (x_1, x_2, \ldots, x_n) = (g_1, g_2, \ldots, g_n)$, i.e., $x_i(t) = g_i(t)$. Then

$$(f \circ g)' = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\frac{dx_i}{dt}, \tag{6.2}$$

where $\frac{dx_i}{dt} = g_i'$. Comparing (6.1) with (6.2), it is seen that they have exactly the same form, except that because f is a function of n variables, in (6.2) the derivatives of f are partial derivatives and we have to sum over the different components x_i. Notice that each $x_i = g_i$ is a function of only one variable (the variable t). Notice also that with this notation, there is no need to write the composition $\circ g$ along with the partial derivative of f: since the coordinates in Ω are denoted by x, and we wrote $\frac{\partial f}{\partial x_i}$, it is implicitly understood that this is evaluated at $x \in \Omega$; but $x = g(t)$.

Sometimes, we shall also need to compute the derivative of the composition of two functions of several variables. Suppose $\Omega_0 \subset \mathbb{R}^m$ and $\Omega_1 \subset \mathbb{R}^n$ are two domains in $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively (notice that m and n can be different). Let $g : \Omega_0 \to \Omega_1$ and $f : \Omega_1 \to \mathbb{R}$, so that $f \circ g$ is well-defined. Notice that $f \circ g$ is a function of m variables, thus, when computing derivatives of $f \circ g$ we have to talk about its *partial* derivatives. To do so, we apply formula (6.2) *for each partial derivative*. More precisely, denote the coordinates in Ω_0 by x, so $x = (x_1, x_2, \ldots, x_m) \in \Omega_0$, and the coordinates in

Ω_1 by y, so $y = (y_1, y_2, \ldots, y_n) \in \Omega_1$. Then we have $y = g(x)$ for each $x \in \Omega_0$, i.e., $y_i = g_i(x)$, $i = 1, 2, \ldots, m$. The i^{th} partial derivative of $f \circ g$ is given by

$$\frac{\partial (f \circ g)}{\partial x_i} = \sum_{j=1}^{n} \frac{\partial f}{\partial y_j} \frac{\partial y_j}{\partial x_i}, \tag{6.3}$$

where $\frac{\partial y_j}{\partial x_i} = \frac{\partial g_j}{\partial x_i}$. Notice that again, by the conventions we adopted to indicate the coordinates in Ω_0 and Ω_1, there is no need to write $\circ g$ along with $\frac{\partial f}{\partial y_j}$ since $y_j = g_j(x)$. The reader should compare (6.3) with (6.2) and realize that the latter is a particular case of the former when $m = 1$.

If $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$, its **gradient**, denoted ∇f, is the n-component vector defined as

$$\nabla f = (\partial_1 f, \partial_2 f, \ldots, \partial_n f).$$

And if $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}^m$, its **Jacobian matrix**, denoted Df, is given by

$$Df = \begin{bmatrix} \partial_1 f_1 & \partial_2 f_1 & \cdots & \partial_n f_1 \\ \partial_1 f_2 & \partial_2 f_2 & \cdots & \partial_n f_2 \\ \vdots & & & \vdots \\ \partial_1 f_m & \partial_2 f_m & \cdots & \partial_n f_m \end{bmatrix}$$

Other notations for Df are

$$\frac{\partial (f_1, f_2, \ldots, f_m)}{\partial (x_1, x_2, \ldots, x_n)} \text{ or } (\partial_i f_j).$$

With these notations, it is instructive to show that (6.2) is given by

$$(f \circ g)' = \langle \nabla f \circ g, g' \rangle,$$

and that (6.3) can be written as

$$D(f \circ g) = \nabla f \circ g \cdot Dg.$$

In this last expression, $\cdot$ is simply the matrix multiplication of the $1 \times n$ matrix $\nabla f \circ g$ by the $n \times m$ matrix Dg. The result is a $1 \times m$ matrix whose i^{th} column is the i^{th} partial derivative of $f \circ g$. The attentive reader has probably already noticed that ∇f equals Df when $f : \mathbb{R} \to \mathbb{R}^n$. Therefore, *all of the above formulas for the chain rule are particular cases of the general formula:*

$$D(f \circ g) = Df \circ g \cdot Dg.$$

7. Integrals

Here we recall some basic facts about integrals of functions of several variables. It will be assumed that the functions satisfy all the required hypotheses to make the integrals involved well-defined. The reader can check the standard literature for a more detailed review of the topics here presented.

The **integral** of a function $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ will be denoted by

$$\int_\Omega f.$$

Notice that we do not write the volume element dV or $d\vec{x}$ for this multidimensional integral, although sometimes it may be convenient to do so (e.g., to stress which ones are the variables of integration), in which case we shall write

$$\int_\Omega f(x)\, dx.$$

Notice, also, that dx represents the volume element in n-dimensions, i.e.,

$$dx = dx_1 dx_2 \cdots dx_n,$$

so $\int_\Omega f(x)\,dx$ can be written more explicitly as

$$\int_\Omega f(x_1, x_2, \ldots, x_n)\,dx_1 dx_2 \cdots dx_n.$$

The above also implies that we avoid using the notations of "several integrals." For example, if Ω is a domain in $\mathbb{R}^2$, we do *not* write,

$$\iint_\Omega f(x_1, x_2)\,dx_1 dx_2$$

since, once it is known that $\Omega \subset \mathbb{R}^2$, it is superfluous to write the integral sign twice.

Consider a domain Ω with smooth boundary $\partial\Omega$. Remember that it is possible to carry out integration of functions defined over $\partial\Omega$, which we write,

$$\int_{\partial\Omega} g,$$

where $g : \partial\Omega \to \mathbb{R}$. As before, we do not use some of the "standard" notation found in calculus books, avoiding writing dA, $d\vec{A}$, etc. for the area element of $\partial\Omega$. When it is necessary to write such an area element, we denote it by ds,

$$\int_{\partial\Omega} g\,ds,$$

or

$$\int_{\partial\Omega} g(x)\,ds.$$

Intuitively, ds is the "restriction" of the n-dimensional volume element dx to the $n-1$-dimensional boundary $\partial\Omega$. For instance, if $\Omega \subset \mathbb{R}^3$ is the upper-half plane,

$$\Omega = \Big\{(x, y, z) \in \mathbb{R}^3 \,\Big|\, z > 0\Big\},$$

then $\partial\Omega$ is the xy-plane,

$$\partial\Omega = \Big\{(x, y, z) \in \mathbb{R}^3 \,\Big|\, z = 0\Big\},$$

and in this case $ds = dxdy$. It is important to remark that even if the boundary is $n-1$-dimensional, ds is still referred to as an *area* element — although sometimes we also use *volume element induced on the boundary*, *induced volume element*, or *boundary volume element*.

Again, the reader should notice that we avoid some of the more involved notation for the boundary integrals. For instance, if Ω is two-dimensional, the $\partial\Omega$ is a curve, and $\int_{\partial\Omega}$ is sometimes written $\oint_{\partial\Omega}$. This notation will *not* be employed here.

In some exceptional situations, one wants to integrate a function of $n+m$ variables with respect to, say, the first n variables. In such cases we write,

$$f = f(x, y) = f(x_1, x_2, \ldots, x_n, y_1, y_2, \ldots y_m),$$

and the integrals

$$\int_\Omega f(x, y)\,dx,$$

and

$$\int_{\partial\Omega} g(x, y)\,ds(x),$$

where $\Omega \subset \mathbb{R}^n$. I.e., even though the functions f and g involve $n+m$ variables, we are considering an integral over a domain Ω that belongs to $\mathbb{R}^n$ — thus, we have integrals over the first n-variables.

In the above, the x in $ds(x)$ is used to emphasize that only the first n variables, encoded in x, enter in the integration.

Consider the domain $\Omega \subset \mathbb{R}^3$ given by

$$\Omega = \Big\{(x, y, z) \in \mathbb{R}^3 \,\Big|\, z \geq 0\Big\}.$$

Then, as in the previous example, $\partial\Omega$ is the xy-plane,

$$\partial\Omega = \Big\{(x, y, z) \in \mathbb{R}^3 \,\Big|\, z = 0\Big\}.$$

Given a function $f : \Omega \to \mathbb{R}$, one naturally gets a function defined on $\partial\Omega$ by simply setting $z = 0$, i.e., the function $f(x, y, 0)$ is a function defined on the boundary $\partial\Omega$. The same idea works for a general domain, as we next explain.

A function $f : \overline{\Omega} \to \mathbb{R}$ naturally defines a function on $\partial\Omega$, called the **restriction** of f to the boundary, denoted by $f|_{\partial\Omega}$, and given by

$$f|_{\partial\Omega}(x) = f(x), \text{ for } x \in \partial\Omega.$$

From this it follows that we can also integrate a function defined on $\overline{\Omega}$ over the boundary , i.e.,

$$\int_{\partial\Omega} f$$

is well-defined.

Another notion that needs to be recalled is that of the *normal derivative*. Given a domain $\partial\Omega$ with smooth boundary and $x \in \partial\Omega$, the **normal vector** to $\partial\Omega$ at x, denoted[13] $\nu(x)$, or ν_x or yet simply ν, is defined as the vector based at x that has unit length, is perpendicular to the tangent plane to $\partial\Omega$ at x, and points towards the "outside" of Ω.

For example, if Ω is the domain

$$\Omega = \Big\{(x, y) \in \mathbb{R}^2 \,\Big|\, y > 0\Big\},$$

then

$$\partial\Omega = \Big\{(x, y) \in \mathbb{R}^2 \,\Big|\, y = 0\Big\},$$

(i.e., $\partial\Omega$ is simply the x-axis) and for any $(x, 0) \in \partial\Omega$, the normal is given by $\nu = (0, -1)$.

The normal to the $n-1$-dimensional sphere S_r at x is given by $\nu = \frac{1}{r}x$. The reader should also check that the normal to $S_r(x_0)$ at x is given by $\nu = \frac{1}{r}(x - x_0)$.

The **normal derivative** of a function $f : \overline{\Omega} \to \mathbb{R}$ is a function on $\partial\Omega$, denoted $\frac{\partial f}{\partial \nu}$ or $\partial_\nu f$, and defined by

$$\frac{\partial f}{\partial \nu} = \langle \nabla f, \nu \rangle, \text{ on } \partial\Omega,$$

or more explicitly,

$$\frac{\partial f}{\partial \nu}(x) = \langle \nabla f(x), \nu(x) \rangle, \, x \in \partial\Omega.$$

The **Laplacian** of a function $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$, denoted Δf, is defined as

$$\Delta f = \frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2} + \cdots + \frac{\partial^2 f}{\partial x_n^2}.$$

[13] Calculus books usually denote the normal vector by N, n, $\vec{N}$, or $\vec{n}$.

The above can be written in several equivalent ways, e.g.,

$$\begin{aligned}\Delta f &= \frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2} + \cdots + \frac{\partial^2 f}{\partial x_n^2} \\ &= \sum_{i=1}^{n} \frac{\partial^2 f}{\partial x_i^2} \\ &= \partial_{11}^2 f + \partial_{22}^2 f + \cdots + \partial_{nn}^2 f \\ &= \sum_{i=1}^{n} \partial_{ii}^2 f \\ &= f_{11} + f_{22} + \cdots + f_{nn} \\ &= \sum_{i=1}^{n} f_{ii} \\ &= f_{x_1x_1} + f_{x_2x_2} + \cdots + f_{x_nx_n} \\ &= \sum_{i=1}^{n} f_{x_ix_i}.\end{aligned}$$

With the above definitions at hand, we can now recall the following **Green's identities:**

$$\int_\Omega \langle \nabla f, \nabla g \rangle = -\int_\Omega f \Delta g + \int_{\partial\Omega} f \frac{\partial g}{\partial \nu},$$

and

$$\int_\Omega (g\Delta f - f\Delta g) = \int_\Omega \left(g\frac{\partial f}{\partial \nu} - f\frac{\partial g}{\partial \nu} \right),$$

where $f, g : \overline{\Omega} \subset \mathbb{R}^n \to \mathbb{R}$. These two formulas can be derived from the formula for **integration by parts** in n-variables:

$$\int_\Omega g\partial_i f = -\int_\Omega f\partial_i g + \int_{\partial\Omega} gf\nu_i,$$

where ν_i is the i^{th} component of the normal vector $\nu = (\nu_1, \nu_2, \ldots, \nu_n)$.

8. Quantifiers and the formation of mathematical sentences

Here we introduce some mathematical symbols that are quite useful to make shorthand notation. We also give some examples of their use, while making some general remarks about the form of certain mathematical statements. Once more, we emphasize that our discussion is informal and lacks proper mathematical rigor[14].

Below is a list of mathematical symbols that are often used, along with their interpretation:

Symbol	Reads as
$\forall$	for all
$\exists$	there exists
$\Rightarrow$	if... then
$\Leftrightarrow$	if and only if

Let us see some examples of how such symbols are employed, and also of how to structure mathematical statements using them.

[14]The reader interested in a thorough discussion can consult [1, 4].

For example,

$$x > 1 \Rightarrow x > 0$$

reads "if x is greater than one, then it is greater than zero." While this statement is true (a number that is greater than one is also greater than zero), the correct use of $\Rightarrow$ has nothing to do with whether the sentence is in fact true. In other words, the sentence $x > 1 \Rightarrow x > 0$ draws a conclusion about x, namely, that $x > 0$, under the assumption that $x > 1$. But whether or not the x in question is in fact greater than one is completely open. Thus,

$$x > 0 \Rightarrow x > 1$$

is also a correct use of the symbol $\Rightarrow$, except that now the sentence is false: from the knowledge that $x > 0$, it cannot be concluded that $x > 1$, since there are numbers that are greater than zero but are not greater than one. Summing up, $\Rightarrow$ is used in form

$$\text{Claim 1} \Rightarrow \text{Claim 2}$$

to state that *if* Claim 1 is true, *then* Claim 2 must also be true. Whether or not Claim 1 is in fact true is not addressed. As a colloquial analogy, the reader can imagine a sentence like "If it rains, then the floor gets wet." It does not say anything about the actual status of the weather, i.e., whether it is raining or not.

Also, the full statement "if .... then ..." does not have to be true in order to carry a correct use of $\Rightarrow$. Although we shall not define what is meant by "correct use," the idea is that it is employed in a sentence that makes sense. The statement $x > 0 \Rightarrow x > 1$ makes sense, i.e., we can read and understand it, even though it is a wrong statement (in fact, had we been unable to understand what it says, we would not even be capable of saying whether it is right or wrong). As another colloquial analogy, consider the statement "If I drink clean water, then I will get very sick." This is a correct use of "if" and "then": the sentence is well-formed, it makes sense, and it is grammatically correct — despite the fact that its content is false.

One important thing to keep in mind is that even if Claim 1 $\Rightarrow$ Claim 2 is true, knowing that Claim 2 is true does not say anything about Claim 1. For instance, consider the statement: *if* a (differentiable) function f has a local maximum at a, then $f'(a) = 0$. This statement is true, as you learned in calculus. However, knowing that $f'(a) = 0$ does not give information about the nature of the point a; it could be a local maximum, a local minimum, or neither (e.g., the derivative of x^3 at zero is zero, but zero is neither a local maximum nor a local minimum). Using once more an analogy, even if the statement "If it rains, the floor gets wet" is true, we cannot conclude that it had rained by noticing that the floor is wet.

$\exists$ is used to indicate that there is at least one element in a set satisfying a certain property. For instance,

$$\exists x \in \mathbb{R},\ x^2 - 1 = 0$$

reads, "there exists an element x in the set of real numbers such that $x^2 - 1 = 0$." Notice that $\exists$ does not say that "there exists only one." For instance, the above statement is true because $x = 1$ satisfies $x^2 - 1 = 0$, but $x = -1$ also satisfies the equation. Had we claimed that there existed only one element satisfying $x^2 - 1 = 0$, then the statement would have been false.

As it happened for $\Rightarrow$, the right use of $\exists$ has nothing to do with whether the statement is correct or not. For instance,

$$\exists x \in \mathbb{R},\ x^2 + 1 = 0$$

is false, since there is no real number x satisfying the equation $x^2 + 1 = 0$. However, this was a perfect legitimate use of $\exists$, as the statement $\exists x \in \mathbb{R},\ x^2 + 1 = 0$, while false, is something that makes sense, in the sense that we can perfectly understand what is being claimed.

$\forall$ is used to indicate that certain statement being made is to be applied for all elements under consideration. For instance,

$$\forall x \in (-1, \infty),\ x + 1 > 0$$

reads "for all x belonging to $(-1, \infty)$, $x+1$ is greater than zero." We sometimes find more convenient to state this in the reverse order[15], i.e.,

$$x + 1 > 0\ \forall x \in (-1, \infty),$$

reading "$x + 1$ is greater than zero for all x belonging to $(-1, \infty)$." We use "for any" as a synonym of $\forall$, so we could also have said "$x+1$ is greater than zero for any x belonging to $(-1, \infty)$." As in the previous examples, notice that the correct use of $\forall$ has nothing to do with whether the statement being made is true or false.

Finally, $\Leftrightarrow$ is used as follows:

$$\text{Claim 1} \Leftrightarrow \text{Claim 2}$$

means that Claim 1 $\Rightarrow$ Claim 2 *and*, reciprocally, Claim 2 $\Rightarrow$ Claim 1. For instance,

$$x^2 = 0 \Leftrightarrow x = 0$$

reads "x^2 is equal to zero if and only if x is equal to zero," and it encodes two statements: that if $x^2 = 0$ then $x = 0$, and also that if $x = 0$ then $x^2 = 0$. Once more, we remark that while the above statement is true, the correct use of $\Leftrightarrow$ is independent of this.

References

[1] Paul R. Halmos. *Naive set theory*. Undergraduate Texts in Mathematics. Springer-Verlag, New York-Heidelberg, 1974. Reprint of the 1960 edition.

[2] James R. Munkres. *Topology*. Prentice Hall, Inc., Upper Saddle River, NJ, 2000. Second edition of [MR0464128].

[3] Yehuda Pinchover and Jacob Rubinstein. *An introduction to partial differential equations*. Cambridge University Press, Cambridge, 2005.

[4] Alfred Tarski. *Introduction to logic and to the methodology of deductive sciences*. Dover Publications, Inc., New York, 1995. Translated from the 1936 Polish original by Olaf Helmer, Reprint of the 1946 translation.

[15] Which, the reader should notice, does not change the meaning of what is being said.