AMATH 731: Applied Functional Analysis Fall 2018

Some important results from real analysis

Many basic results from real analysis will be important in this course, not only in their own right, but also because of their analogues in metric spaces (e.g., convergence, Cauchy convergence). In what follows, we summarize some of these basic and important results. Much of this section follows the presentation of background material in Chapter 1 of the book, *Functional Analysis: Applications in Mechanics and Inverse Problems*, Second Edition, by L.P. Lebedev, I.I. Vorovich and G.M.L. Gladwell (Kluwer 2002). Another helpful source, in particular for the discussion on the construction of the real numbers, was *Analysis By Its History*, by E. Hairer and G. Wanner (Springer 1996).

Caution! In an effort to limit the length of this review, some – but not all – portions of this section are presented rather dryly, with little discussion, explanation, or motivation, i.e., "Theorem," "Theorem," "Definition," "Theorem," "Remark," "Theorem," etc.. Reader discretion – and tolerance – is advised.

One other point that should be mentioned: In no way do we pretend that the following discussion is complete – we simply present some of the most important results from real analysis. Furthermore, the presentation is not perfectly ordered, i.e. it does not necessarily follow a logical progression of concepts (especially with regard to the idea of "sets") as would be done in a formal course on real analysis. This should not be, however, a serious drawback.

Let's start with one of the simplest results of real analysis, the *triangle inequality*:

$$|x+y| \ \leq |x| \ + \ |y|, \quad x, y \in \mathbf{R}\,. \tag{1}$$

A slight modification produces one of the most fundamental results in analysis (and probably one of the most often employed results, when you include its generalizations/analogues in other spaces). First replace y with $-y$,

$$|x-y| \ \leq |x| \ + \ |y|, \quad x, y \in \mathbf{R}\,. \tag{2}$$

(since $|y| = |-y|$) and replace x and y with $x-z$, $y-z$ for any $z \in \mathbf{R}$ to obtain

$$|(x-z)-(y-z)| \ \leq |x-z| \ + \ |y-z|, \quad x, y, z \in \mathbf{R}\,, \tag{3}$$

which reduces to

$$|x-y| \ \leq \ |x-z| \ + \ |z-y|. \tag{4}$$

Keeping in mind that $|x-y|$ measures the distance between x and y on the real line, the above inequality may be interpreted as follows:

> The distance between any two points x and y is less than the sum of their respective distances to a third point z.

Of course, we know that this property is true for points $x, y \in \mathbf{R}^N$ in the case of the Euclidean distance in $\mathbf{R}^N$. In general, however, Eq. (4) expresses of the fundamental properties of a *metric* or distance function between elements of a *metric space*, one of the topics of this course (which you have most probably seen in an earlier course). In this context, Eq. (2) is referred to as the *triangle inequality*.

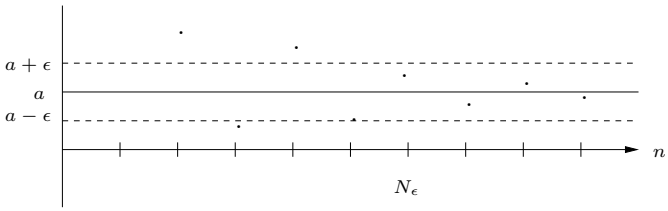
There is actually something even deeper here. Eq. (1) represents a fundamental property of the *norm*, $|x|$, which characterizes the magnitude of a real number. In a *normed vector space*, e.g., the real line $\mathbf{R}$ (and $\mathbf{R}^N$), we can use the norm to define a distance, between two elements of the space. We're very much used to this idea because of our acquaintance with the spaces $\mathbf{R}^N$. But it also applies to other normed spaces, for example, spaces of functions, as we'll see in this course.

1 Convergence and Cauchy sequences

Definition 1 *(Convergence of a sequence to a limit; D'Alembert 1765, Cauchy 1821) The (infinite) sequence of real numbers $x_1, x_2, \cdots$, which shall be denoted as $\{x_n\}$, is said to converge to (the limit) a if, given any $\epsilon > 0$, there exists an integer $N_\epsilon > 0$ (which generally depends on ϵ) such that*

$$|x_n - a| < \epsilon \quad \textit{for all} \;\; n > N_\epsilon \,. \tag{5}$$

Remark: Sometimes, the phrase "to (the limit) a" is omitted in discussions, e.g., "Let $\{x_n\}$ be a convergent sequence." Whenever a sequence is said to "converge" or be "convergent", the existence of a limit (in the appropriate set) is understood.



Graphical representation of the "ϵ-N_ϵ" definition of "$\lim_{n\to\infty} x_n = a$."

Working with the mathematical statement of convergence and its converse

We very accustomed to the above limit of convergence/limit, since our exposure to the "ϵ-δ" idea of limits goes back to our first course in Calculus. For example, one can easily use the above definition, i.e. find an N_ϵ, to show that the sequence,

$$x_n = \frac{1}{n}\,, \quad n = 1, 2, \cdots\,, \tag{6}$$

converges to the limit $a = 0$:

If $\lim_{n\to\infty} \frac{1}{n} = 0$, then given an $\epsilon > 0$, there exists an $N_\epsilon > 0$ such that

$$\left|\frac{1}{n} - 0\right| < \epsilon \quad \text{for all} \quad n > N_\epsilon\,. \tag{7}$$

But the first inequality may be rewritten as follows (using the fact that $n > 0$),

$$\left|\frac{1}{n}\right| = \frac{1}{n} < \epsilon \quad \Longrightarrow \quad n > \frac{1}{\epsilon}\,. \tag{8}$$

So, like can we, like, find an N_ϵ so that, like,

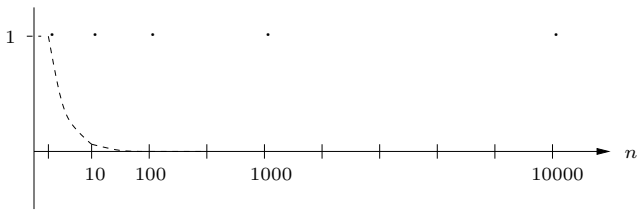
$$n > \frac{1}{\epsilon} \quad \text{for all} \quad n > N_\epsilon\,? \tag{9}$$

Like, yes! We have that $N_\epsilon = \frac{1}{\epsilon}$. (Oh, if life were always this simple!)

Clearly, the "ϵ-N_ϵ" definition of convergence is a strong one, placing a very strict requirement on a sequence $\{x_n\}$. Perhaps too strict? For example, what about the slightly modified sequence,

$$x_n = \begin{cases} \frac{1}{n}\,, & n \in \{1, 2, 3, \cdots\} - \{10, 10^2, 10^3, \cdots\} \\ 1 & n \in \{10, 10^2, 10^3, \cdots\} \quad ? \end{cases} \tag{10}$$

A rough graphical depiction of the nature of this sequence is given in the figure below.



In some situations, might not one be willing to overlook the increasingly sparse "spiking" of the x_n to the value 1 and, perhaps with some "gulping", state that the above sequence "converges," in some sense, to the limit 0 ? Such modifications of the nature of convergence do exist (for example, the so-called "Cesaro limit") but we won't consider them here. We shall continue to work with the strict mathematical definition of limit.

But that being said, we would probably be very quick to dismiss the above sequence as not being convergent. It doesn't look like it has a limit. But that's not a proof! How do we show this suspected nonconvergence mathematically?

Remark: It is very instructive to perform this exercise, since the determination of the negations of mathematical statements is an important tool in constructing proofs. Unfortunately, this subject often receives rather small attention in courses, thereby contributing to the frustration of students when they have to construct proofs.

So, like, what is the negation of the mathematical statement that a sequence $\{x_n\}$ has a limit a? (And please don't reply, "The sequence $\{x_n\}$ doesn't have a limit a.")

Mathematical statement of $\lim_{n\to\infty} x_n = a$:

Given an $\epsilon > 0$ (which really means "given any $\epsilon > 0$" or "for any $\epsilon > 0$"), there exists an $N_\epsilon > 0$ such that

$$|x_n - a| < \epsilon \quad \text{for all } (n \text{ such that}) \quad n > N_\epsilon\,. \tag{11}$$

Before we proceed to construct the negation of the above statement, let's go back to the crazy "spiking" sequence above. In terms of ϵ's and N_ϵ's, why do we think that it doesn't have a limit? After all, if we choose $\epsilon = 2$, then the elements x_n lie entirely inside the interval $[-2, 2]$ for all $n > 0$. The same for $\epsilon = 1.5$. But when we get to $\epsilon = 1$, we're in trouble. And just to avoid any complications with strict inequalities vs. equalities, we have the same problem when $\epsilon = \frac{1}{2}$. Do we keep having to consider other ϵ values, e.g., $\epsilon = \frac{1}{4}$? $\epsilon = \frac{1}{5}$? $\frac{\pi}{14}$? The answer is, fortunately, NO! All we have to do is to produce ONE example of an $\epsilon > 0$ for which the mathematical statement of the limit is FALSE. We've found one, e.g., $\epsilon = \frac{1}{2}$.

Returning to the statement in (11), we start writing,

"There exists an $\epsilon > 0$"

Negating the next part of the statement may appear somewhat complicated. Do we simply write "there exists no $N_\epsilon > 0$ such that ..."? It seems that this is the case for the "spiky sequence" example, i.e.,

"there exists no $N_\epsilon > 0$ such that $|x_n - a| < \epsilon$ for all n such that $n > N_\epsilon$."

This seems to explain the problem with the "spiky sequence." But it may be beneficial to actually state why the above statement is true, i.e., why there exists no such N_ϵ. Clearly, the problem lies with the fact that the above inequality is violated at an infinity of n values, i.e.,

"there exists a sequence of positive integers, $\{n_k\}$, with $n_k \to \infty$ as $k \to \infty$, such that

$$|x_{n_k} - a| \geq \epsilon \quad \text{for all} \quad k\,." \tag{12}$$

It seems that we've lost our N_ϵ, however. But we can recover it by rewriting the inequality in (12) in terms of not a single N_ϵ but an infinity of them. "There exists an $N_\epsilon > 0$ such that ..." will be negated as follows,

For all $N_\epsilon > 0$, there exists an $n > N_\epsilon$ such that

$$|x_n - a| \geq \epsilon\,. \tag{13}$$

We have arrived at the negation of the mathematical statement of the existence of a limit of a sequence:

There exists an $\epsilon > 0$ such that for any $N_\epsilon > 0$, there exists an $n > N_\epsilon$ such that

$$|x_n - a| \geq \epsilon \,. \tag{14}$$

(Technically, however, we've negated the statement that the limit of the sequence is a.)

Informally, the above states that there exists an $\epsilon > 0$ for which spiking – with deviation of at least ϵ from the value a – occurs over the tail of the sequence $\{x_n\}$ an infinite number of times.

Problem 1 *Show that a convergent sequence $\{x_n\}$ (i.e., one that has a limit), has a unique limit, i.e., it cannot converge to two different limits.*

Idea of Proof: Assume that two distinct limits, i.e., $a_1 \neq a_2$ exist and establish a contradiction. (You most probably did this in first-year Calculus to prove the uniqueness of the limit L of a function $f(x)$ at a.)

Cauchy sequences

The result in Definition 1 is fine, if you happen to know the limit of the sequence. But what if you don't? Is there any hope of establishing the convergence of a sequence from some knowledge about its behaviour? As probably everyone knows, it is not sufficient that the consecutive elements x_n and x_{n+1} of a sequence get closer to each other, i.e.,

$$\lim_{n\to\infty} |x_n - x_{n+1}| = 0 \,. \tag{15}$$

To illustrate, consider the sequence $\{S_n\}$ of partial sums of the harmonic series,

$$S_n = \sum_{k=1}^{n} \frac{1}{k}, \quad n \geq 1 \,. \tag{16}$$

Then

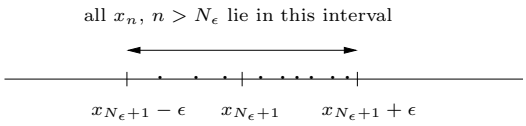
$$S_{n+1} - S_n = \frac{1}{n+1} \to 0 \quad \text{as} \quad n \to \infty \,. \tag{17}$$

The terms S_n and S_{n+1} are getting closer and closer to each other. However, as we know from Calculus, the harmonic series diverges, i.e., $S_n \to \infty$ as $n \to \infty$.

Cauchy struggled with the problem of establishing the convergence of a sequence and came up with the following definition.

Definition 2 *(Cauchy sequence; Cauchy 1821) A sequence $\{x_n\}$ is said to be a Cauchy sequence if, given any $\epsilon > 0$, there exists an $N_\epsilon > 0$ such that*

$$|x_n - x_m| < \epsilon \quad \textit{for all} \;\; m, n > N_\epsilon \,. \tag{18}$$



As in the case of the definition of the limit, this is a strong requirement on the elements of a sequence. Given an $\epsilon > 0$, Eq. (18) is true for the particular case $m = N_\epsilon + 1$, i.e.,

$$|x_n - x_{N_\epsilon+1}| < \epsilon \quad \text{for all} \quad n > N_\epsilon \,. \tag{19}$$

In other words, the distances between $x_{N_\epsilon+1}$ and ALL elements of the "tail" of the sequence $\{x_n\}$, i.e., x_n, $n > N_\epsilon$, are less than ϵ, as sketched in the figure above.

Theorem 1 *All convergent sequences are Cauchy sequences.*

A somewhat annotated proof: Let $\{x_n\} \subset \mathbf{R}$ be a convergent sequence. We'd like to show that for any $\epsilon > 0$, there exists an N_ϵ such that

$$|x_n - x_m| < \epsilon \quad \text{for all} \quad n, m > N_\epsilon \,. \tag{20}$$

We're given that the sequence $\{x_n\}$ is convergent, that is, it has a limit $a \in \mathbf{R}$. From the definition of limit, it follows that for any $\delta > 0$, there exists an $N_\delta > 0$ such that

$$|x_n - a| < \delta \quad \text{for all} \quad n > N_\delta \,. \tag{21}$$

We now try to "import" this information into the Cauchy sequence inequality. Once again, if in doubt, try the triangle inequality, i.e.,

$$\begin{aligned} |x_n - x_m| &= |(x_n - a) - (x_m - a)| \\ &\leq |x_n - a| + |x_m - a| \\ &< 2\delta \quad \text{for all} \quad n, m > N_\delta \,. \end{aligned} \tag{22}$$

We have arrived at the desired result: Setting $\epsilon = 2\delta$ and $N_\epsilon = N_\delta$, we have

$$|x_n - x_m| < \epsilon \quad \text{for all} \quad n, m > N_\epsilon \,, \tag{23}$$

which proves that the sequence is Cauchy.

Cauchy went further and proved the following result.

Theorem 2 *(Cauchy 1821) A sequence $\{x_n\}$ of real numbers is convergent (with a real number as limit) if and only if it is a Cauchy sequence.*

Returning to the previous figure, the limit a of the Cauchy sequence $\{x_n\}$ must lie in the interval $[x_{N_\epsilon+1} - \epsilon, x_{N_\epsilon+1} + \epsilon]$.

Note that the above theorem applies to Cauchy sequences of real numbers. If we restrict our attention to "incomplete" sets, the statement "All convergent sequences are Cauchy sequences" is still true, since the existence of a limit is assumed. However, the converse, "All Cauchy sequences are convergent sequences" is false, since we are not guaranteed the existence of a limit. A "cheap" example is provided if we restrict our attention to the set $S = (0, 1)$. Then the sequence $x_n = \frac{1}{n}$ is Cauchy but it does not have a limit in the set S. Of course, we can "complete" or "close" the set S to include this and all other limit points of Cauchy sequences, hence ensuring that all Cauchy sequences are convergent. In this case, the "closure" of S is the set $\bar{S} = [0, 1]$.

Another more fundamental example, which will be useful for a later discussion on the completion of metric spaces, is provided if we restrict our attention to the set $S = \mathbf{Q}$, the set of rational numbers. This was, in fact, important in the construction of the of real number line. The sequence of rational numbers formed by truncating the infinite decimal expansion of $\sqrt{2}$, i.e.,

$$x_1 = 1, \quad x_2 = 1.4 = \frac{14}{10}, \quad x_3 = 1.41 = \frac{141}{100}, \quad \cdots , \tag{24}$$

is Cauchy, but it is not convergent on $\mathbf{Q}$. Once again, we can "complete" or "close up" the set $\mathbf{Q}$ to form the real line $\mathbf{R}$ so that all Cauchy sequences are convergent. An important idea in the "completion" of such incomplete spaces is that of *equivalent* sequences:

Definition 3 *Two Cauchy sequences $\{x_n\}$ and $\{y_n\}$ are said to be* **equivalent***, written as*

$$\{x_n\} \equiv \{y_n\} ,$$

if $\lim_{n\to\infty} |x_n - y_n| = 0$.

Note: This does *not* imply that either or both of the sequences have limits. For example, consider the sequence $\{y_n\}$ defined on $\mathbf{Q}$ as follows,

$$y_1 = 1 , \qquad y_{n+1} = \frac{y_n}{2} + \frac{1}{y_n} \quad n \geq 1 . \tag{25}$$

The first few elements of this sequence are given by

$$y_1 = 1, \quad y_2 = \frac{3}{2}, \quad y_3 = \frac{17}{12}, \quad y_4 = \frac{577}{408}, \quad \cdots . \tag{26}$$

We state here, without proof, the following:

- The sequence $\{y_n\}$ is Cauchy.
- With reference to the sequence $\{x_n\}$ in (24), we have that

  $$\lim_{n\to\infty} |x_n - y_n| = 0 . \tag{27}$$

  From this result, it follows that the sequences $\{x_n\}$ an $\{y_n\}$ in (24) and (26) are equivalent.
- Both sequences $\{x_n\}$ and $\{y_n\}$ converge to the real number $\sqrt{2}$ which is not in $\mathbf{Q}$. From this, we may conclude that neither sequence is convergent in $\mathbf{Q}$.

Problem 2 *Show that the sequence $\{y_n\}$ defined above arises from the application of Newton's method to the polynomial $f(x) = x^2 - 2$. (Newton's method will be examined in more detail in this course.)*

From the definition of equivalence, it is quite straightforward to show the following:

Problem 3 *Show that (a) $\{x_n\} \equiv \{x_n\}$ (reflexivity), (b) if $\{x_n\} \equiv \{y_n\}$ then $\{y_n\} \equiv \{x_n\}$(symmetry), (c) if $\{x_n\} \equiv \{y_n\}$ and $\{y_n\} \equiv \{z_n\}$, then $\{x_n\} \equiv \{z_n\}$ (transitivity).*

From this result, it is possible to partition the set of Cauchy sequences into *equivalence classes.*

Definition 4 *Associated with the sequence $\{x_n\}$ is the* **equivalence class***, denoted $\bar{x}$ and defined as follows,*

$$\bar{x} = \left\{ \{y_n\} \;\middle|\; \{y_n\} \text{ is a Cauchy sequence and } \{y_n\} \equiv \{x_n\} \right\}. \tag{28}$$

A particular Cauchy sequence $\{x_n\}$ in an equivalence class $\bar{x}$ is called a **representative** *of that class.*

Equivalence classes divide all Cauchy sequences into separate groups: **A sequence $\{x_n\}$ cannot belong to two different equivalence classes.** This is a consequence of the definition of equivalent sequences.

If $\{x_n\}$ and $\{y_n\}$ belong to **different** equivalence classes, $\bar{x}$ and $\bar{y}$, respectively, then they are not equivalent, which implies that the statement $\lim_{n\to\infty} |x_n - y_n| = 0$ is not true. Therefore (by negation of this statement), there is an $\epsilon > 0$ such that for any $N > 0$, there exists an $n > N$ such that $|x_n - y_n| \geq \epsilon$.

This implies that two different equivalence classes are **separated** from each other in the sense stated above. Using the definition of Cauchy sequences, it can then be shown that if $\{x_n\}$ and $\{y_n\}$ belong to difference equivalence classes, there is an $N > 0$ such that for all $m, n > N$, either $x_n < y_m$ or $x_n > y_m$. In the former case, we shall write "$\bar{x} < \bar{y}$" and in the latter case, "$\bar{x} > \bar{y}$". This implies the equivalence classes, like (rational) numbers can be **ordered**, i.e., if $\bar{x}$ and $\bar{y}$ are two classes, then either $\bar{x} < \bar{y}$ or $\bar{x} > \bar{y}$ or $\bar{x} = \bar{y}$. (In the case $\bar{x} = \bar{y}$, the classes are equivalent.) This leads to the following result.

Definition 5 *A* **real number** *is an equivalence class of Cauchy sequences of rational numbers, i.e.,*

$$\mathbf{R} = \{\bar{x} \mid \{x_n\} \text{ is a rational Cauchy sequence}\}. \tag{29}$$

The set $\mathbf{Q}$ or rational numbers can be interpreted as a subset of $\mathbf{R}$ as follows: If $r \in \mathbf{Q}$, then it is associated with the constant, or *stationary*, rational Cauchy sequence $\{r, r, \cdots\}$. Thus the rational number r is identified with the real number $\overline{\{r, r, \cdots\}}$.

There can be, of course, other rational sequences that are equivalent to r, e.g., the sequence

$$x_n = \frac{n}{n+1} r \in \mathbf{Q}\,.$$

There are a number of additional results involving Cauchy sequences - specifically representatives of equivalence classes - that allow us to treat real numbers in the same way as rational numbers, i.e., they can be added, subtracted multiplied and divided, as well as *ordered.*

The final result, which implies the *completeness* of the of real numbers $\mathbf{R}$, involves the idea of Cauchy sequences $\{s_n\}$ of *real numbers.* It has already been expressed in Theorem 2.

2 Sets of points in R

Here we consider briefly some important properties of sets of points in $\mathbf{R}$, i.e., subsets of the real line $\mathbf{R}$. Typically, the most interesting cases are infinite sets, i.e., sets with an infinite number of elements.

Definition 6 *An* **open interval** (a, b) *on* $\mathbf{R}$ *is a set of numbers* x *satisfying* $a < x < b$. *An* **closed interval** $[a, b]$ *on* $\mathbf{R}$ *is a set of numbers* x *satisfying* $a \leq x \leq b$.

The main point regarding an open interval (a, b) is not that the endpoints a and b are excluded, but rather that any point $x \in (a, b)$ is itself the *center* of an open interval, say (c, d), that lies entirely in (a, b).

Definition 7 *A set* $S \subset \mathbf{R}$ *is* **open** *if every point in* S *is the center of an open interval lying entirely in* S.

Alternate definition: A set $S \subset \mathbf{R}$ is **open** if for every point $x \in X$, there exists an $\epsilon > 0$ such that the (open) interval $(x - \epsilon, x + \epsilon)$ lies entirely in S.

Definition 8 *A set* $S \subset \mathbf{R}$ *is* **closed** *if every convergent sequence* $\{x_n\} \subset S$ *converges to a point in* S.

Problem 4 *Show that if* S *is a closed set in* $\mathbf{R}$, *then every Cauchy sequence* $\{x_n\} \subset S$ *converges to a point* $x \in S$.

Definition 9 *A set* $S \subset \mathbf{R}$ *is said to be* **bounded** *if there is a number* $M \geq 0$ *such that all* $x \in S$ *satisfy the relation* $|x| \leq M$.

Note: The above implies that $S \subset [-M, M]$.

Theorem 3 *A Cauchy sequence* $\{x_n\} \subset \mathbf{R}$ *is bounded.*

Proof 1: Let $\{x_n\} \subset \mathbf{R}$ be a Cauchy sequence. From Cauchy's Theorem, it is convergent, i.e., it has a limit $a \in \mathbf{R}$. Therefore, given an $\epsilon > 0$, there exists an $N_\epsilon > 0$ such that

$$|x_n - a| < \epsilon \quad \text{for all} \quad n > N_\epsilon \,. \tag{30}$$

For such a fixed $\epsilon > 0$, the above inequality implies that $x_n \in [a - \epsilon, a + \epsilon]$ for all $n > N_\epsilon$. This implies that for $n > N_\epsilon$, $|x_n| < M_1$, where $M_1 = \max\{|a - \epsilon|, |a + \epsilon|\} > 0$. Now let

$$M_2 = \max\{|x_1|, |x_2|, \cdots, |x_{N_\epsilon}|\} \,. \tag{31}$$

It then follows that $|x_n| \leq M$, where $M = \max\{M_1, M_2\}$. Therefore, the sequence is bounded and the proof is complete.

Remark: Here is a "nicer" proof which does not rely on the fact that the Cauchy sequence has a limit a. Such a proof can be used in more general cases where it may not be true that all Cauchy

sequences are convergent. (More on this later.)

Proof 2: Recall from the definition of a Cauchy sequence (Definition 2 above), that for a given an $\epsilon > 0$, it follows that $x_n \in [x_{N_\epsilon+1} - \epsilon, x_{N_\epsilon+1} + \epsilon]$. (Here, we have set $m = N_\epsilon + 1$.) This implies that for all $n > N_\epsilon$, $|x_n| < M_1$, where $M_1 = \max\{|x_{N_\epsilon+1} - \epsilon|, |x_{N_\epsilon+1} + \epsilon|\}$. Now let M_2 be defined as in Proof 1 and complete the proof.

Problem 5 *Suppose that the sequence $\{x_n\} \subset \mathbf{R}$ converges to x. Show that any subsequence $\{x_{n_k}\}$ of $\{x_n\}$ converges to x. Conversely, show that if $\{x_n\} \subset \mathbf{R}$ is a convergent sequence, and a subsequence $\{x_{n_k}\}$ converges to x, then $\{x_n\}$ must converge to x.*

Definition 10 *A set $S \subset \mathbf{R}$ is said to be* **compact** *if every sequence $\{x_n\} \subset S$ contains a subsequence that converges to a point $x \in S$.*

Theorem 4 *(Bolzano-Weierstrass (BW) Theorem) A set $S \subset \mathbf{R}$ is compact iff it is closed and bounded.*

Problem 6 *Prove the above theorem. First prove that if S is compact, then it is closed and bounded. Then prove that if S is closed and bounded, then it is compact. (Hint: For the latter, use a method of bisection. Bisect the interval $I = [-M, M]$ into two closed subintervals. One of these subintervals must contain an infinite number of points. Repeat the procedure ...)*

The BW theorem applies to a set which is closed and bounded. What about a set S that is just bounded? It may not be closed, but we can "close it" as follows:

Definition 11 *The* **closure** *of a set S, denoted as $\overline{S}$, is the set obtained by adding to S all limit points of all convergent sequences $\{x_n\} \subset S$.*

By construction $\overline{S}$ is closed.

Simple examples:

1. If $S = (0, 1)$, then $\overline{S} = [0, 1]$.
2. If $S = [0, 1] \cap \mathbf{Q} = \{x \in [0, 1] \mid x \in \mathbf{Q}\}$, then $S = [0, 1]$.

Finite sets vs. infinite sets

Let $S \subset \mathbf{R}$ be a finite set of real numbers, i.e., $S = \{y_1, y_2, \cdots, y_N\}$. The finite set S has a greatest (maximum) and least (minimum) value, to be denoted as

$$M = \max_{i=1,2,\cdots,N} y_i \quad \text{and} \quad m = \min_{i=1,2,\cdots,N} y_i, \tag{32}$$

respectively. The finite set S is bounded since

$$|y_i| \leq \max\{|M|, |m|\} \quad 1 \leq i \leq N\,. \tag{33}$$

Now consider an *infinite* set $S \subset \mathbf{R}$ of real numbers $x_1, x_2, \cdots$. Even if this set is bounded, it may not have a maximum or minimum value.

Example: The set $S = \{0, 1/2, -1/2, 2/3, -2/3, \cdots\}$. Here, $S \subset [-1, 1]$, so it is bounded. The even-indexed subsequence $\{y_{2n}\}$ is approaching -1 and the odd-indexed subsequence $\{y_{2n-1}\}$ is approaching 1, yet neither of these values belongs to the set.

Supremum of a set $S \subset \mathbf{R}$

The method of bisection used to prove the BW theorem may be employed to show that a set $S \subset \mathbf{R}$ which is bounded above has a *least upper bound* or *supremum*, denoted as

$$M = \sup_{x \in S} x \quad \text{or simply} \quad \sup S. \tag{34}$$

The supremum satisfies the following properties:

1. if $x \in S$, then $x \leq M$
2. if $c < M$, then there exists an $x \in S$ such that $c < x$.
3. There is a sequence $\{y_k\} \subset S$ which converges to M.

Given the importance of the supremum, it is instructive to go through a proof of its existence.

Proof: From the assumption the set $S \subset \mathbf{R}$ is bounded from above, there exists an $B_1 \in \mathbf{R}$ such that $x < B_1$ for all $x \in S$. Now choose any element $A_1 \in S$ and define $y = \frac{1}{2}(A_1 + B_1)$.

1. If $y \geq x$ for all $x \in S$, then define $B_2 = y$ and $A_2 = A_1$.
2. Otherwise, define $A_2 = y$ and $B_2 = B_1$.

Now iterate this procedure: For $n \geq 2$:

1. Define $y = \frac{1}{2}(A_n + B_n)$.
2. If $y \geq x$ for all $x \in S$, then define $B_{n+1} = y$ and $A_{n+1} = A_n$.
3. Otherwise, define $A_{n+1} = y$ and $B_{n+1} = B_n$.

This iterated bisection procedure produces (i) a monotonically increasing set of points, $A_1 \leq A_2 \leq A_3 \cdots$, (ii) a monotonically decreasing set of points, $B_1 \geq B_2 \geq B_3 \cdots$. Furthermore, $A_i \leq B_j$ for all $i, j \geq 1$, i.e.,

$$A_1 \leq A_2 \leq A_3 \leq \quad \cdots \quad \leq B_3 \leq B_2 \leq B_1 \,. \tag{35}$$

Also note that the lengths of the intervals $[A_n, B_n]$ tend to zero as $n \to \infty$ since $|B_n - A_n| = 2^{-n+1}|B_1 - A_1|$. This implies that

$$\lim_{n \to \infty} A_n = \lim_{n \to \infty} B_n = M \ . \tag{36}$$

Note that by construction, for each $x \in S$, $x \leq B_n$ for all $n \geq 1$. We may take the limit of both sides of this inequality to arrive at the result that

$$x \leq M \quad \text{for all} \quad x \in S\,, \tag{37}$$

thus proving the first property of the supremum listed above.

Also note that by construction, every interval $[A_n, B_n]$, $n \geq 1$, contains at least one point $y_n \in S$ so that $A_n \leq y_n \leq B_n$. From Eq. (36), it easily follows that $y_n \to M$ as $n \to \infty$, proving the third property listed above.

To prove the second property listed above, let $c < M$. From Eq. (36), there exists an N such that $c < A_N < M$. From the previous paragraph, the interval $[A_N, B_N]$ contains a point $x \in S$. Therefore $c < x$ and the proof is complete.

Infimum of a set $S \subset \mathbf{R}$

In a similar fashion, a set $S \subset \mathbf{R}$ which is bounded below has a *greatest lower bound* or *infimum*, denoted as

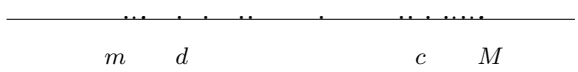
$$m = \inf_{x \in S} x \quad \text{or simply} \quad \inf\ S. \tag{38}$$

The infimum satisfies the following properties:

1. if $x \in S$, then $x \geq m$
2. if $d > m$, then there exists an $x \in S$ such that $x < d$.
3. There is a sequence $\{y_k\} \subset S$ which converges to m.

The existence of an infimum of a set S bounded from below may be proved by means of a bisection procedure similar to that used for the supremum.

It follows that if a set $S \subset \mathbf{R}$ is bounded (i.e., bounded from above and below), then it has both an infimum m and a supremum M, as sketched in the figure below.



The set $S \subseteq [m, M]$

Monotonic sequences

Definition 12 *A sequence* $\{x_n\} \subset \mathbf{R}$ *is said to be* **monotonically increasing (decreasing)** *if* $x_n \leq x_{n+1}$ *(or* $x_n \geq x_{n+1}$*) for* $n = 1, 2, \cdots$. *It is said to be* **strictly monotonic** *if the inequality holds, i.e.,* $x_n < x_{n+1}$ *or* $x_n > x_{n+1}$.

The argument used to prove the B-W theorem may be adapted to prove the following well-known results:

Theorem 5 *A monotonically increasing sequence $\{x_n\} \subset \mathbf{R}$ that is bounded above by b converges to a limit $x \leq b$. Similarly, a monotonically decreasing sequence $\{x_n\} \subset \mathbf{R}$ that is bounded below by a converges to a limit $x \geq a$.*

A final and important comment about the real number line R as an "ordered set"

The concepts of infimum, supremum and monotonically increasing/decreasing for sets $S \subset \mathbf{R}$ rely on a particular property of the real number line $\mathbf{R}$, namely that it is **ordered**. The ordering relation is defined in terms of operations of "greater than" and "less than," which we take for granted when working with the real numbers. When it is stated that $a < b$ for two real numbers $a, b \in \mathbf{R}$, we can immediately picture the situation as the point a lying to the left of b on the real number line $\mathbf{R}$. In this way, the operation "$<$" defines a natural ordering of the real numbers.

This is not the case for sets of higher dimension, e.g., subsets of $\mathbf{R^2}$. Given any two distinct points $(a_1, a_2), (b_1, b_2) \in \mathbf{R^2}$, can we define unambigiously the conditon that one point be "greater than" the other? The answer is in the negative – the set $\mathbf{R^2}$ is not an ordered set. But it is a **partially ordered** set – we can define an ordering involving subsets. We may come back to this idea later in the course.

3 Generalization to sets in $\mathbf{R}^N$

With a little work, the above ideas for closed, open bounded and compact sets on $\mathbf{R}$, as well as the idea of convergence, may be generalized to $\mathbf{R}^N$. (Of course, this entire analysis could be performed conveniently and compactly in the context of **metric spaces**, but that is the subject of a later lecture.)

Definition 13 *The* **Euclidean distance** *between points $x = (x_1, x_2, \cdots, x_n)$ and $y = (y_1, y_2, \cdots, y_n)$ in $\mathbf{R}^N$, to be denoted as $d_E(x, y)$, is given by*

$$d_E(x, y) = \left[\sum_{i=1}^{N} (x_i - y_i)^2 \right]^{1/2}. \tag{39}$$

In the definiton of an open set, *open interval* must be replaced by *open ball*:

Definition 14 *The* **open ball** *with center x_0 and radius $a > 0$ in $\mathbf{R}^N$ is the set*

$$B(x_0, a) = \{x \in \mathbf{R}^N \mid d_E(x, x_0) < a\}. \tag{40}$$

Definition 15 *A set $S \subset \mathbf{R}^N$ is* **open** *if every point of S is the center of an open ball lying entirely in S.*

Definitions involving convergence in $\mathbf{R}^N$ must now employ the Euclidean distance.

Definition 16 *The sequence $\{x_n\} \subset \mathbf{R}^N$* **converges** *to (the limit) a if, given any $\epsilon > 0$, there exists an integer $N_\epsilon > 0$ (which generally depends on ϵ) such that for all $n > N_\epsilon$, $d_E(x_n, a) < \epsilon$.*

Definition 17 *A sequence $\{x_n\} \subset \mathbf{R}^N$ is a* **Cauchy sequence** *if, given any $\epsilon > 0$, there exists an $N_\epsilon > 0$ such that for all $m, n > N_\epsilon$, $d_E(x_m, x_n) < \epsilon$.*

In order to characterize the boundedness of sets, we must consider the magnitude of a point $x \in \mathbf{R}^N$.

Definition 18 *The magnitude or* **norm** *of a point $x = (x_1, x_2, \cdots, x_N) \in \mathbf{R}^N$, denoted by $N(x)$, is given by*

$$N(x) = \left[\sum_{i=1}^{N} x_i^2 \right]^{1/2}. \tag{41}$$

Note that $N(x) = d_E(x, 0)$.

Definition 19 *A set $S \subset \mathbf{R}^N$ is* **bounded** *if there is a number $M \geq 0$ such that all $x \in S$ satisfy the relation $N(x) \leq M$.*

The definitions of closed and compact sets in $\mathbf{R}^N$ follow naturally from the earlier definitions on $\mathbf{R}$. Finally, the Bolzano-Weierstrass theorem holds for sets $S \subset \mathbf{R}^N$:

Theorem 6 *A set $S \subset \mathbf{R}^N$ is compact iff it is closed and bounded.*

4 Functions

In what follows, we let Ω denote a non-empty open set in $\mathbf{R}^N$ and refer to it as a *domain*.

Definition 20 *A rule which assigns a unique real number to every $x \in \Omega$ is said to define a real (or real-valued) function $f(x)$ on Ω. One distinguishes between a function f and its value $f(x)$ at a point $x \in \Omega$.*

Definition 21 *The* **support** *of $f(x)$ on Ω, denoted as* **supp** *f, is defined as*

$$supp\ f = \overline{\{x \in \Omega \mid f(x) \neq 0\}}, \tag{42}$$

where the overbar denotes closure in $\mathbf{R}^N$. The function f is said to have **compact support** *if supp f is* **bounded**. *It is said to have compact support in Ω if supp $f \subset \Omega$.*

Note that, by definition, the support of a function is a closed set.

Definition 22 *Let f be a real function on Ω. Let $x_0 \in \Omega$. The function f is said to be* **continuous at** *x_0 if, given any $\epsilon > 0$, there exists a δ_ϵ (also depending on x_0) such that if $d_E(x, x_0) < \delta_\epsilon$, then $|f(x) - f(x_0)| < \epsilon$. The function f is said to be* **continuous on** *Ω if it is continuous at every $x \in \Omega$. In such a case, we write that $f \in C(\Omega)$, the space of continuous functions on Ω.*

Theorem 7 *A real-valued function that is continuous on a* **closed** *and* **bounded**, *i.e., a* **compact** *region $\overline{\Omega} \subset \mathbf{R}^N$ is bounded. Moreover, it achieves its supremum and infimum in $\overline{\Omega}$.*

Definition 23 *A real-valued function $f \in C(\Omega)$ is said to be* **uniformly continuous** *on Ω if, given any $\epsilon > 0$, we can find a $\delta_\epsilon > 0$ such that if $x, y \in \Omega$ and $d_E(x, y) < \delta_\epsilon$, then $|f(x) - f(y)| < \epsilon$.*

Remark: The difference between f being continuous and uniformly continuous on Ω is that in the latter case, δ is dependent only on ϵ and not on x or y in Ω. In other words, given an $\epsilon > 0$, one δ_ϵ works at all points in Ω.

Theorem 8 *If f is continuous on a* **compact** *region $\overline{\Omega} \subset \mathbf{R}^N$, then it is* **uniformly continuous** *on $\overline{\Omega}$.*

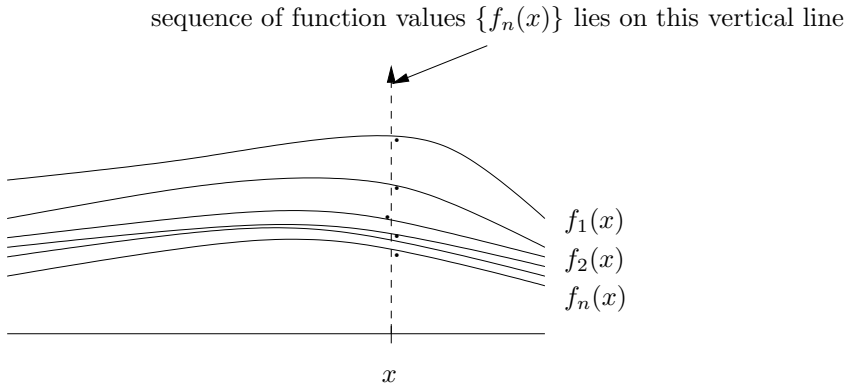
A proof of this theorem is given in the Appendix at the end of this section.

Sequences of functions

Now let $f_1, f_2, \cdots$ be a sequence of functions on $\Omega \subset \mathbf{R}^N$. For any particular $x \in \Omega$, we consider the sequence $\{f_n(x)\} \subset \mathbf{R}$.

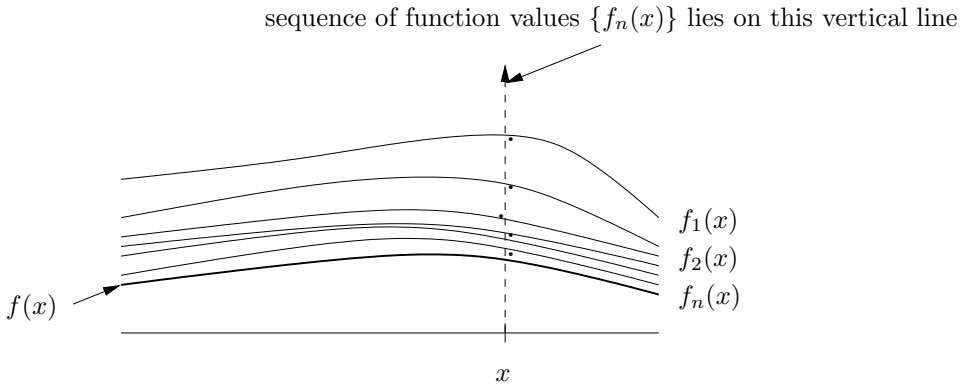
1. For this value of x, the sequence will be a Cauchy sequence if, given an $\epsilon > 0$, there exists an $N_{\epsilon,x}$ (depending on ϵ and x) such that $|f_n(x) - f_m(x)| < \epsilon$ for all $m, n \geq N_{\epsilon,x}$.

   If, for any $\epsilon > 0$, it is possible to find an $N_\epsilon > 0$, depending on ϵ but not on x, which will work for all $x \in \Omega$, then the sequence $\{f_n(x)\}$ is said to be *uniformly Cauchy.*



2. Similarily, for a particular value of x, the sequence $\{f_n(x)\}$ is said to converge to $f(x)$ if, given any $\epsilon > 0$, there exists an $M_{\epsilon,x}$ (depending on ϵ and x) such that $|f_m(x) - f(x)| < \epsilon$ for all $m \geq M_{\epsilon,x}$.

   If, for any $\epsilon > 0$, it is possible to find an M_ϵ, depending on ϵ but not on x, which will work for all $x \in \Omega$, then the sequence $\{f_n(x)\}$ is said to be *uniformly convergent*, or *converge uniformly*, to $f(x)$.



Theorem 9 *(Weierstrass' theorem) A uniformly Cauchy sequence $\{f_n\}$ of functions which are uniformly continuous on a compact region $\overline{\Omega} \subset \mathbf{R}^N$ converges to a uniformly continuous function f.*

Weierstrass polynomial approximation theorem

This very important result deserves separate billing. We'll return to this result later in the course.

Theorem 10 *Any function which is continuous on a closed and bounded region $\overline{\Omega} \subset \mathbf{R}^N$ (hence uniformly continuous on $\overline{\Omega}$) may be uniformly approximated arbitrarily closely by a polynomial.*

Appendix: Proof of Theorem 8

Theorem 8: *If f is continuous on a* **compact** *region $\overline{\Omega} \subset \mathbf{R}^N$, then it is* **uniformly continuous** *on $\overline{\Omega}$.*

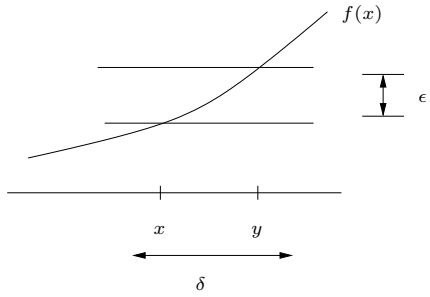
Proof: We shall prove this result by contradiction, assuming that f is continuous on the compact region $\overline{\Omega}$ but that it is **not** uniformly continuous. Let's first recall the definition of uniformly continuous functions (Definition 23 of handout):

> A function $f \in C(\Omega)$ is said to be **uniformly continuous** on Ω if, given any $\epsilon > 0$, we can find a $\delta > 0$ such that if $x, y \in \Omega$ and $d_E(x, y) < \delta$, then $|f(x) - f(y)| < \epsilon$.

If f is not uniformly continuous then we have to negate the above definition. It means that there exists an $\epsilon > 0$ such that for *every* $\delta > 0$, we can find a pair of points $x, y \in \overline{\Omega}$ such that $d_E(x, y) < \delta$ but $|f(x) - f(y)| \geq \epsilon$.

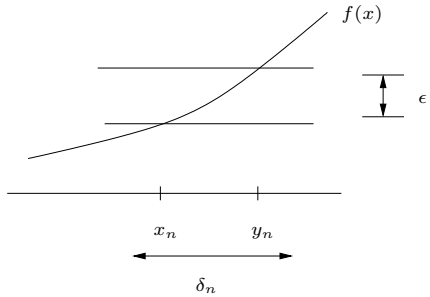
(**Remark:** At first glance, it might appear that the above statement implies that f is not continuous. But this is not the case, since x and y can vary with δ. If the above were true for the same x and y values for all δ, then f would be discontinuous at x and y.)

Continuing the proof, for such an $\epsilon > 0$, let us select a sequence of δ values, $\delta_n = 1/n$, $n = 1, 2, \cdots$. Then, from the previous paragraph, for each δ_n, we can find a pair of points $x_n, y_n \in \overline{\Omega}$ such that



$d_E(x_n, y_n) < \delta_n$ but

$$|f(x_n) - f(y_n)| \geq \epsilon. \tag{43}$$



The following paragraph is purely explanatory, and not necessary in the formal proof of the Theorem. As such, it is enclosed in parentheses.

(Somewhere, the fact that $\overline{\Omega}$ is compact is going to have to play a role, and here it is: Note that both (infinite) sequences $\{x_n\}$ and $\{y_n\}$ lie in the compact set $\overline{\Omega}$. From the definition of compactness, it follows that each of these sequences contains a convergent subsequence, $\{x_{n_k}\}$ and $\{y_{m_k}\}$, with limits $\bar{x}$ and $\bar{y}$, respectively, i.e.,

$$x_{n_k} \to \bar{x}, \qquad y_{m_k} \to \bar{y} \quad \text{as} \quad k \to \infty\,. \tag{44}$$

Note that n_k is not necessarily equal to m_k. For reasons that will become clear below, it is necessary to extract two subsequences with the same indices. To do this, we proceed as follows.)

Starting with the convergent subsequence $\{x_{n_k}\}$ – with limit $\bar{x}$ – consider the sequence $\{y_{n_k}\}$ (which is not necessarily convergent). From the compactness of Ω, it follows that the sequence $\{y_{n_k}\}$ contains a convergent subsequence, to be denoted as $\{y_{n_{k_l}}\}$. We shall let $\bar{y}$ denote the limit of this sequence. We now have that

$$x_{n_{k_l}} \to \bar{x}, \qquad y_{n_{k_l}} \to \bar{y} \quad \text{as} \quad l \to \infty\,. \tag{45}$$

(The first line follows from the uniqueness the limit of a convergent sequence.) Recall that $d_E(x_n, y_n) < \delta_n = 1/n$, which implies that

$$d_E(x_{n_{k_l}}, y_{n_{k_l}}) \to 0 \quad \text{as} \quad l \to \infty\,.$$

Once again from the (incredibly useful!) triangle inequality,

$$d_E(\bar{x}, \bar{y}) \leq d_E(\bar{x}, x_{n_{k_l}}) + d_E(x_{n_{k_l}}, y_{n_{k_l}}) + d_E(y_{n_{k_l}}, \bar{y}) \to 0 \quad \text{as} \quad l \to \infty\,. \tag{46}$$

Therefore $\bar{x} = \bar{y}$, which also implies that $f(\bar{x}) = f(\bar{y})$.

There are a few ways to proceed from here, all of which are essentially equivalent, e.g.,

1. Recall that f was assumed to be continuous on $\overline{\Omega}$. From (44), it follows that

$$f(x_{n_k}) \to f(\bar{x}) \quad \text{and} \quad f(y_{m_k}) \to f(\bar{y}). \tag{47}$$

   From this result, and taking limits on both sides of in (43), we have

$$|f(\bar{x}) - f(\bar{y})| \geq \epsilon > 0, \tag{48}$$

   which contradicts the previously established result that $f(\bar{x}) = f(\bar{y})$. Thus the theorem is proved.

2. Once again, f was assumed to be continuous on $\overline{\Omega}$. It follows that

$$|f(x_{n_k}) - f(y_{m_k})| \leq |f(x_{n_k}) - f(\bar{x})| + |f(\bar{x}) - f(\bar{y})| + |f(\bar{y}) - f(y_{m_k})| \to 0, \tag{49}$$

   which contradicts (43). Thus, the theorem is proved.

A return to sets: Countable and uncountable sets in R

This section may be considered as a supplement to Section 2, "Sets of points in **R**," in particular to the subsection entitled "Finite vs. infinite sets." We shall be encountering the notion of "countable" and "uncountable" (infinite) sets in several places throughout this course, so it is probably a good idea to provide a brief discussion and some definitions at the outset. (Once again, this discussion is certainly not complete – far from it!)

First of all, finite sets of the form $\{x_1, x_2, \cdots, x_N\}$ are quite trivial because of their finiteness. We can also compare the "sizes" of two finite sets in terms of their **cardinalities**, i.e., the number of elements in each of the finite sets. But what about infinite sets? For example, consider the following subsets of **R**,

$$S_1 = \{1, 2, 3, \cdots\}\,, \quad S_2 = \{1, 4, 9, \cdots\}\,, \quad S_3 = [1, 2]\,, \quad S_4 = [1, 3]\,. \tag{50}$$

How do the "sizes" of these sets compare, in terms of numbers of elements? For example, $S_2 \subset S_1$, so we might expect that S_1 is "larger" than S_2, i.e., S_1 contains more points than S_2. And $S_3 \subset S_4$, so we might expect that S_4 is "larger" than S_3. However, as will be discussed very shortly, when it comes to infinite sets of points, we have to discard the ideas of cardinality that we learned for finite sets. For example, in terms of the set-valued concept of cardinality, S_1 and S_2 have the same cardinality, the cardinality of the natural numbers, which is traditionally denoted as $\aleph_0$ ($\aleph$, pronounced

"aleph," is the first letter of the Hebrew alphabet.) And S_3 and S_4 also have the same cardinality, even though the lengths of the two intervals – another reasonable measure of the "sizes" of these sets – are different. The cardinality of these two sets is $\aleph_1$, the "cardinality of the continuum." For now, we simply mention that $\aleph_0 < \aleph_1$. (Even though both numbers are infinite, there is still an ordering of such **transfinite numbers**.)

Some definitions are now in order. These definitions will actually apply to sets in general, but we shall be particularly concerned with infinite sets in $\mathbf{R}$.

Definition 24 *Let M and N be subsets of $\mathbf{R}$. We say that M and N are* **equivalent** *to be denoted as $M \sim N$, if there exists a one-to-one correpondence ϕ between elements of M and N, i.e., for every $x \in M$, there exists a unique $y \in N$ such that $y = \phi(x)$ and $x = \phi^{-1}(y)$.*

Theorem 11 *If M, N and O are sets, and $M \sim N$ and $N \sim O$ then $M \sim O$.*

Returning to the sets $S_1 \cdots, S_4$, we can easily establish the following results:

1. $S_1 \sim S_2$. If we denote the elements of S_1 as x_n and those of S_2 as y_n, $n = 1, 2, \cdots$, then an obvious 1-1 correspondence is $y_n = \phi(x_n)$, where $\phi(x) = x^2$.

2. $S_3 \sim S_4$. For an $x \in S_3 = [1, 2]$, there is a unique element $y \in S_4 = [1, 3]$ given by $y = \phi(x) = 2x - 1$.

Later, we'll show that S_1 and S_3 are not equivalent, i.e., we cannot establish a one-to-one correspondence between elements of S_1 (or S_2) and elements of S_3 (or S_4).

Countable sets

Definition 25 *A set S, the elements of which can be placed in a 1-1 correspondence with the set of natural numbers $\mathbf{N} = \{0, 1, 2, \cdots\}$ is said to be* **enumerable**, **denumerable** *or* **countable**. *In this section, we shall usually use the term "countable", but the other terms are perfectly valid.*

To show that an infinite set is countable, we must simply indicate how its elements can be presented or arranged (without repetitions) as an "infinite list": The first element of the list corresponds to the digit 0, the second to 1, and so on. A particular infinite list of the elements of an infinite set, or a 1-1 correspondence between the elements and the natural numbers, is called an **enumeration** of the set. The natural number corresponding to a given member in the list is the **index** of the member in the enumeration.

The elements of a finite set can also be represented by a list, namely a finite list. Finite sets are therefore countable. Sometimes, it is desired to emphasize that a particular set of interest which is countable is also infinite – in such cases, we say that the set is **countably infinite** (or **denumerably infinite** or **enumerably infinite**).

Examples:

1. The sets $S_1 = \{1, 2, 3, \cdots\}$ and $S_2 = \{1, 4, 9, \cdots\}$ are countable (or countably infinite). This may seem rather trivial, but in each case, the elements of the set can be presented as a list – an infinite list. We could go one further step and establish a formula between the elements of each set and the natural numbers, e.g., $x_n = n + 1$, $n = 0, 1, 2, \cdots$, for S_1, but it is really not necessary.

2. The set of integers, $\mathbf{Z}$, is countable since its elements may be listed in the following order,

$$0,\ 1,\ -1,\ 2,\ -2,\ 3, \cdots .$$

3. The set of rational numbers, $\mathbf{Q}$, is also countable, which may seem surprising since we know that the set $\mathbf{Q}$ is dense in $\mathbf{R}$. Here, we show that the set of positive rational numbers is countable – the extension to all rational numbers is straightforward using the method employed above for the integers. Recalling that any rational number $x \in \mathbf{Q}$ may be written in the form m/n, where m and n are integers, we construct the following infinite array of ratios of integers.

$$\begin{array}{ccccccccc}
1/1 & & 1/2 & \rightarrow & 1/3 & & 1/4 & & \cdots \\
\downarrow & \nearrow & & \swarrow & & \nearrow & & & \\
2/1 & & 2/2 & & 2/3 & & 2/4 & & \cdots \\
 & \swarrow & & \nearrow & & & & & \\
3/1 & & 3/2 & & 3/3 & & 3/4 & & \cdots \\
\downarrow & \nearrow & & & & & & & \\
4/1 & & 4/2 & & 4/3 & & 4/4 & & \cdots \\
\vdots & & \vdots & & \vdots & & \vdots & &
\end{array}$$

Starting at the upper left entry $1/1$, we travel through the array in the zig-zag-like manner shown in the figure. As we travel in this manner, we remove each fraction that are equal in value to some member that has been visited earlier. For example, the element $2/2$ is removed since it is equal in value to $1/1$. The net result is the following enumeration of the positive rational numbers,

$$1,\ 2,\ 1/2,\ 1/3,\ 3,\ 4,\ 3/2,\ 2/3,\ 1/4,\ 1/5, \cdots .$$

4. If we restrict ourselves to only the elements of the above array which lie above the diagonal, i.e., ratios m/n for which $m < n$, then we produce an enumeration of the rational numbers which lie in the interval $(0, 1)$, i.e., $\mathbf{Q} \cap (0, 1)$, i.e.,

$$1/2,\ 1/3,\ 2/3,\ 1/4,\ 1/5, \cdots .$$

This set is countable. Only two more elements need be added to this list to yield the rational numbers which lie in $[0, 1]$, i.e., $\mathbf{Q} \cap [0, 1]$, i.e.,

$$0,\ 1,\ 1/2,\ 1/3,\ 2/3,\ 1/4,\ 1/5, \cdots .$$

This set is therefore countable.

5. Another enumeration of the set $\mathbf{Q} \cap (0,1)$ may be produced by systematically listing, for each $n = 2, 3, 4, \cdots$, the elements $1/n, 2/n, \cdots, (n-1)/n$, once again not considering any ratio that is equal in value to a previous member in the list, i.e.,

$$1/2,\ 1/3,\ 2/3,\ 1/4,\ 3/4,\ 1/5,\ 2/5,\ 3/5,\ 4/5,\ 1/6,\ 5/6, \cdots .$$

   Once again, this set is countable.

6. The matrix-based method used above can be generalized to establish that the set of ordered pairs of an enumerable set, e.g., the ordered pairs of natural numbers, ordered pairs of integers and even ordered pairs of rational numbers, is enumerable. The rows of the matrix are the enumerations of the pairs with the first member of the pair fixed.

7. The set of ordered triples of an enumerable set then be established by taking as the rows the enumerations of the triples with the first members of the triples fixed (the remaining pairs are enumerable).

8. Successive applications of the matrix method yields that ordered n-tuples of an enumerable set form an enumerable set.

9. From the above result, it follows that the set of all **algebraic equations with integral coefficients**, i.e.,

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 = 0\,, \quad a_n \neq 0\,, \tag{51}$$

   is enumerable, since each equation is defined by the ordered n-tuple,

$$(a_0, a_1, \cdots, a_{n-1}, a_n)\,. \tag{52}$$

   A (real) algebraic number is a real root of such an equation. Since each equation has at most n different roots, it follows that the set of all algebraic numbers is enumerable. (Note that the a_k's could have been assumed to be rational numbers, i.e., $a_k = p_k/q_k$, $p_k, q_k \in \mathbf{Z}$, in which case we simply multiply both sides of the equation by the least common multiple of the q_k.)

A note on the importance of countability

Countability, i.e., the ability to assign a 1-1 correspondence between a given infinite set and the natural numbers, is an extremely important property, and most likely one which we take for granted because of its use in so many applications. Perhaps the simplest example is that of (infinite) sequences of real numbers having the form $\{a_0, a_1, a_2, \cdots\}$. A sequence $\{a_n\}_{n=0}^{\infty}$ is obviously a countable set.

As discussed in a previous lecture, one common method of generating sequences is iteration, i.e.,

$$x_{n+1} = T x_n\,. \tag{53}$$

The production of one iterate at a time is a quite natural process. The infinite set of iterates $\{x_n\}$ produced by this procedure is a countable set.

Another example may be found in the well-known Fourier series expansion for a function $f(x)$ defined on the interval $(-\pi, \pi)$,

$$f(x) = a_0 + \sum_{k=1}^{\infty} [a_k \sin kx + b_k \cos kx]\,. \tag{54}$$

As we shall discuss in much more detail later in this course, the countably infinite set of functions, $\{1, \sin x, \cos x, \sin 2x, \cos 2x, \cdots\}$, forms a **basis** (in fact, an orthogonal one) in an appropriate Hilbert space of functions defined on $(-\pi, \pi)$ – the space denoted as $L^2(-\pi, \pi)$. The countability of this set permits the use of the summation over basis elements in Eq. (54) as opposed to an integration. Partial sums of the infinite series conveniently yield approximations to a given function $f(x)$ in this space to arbitrary accuracy: Given an $\epsilon > 0$, there exists an $N_\epsilon > 0$ such that for all $n > N_\epsilon$, the partial sums of the Fourier series,

$$S_n(x) = a_0 + \sum_{k=1}^{n} [a_k \sin kx + b_k \cos kx] \,, \tag{55}$$

satisfy the inequality,

$$\|f - S_n\|_2 < \epsilon \,, \tag{56}$$

where $\|\cdot\|_2$ denotes the L^2 norm on this space.

Some important properies of countable sets

Since we shall be dealing with countable sets at several places in this course, it is useful to state a few important properties. Most appear to be quite straightforward and even obvious. That being said, their proof still requires a little work using set theory.

Theorem 12 *Every infinite set has a countably infinite subset.*

Examples: (i) $\mathbf{N} \subset \mathbf{R}$, (ii) $\mathbf{Q} \subset \mathbf{R}$, (iii) $\mathbf{N} \subset \mathbf{Q}$, (iv) $\{2, 4, 6, \cdots\} \subset \{1, 2, 3, \cdots\}$.

Theorem 13 *If a is any infinite cardinal number, i.e., the cardinal number of an infinite set, then $\aleph_0 \leq a$.*

From this result, we conclude that the countable infinity of the natural numbers $\mathbf{N}$ is the "smallest infinity" that a set can possess – there is no lower type or degree of infinity that an infinite set can have.

Theorem 14 *Any subset of a countable set is countable.*

Theorem 15 *The union of any countable family of countable sets is a countable set, i.e., if $\{A_i\}_{i \in I}$ is a family of sets such that I (the index set) is countable and each A_i is countable, then $A = \cup_{i \in I} A_i$ is countable.*

Uncountable sets

The mathematician G. Cantor (1874) proved that there are infinite sets whcih are not countable (countably infinite). Such sets are said to be **uncountable** (or nonenumerable, nondenumerable). The set of real numbers $\mathbf{R}$ is uncountable. Cantor showed this by means of his famous "diagonal method," which we sketch below.

First consider the set of real numbers in the interval $(0, 1]$. Each real number $x \in (0, 1]$ possesses a unique non-terminating decimal expansion of the form $.d_0 d_1 d_2, \ldots,$ where $d_i \in \{0, 1, 2, \cdots, 9\}$. A

real number $x \in (0,1]$ may have a terminating expansion, e.g., $0.5278 = 0.5278000000\ldots$, but this expansion can be replaced by a non-terminating one, i.e., $0.5277999999999\ldots$, which represents the same number x. In this way, we can associate a unique non-terminating decimal expansion to each $x \in (0,1]$.

Now suppose that

$$x_0,\ x_1,\ x_2,\ x_3, \cdots \tag{57}$$

is an infinite list of real numbers which belong to the interval. (Whether or not this list will include all real numbers in $(0,1]$ is not known at this time. This is, in fact, what we wish to determine.) Now write down, one below another, the non-terminating decimal expansions of these numbers as follows,

$$\begin{array}{ccccccccc}
.d_{00} & & d_{01} & & d_{02} & & d_{03} & & \cdots \\
 & \searrow & & & & & & & \\
.d_{10} & & d_{11} & & d_{12} & & d_{13} & & \cdots \\
 & & & \searrow & & & & & \\
.d_{20} & & d_{21} & & d_{22} & & d_{23} & & \cdots \\
 & & & & & \searrow & & & \\
.d_{30} & & d_{31} & & d_{32} & & d_{33} & & \cdots \\
 & & & & & & & \searrow & \\
\vdots & & \vdots & & \vdots & & \vdots & &
\end{array} \tag{58}$$

It is probably useful to recall that each of these infinite, non-terminating expansions is unique and appears only once in the complete list.

Now select the diagonal elements of this array, d_{kk}, $k \geq 0$, as shown by the arrows. We shall now use the infinite sequence $\{d_{kk}\}$ to construct a new infinite sequence which is not in the list of decimal expansions. To do this, we change each digit d_{nn} to a different digit d'_{nn}, but careful not to produce a terminating fraction, i.e., an infinite tail of 0's. For example, if $d_{00} = 1$, d'_{00} could be any of the other nine digits not equal to 1, e.g., $d'_{00} = 5$.

We now claim that the resulting fraction,

$$.d'_{00}\, d'_{11}\, d'_{22}\, d'_{33}\ \cdots\ , \tag{59}$$

represents a real number x' which belongs to the interval $(0,1]$ but not to the infinite list in (58). To see this – and to appreciate the ingenuity of the construction – we simply compare the decimal expansion (i.e., the list of decimal digits) to each expansion in (58):

1. By construction, the expansion of x' differs from the first expansion in the first digit, i.e., $d_{00} \neq d'_{00}$. Therefore, x' cannot be equal to the real number x_0 defined by the first expansion.

2. Once again by construction, the expansion of x' differs from the second expansion in the second digit, i.e., $d_{11} \neq d'_{11}$. Therefore, x' cannot be equal to the real number x_1 defined by the second expansion.

The reader should now see the pattern: x' cannot be equal to any of the real numbers x_n, $n = 0, 1, 2, \cdots$, defined by the expansions which comprise the original infinite list in (57). In other words,

this original enumeration is not an enumeration of all real numbers in the interval $(0, 1]$. We conclude, therefore, than an enumeration of all real numbers in the interval $(0, 1]$ is not possible, i.e., the set $(0, 1]$ is uncountable (which implies that the set $[0, 1]$ is uncountable).

> **Note:** At this point, the reader may be thinking – and quite understandably so – the following: "OK, I see that since, say, the second digit of the new decimal sequence, d'_{11}, is different that the original second digit, d_{11}, of the second row, the new sequence cannot agree with the second row. But maybe there is a sequence down the road, say the Nth row, with digits that are identical to the second row after its second element has been modified, i.e.,
>
> $$.d_{N0}\, d_{N1}\, d_{N2} \cdots \;=\; d_{10}\, d'_{11}\, d_{12}\, \cdots . \tag{60}$$
>
> That may well be, and most probably is, the case. But recall that by (clever) construction, the Nth digit of our new sequence, d'_{NN}, has been selected **not** to agree with the Nth element, d_{NN}, of this original decimal sequence.

The extension to the entire real number line $\mathbf{R}$ is straightforward. Any real number can be written in the form $x = i + f$, where i denotes the integer part of x and $f \in (0, 1]$ its fractional part, e.g., $3.14159265\cdots = 3 + .14159265\cdots$. We are simply "translating" the result from $(0, 1]$ to cover the entire real line. As such, we may conclude that the set of real numbers $\mathbf{R}$ is uncountable/non-denumerable.

A final note on Cantor's diagonal proof: Instead of decimal expansions using digits $d_{ij} \in \{0, 1, 2, \cdots, 9\}$, binary expansions of real numbers using the digits $b_{ij} \in \{0, 1\}$ could also have been used.

Cardinal numbers, power sets and the "continuum"

From the above result, it follows that the set of real numbers $\mathbf{R}$, or even a subset $(a, b) \in \mathbf{R}$, cannot be equivalent to the set of natural numbers $\mathbf{N}$, i.e., no 1-1 correspondence exists between the reals (or a subset) and the natural numbers $\mathbf{N}$. One might say that the set of real numbers $\mathbf{R}$ is "too large" to be countable. One might also say that the set $\mathbf{R}$ is, in some way, larger than the set $\mathbf{N}$. But the set of natural numbers, $\mathbf{N}$, is already infinite, i.e., it contains an infinite number of elements. In other words, its cardinality is infinite. How can you have another set, $\mathbf{R}$, the cardinality of which is larger than infinity?

These thoughts did perplex mathematicians in the 1800's, which led to the development of **set theory** well beyond the elementary set theory that existed up to that time. The development was not smooth, however, and there are still many unresolved issues and even different schools of thought regarding how sets can be defined and whether or not infinite sets actually exist! These problems, however, lie well beyond the scope of this course. Very fortunately, we can safely use some quite standard concepts of set theory in our discussion of sets that are encountered in classical analysis.

That being said, for an excellent discussion of the development of set theory and the foundations of mathematics, the reader is referred to the book, *Introduction to Metamathematics* by S. Kleene (first published in 1952, seventh edition published in 1974).

First of all, recall that in the finite case, the size of a set S is usually characterized by the number of elements it contains – the co-called **cardinality** of S, which is commonly denoted as $\overline{\overline{S}}$. The cardinality of a set of N elements, i.e.,

$$S_N = \{a_1, a_2, \cdots a_N\}\,, \tag{61}$$

is obviously N. Of course, as the number N of elements in S_N increases, its cardinality increases. In the limit $N \to \infty$, e.g., the natural numbers $\mathbf{N} = \{0, 1, 2, \cdots\}$, we could simply state that the cardinality of $\mathbf{N}$ is "∞." But what about the set of real numbers $\mathbf{R}$ which, as we stated earlier, cannot be put into a 1-1 correspondence with the set $\mathbf{N}$? Should we use another symbol to denote an infinity which, in some way, is greater than ∞? This is actually accomplished by means of **transfinite numbers**, infinite numbers which represent different cardinalities of sets and which obey an ordering relation which, in some sense, may be viewed as "less than/greater than". As mentioned at the beginning of this section, the cardinality of the natural numbers $\mathbf{N}$ is denoted as $\aleph_0$ and that of the real numbers $\mathbf{R}$ as $\aleph_1$. Here we briefly sketch the ideas that lead to the relation $\aleph_0 < \aleph_1$ and finally to another, more concrete relation.

In order to do so, we must recall the following idea from elementary set theory:

Definition 26 *Let S be a set. The* **power set** *of S, denoted as $\mathcal{P}(S)$, is the set of all subsets (including the null set ϕ) of S.*

Example: Let $S = \{1, 2, 3\}$. Then $\mathcal{P}(S) = \{\phi, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$.

In the above example, $\overline{\overline{S}} = 3$ and $\overline{\overline{\mathcal{P}(S)}} = 8 = 2^3$. In general,

$$\text{If } \overline{\overline{S}} = N, \text{ then } \overline{\overline{\mathcal{P}(S)}} = 2^N\,. \tag{62}$$

Let $S = \{a_1, a_2, \cdots, a_N\}$. In order to generate all possible subsets, we simply consider all possibilities in which an element a_k, $1 \le k \le N$, is either in a subset or not in it. The total number of subsets must therefore be 2^N. (The one case in which all elements are absent corresponds to the null set ϕ. The one case in which all elements are present corresponds to the set S itself.)

The natural question is, "What happens in the limit that $N \to \infty$?" In other words, what is $\overline{\overline{\mathcal{P}(\mathbf{N})}}$, i.e., the cardinality of the set of all subsets of the natural numbers? Is it 2^∞, whatever that means? In fact, the answer is "Yes". The relation in Eq. (62) holds for transfinite cardinal numbers as well. Recalling that the cardinality of the set of natural numbers $\mathbf{N}$ is denoted as $\aleph_0$, we have that

$$\overline{\overline{\mathcal{P}(\mathbf{N})}} = 2^{\aleph_0}\,, \tag{63}$$

(whatever this means).

We now examine $\mathcal{P}(\mathbf{N})$, the set of all subsets of the natural numbers. We can represent a (finite) set of natural numbers, S, by an infinite sequence, Σ, of 0's and 1's (hence a nonterminating sequence, in which the 0's play the role of the 9's in our earlier decimal sequences) as follows: For $n \ge 0$, if the integer n lies in S, then the nth digit of the sequence Σ is a "1"; otherwise it is a "0". For example, the set $\{1, 2, 6, 10\}$ is represented by the sequence

$$0\,1\,1\,0\,0\,0\,1\,0\,0\,0\,1\,0\,0\,\cdots\,. \tag{64}$$

Corresponding to the null set $\phi \in \mathcal{P}(\mathbf{N})$ is the unique sequence,

$$0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,\cdots\,. \tag{65}$$

Now suppose that

$$S_0,\, S_1,\, S_2,\, \cdots\,, \tag{66}$$

is an infinite list of distinct subsets of $\mathbf{N}$. The reader should see where we are going with this. Write down, one below another, the non-terminating binary sequences Σ_k associated with the subsets S_k. These sequences may be viewed once again as rows (of infinite length) of an infinite matrix. We now perform Cantor's "diagonal method" on this matrix of squences, i.e., take the diagonal elements of this matrix, $\{\sigma_{kk}\}$, and use them to construct a new infinite sequence which is not in the list of sequences Σ_k. If for a $k \geq 0$, $\sigma_{kk} = 0$, then define $\sigma'_{kk} = 1$. If $\sigma_{kk} = 1$, then define $\sigma'_{kk} = 0$. Once again, by construction, the sequence

$$\sigma'_{00},\, \sigma'_{11},\, \sigma'_{22},\, \cdots\,, \tag{67}$$

does not belong to the original list of sequences Σ_k for the same reason as before. We may therefore conclude that no infinite list of subsets S_k of the natural numbers can produce all subsets of $\mathbf{N}$, i.e., $\mathcal{P}(\mathbf{N})$, the set of subsets of $\mathbf{N}$, is uncountable.

Both in the case of $\mathcal{P}(\mathbf{N})$ as well as the set of real numbers $\mathbf{R}$, Cantor's diagonal method was employed on an infinite list of (infinite) sequences to show that the list did not contain all elements, i.e., that the original set was uncountable. One might naturallly wonder if the two sets are equivalent, i.e., if there is a 1-1 correspondence between elements of $\mathbf{R}$ and $\mathcal{P}(\mathbf{N})$. The answer is "Yes." In the case of $\mathbf{R}$, instead of employing decimal expansions of real numbers, which are composed of the digits $0 - 9$, we can use their binary expansions, which are composed of 0's and 1's. Modulo a couple of technical points, the two problems then appear identical. Without going into any further discussion, we simply state the final result,

$$\overline{\overline{\mathcal{P}(\mathbf{N})}} = \overline{\overline{\mathbf{R}}} = \aleph_1\,. \tag{68}$$

From Eq. (63), we have that

$$2^{\aleph_0} = \aleph_1\,. \tag{69}$$

The transfinite number $\aleph_1$ is also called the **power of the continuum**, the term "continuum" referring to the real line $\mathbf{R}$.

Of course we don't have to stop here! What about $\mathcal{P}(\mathcal{P}(\mathbf{N})))$, the set of all subsets of the power set of $\mathbf{N}$? The cardinality of this set is the transfinite number $\aleph_2$ which is related to $\aleph_1$ as follows,

$$2^{\aleph_1} = \aleph_2\,. \tag{70}$$

The reader can see that an iteration of this process produces an infinite sequence of transfinite cardinal numbers $\aleph_0 < \aleph_1 < \aleph_2 < \cdots$ which satisfy the relation,

$$2^{\aleph_n} = \aleph_{n+1}\,, \quad n = 0, 1, 2, \cdots\,. \tag{71}$$

Finally, we mention that the famous **Continuum Hypothesis** states that there are no intermediate cardinal values between $\aleph_0$ and $\aleph_1$.

References:

1. S. Kleene, *Introduction to Metamathematics*, Wolters-Noordhoff Publishing, Seventh Edition, 1974.

2. R. Rucker, *Infinity and the Mind: The Science and Philosophy of the Infinite*, Princeton University Press, 2005.

3. E. Hewitt and K. Stromberg, *Real and Abstract Analysis: A modern treatment of the theory of functions of a real variable*, Springer-Verlag, 1965. Chapter 1, Set Theory and Algebra.