MATH 6701
Mathematical Methods of Applied Science I
Fall Semester 2026

John McCuan

August 25, 2026

Contents

Preface

These are notes for MATH 6701, Mathematical Methods of the Physical Sciences I, taught in Fall 2026 at Georgia Tech. The course is a review of linear algebra and ordinary differential equations for graduate students with the addition of an introduction to complex analysis. The course is also the first part in a two course sequence with MATH 6702, Mathematical Methods of the Physical Sciences II, giving a review of multivariable calculus and an introduction to partial differential equations.

The notes (and the instructor) aim to give the student an opportunity to view the reviewed topics from a more advanced, or mathematical, perspective than is perhaps customary.

Part I

Introduction/Overview

Chapter 1

Introduction

1.1 MATH 6701

The course MATH 6701, Mathematical Methods of Applied Scince I, nominally consists of a review of linear algebra and ordinary differential equations (ODE) along with an introduction to complex analysis. The course is for graduate students, and it is assumed each student has taken a dedicated undergraduate course in linear algebra and also an introductory undergraduate course in ordinary differential equations. It is not expected that a student in the course has taken a previous course in complex analysis, though most students have some familiarity with complex numbers.

The text *Mathematical Methods in the Physical Scineces* by Mary Boas contains the following chapters

- Chapter 2 Complex Numbers
- Chapter 3 Linear Algebra
- Chapter 8 Ordinary Differential Equations
- Chapter 14 Functions of a Complex Variable

and to this extent provides an organizational structure for the course as well as explanations and problems which may be useful to a student. Reference may also be made to the first chapter of Boas' book which contains material on sequences and series as some of this material is central to complex analysis as pointed out by Boas in section 6 of her Chapter 2. The aims of the text

differ, however, to a certain extent from the instructor's objective to offer the student a broader and more mathematical perspective from which to view the subjects. In particular, I will attempt to introduce a framework having at its core the careful understanding of the definition and role played by the concept of **function** and especially the importance of understanding the sets, domain and codomain, associated with a function. This is both generally absent from the text and from the standard undergraduate courses taken as review material for this course. Elaborating this perspective is the purpose of these notes. In addition, I include some brief preliminary material on mathematical aspects of **sets** and additional aspects of **algebra** or what might be called abstract algebra putting especially the notion of a vector space and linear algebra in a broader context. It is hoped that this "new" point of view or "new" material will provide a stimulating and motivating opportunity for a student to review topics with which he may have some familiarity and also a perspective that may be compared and contrasted with that of the text.

To illustrate the basic point of view taken, I offer (re)definitions of each subject included in the course:

- Complex analysis is the study of complex differentiable functions.
- Linear algebra is the study of linear functions.
- Ordinary differential equations is the study of solutions (which are functions) and the structure of solution sets.[1]

In each case, some cursory study of the subjects may be undertaken without special regard for the sets associated with the functions (the domains and codomains) or the consideration of the underlying structure of various sets involved and the structures associated with various sets of functions. This point of view is taken in the text. These notes offer something "new" and different.

[1] Of course, the study of ordinary differential equations (ODE) is very properly viewed as the study of certain equations, but these are equations for functions. If the role played by the functions in the study of the equations is not appreciated, which I suggest is often the case, then a different perspective may well prove desirable.

1.2 Sets

The topic with which I wish to start is **complex analysis**, but to approach this subject and the others (linear algebra and ordinary differential equations) from the perspective I wish to present requires some preliminary discussion. Some aspects of this discussion will be familiar. Let me emphasize again that the notion of **function** will play the central role, so the student can be on the alert for the preliminary discussion of functions. Also, various aspects of "abstract algebra" will provide a background framework and it may be useful for the student to look for the elements of the following outline as they appear in the course of these notes:

1. Basic Algebraic Structures
   (a) groups[2]
   (b) rings
   (c) fields
2. Secondary/Derived Algebraic Structures
   (a) linear space
   (b) vector space (normed linear space)
   (c) Banach space
   (d) inner product space
   (e) Hilbert space

The student is not expected to know any of these terms. They are defined in the notes below. We begin with some basic notation (and perhaps a few ideas).

1.2.1 In or out?

The central defining characteristic of a **set** is **inclusion**. Given a set A, one should be able to tell if some "object" x is either in A or is not in A. In the former case I will write

$$x \in A$$

[2]Boas does offer a brief introduction to groups in section 13 of her Chapter 3.

read "x is in A." In the latter case

$$x \notin A$$

which is read "x is not in A." Generally, elements enclosed in curly brackets indicate the consideration of a set, and numbers tend to make good objects to either be in or to not be in a set. For example, the number 2 is not in the set containing the numbers 1, 3, and 4:

$$2 \notin \{1, 3, 4\}.$$

Some sets have special names. The positive integers may be denoted by

$$\mathbb{N} = \{1, 2, 3, \ldots\}.$$

Sometimes it is not easy to list out all the elements in a set, and it is convenient to combine curly brackets with a colon and a description to determine which elements are in a set and which are not. The set of positive integers divisible by 3 is

$$D = \{n \in \mathbb{N} : n \text{ is divisible by } 3\}.$$

We know then that $4 \notin D$ and $6 \in D$.

It turns out some limitations are required on the "objects" under consideration, and a good number of sets and objects seem to be available, but one can't go overboard and start talking about "the set of all sets" and such things. One can take the point of view that such trouble is not likely to arise, so we will not go into much detail about what "sets" are unacceptable, but I'll indicate briefly that there could conceivably be problems. Say, we allow the set of all sets X. Then one can consider the set of all sets A for which $A \notin A$, that is A is not an element of A. Let's call this secondary set B. The condition $A \notin A$ is perhaps a bit unusual, but it does seem to hold for most sets. The set containing the number 1 is not an element of that set, or in symbols

$$\{1\} \notin \{1\}.$$

I guess this means $\{1\} \in B$. One might imagine there is no set for which $A \in A$ so

$$B = \{A \in X : A \notin A\}.$$

is just another name for X. In any case, the condition $A \notin A$ makes sense. Recalling the key defining characteristic that one should be able to determine

inclusion, we might then ask about B itself. To say $B \notin B$ would mean $B \in X$ and B satisfies the condition after the colon in the definition of B. Thus, we would have $B \in B$. This is not okay because we can't determine if some object, namely B is in B or not in B.

One way to avoid this mess is to always assume there is some well-defined (and properly restricted) collection of elements U containing everything under consideration for the discussion, but certainly not everything possible one might be able to imagine, like the set of all sets. It is the details of constructing the collection U of everything under consideration and/or showing such a construction is or might be possible that I will ignore and avoid. I will just assume such a set U exists and that something like the set of all sets can be safely banished from the discussion. I suggest you do the same. If you are not satisfied with that, I commend to you a more serious study of set theory. There are books; there are courses; or you can just try to figure it out yourself.

1.2.2 subsets

Since I've just mentioned **subsets** I'll record the notation for that:

$$A \subset U$$

means every element of A is an element of U. Notice this allows the possibility that $A = U$, though usually I will not take the trouble to type the more complicated symbol in

$$A \subseteqq U.$$

If I happend to know (or require) $A \subset U$ and $A \neq U$, then I'll have to go to the trouble of typing the complicated symbol in

$$A \subsetneqq U.$$

There is also a special name for this. If $A \subset U$ and $A \neq U$, then A is said to be a **proper** subset of U.

Incidentally, my name for one set being a subset of another is **containment**. When one element is in a set, that is inclusion. These are nominally different things:

$$a, b, c \in \{a, b, c, d\} \qquad \text{but} \qquad \{a, b\} \subset \{a, b, c, d\}.$$

In the former case "a, b, and c are included in the set $\{a, b, c, d\}$." This is inclusion of elements. In the latter case "the set $\{a, b\}$ is contained in the set $\{a, b, c, d\}$." If you like you can remember that inclusion/exclusion (in or out) is for elements and this is the defining characteristic of a set (or more properly of sets). What is the opposite of containment?

1.2.3 cross product

Given two sets X and Y, I may (and often do) consider ordered pairs (x, y) with $x \in X$ and $y \in Y$. The set of all such ordered pairs has a special name and notation:

$$X \times Y = \{(x, y) : x \in X,\ y \in Y\}$$

is called the **cross product** of X and Y. This completes our (short and easy) discussion of sets. There are only three main things:

1. inclusion $x \in A$ or $x \notin A$.
2. specification (or building sets): $\{x \in U : S(x)\}$ read "the set of all elements x in U for which the condition/statement $S(x)$ holds."
3. cross product.

I use these words and notations very often when I'm discussing mathematics, and many people find them part of a convenient language for that kind of communication. Welcome to the discussion.

1.3 Functions

Now I come to the central player, at least in my mind. The most careful and precise definition of a function is as a subset of the cross product of two sets.

1.3.1 relation

Given two sets X and Y a **relation** between X and Y is (just) a subset of $X \times Y$.

There's not much more to say about that. Perhaps an example:

$$R = \{(x, y) : x^2 + y^2 = 1\}$$

is a relation. I didn't say it, but I'm assuming here the ordered pair (x, y) consists of two real numbers, that is (x, y) is an element of the cross product of the real numbers with itself. We'll come back and talk a bit more about the real numbers later, but I'll just throw out the symbol $\mathbb{R}$ for now. Thus, the circle above is

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

where $\mathbb{R}^2$ is another notation for $\mathbb{R} \times \mathbb{R}$.

Exercise 1.1. Draw and color the relations

$$\{(x, y) : x^2 + y^2 = 4\}$$

and

$$\{(x, y) : x^2 + y^2 \leq 4\}.$$

In summary **relation** is a special word used merely to refer to a subset of $X \times Y$. This word takes us in the direction of the notion of a function.

1.3.2 function

Given two sets X and Y a **function** from X to Y is a relation F satisfying the two conditions

1. $\{x \in X : (x, y) \in F\} = X$ and
2. If $(x, y_1) \in F$ and $(x, y_2) \in F$, then $y_1 = y_2$.

I believe defining a function as a set in this way gives the most precise mathematical definition of a function. This is not the most commonly used definition of the term function. It is also probably not the most intuitive or useful definition of a function. The definition suggested by Euler is probably more useful, though one should note the relation (or equivalence) to the set definition above.

Perhaps a worksheet is in order here.

worksheet on the definition of a function

Write down a definition of the term function (from a set X to a set Y) in more informal, colorful, or informative language. Try to be reasonably precise and capture the essentials of the set definition above.

Here is a place for you to write down Euler's definition of a function. You can compare what Euler suggested to what you have written.

1.3.3 function (part two)

Given two sets X and Y, a **function** f from X to Y is a rule or correspondence which assigns to each $x \in X$ a unique $y \in Y$.

The most imprecise part centers around the words "rule or correspondence." If one wished, one might object that the words "rule" and/or "correspondence" have no solid mathematical definition. If one wishes to say that, then perhaps one should be referred to the set definition which does not contain these words. In any case, we call such a rule by a name, like f and write

$$f : X \to Y$$

which is read "f is a function from X to Y." The set associated with a function, as considered in the previous section, is then given by the **graph** of the function

$$\mathcal{G} = \{(x, f(x)) : x \in X\}.$$

The set X in this case is called the **domain** and the set Y is called the **codomain** or target. Whenever one wishes to consider a function, one should at least make an effort to identify the domain and the codomain or at least some sets which could possibly be the domain and codomain if such sets are not specified.

The codomain is not necessarily the **range**, but the range is

$$\{y \in Y : (x, y) \in \mathcal{G} \text{ for some } x \in X\} \subset Y.$$

The range can also be written more simply as

$$\{f(x) : x \in X\}.$$

The range is also often denoted by $f(X)$ and is also called the **image** of the function f. Note carefully the notation:

$$f(A) = \{f(x) : x \in A\}$$

which is most often used when $A \subset X$ and X is the domain of a function f so that $f : X \to Y$. The image of any set will always be a subset of the codomain.

As mentioned above complex analysis is the study of "complex differentiable" functions, and I will make it one of my first priorities to explain what

"complex differentiable" means. Before one gets to that it may be noted that the domain of a complex differentiable function is either the complex numbers $\mathbb{C}$ or some suitable subset of $\mathbb{C}$. Some attention will be given to explaining things about the complex numbers $\mathbb{C}$ at the beginning of the next chapter in these notes. The codomain of a complex differentiable function can essentially always be taken to be $\mathbb{C}$.

Linear algebra is the study of linear functions, and perhaps it is useful to understand carefully some things about the domain and codomain of such a function. For the moment I wish only to point out that this description of linear algebra is almost certainly very different from the manipulations of systems of equations and matrices found in most courses on linear algebra. You can also see this in the first twelve out of fifteen sections of Boas' Chapter 3.

By the time we get to our "review" of ordinary differential equations (ODE) hopefully you will appreciate the role played by functions in making sense of the equations (DE).

1.3.4 functions in the background

Many functions go unnamed. In the next chapter I write about "operations." You should be familiar with operations from first or second grade:

$$1 + 1 = 2, \qquad 3 + 5 = 8.$$

These are operations on the set of positive integers $\mathbb{N}$ mentioned above. That is to say, there is a function "combining" two positive integers and giving back a third as "output." I hope you agree this "combining" constitutes a "rule or correspondence." We might call this function the **addition** function on the positive integers and name it "+" so that $+ : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$:

$$+(1, 1) = 2, \qquad +(3, 5) = 8.$$

You may notice some missing parentheses here: If an ordered pair in $\mathbb{N}\times\mathbb{N}$ is given by two positive integers hemmed in by parentheses: $(n_1, n_2) = (3, 5)$; and function values are given by the name of the function with its argument in parentheses: $y = f(x)$; then we should technically write

$$+(\ (3, 5)\ ) = 8.$$

One could say we are just leaving out some "extra" parentheses, and that's fine. But one may also say that such identifications (like a rule or correspondence) involve some unnamed functions in the background.

Another similar example is the following: Say we start with the operation-like function of "subtraction on the positive integers" $- : \mathbb{N} \times \mathbb{N} \to \mathbb{Z}$ with $-(3,5) = -2$. The domain of this function is the cross product $\mathbb{N}\times\mathbb{N}$ and the codomain is $\mathbb{Z}$. The graph should be a subset of the cross product $(\mathbb{N}\times\mathbb{N})\times\mathbb{Z}$ the elements of which look like

$$((n_1, n_2), n_3).$$

We will very often use instead the elements of a slightly different set $\mathbb{N}\times\mathbb{N}\times\mathbb{Z}$ which are ordered triples (n_1, n_2, n_3). Again, in the background there is a special kind of function which "identifies" $(\mathbb{N} \times \mathbb{N}) \times \mathbb{Z}$ with $\mathbb{N} \times \mathbb{N} \times \mathbb{Z}$.

We will mostly not be interested in (or mention) these many unnamed functions, but they are there for our reference should we need them.

1.3.5 kinds of functions

Another name for function is **morphism**. In particular, a function with the same domain and codomain is called an **automorphism**. A function $f : \mathbb{C} \to \mathbb{C}$ in complex analysis or a function $f : B_1(0) \to B_1(0)$ where $B_1(0) = \{x + iy : x^2 + y^2 < 1\} \subset \mathbb{C}$ is an automorphism. Linear automorphisms $L : V \to V$ will be of special interest in linear algebra, and these are also called **operators**.

Everything in the previous paragraph is just about names, but there are also some concepts going along with names that are useful to know. We know that for a function $f : X \to Y$ one has the property

$$\text{If } y_1 = f(x) \text{ and } y_2 = f(x), \text{ then } y_1 = y_2.$$

That is, for each $x \in X$, there is a unique $y \in Y$ assgined to x, that is with $y = f(x)$. This is one of the defining properties of a function. It is not always true for every function that

$$\text{If } f(x_1) = f(x_2), \text{ then } x_1 = x_2.$$

If this holds for a function, we say the function is **one-to-one** or **injective**.

Similarly, it is not always true that

$$f(X) = \{f(x) : x \in X\} = Y,$$

but if the range is equal to the codomain, then the function is said to be **surjective** or **onto**.

A function $f : X \to Y$ which is both injective and surjective is said to be **bijective**. Such a function is also often referred to as a **one-to-one correspondence**.

Injective, surjective, and bijective are adjectives. The corresponding nouns are **injection**, **surjection**, and **bijection**.

An example of a bijection is the **identity function** on any set. This function is often denoted by $\mathrm{id}_X : X \to X$, and has values given by $\mathrm{id}_X(x) = x$ for each $x \in X$.

More generally, a function $f : X \to Y$ is said to be **invertible** or to have an **inverse** if there exists a function $g : Y \to X$ with the functions

$$f \circ g : Y \to Y \qquad \text{by} \qquad f \circ g(y) = f(g(y))$$

and

$$g \circ f : X \to X \qquad \text{by} \qquad g \circ f(x) = g(f(x))$$

satisfying $f \circ g = \mathrm{id}_Y$ and $g \circ f = \mathrm{id}_X$.

The functions $f \circ g$ and $g \circ f$ are called **compositions**, and the function g in the definition of invertible, when it exists, is called the **inverse** of f.

Exercise 1.2. A function $f : X \to Y$ is invertible if and only if f is a bijection. (Show it.)

The inverse of a function $f : X \to Y$, when it exists is denoted by $f^{-1} : Y \to X$.

More general **compositions** are also possible: If $f : X \to Y$ and $g : Y \to Z$, then there is a composition function $g \circ f : X \to Z$ with $g \circ f(x) = g(f(x))$.

I'm probably missing some definitions related to sets and functions which should be recorded in this section of the notes, but the terms and conditions listed above are a pretty good start.

1.3.6 extra stuff

I will add a subsection here to contain additional information about sets and functions when I happen to want to use that information in the notes below. You are encouraged to skip this section initially and come back to it when the need arises. I guess this section can be expected to change as the notes expand.

There is a special set with no elements. This set is called the **empty set**, and I will generally denote the empty set using the symbol ϕ. The symbol ϕ may be used for other things as well (functions, angles, etc.), but if I am worried about other uses becoming confusing with the special use for the empty set, I may resort to the symbol $\emptyset$ for the empty set.

The **union** of two sets A and B is the collection of all elements which are either in A or B. The union of two sets is denoted $A \cup B$.

The **intersection** of two sets A and B is the collection of all elements belonging to both A and B. The intersection is denoted $A \cap B$.

The word "collection" is just another word for "set," but it sometimes makes mathematical writing easier to understand when such synonyms are used. "Belonging" is the same as "inclusion."

Given a function $f : X \to Y$ and a subset $A \subset X$, the **restriction** of f to the set A is the function with the same values as f on A, but with smaller domain. This function is denoted by

$$f\big|_A : A \to Y$$

which one reads as "the restriction of f to A." As mentioned, the values are the same:

$$f\big|_A(x) = f(x) \qquad \text{for} \qquad x \in A.$$

1.3.7 an exercise

In the next chapter I will write something about the definition and properties of the complex numbers. I assume however you are at least a bit familiar with complex arithmetic. If that is not the case, the first five sections of Chapter 2 in Boas' book contain some useful information and problems at which you may want to have a look.

I want to consider here a certain function $f : \mathbb{C} \to \mathbb{C}$. This is the squaring function with values given by

$$f(x + iy) = x^2 - y^2 + 2xy\, i.$$

This function is not one-to-one, but it is surjective. The complex numbers

$$\mathbb{C} = \{x + iy : x, y \in \mathbb{R}\}$$

are in some ways very like the cross product $\mathbb{R}^2$. In fact, $\mathbb{R}^2$ is often called a (Cartesian) plane, and $\mathbb{C}$ is called the complex plane. The exercise I suggest for you is the following:

Exercise 1.3. Find at least two different (proper) subsets A of the complex plane for which the squaring function $f : A \to \mathbb{C}$ is invertible.

Perhaps you already know how to do this exercise, but if not I expect you may have some difficulty getting started. It might be good to "struggle" for a while on your own in order to get started. If you are capable of that, and you want to do that, feel free. Alternatively, I will try to suggest some ideas that might help someone who has never thought too much about complex numbers get started.

I mentioned that $f(z) = z^2$, that is the squaring function on $\mathbb{C}$, is not injective. In fact, there is only one unique complex number z for which $f(z) = 0$. That number is $z = 0$ as you can see by considering the system of equations associated with $f(x + iy) = 0$ or

$$\begin{cases} x^2 - y^2 = 0 \\ 2xy = 0. \end{cases}$$

From the second equation either $x = 0$ or $y = 0$. The first equation then says the other must be zero. For example, if $x = 0$ in the second equation, then the first equation becomes $-y^2 = 0$ which for a real number y has only the solution $y = 0$. Thus, the unique solution of $(x + iy)^2 = 0$ is $x + iy = 0 + 0\,i$ or $x = y = 0$, so $x + iy = 0$.

Now consider moving away from $z = 0$ in the complex plane. Any other complex number $z = x + iy$ determines a unique angle θ with $0 \leq \theta < 2\pi$ for which

$$\begin{cases} \cos\theta = x/\sqrt{x^2 + y^2} \\ \sin\theta = y/\sqrt{x^2 + y^2}. \end{cases}$$

This angle θ is sometimes called the **argument** of $z \in \mathbb{C}\backslash\{0\}$. We see then that $z = x + iy$ can be written as

$$z = \sqrt{x^2 + y^2}\ (\cos\theta + i\sin\theta).$$

Incidentally, the radius associated with the point $(x, y) \in \mathbb{R}^2$ is also called the **modulus** of the complex number $z = x + iy$. If we square z in this form and remember the trigonometric identities

$$\begin{cases} \cos^2\theta - \sin^2\theta = \cos(2\theta), & \text{and} \\ 2\cos\theta\sin\theta = \sin(2\theta), \end{cases}$$

we find

$$\begin{aligned} z^2 &= (x^2+y^2)\left[\cos^2\theta - \sin^2\theta + 2\cos\theta\sin\theta\ i\right] \\ &= (x^2+y^2)\left[\cos(2\theta) + \sin(2\theta)\ i\right]. \end{aligned}$$

It will be noted that this is essentially the same form

$$w = r(\cos\tau + i\sin\tau)$$

for $r = x^2+y^2$ and $\tau = 2\theta$. Thus, given a complex number

$$w = a + bi = r(\cos\tau + i\sin\tau) \in \mathbb{C}\backslash\{0\}$$

we can look for complex numbers $x+iy$ on the circle with

$$x^2 + y^2 = r,$$

that is on the circle with radius

$$\sqrt{x^2+y^2} = \sqrt{r} = \sqrt[4]{a^2+b^2},$$

and having argument θ for which

$$\begin{cases} \cos(2\theta) = \cos\tau \\ \sin(2\theta) = \sin\tau. \end{cases} \tag{1.1}$$

In these argument equations the angle τ is fixed, and we can remember that $\cos\theta$ and $\sin\theta$ are 2π periodic. More properly, the functions sine and cosine are 2π periodic. The argument appearing in the sine and cosine however is 2θ. This means that when θ runs from 0 to 2π the quantity 2θ is running from 0 to 4π. Consequently, the sine and cosine appearing on the left of the argument system (1.1) are seeing two periods of values. Put another way, each pair of desired values $(\cos\tau, \sin\tau)$ is hit twice.

There is a unique angle 2θ with $0 \le 2\theta < 2\pi$, that is with

$$0 \le \theta < \pi$$

for which $(\cos(2\theta), \sin(2\theta)) = (\cos\tau, \sin\tau)$. And then there is another angle $2\psi = 2\theta + 2\pi$ with

$$2\pi \le 2\theta + 2\pi < 4\pi \qquad \text{so that} \qquad \pi \le \psi < 2\pi$$

for which $(\cos(2\psi), \sin(2\psi)) = (\cos\tau, \sin\tau)$ as well.

The points

$$\sqrt[4]{a^2+b^2}\,(\cos\theta + i\sin\theta) \in \mathbb{C}\backslash\{0\} \tag{1.2}$$

and

$$\sqrt[4]{a^2+b^2}\,(\cos(\theta+\pi) + i\sin(\theta+\pi)) = \sqrt[4]{a^2+b^2}\,(-\cos\theta - i\sin\theta) \tag{1.3}$$

are **antipodal** points in the complex plane; in particular, they are different from one another, and they are two points having the property that when they are squared, the result is $w = a+bi$. This means that with the exception of $w = a + bi = 0$, the squaring function is not injective (one-to-one). You should be able to see at this point, more or less, precisely how the complex squaring function works.

Exercise 1.4. Start with the point $w = 2 + i\,\sqrt{3}$.

(a) Find two points $z \in \mathbb{C}$ for which $z^2 = w$. Hint: Find the modulus and argument of w.

(b) Draw the two points you found in part **(a)** along with their arguments.

Exercise 1.5. Show that given any point $w = a + bi \in \mathbb{C}\backslash\{0\}$ there are **exactly two** points z_1 and z_2 in $\mathbb{C}$ with $z_2 = -z_1$ and $z_j^2 = w$ for $j = 1, 2$.

Exercise 1.6. Make a "mapping drawing" indicating how the complex function $f : \mathbb{C} \to \mathbb{C}$ with $f(z) = z^2$ "works."

(a) Start by drawing a copy of the complex plane on the left of your paper and indicating (by shading or some other means) the proper subset

$$\Omega = \{z \in \mathbb{C}\backslash\{0\} : 0 \le \arg(z) < \pi\}.$$

(b) Draw an arrow to the right with the symbol f over the top of it, and to the right of that arrow draw the image of Ω, that is the set

$$f(\Omega) = \{z^2 \in \mathbb{C} : z \in \Omega\}.$$

(c) Draw a particular point $z \in \Omega$ and label the modulus $|z|$ and argument $\arg(z)$ for that point. Then draw $w = f(z)$ in the image plane on the right of your drawing with appropriate expressions and labels for $|w| = |z|^2$ and $\arg(w) = 2\arg(z)$.

(d) You should still (hopefully) have some room on your paper under the mapping drawing you have so far. Draw a second arrow specifically from $\mathbb{C}\backslash\Omega$ on the left to the image $f(\mathbb{C}\backslash\Omega)$ that you draw below your picture of the image $f(\Omega)$.

(e) Draw and label the point $-z$ in your drawing. Note the point $-z$ should be in $\mathbb{C}\backslash\Omega$. Similarly, draw $f(-z)$ in your picture of the image $f(\mathbb{C}\backslash\Omega)$ on the right. You may wish to use a different color for your drawing of $-z$ and $f(-z)$.

(f) Draw a semicircle with center $z = 0$ connecting the labeled points z and $-z$ in the domain in your mapping picture; draw the image of this semicircle as well. (Change the color of the semicircle when it exits A and enters $\mathbb{C}\backslash A$ and put appropriate arrows on that semicircle and its image.)

The mapping drawing from Exercise 1.6 should have two "copies" of $\mathbb{C}$ on the right. More precisely, there should be a copy of $\mathbb{C}\backslash\{0\}$ on the top right and a copy of $\mathbb{C}$ on the bottom right. Something interesting and important is happening in each of these copies along the positive real axis

$$A = \{w \in \mathbb{C} : \arg(w) = 0\}.$$

In fact, if these two copies are "sewn" together correctly their union[3] makes a single object consisting of two copies of $\mathbb{C}\backslash\{0\}$ along with the single point $\{0\}$. In this case some terminology and accompanying symbols may be introduced. The image set consisting of the two sewn together pieces is called a Riemann surface for the complex squaring function. Let's call this Riemann surface Σ. Aside from the origin $0 \in \mathbb{C}$, the Riemann surface Σ may be viewed as consisting of the two "sheets," one on the top right of the mapping drawing and one on the bottom right. We can denote these two sheets by Σ_1 and Σ_2 so that

$$\Sigma = \Sigma_1 \cup \Sigma_2 \cup \{0\}.$$

Notice that the squaring function considered with codomain the Riemann surface Σ rather than codomain $\mathbb{C}$ is invertible. In particualar, the restriction[4]

$$f\big|_{\Omega} : \Omega \to \Sigma_1$$

[3]Take a look back at the extra stuff in the material on sets above.

[4]See the section on extra stuff again.

has an inverse: For each $w \in \Sigma_1$ where Σ_1 is essentially just a copy of $\mathbb{C}\backslash\{0\}$, there exitst a unique

$$z \in \Omega = \{z \in \mathbb{C}\backslash\{0\} : 0 \leq \arg(z) < \pi\}$$

for which $f(z) = z^2 = w$.

You should now be able to return to the main exercise, Exercise 1.3, and complete the exercise in a variety of ways. In fact, given any argument θ_0 there is a region/set

$$\Omega_0 = \{z \in \mathbb{C}\backslash\{0\} : \theta_0 \leq \arg(z) < \theta_0 + \pi\} \cup \{0\}$$

for which the restriction

$$g = f\big|_{\Omega} : \Omega_0 \to \mathbb{C}$$

of the squaring function f with $f(z) = z^2$ is a bijection. In this case, the image of the line $\arg(z) = \theta_0$ is called a **branch cut** and the inverse of the restriction g is called a **branch of the square root**. Note that $g^{-1} : \mathbb{C} \to \mathbb{C}$ is injective but not surjective since the image is $\Omega_0 \subsetneqq \mathbb{C}$. We will see later that the branch cut is important because the branch g^{-1} is not even continuous across the branch cut much less differentiable[5] there. Using the Riemann surface for the domain of the square root resolves this problem to a certain extent though there is still a problem with differentiability at $w = 0$ which is a **branch point**.

[5]Remember complex analysis is the study of complex differentiable functions, though you very well may not know what that means at the moment. We'll get there.

Chapter 2

Complex Numbers

As suggested by the table of contents in the book *Mathematical Methods in the Pyysical Scineces* by Mary Boas, I think it's nice to start the course with some discussion of complex numbers before discussing too much linear algebra or ordinary differential equations. My reasons are probably a bit different from hers, but still that's what I suggest.

Interest in the complex numbers centers largely around the fact that the complex numbers are an algebraically complete field containing the real numbers. My first objective is to explain to some extent what that means. My explanation will come partially in the form of putting the notion of "field" in the context of a more general algebraic framework. While I am at it, I will briefly describe some additional structures relevant to the study of linear algebra. All of this material may be viewed as falling under the heading of **abstract algebra**.

On the next page is an outline, with space to write down definitions, of the list of algebraic structures on sets which comprise the excerpt of abstract algebra relevant to this course.

algebraic structures on sets

Basic

(a) group

(b) ring

(c) field

Secondary or derived "spaces"

(a) linear space over a field

(b) normed space (vector space)

(c) Banach space

(d) inner product space

(e) Hilbert space

2.1 Groups, rings, and fields

2.1.1 group structure

One of the simplest algebraic structures that may be considered on a set is that of a **group**. A **group** is a set G along with an "operation"

$$* : G \times G \to G.$$

The usual notation for function values would be $*(g_1, g_2)$, but in the case of groups (and "operations") one usually uses the alternative notation

$$g_1 * g_2.$$

In applications, the symbol "$*$" is often addition or multiplication or something similar, so $g_1 * g_2$ may appear as $g_1 + g_2$ or $g_1 g_2$, depending on the context. In all cases, however, the "structure" associated with a group is expressed by these three conditions:

(i) $(g_1 * g_2) * g_3 = g_1 * (g_2 * g_3)$ whenever $g_1, g_2, g_3 \in G$.

(ii) There is some element $e \in G$ for which $g * e = e * g = g$ whenever $g \in G$.

(iii) Whenever $g \in G$, there is an element $h \in G$ for which $g * h = h * g = e$.

Notes:

- Of course there could be simpler algebraic structures. One could leave off condition **(iii)** for example. Sometimes sets with such structures will be considered, but I won't introduce special names for such structures. Except for such simpler structures one might take these three properties as practially the definition of an operation.

- Condition **(i)** is called the associative property. This should be familiar from arithmetic:

$$\begin{aligned} (2 + 6) + 3 &= 8 + 3 = 11, \qquad \text{and} \\ 2 + (6 + 3) &= 2 + 9 = 11. \end{aligned}$$

Also,

$$((2)(6))(3) = (12)(3) = 36, \qquad \text{and}$$
$$(2)((6)(3)) = (2)(18) = 36.$$

These elements $2, 3, 6, 9, 11, 12, 18$, and 36 are elements of the natural numbers or positive integers

$$\mathbb{N} = \{1, 2, 3, \ldots\}.$$

This is an example of one of those simpler structures. See Exercise 2.1 below.

- When an associative property holds for three group elements $g_1, g_2, g_3 \in G$, as it always does, so that

$$(g_1 * g_2) * g_3 = g_1 * (g_2 * g_3), \tag{2.1}$$

then the expression $g_1 * g_2 * g_3$ may be used to represent either expression left or right in (2.1).

- Condition **(ii)** is called the (group) identity property, that is, there exists an **identity** for the (group) operation. The set

$$\mathbb{N}_0 = \{0, 1, 2, 3, \ldots\},$$

that is, the natural numbers with zero, has identity elements for its two operations of addition and miltiplication:

$$0 + n = n + 0 = n \text{ for } n \in \mathbb{N}_0,$$
$$1 \cdot n = n \cdot 1 = n \text{ for } n \in \mathbb{N}_0.$$

- Condition **(iii)** is the existence of **inverses**.

Only one element in $\mathbb{N}_0$ has an additive inverse and only one element of $\mathbb{N}_0$ has a multiplicative inverse. Consequently, neither the natural numbers nor $\mathbb{N}_0$ is a group under either of the familiar operations.

Exercise 2.1. (a) Which of properties **(i)-(iii)** do the positive integers have under addition?

(b) Which of properties **(i)-(iii)** do the positive integers have under multiplication?

Example 1. The **integers**

$$\mathbb{Z} = \{0, \pm 1, \pm 2, \pm 3, \ldots\}$$

is a group under addition.

Example 2. The **rational numbers**

$$\mathbb{Q} = \left\{ \frac{p}{q} : p \in \mathbb{Z},\ q \in \mathbb{N} \right\}$$

is a group under addition.

Example 3. The **positive rational numbers**

$$\mathbb{Q}^+ = \left\{ \frac{p}{q} : p, q \in \mathbb{N} \right\}$$

is a group under multiplication.

Exercise 2.2. Find a subset of $\mathbb{Q}$ containing the negative integers

$$\mathbb{N}^- = \{-1, -2, -3, \ldots\}$$

which is a group under multiplication. **Exclusion notation** may be helpful:

$$A \backslash B = \{x \in A : x \notin B\}.$$

The real numbers are obtained from the rational numbers by adjoining the limits of (all) convergent sequences:

$$\sqrt{2} = \lim_{n \to \infty} x_n$$

where $x_1 = 1$, $x_2 = 4/3$, $x_3 = 24/17$, $\ldots$;

$$x_{j+1} = \frac{4x_j}{x_j^2 + 2},$$

$$\pi = \lim_{n \to \infty} y_n$$

where $y_1 = 4$, $y_2 = 8/3$, $y_3 = 52/15$, $\ldots$;

$$y_{j+1} = y_j + \frac{4(-1)^j}{2j+1},$$

etc.

Example 4. The **real numbers** $\mathbb{R}$ is a group under addition.

The real numbers have a number of additional properties with which you are probably familiar—at least somewhat familiar. These (many) properties make the real numbers the most used set for mathematical modeling and applications. In order to see more clearly how the use of the real numbers works in application, it is often useful to consider the complex numbers which can also be constructed using the real numbers:

$$\mathbb{C} = \{a + bi : a, b \in \mathbb{R}\}.$$

Example 5. The **complex numbers** $\mathbb{C}$ is a group under addition with

$$(a + bi) + (c + di) = a + c \ + (b + d)i$$

where $a, b, c, d \in \mathbb{R}$.

The sets $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ mentioned above are all **fields**, but that is getting ahead of ourselves.

There are many interesting things one can "show" or verify simply from the three properties defining a group. Any such verification applies to all groups. For example, condition **(ii)** says there exists an identity element, but one might wonder "Could there be two different identity elements?" The answer is "no" because if e_1 and e_2 are both identity elements so that $ge_1 = e_1g = g$ and $ge_2 = e_2g = g$ for every element $g \in G$, then we have $e_2 = e_2e_1$ because e_1 is an identity element. On the other hand, $e_2e_1 = e_1$ because e_2 is an identity element. That is,

$$e_2 = e_2e_1 = e_1,$$

so $e_1 = e_2$, and there is exactly one identity element. Once this fact is known, perhaps it is not so interesting, but still it is a fact that might be useful to know. Here is another similar fact:

Exercise 2.3. Condition **(iii)** gives the existence of inverse elements, but could there be two different inverse elements associated with a particular element g in a group G? Assume $gh_1 = h_1g = e$ and $gh_2 = h_2g = e$ for some elements $h_1, h_2 \in G$. Show $h_1 = h_2$.

If you enjoy making such verifications, here is one more:

Exercise 2.4. Assume you have a set with an associative operation $*$ and an identity element e. Show, without using property **(iii)**, that if an element g has a **left inverse** h_ℓ for which

$$h_\ell g = e,$$

and g has a right inverse h_r for which

$$g h_r = e,$$

then $h_\ell = h_r$ and g has an inverse in the sense of condition **(iii)**.

As a last topic in my short discussion of groups[1] I will introduce a property satisfied by some groups but not all groups.

A group G is said to be **commutative** or **Abelian** if $g_1 g_2 = g_2 g_1$ whenever $g_1, g_2 \in G$. We have not seen an example of a noncommutative group, so let's remedy that. We will start with a kind of operation given by multiplication on matrices. The set $M_2(\mathbb{R})$ of 2×2 matrices with real entries will be adequate. This set is also denoted by $\mathbb{R}^{2\times 2}$ and various other names and symbols. In any case, multiplication is given by the familiar formula

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{pmatrix}.$$

Note that in general a_{ij} is the entry in the i-th row and the j-th column, and we can also write/express the first factor in the matrix product above by (a_{ij}) so that

$$(a_{ij}) = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

Matrix multiplication is associative, so condition **(i)** for a group is satisfied. It may not be immediately obvious that matrix multiplication is associative. The entry in the first row and first column of the matrix product $[(a_{ij})(b_{ij})](c_{ij})$ is

$$(a_{11}b_{11} + a_{12}b_{21})c_{11} + (a_{11}b_{12} + a_{12}b_{22})c_{21},$$

[1] In section 13 of her Chapter 3 on linear algebra Boas also gives an introduction to groups. It seems she uses this introduction primarily as a motivation for the "abstract" definition of "general vector spaces" in the next section. You can have a look and compare the two approaches. I notice that Boas gives four conditions defining a group, so you can contemplate why she has, at least superficially, one more condition that I do.

and the corresponding $(1,1)$ entry in $(a_{ij})[(b_{ij})(c_{ij})]$ is

$$a_{11}(b_{11}c_{11} + b_{12}c_{21}) + a_{12}(b_{21}c_{11} + b_{22}c_{21}).$$

You can see these are the same expressions, but the terms have been somewhat rearranged and factored differently.

I will now attempt to generalize the above calculation to apply to all entries instead of just the first row and first column of the triple products. Writing $A = (a_{ij})$, $B = (b_{ij})$ and $C = (c_{ij})$ for the three matrices involved, I wish to show

$$(AB)C = A(BC).$$

The entry in the i-th row and the k-th column, the (i,k) entry, of the product AB is (as you should check)

$$p_{ik} = \sum_{\ell=1}^{2} a_{i\ell}b_{\ell k}.$$

Similarly, in the product $(AB)C$ the (i,j) entry is

$$\sum_{k=1}^{2} p_{ik}c_{kj} = \sum_{k=1}^{2}\left(\sum_{\ell=1}^{2} a_{i\ell}b_{\ell k}\right)c_{kj} = \sum_{k=1}^{2} c_{kj}\left(\sum_{\ell=1}^{2} a_{i\ell}b_{\ell k}\right) = \sum_{k=1}^{2} c_{kj}\sum_{\ell=1}^{2} a_{i\ell}b_{\ell k}.$$

I have used the commutative property of multiplication in $\mathbb{R}$, though you may recall that $\mathbb{R}$ is not a group under multiplication. Still there is an operation of multiplication on $\mathbb{R}$, and that operation is associative and commutative. There is also a distributive property

$$a(b + c) = ab + ac$$

for $a, b, c \in \mathbb{R}$. This property involves the interaction of addition and multiplication of real numbers and extends by induction to sums:

$$a\sum_{\ell=1}^{m} b_\ell = \sum_{\ell=1}^{m} ab_\ell.$$

Applying this extended distributive property to each term in the outer sum of the quantity at hand, we get

$$\sum_{k=1}^{2} p_{ik}c_{kj} = \sum_{k=1}^{2}\sum_{\ell=1}^{2} c_{kj}(a_{i\ell}b_{\ell k}) = \sum_{k=1}^{2}\sum_{\ell=1}^{2}(a_{i\ell}b_{\ell k})c_{kj} = \sum_{k=1}^{2}\sum_{\ell=1}^{2} a_{i\ell}b_{\ell k}c_{kj}.$$

Note that when an associative property holds for group elements g_1, g_2, g_3 under any operation "$*$" one may write

$$g_1 * g_2 * g_3 = (g_1 * g_2) * g_3 = g_1 * (g_2 * g_3)$$

without any difficulty caused by the "order of operations." This same thing is true even when one does not have a group but is dealing with an associative operation like multiplication of real numbers. This explains the last expression

$$\sum_{k=1}^{2} \sum_{\ell=1}^{2} a_{i\ell} b_{\ell k} c_{kj} \tag{2.2}$$

in our calculation above.

The calculation of the (i, j) entry in the triple product $(AB)C$ has become somewhat cumbersome to describe precisely partially because you are so familiar with the arithmetic properties of real numbers, but since one of the main objectives of the consideration of groups (and rings and fields) is to give you the opportunity[2] to consider carefully the abstract framework into which those arithmetic properties fit—or alternatively the abstract framework from which those arithmetic properties "come" or are derived—I am trying to be precise and careful.

Exercise 2.5. Carefully calculate the (i, j) entry of the triple matrix product $A(BC)$, and show that entry can be put in the form (2.2). This completes the verification of the associative property for matrix multiplication.

A nominally different approach to the verification of the associative property for matrix multiplication is possible. Essentially the same calculation is at the heart of this different approach, but the phrasing is rather different and ultimately quite important as I will try to explain later. I view this phrasing as so important that I will cover it twice in this course: I will explain it now as a digression concerning matrix multiplication and again from an opposite point of view in the notes on linear algebra below.

[2]Boas gives essentially the same calculation on page 139 of section 9 of her Chapter 3 on linear algebra. You may read her presentation and note that aside from no careful explanation of the real arithmetic involved, she apparently gives no independent definition of the triple matrix product ABC used in her explanation of the associative property of matrix multiplication. Without associativity of matrix multiplication the triple product ABC is not well-defined.

There is a connection between matrices and certain functions. You may recall the assertion above that linear algebra is the study of linear functions. This also may have seemed strange to you because you took a course on linear algebra which didn't really seem to involve anything about functions but was rather all about certain systems of equations and matrix manipulations. One connection is the following: For each matrix (a_{ij}), for example a 2×2 matrix $(a_{ij}) \in M_2(\mathbb{R})$, there is a function $L : \mathbb{R}^2 \to \mathbb{R}^2$ with values given by

$$L(x, y) = (a_{11}x + a_{12}y, a_{21}x + a_{22}y). \tag{2.3}$$

Note the components of the image here correspond to what one would get by usual matrix multiplication if both the domain element (x, y) and the codomain image are considered as column vectors:

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{pmatrix}.$$

On the other hand there is a set of functions, specifically **linear** functions sometimes denoted $L(\mathbb{R}^2)$ or $\mathrm{Aut}(\mathbb{R}^2)$. I will temporarily use the second notation arising from the fact which you may recall that a function with the same domain and codomain is an **automorphism**. It should be kept in mind, however, that we are only considering here **linear automorphisms** of $\mathbb{R}^2$. To each such function $L : \mathbb{R}^2 \to \mathbb{R}^2$ with $L \in \mathrm{Aut}(\mathbb{R}^2)$ there corresponds a matrix. In this case, the fact that L is **linear** is expressed by the condition(s)

$$L(\mathbf{x} + \mathbf{y}) = L(\mathbf{x}) + L(\mathbf{y})$$

for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$ and

$$L(c\mathbf{x}) = cL(\mathbf{x})$$

for any $c \in \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^2$. Separately these conditions are called additivity and homogeneity. Equivalently, one may say

$$L(c_1\mathbf{v} + c_2\mathbf{w}) = c_1L(\mathbf{v}) + c_2L(\mathbf{w})$$

whenever $c_1, c_2 \in \mathbb{R}$ and $\mathbf{v}, \mathbf{w} \in \mathbb{R}^2$. In this formulation the expression $c_1\mathbf{v} + c_2\mathbf{w}$ is called a **linear combination** of $\mathbf{v}$ and $\mathbf{w}$. Notice also that $\mathbb{R}^2$ is an example of an additive group with componentwise addition:

$$(x_1, x_2) + (y_1, y_2) = (x_1 + y_1, x_2 + y_2),$$

and one is using the additive group structure to form the linear combination.

Exercise 2.6. Verify that (2.3) defines a linear function.

The relation in which I am primarily interested at the moment may be considered to be at a secondary level. Specifically, I have defined a function $\Phi : M_2(\mathbb{R}) \to \mathrm{Aut}(\mathbb{R}^2)$ which assigns to each 2×2 matrix (a_{ij}) with real entries a linear function $L : \mathbb{R}^2 \to \mathbb{R}^2$. The linear function, as described above, is essentially given by matrix multiplication, that is, the formula in (2.3). If the matrices (a_{ij}) and (b_{ij}) are both assigned to the same linear function $L \in \mathrm{Aut}(\mathbb{R}^2)$, then I claim the matrices are the same: $a_{ij} = b_{ij}$ for $i, j = 1, 2$. To see this, note that $L(1, 0) = (a_{11}, a_{21})$. This comes from substituting the domain element $(1, 0) \in \mathbb{R}^2$ into the formula from (2.3). On the other hand, the hypothesis is that if we replace the matrix (a_{ij}) in (2.3) with the matrix (b_{ij}), then we get the same function, and the same function values in particular. From this we also know $L(1, 0) = (b_{11}, b_{21})$. Hence

$$(a_{11}, a_{21}) = (b_{11}, b_{21}),$$

and this says the first column of (a_{ij}) is the same as the first column of (b_{ij}), that is $b_{i1} = a_{i1}$ for $i = 1, 2$. You can check that the second columns match as well by looking at the function value $L(0, 1)$. The elements $(1, 0)$ and $(0, 1)$ in $\mathbb{R}^2$ having one entry the additive identity (zero) from $\mathbb{R}$ and the other entry the multiplicative identity (one) have some special significance and are often denoted by

$$\begin{aligned} \mathbf{e}_1 &= (1, 0), \text{ and} \\ \mathbf{e}_2 &= (0, 1). \end{aligned}$$

We will very often use this notation. The condition we have verified:

$$\Phi(a_{ij}) = \Phi(b_{ij}) \qquad \Longrightarrow \qquad (a_{ij}) = (b_{ij})$$

is precisely what it means for $\Phi : M_2(\mathbb{R}) \to \mathrm{Aut}(\mathbb{R}^2)$ to be injective or one-to-one.[3]

Exercise 2.7. Show $\Phi : M_2(\mathbb{R}) \to \mathrm{Aut}(\mathbb{R}^2)$ given by (2.3) is surjective.

This is a key fact: Φ is a one-to-one correspondence between 2×2 matrices and linear functions in $\mathrm{Aut}(\mathbb{R}^2)$. The set $\mathrm{Aut}(\mathbb{R}^2)$ also has something like an operation.

[3]The symbol " $\Longrightarrow$ " is read "implies." The statement is equivalent to "If $\Phi(a_{ij}) = \Phi(b_{ij})$, then $(a_{ij}) = (b_{ij})$."

Exercise 2.8. If L_1 and L_2 are both linear automorphisms, show the composition $L_2 \circ L_1$ with values $L_2 \circ L_1(v) = L_2(L_1(v))$ is also a linear automorphism.

Recalling the original question/motivation concerning the associativity of matrix multiplication, I now make two observations:

1. It is very easy to see that function composition on $\mathrm{Aut}(\mathbb{R}^2)$ is associative. In fact,

   $$(L_3 \circ L_2) \circ L_1(v) = (L_3 \circ L_2)(L_1(v)) = L_3(L_2(L_1(v))) = L_3((L_2 \circ L_1)(v)).$$

   In fact, function composition of automorphisms is always associative; this important observation has nothing to do with linearity.

2. The operation of matrix multiplication in $M_2(\mathbb{R})$ corresponds to function evaluation in $\mathrm{Aut}(\mathbb{R}^2)$. More precisely what I mean to say is that the definition of $L(x, y)$ given in (2.3) corresponds to multiplication of the matrix (a_{ij}) to the column "matrix"

   $$\begin{pmatrix} x \\ y \end{pmatrix}.$$

   This has been mentioned above, and is also quite obvious.

Let us also assume we can establish a correspondence between matrix multiplication in $M_2(\mathbb{R})$ and function composition in $\mathrm{Aut}(\mathbb{R}^2)$. Specifically, let us assume

$$\Phi(AB) = \Phi(A) \circ \Phi(B) \tag{2.4}$$

where A and B are any matrices in $M_2(\mathbb{R})$. Applying this assumption to the product $(AB)C$ with three matrices $A, B, C \in M_2(\mathbb{R})$,

$$\Phi((AB)C) = \Phi(AB) \circ \Phi(C) = (\Phi(A) \circ \Phi(B)) \circ \Phi(C) = \Phi(A) \circ \Phi(B) \circ \Phi(C)$$

where the last expression has meaning given by our observation above that function composition as an "operation" is always associative. On the other hand, we may also apply the assumed correspondence to the product $A(BC)$ to obtain

$$\Phi(A(BC)) = \Phi(A) \circ \Phi(BC) = \Phi(A) \circ (\Phi(B) \circ \Phi(C)) = \Phi(A) \circ \Phi(B) \circ \Phi(C).$$

Thus, since $\Phi((AB)C)$ and $\Phi(A(BC))$ correspond to the same function in $\mathrm{Aut}(\mathbb{R}^2)$ and $\Phi : M_2(\mathbb{R}) \to \mathrm{Aut}(\mathbb{R}^2)$ is a bijection we must have

$$(AB)C = A(BC).$$

That is, the associative property for matrix multiplication holds.

It remains to verify matrix multiplication in $M_2(\mathbb{R})$ corresponds to function composition in $\mathrm{Aut}(\mathbb{R}^2)$ as expressed in (2.4). This is an equality of functions in $\mathrm{Aut}(\mathbb{R}^2)$, and equality of automorphisms with the same domain (and codomain) is essentially equality of values. Thus, we wish to show for each $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$ there holds

$$\Phi(AB)(\mathbf{x}) = \Phi(A) \circ \Phi(B)(\mathbf{x}).$$

In order to "unpack"this equality in terms of matrix multiplication, it is convenient to use column "matrices" (often called column vectors) associated with elements of $\mathbb{R}^2$ and the element $\mathbf{x}$ in particular. There is a formal and proper way to do this using the operation of **transpose**. I will be rather formal and proper for the moment, though often I am a bit more sloppy, and very well may often be sloppy about this in the future.

One writes

$$(x_1, x_2)^T = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

to indicate the transpose of the element $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$. The transpose function involved here is often left in the background. In order to be fully formal and proper, the domain and codomain of the transpose should be considered. The domain is relatively simple. That is $\mathbb{R}^2$, the set of ordered pairs of real numbers:

$$\mathbb{R}^2 = \{(x_1, x_2) : x_1, x_2 \in \mathbb{R}\}.$$

The codomain is

$$\left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : x_1, x_2 \in \mathbb{R} \right\}.$$

To the extent that these are two different sets, the transpose is not an automorphism. On the other hand, the transpose as presented here is a very simple one-to-one correspondence. Extending my commitment to be formal and proper, I will refer to elements of $\mathbb{R}^2$ as "points" and distinguish elements

of the codomain of the transpose as "columns." I will also give the codomain a name:

$$\mathbb{R}^2_c = \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : x_1, x_2 \in \mathbb{R} \right\},$$

and this set/codomain may be referred to as the column version of $\mathbb{R}^2$ or simply "column $\mathbb{R}^2$." There is a nominally different function $T_c : \mathbb{R}^2_c \to \mathbb{R}^2$ exchanging columns for rows, that is points:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}^{T_c} = (x_1, x_2).$$

This should be enough formality to get us through the discussion of condition (2.4) above, though I will return to to the transposes T and T_c with some additional comments below.

If $L(x_1, x_2)$ is the linear function value associated with (the value) $L = \Phi(A) = \Phi(a_{ij})$, then

$$L(x_1, x_2) = [(a_{ij})(x_1, x_2)^T]^T. \tag{2.5}$$

Notice the notation for the transpose function is "exponential." Accordingly, we are employing here a tacit "order of operations" so that $A\mathbf{x}^T$ is really $A(\mathbf{x}^T)$.

Exercise 2.9. In the identity (2.5) I have used the symbol "T" to represent both transposes T and T_c. This is an abuse of notation. Rewrite (2.5) properly using T_c.

Recall the two linear functions $\Phi(A) \circ \Phi(B)$ and $\Phi(AB)$ under consideration here. For $\Phi(AB)$, the value $\Phi(AB)(\mathbf{x})$ is

$$[(AB)\mathbf{x}^T]^T.$$

For the composition $\Phi(A) \circ \Phi(B)$, the corresponding value is

$$[A((B\mathbf{x}^T)^T)^T]^T = [A(B\mathbf{x}^T)]^T \tag{2.6}$$

where we have used the observation that applying the transpose twice to a column gives back the same column or more properly applying T_c the transpose of a column followed by T the transpose of a point gives back the same column:

$$\begin{pmatrix} c_1 \\ c_2 \end{pmatrix}^{T_c} = (c_1, c_2) \in \mathbb{R}^2 \qquad \text{and} \qquad \left(\begin{pmatrix} c_1 \\ c_2 \end{pmatrix}^{T_c} \right)^T = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}.$$

Exercise 2.10. Identify the column $\mathbf{c} \in \mathbb{R}^2_c$ in (2.6) to which the identity $(\mathbf{c}^{T_c})^T = \mathbf{c}$ has been applied.

We have reduced verification of the identity (2.4) to showing that for each pair of matrices $A, B \in M_2(\mathbb{R})$ and each $\mathbf{x} \in \mathbb{R}^2$ there holds

$$[(AB)\mathbf{x}^T]^T = [A(B\mathbf{x}^T)]^T.$$

The outer transpose is of course T_c the transpose of a column, so this is an equality in $\mathbb{R}^2$. We can dispense with the outer column transpose, and it is enough to verify the identity

$$(AB)\mathbf{x}^T = A(B\mathbf{x}^T)$$

of columns in $\mathbb{R}^2_c$. Finally, we can replace the column $\mathbf{x}^T$ by an arbitrary column $\mathbf{c} \in \mathbb{R}^2_c$ and seek to show that for every pair of matrices $A, B \in M_2(\mathbb{R})$ and every column $\mathbf{c} \in \mathbb{R}^2_c$ there holds

$$(AB)\mathbf{c} = A(B\mathbf{c}). \tag{2.7}$$

This at least looks much simpler.

Recall from the definition of matrix multiplication on page 33

$$AB = \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{pmatrix}.$$

Thus if $\mathbf{c} = (c_1, c_2)^T$, then

$$(AB)\mathbf{c} = \begin{pmatrix} (a_{11}b_{11} + a_{12}b_{21})c_1 + (a_{11}b_{12} + a_{12}b_{22})c_2 \\ (a_{21}b_{11} + a_{22}b_{21})c_1 + (a_{21}b_{12} + a_{22}b_{22})c_2) \end{pmatrix}. \tag{2.8}$$

On the other hand,

$$B\mathbf{c} = \begin{pmatrix} b_{11}c_1 + b_{12}c_2 \\ b_{21}c_1 + b_{22}c_2 \end{pmatrix}$$

and

$$A(B\mathbf{c}) = \begin{pmatrix} a_{11}(b_{11}c_1 + b_{12}c_2) + a_{12}(b_{21}c_1 + b_{22}c_2) \\ a_{21}(b_{11}c_1 + b_{12}c_2) + a_{22}(b_{21}c_1 + b_{22}c_2) \end{pmatrix}. \tag{2.9}$$

A little algebraic rearrangement using the distributive property now makes it clear that (2.8) and (2.9) are the same thing. That is, we have established the grouping identity

$$(AB)\mathbf{c} = A(B\mathbf{c})$$

stated in (2.7) and this establishes the identity

$$\Phi(AB) = \Phi(A) \circ \Phi(B)$$

for linear functions. That was a long digression elements of which will be continued in some additional comments below. Let us return to the main task of finding a noncommutative group.

So far we have a set $M_2(\mathbb{R})$ of 2×2 matrices and a way to combine two matrices in $M_2(\mathbb{R})$ with matrix multiplication. The matrix multiplication is associative, but $M_2(\mathbb{R})$ is not a group. By way of our digression above we also have a set $L(\mathbb{R}^2)$ or $\mathrm{Aut}(\mathbb{R}^2)$ of linear automorphisms of interest in the subject of linear algebra. There is a seemingly very similar operation on the linear functions given by function composition $L_2 \circ L_1$.

I have stated that $M_2(\mathbb{R})$ is not a group. This is not because $M_2(\mathbb{R})$ fails to have an identity element. You can check very easily (and I'm sure you are aware) that the matrix

$$I = I_{2\times 2} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

happily satisfies condition **(ii)** on page 29 concerning the existence of an identity element in a group. Correspondingly, $\Phi(I) = \mathrm{id}_{\mathbb{R}^2}$ with values $\mathrm{id}_{\mathbb{R}^2}(\mathbf{x}) = \mathbf{x}$ is the identity function, which is linear and provides an identity element for functions in $\mathrm{Aut}(\mathbb{R}^2)$. The failure of $M_2(\mathbb{R})$ and the corresponding collection $L(\mathbb{R}^2) = \mathrm{Aut}(\mathbb{R}^2)$ of functions is with regard to inverses. There is no matrix B for which

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} B = B \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

And the zero matrix is not the only problem.

Exercise 2.11. Show there is no matrix B for which

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} B = B \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

There are on the other hand some matrices in $M_2(\mathbb{R}^2)$ which do have inverses:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} I = I \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

That is,

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

is its own inverse. You may already be familiar with conditions according to which the matrices with inverses in $M_2(\mathbb{R})$ can be distinguished from those not having inverses. This question is perhaps most naturally addressed in linear algebra, and we will set it aside for now, but I will include a couple preliminary exercises at the end of this section.

we are almost in a position to finish up giving an example of a noncommutative group. Say two matrices $A, B \in M^2(\mathbb{R})$ both have inverses. Then there is a matrix $B^{-1}A^{-1} \in M^2(\mathbb{R})$ and

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1}A^{-1} = AIA^{-1} = I = (B^{-1}A^{-1})(AB).$$

Here we have used the associativity of matrix multiplication nominally justified in two different ways above. The result is that the product AB is a matrix in $M_2(\mathbb{R})$ having an inverse. This is important because it means matrix multiplication is a well-defined operation on the smaller collection of matrices with inverses under matrix multiplication. One might also say this set is "closed" under matrix multiplication. This set is sometimes denoted by $M_2^*(\mathbb{R})$.

The claim is that the set of matrices $M_2^*(\mathbb{R})$ is a group under matrix multiplication.[4] Let us summarize: We have a collection of matrices $M_2^*(\mathbb{R})$ which is a proper subset of the collection of all 2×2 matrices. This subset contains the identity matrix but does not contain the zero matrix. We have a way to combine matrices A and B with inverses (matrix multiplication) to obtain another matrix AB with an inverse. This is a well-defined operation on $M_2^*(\mathbb{R})$. The operation was associative on the bigger set $M_2(\mathbb{R})$, so it is still associative on the smaller set. The identity element/matrix I is, first of all, still in $M_2^*(\mathbb{R})$, and naturally I still functions as an identity element. Finally, by definition, each element in $M_2^*(\mathbb{R})$ has an inverse. Indeed $M_2^*(\mathbb{R})$ is a group (under matrix multiplication).

The second claim is that $M_2^*(\mathbb{R})$ is a noncommutative group. An example of two elements in $M_2^*(\mathbb{R})$ not commuting with each other is adequate to

[4]This is a special case of something one can see again more generally in the next two sections.

complete the discussion of our example. Let's try

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \qquad \text{and} \qquad B = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

You can check that

$$A^{-1} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

and B is its own inverse, or $B^2 = I$. Thus, we have $A, B \in M_2^*(\mathbb{R})$.

For the product AB we get

$$AB = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

On the other hand

$$BA = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}.$$

Thus, $AB \neq BA$, and we have here a noncommutative group.

2.1.2 homomorphism

The condition

$$\Phi(AB) = \Phi(A) \circ \Phi(B)$$

is an example of what is meant by a group homomorphism. More generally given two groups G_1 and G_2 and a function $\phi : G_1 \to G_2$, the function ϕ is called a **homomorphism** if

$$\phi(g \ *_1 \ h) = \phi(g) \ *_2 \ \phi(h)$$

for each $g, h \in G_1$ where "$*_1$" denotes the operation on G_1 and "$*_2$" denotes the operation on G_2.

It might be said that even a cursory discussion of groups should include mention of the main function relevant to studying groups, namely the group homomorphism. I will also mention that if a group homomorphism is a bijection, as in the case of the function $\Phi : M_2^*(\mathbb{R}) \to \Phi(M_2^*(\mathbb{R}))$, then the domain group and the codomain group are equivalent in every way pertaining to the group structure. Such a group homomorphism is called a **group isomorphism**.

The isomorphic image $\Phi(M_2^*(\mathbf{r})) \subset L(\mathbb{R}^2)$ in the example given above has a special name and notation. The name is the **general linear group** (of automorphsims on $\mathbb{R}^2$), and the notation is

$$\Phi(M_2^*(\mathbf{r})) = GL(\mathbb{R}^2).$$

The notation $GL_2(\mathbb{R})$ is also used. Below are the promised additional exercises relating matrices more fully to the study of linear functions (linear algebra).

Exercise 2.12. Given a matrix $A \in M_2(\mathbb{R})$, if there exists a matrix $B \in M_2(\mathbb{R})$ for which

$$AB = BA = I,$$

then $\Phi(A) \in \mathrm{Aut}(\mathbb{R}^2)$ is a bijection. (Show it.)

Solution: Let $L = \Phi(A)$. By definition $L(bx) = A\mathbf{x}$ or more properly $L(\mathbf{x}) = (\mathbf{x}^T)^T$. Or even more properly $L(\mathbf{x}) = (\mathbf{x}^T)^{T_c}$. In any case, if there is some $\mathbf{y} \in \mathbb{R}^2$, then $\mathbf{y}B \in \mathbb{R}^2$ and... wait $\mathbf{y}B$? (Why B?)

Notice the product $\mathbf{y}B$ makes good sense using the same "kind" of matrix multiplication introduced above:

$$(y_1, y_2)\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} y_1b_{11} + y_2b_{21} & , & y_1b_{12} + y_2b_{22} \end{pmatrix}.$$

That is, $\mathbf{y}B = (B\mathbf{y}^T)^{T_c} = \Phi(B)(\mathbf{y})$. So that's why B... or rather $\mathbf{y}B$.

In fact, $L(\mathbf{x}) = \Phi(A)(\mathbf{x}) = \mathbf{x}A$ and

$$L(\mathbf{y}B) = (\mathbf{y}B)A = \mathbf{y}(BA) = \mathbf{y}I = \mathbf{y}.$$

This means L is surjective. Furthermore, if $L(\mathbf{x}_1) = L(\mathbf{x}_2)$, then

$$\mathbf{x}_1A = \mathbf{x}_2A$$

so that $\mathbf{x}_1AB = \mathbf{x}_2AB$ and $\mathbf{x}_1 = \mathbf{x}_2$. This means L is injective. (This completes the solution of the exercise.)

The function $L = \Phi(A)$ is a bijection when A has a multiplicative inverse $B = A^{-1}$. This means over in the set of functions $\mathrm{Aut}(\mathbb{R}^2)$ one can apply Exercise 1.2 to conclude there exists an inverse function L^{-1}, and indeed it is pretty easy to see that the inverse function $L^{-1} : \mathbb{R}^2 \to \mathbb{R}^2$ is linear as well.

For example, $L^{-1}(\mathbf{x}+\mathbf{y})$ is the unique element $\mathbf{z} \in \mathbb{R}^2$ for which $L(\mathbf{z}) = \mathbf{x}+\mathbf{y}$. That element is

$$L^{-1}(\mathbf{x}) + L^{-1}(\mathbf{y})$$

because (using only the linearity of L)

$$L(L^{-1}(\mathbf{x}) + L^{-1}(\mathbf{y})) = L \circ L^{-1}(\mathbf{x}) + L \circ L^{-1}(\mathbf{y}) = \mathbf{x} + \mathbf{y}.$$

That is

$$L^{-1}(\mathbf{x} + \mathbf{y}) = \mathbf{z} = L^{-1}(\mathbf{x}) + L^{-1}(\mathbf{y}).$$

Exercise 2.13. Show $L^{-1}(a\mathbf{x}) = aL^{-1}(\mathbf{x})$ for any $a \in \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^2$.

Exercise 2.14. If $L \in \mathrm{Aut}(\mathbb{R}^2)$ is any linear bijection, then the unique matrix A with $\Phi(A) = L$ has an inverse in $M_2(\mathbb{R})$ under matrix multiplication.

Solution: I've already shown above (modulo Exercise 2.13 showing homogeneity) that L^{-1} is linear. This means there is also a unique matrix $B \in M_2(\mathbb{R}^2)$ with $L^{-1}(\mathbf{y}) = \mathbf{y}B$ for every $\mathbf{y} \in \mathbb{R}^2$. Remember Φ is a one-to-one correspondence and

$$\Phi(B)(\mathbf{y}) = [B\mathbf{y}^T]^T = \mathbf{y}B.$$

By the relation (2.4) satisfied by Φ we can say

$$\Phi(AB) = \Phi(A) \circ \Phi(B) = L \circ L^{-1} = \mathrm{id}_{\mathbb{R}^2}.$$

This means $AB = I$. Similarly, applying the composition relation to $\Phi(BA)$ leads to the conclusion $BA = I$. (This completes the solution of the exercise.)

2.1.3 associativity and transposes

You may recall that in our alternative proof of the associativity relation

$$(AB)C = A(BC)$$

for matrix multiplication the final verification was reduced to showing

$$(AB)\mathbf{c} = A(B\mathbf{c})$$

where $\mathbf{c} \in \mathbb{R}^2_c$ is a column (of length 2 in this case). This is certainly a grouping identity, but it is not properly an associative property since the

operations are not a single operation on a specified set. The mechanics of verification in each case, however, are quite similar.

We also considered two different transpose functions $T : \mathbb{R}^2 \to \mathbb{R}^2_c$ and $T_c : \mathbb{R}^2_c \to \mathbb{R}^2$. In this case, we can define a larger set, say $\mathcal{M} = \mathcal{M}(\mathbb{R})$ of all matrices (a_{ij}) with real entries of any arbitrary size. On this set a transpose operation is well-defined taking an $m \times n$ matrix to an $n \times m$ matrix by exchanging the rows and columns.

Some of these matrices A and B in $\mathcal{M}$ can be multiplied, but not all of them. Consequently, there is no well-defined "operation" of matrix multiplication on $\mathcal{M}$. The condition required for multiplication is that if A has m rows and n columns and B has n rows, then the product AB is well-defined. In this context one can have various identities and grouping rules. One Example is

$$(A^T)^T = A.$$

This is always true no matter what dimensions A might have, as long as $T : \mathcal{M} \to \mathcal{M}$. You don't formally need the special cases T and T_c introduced above.

Exercise 2.15. Explain why the grouping identity

$$(AB)C = A(BC)$$

is always valid for $A, B, C \in \mathcal{M}$ as long as the dimensions of A, B, and C allow meaning for the products AB and BC. (Hint: Give the dimensions of A, B, and C names and explain what is required in order to form the products.)

Exercise 2.16. Explain why $(AB)^T = B^T A^T$ whenever the product AB is well-defined for matrices A and B in $\mathcal{M}$.

2.1.4 review of nongroups

The natural numbers $\mathbb{N} = \{n \in \mathbb{Z} : n > 0\}$ has an operation of addition. This operation is associative, but there is no zero element and there are no additive inverses.

The natural numbers with zero $\mathbb{N}_0 = \{n \in \mathbb{Z} : n \geq 0\}$ has also an associative (and commutative) operation of addition and a zero element, but still most elements have no additive inverse.

The 2×2 real matrices $M_2(\mathbb{R})$ has an operation of multiplication of matrices. This operation is associative. There is an identity element

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Some elements have (multiplicative) inverses, and some do not. This operation of multiplication on this nongroup is marginally notable because it is noncommutative. Most notably, there is a smaller set $M_2^*(\mathbb{R}) \subset M_2(\mathbb{R})$ of elements with inverses that is a group, and enough matrices are retained that this group is nonAbelian.

2.1.5 Rings

I've spent a fair amount of time introducing and discussing the concept of a group. Most other algebraic structures are built on the notion of a group.

As you may have noted many of the sets considered as groups above have two operations. Examples include $\mathbb{R}$, $\mathbb{C}$, $\mathbb{Q}$, $\mathbb{Z}$, and $M_2(\mathbb{R})$.

Exercise 2.17. Show $M_2(\mathbb{R})$ is a group under matrix addition, i.e., the addition of corresponding elements.

Of course, not all of these operations are full group operations, but some are simply associative. This kind of algebraic structure, though more complicated than that of a group, occurs so frequently that it has a special name. That name is "ring."

Definition 1. (ring) A ring R is a set with two operations. The set is required to be an Abelian group under one of the operations, and this operation is denoted by "+" (i.e., addition). The set R is not required to be a group under the second operation,[5] but the operation is required to satisfy the first two conditions for a group, namely that the operation is associative and that there exists an identity element. The second operation is denoted by "$\cdot$" (or adjacency, i.e., multiplication), so

$$(ab)c = a(bc)$$

[5]Some authors reserve the notion of an operation for the operation in a group and might use different terminology like a "law of composition" or a "way to combine elements" for the multiplication in a ring.

for all ring elements $a, b, c \in R$, and there is some element $1 \in R$ for which

$$1 \cdot a = a \cdot 1 = a$$

for every $a \in \mathbb{R}$. Finally, one substantially new condition is required in a ring relating the two operations, and this is called the distributive property:

$$a(b + c) = ab + ac \tag{2.10}$$

whenever $a, b, c \in R$. In summary,

(i) R is an Abelian group under addition.

(ii) R has multiplication which is associative and there exists a multiplicative identity.

(iii) Multiplication distributes across addition (2.10).

You may recall that the multiplicative identity in such a set is unique in the sense that if 1_1 and 1_2 are both multiplicative identities in a ring, then $1_1 = 1_2$. This does not mean the additive identity $0 \in R$ (which is also unique as an additive identity with respect to the additive group structure) has to be distinct from the multiplicative identity $1 \in \mathbb{R}$, but that is usually the case.

Exercise 2.18. Characterize all rings for which the additive identity and the multiplicative identity are the same element: $1 = 0$. By "characterize" I mean find every such ring and explain the algebraic structure of every such ring completely.

Note carefully that the multiplication in a ring is not required to be Abelian or commutative. In fact, $M_2(\mathbb{R})$ is a ring. As we know also the function $\Phi : M_2(\mathbb{R}) \to L(\mathbb{R}^2)$ provides a strong connection between these two additive groups. In fact $L(\mathbb{R}^2)$ is also a ring.

Exercise 2.19. Define the additive (group) structure on $L(\mathbb{R}^2)$ and recall that the multiplication operation is actually function composition. Show

$$\Phi(A + B) = \Phi(A) + \Phi(B).$$

Any ring is called a **commutative ring** if the ring is commutative under multiplication.

Here is a definition I will mention mostly because it should be mentioned; I don't know that we'll use it much except in the special case of the relation between matrices and linear functions.

Definition 2. Given two rings R_1 and R_2 a function $\phi : R_1 \to R_2$ is a **ring homomorphism** if the following holds:

$$\begin{aligned} &\phi(a \ +_1 \ b) = \phi(a) \ +_2 \ \phi(b) && \text{(additive group homomorphism)} \\ &\phi(a \ *_1 \ b) = \phi(a) \ *_2 \ \phi(b), && \text{and} \\ &\phi(1_1) = 1_2. \end{aligned}$$

A bijective ring homomorphism is a **ring isomorphism**.

Though function composition is always associative, many collections of functions fail to satisfy the distributive property and thus fail to be rings. The distributive property does work for linear automorphisms:

$$L \circ (L_1 + L_2) = L \circ L_1 + L \circ L_2$$

or correspondingly

$$\Phi(A) \circ (\Phi(B) + \Phi(C)) = \Phi(A(B + C)) = \Phi(A) \circ \Phi(B) + \Phi(A) \circ \Phi(C).$$

Notice $\mathbb{R}^n$ is not generally a ring due to the absence of a (natural) product of elements.

Collections of functions taking values in a ring are often rings themselves where the addition and multiplication are derived from the codomain ring:

$$(f + g)(x) = f(x) + g(x) \qquad \text{and} \qquad (fg)(x) = f(x)g(x).$$

There is another nice example of a pair of rings related by a ring isomorphism. On the one hand, there are polynomials considered as formal expressions

$$a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

where $n \in \mathbb{N}_0$ and $a_n \neq 0$. Here the coefficients $a_0, a_1, \ldots, a_n$, are taken to be in some given ring R. The highest power in a certain given polynomial of the form above is n, and that number is called the **degree** of the polynomial. The

polynomials of degree less than or equal to a fixed n and having coefficients in a given ring R is denoted by

$$R_n[x] = \left\{ \sum_{j=0}^{n} a_j x^j : n \in \mathbb{N}_0,\ a_0, \ldots, a_n \in R, a_n \neq 0 \right\}.$$

Though $R_n[x]$ is an additive group multiplication is not well defined when $n > 0$. For example $x \in R_n[x]$ when $n > 0$ and $x^n = 1$ $x^n \in R_n[x]$ but

$$x \cdot x^n = x^{n+1} \notin R_n[x],$$

so $R_n[x]$ is not a ring. This is remedied by taking

$$R[x] = \bigcup_{n=0}^{\infty} R_n[x]$$

the collection of all (formal) polynomials. There is a small problem in the discussion above which was manifest in the first line: The zero polynomial was not included.

Exercise 2.20. Correct the introduction of the polynomial symbols above by first setting

$$R_n^+[x] = \left\{ \sum_{j=0}^{n} a_j x^j : n \in \mathbb{N}_0,\ a_0, \ldots, a_n \in R, a_n \neq 0 \right\}$$

and

$$R_n[x] = R_n^+[x] \bigcup \{0\}.$$

Examine carefully how each of the following statements and definitions is affected, and show (your correctly defined) $R[x]$ is a ring.

On the other hand we may let $\mathcal{P}(R)$ denote the collection of all **polynomial functions**, i.e., functions $p : R \to R$ (automorphisms actually) with values given by

$$p(x) = \sum_{j=0}^{n} a_j x^j$$

for some $n \in \mathbb{N}_0$ and $a_0, a_1, \ldots, a_n \in R$. Here x is not considered a formal symbol but rather the argument of the function p in the domain R. There is no problem here with the zero polynomial.

Exercise 2.21. Show the coefficients $a_0, a_1, \ldots, a_n$ are, in some appropriate sense, uniquely determined.

Exercise 2.22. Show $\mathcal{P}(R)$ is a ring.

Exercise 2.23. Show $\phi : R[X] \to \mathcal{P}(R)$ by

$$\phi\left(\sum_{j=0}^{n} a_j X^j\right) = p$$

where $p : R \to R$ by

$$p(x) = \sum_{j=0}^{n} a_j x^j$$

when

$$\sum_{j=0}^{n} a_j X^j \in R[X] \backslash \{0\}$$

and $\phi(0)$ is the (constant) zero function when 0 is the zero symbol is a ring isomorphism.

Given an element $\alpha \in R$, one can calculate the **evaluation** of α with respect to either a formal polynomial symbol or with respect to a polynomial function:

$$p(\alpha) = \sum_{j=0}^{n} a_j \alpha^j.$$

An element $\alpha \in R$ is said to be a **zero** or **root** of a polynomial (or polynomial function) if the evaluation $p(\alpha) = 0 \in R$. Here one takes of course $p = \phi(P)$ if $P \in R[X]$.

It may be that a polynomial has no zeros. For example $X^2 + 1$ has no zeros in $\mathbb{R}[X]$.

Exercise 2.24. A ring R is not required to be a group under multiplication, but sometimes that happens. Characterize all rings that are groups under multiplication.

The collection R^* of elements in the ring with multiplicative inverses is always a group under multiplication. .

Exercise 2.25. What happens if $0 \in R^*$?

R^* is called the **multiplicative group** of the ring R.

2.1.6 Fields

A **field** is a ring F satisfying the following three conditions

(i) F is a commutative ring.

(ii) $1 \neq 0$, that is the zero (additive identity) and the multiplicative identity are not the same element.

(iii) The multiplicative group F^* is $F\backslash\{0\}$.

Another way to formulate condition **(ii)** is that $F \supsetneq \{0\}$ or $F \neq \{0\}$. This gives part of the broader algebraic context for the complex numbers. In particular, $\mathbb{C}$ has all the properties of a field. In fact, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are all commutative rings with zero element distinct from the multiplicative identity 1. Furthermore given any rational number p/q with $p \in \mathbb{Z}\backslash\{0\}$ and $q \in \mathbb{N}$, one can consider the number $q/p \in \mathbb{Q}$ when $p > 0$ and $(-q)/(-p)$ when $p < 0$ to see p/q has a multiplicative inverse making $\mathbb{Q}$ a field. A nonzero real number a has multiplicative inverse $1/a \in \mathbb{R}$, and as long as $a + bi \in \mathbb{C}\backslash\{0\}$, the complex numbers

$$\frac{a}{a^2+b^2} - i\,\frac{b}{a^2+b^2} \in \mathbb{C}$$

is a well-defined multiplicative inverse. (Check it.) Thus $\mathbb{R}$ and $\mathbb{C}$ are fields as well.

Here is another example of a field you may not have considered. Let $F = \{0, 1\}$ with addition determined by the "table"

$$\begin{aligned} 0 + 0 &= 0 \\ 0 + 1 &= 1 \\ 1 + 0 &= 1 \\ 1 + 1 &= 0 \end{aligned}$$

and multiplication table

$$\begin{aligned} 0 * 0 &= 0 \\ 0 * 1 &= 0 \\ 1 * 0 &= 0 \\ 1 * 1 &= 1. \end{aligned}$$

This field is called $\mathbb{Z}_2$ or $\mathbb{F}_2$, and it is the only field with two elements.

2.2 Linear space

The discussion of linear functions $L : \mathbb{R}^2 \to \mathbb{R}^2$ and the definition of linearity on page 36 in particular is somewhat deficient. A complete definition requires a more careful consideration of the algebraic structure on the domain and codomain. Such a definition is much easier to present now with the groundwork involving the definition of a field given above. Here is the relevant algebraic structure for linearity:

Definition 3. (linear space) Given a field F, a set V is a **linear space** over F if

(i) V is an Abelian group (under addition), and...

The field F is considered for the set V a field of **scalars** in the sense that there exists a **scaling**, i.e., a function with domain $F \times V$ and codomain V satisfying the following conditions:

(ii) $a(bv) = (ab)v$ for $a, b \in F$ and $v \in V$.

(iii) $1\ v = v$ for al $v \in V$ where $1 \in F$ is the multiplicative identity.

(iv) $(a + b)v = av + bv$ for all $a, b \in F$ and $v \in V$.

(v) $a(v + w) = av + aw$ for all $a \in F$ and $v, w \in V$.

Conditions **(iv)** and **(v)** are called the distributive properties.

Now the homogeneity condition

$$L(cv) = cL(v)$$

for $c \in F$ and $v \in V$ appearing on page 36 has a context, and a relatively broad context, in which to make sense.[6]

Exercise 2.26. Show $\mathbb{R}^2$ is a linear space over the field $\mathbb{R}$, and describe the scaling.

[6]Boas gives this definition in section 14 of her Chapter 3; also she gives a separate treatment of the Euclidean spaces $\mathbb{R}^n$ in section 10 of the same chapter.

Exercise 2.27. It is a bit unusual to do so, but $\mathbb{R}^2$ may be considered as a set with two operations in the following way: Denote by $\mathbb{R}^2_c$ the set $\mathbb{R}^2$ with the usual operation of componentwise addition and the operation of multiplication given by

$$(x, y)(\xi, \eta) = (x\xi - y\eta, x\eta + y\xi).$$

(a) Show $\mathbb{R}^2_c$ is a field.

(b) Show $\mathbb{R}^2_c$ is a vector space over $\mathbb{R}$.

In a certain sense, the point of this section is to mention that in this course I will primarily be starting with linear spaces over the field of complex numbers instead of the (probably) more familiar real linear spaces like $\mathbb{R}^n$. As mentioned at the beginning of this chapter, this is primarily because $\mathbb{C}$ is an algebraically complete field containing $\mathbb{R}$. I have still not explained what these words mean, but I will by the end of the chapter.

An example of a complex linear space is $\mathbb{C}^n$ consisting of ordered n-tuples of complex numbers. These will be linear spaces of primary interest.

2.3 Complex arithmetic

In sections 1-5 of Boas' Chapter 2 there is a nice (and in some context very useful) presentation of complex arithmetic. It may be very useful for you to take the time to go through these sections carefully and work many problems in complex arithmetic becoming both skilled in the arithmetic manipulations, terminology, and geometry of complex numbers. I will assume you can take responsibility for working on this material as needed. I assume many of you have mastered almost all of the material on complex arithmetic in sections 1-5. I will only make a few comments below and some additional parenthetical comments in various places when it might be reasonable to do so and include a few "straight arithmetic" problems in the first assignment. Otherwise, I will mostly pick up around Boas' section 6 of her Chapter 2.

The arctangent function (the real arctangent function) has domain $\mathbb{R}$ and is surjective onto the codomain interval $(-\pi/2, \pi/2)$. This function is increasing with positive derivative

$$\frac{d}{dx} \tan^{-1} x = \frac{1}{x^2 + 1}.$$

Exercise 2.28. Plot $\tan : (-\pi/2, \pi/2) \to \mathbb{R}$ and $\tan^{-1} : \mathbb{R} \to (-\pi/2, \pi/2)$, i.e., draw the graphs of these functions.

Consequently, the **argument** of a point (or vector)

$$(x, y)$$

in $\mathbb{R}^2$ or a point/number

$$x + iy$$

in $\mathbb{C}$ is given correctly (or accurately) by

$$\theta = \tan^{-1}\left(\frac{y}{x}\right)$$

only if $x > 0$. More generally, one can say

$$\theta = \tan^{-1}\left(\frac{y}{x}\right) + 2\pi k$$

as an argument for $x + iy$ if $x = \mathrm{Re}(x + iy) > 0$ and $k \in \mathbb{Z}$ is any integer. I will write down something rather more precise soon. Note now, however that a unique argument with $0 \leq \theta_0 < 2\pi$ is determined by

$$\begin{cases} \cos\theta_0 = x/\sqrt{x^2 + y^2} \\ \sin\theta_0 = y/\sqrt{x^2 + y^2} \end{cases} \tag{2.11}$$

as long as

$$\sqrt{x^2 + y^2} = |(x, y)| = |x + iy| \neq 0.$$

Every argument of $z = x + iy \in \mathbb{C}\backslash\{0\}$ has the form

$$\theta = \theta_0 + 2\pi k$$

for some $k \in \mathbb{Z}$ and θ_0 determined by (2.11). Also the restriction/specification $-\pi \leq \theta_0 < \pi$, or alternatively $-\pi < \theta_0 \leq \pi$, with

$$\begin{cases} \cos\theta_0 = x/|z| \\ \sin\theta_0 = y/|z| \end{cases}$$

may be used to obtain a similar conclusion.

Exercise 2.29. Use the real arctangent function $\tan^{-1} : \mathbb{R} \to (-\pi/2, \pi/2)$ to express the correct principle argument θ_0 with $0 \leq \theta_0 < 2\pi$ of any $z \in \mathbb{C}\backslash\{0\}$. Hint: Consider cases for $x = \mathrm{Re}\, z$ and $y = \mathrm{Im}\, z$.

2.3.1 some complex functions

We considered the complex square function $f : \mathbb{C} \to \mathcal{R}_2$ where $f(z) = z^2$ and $\mathcal{R}_2 = \Sigma_0 \cup \Sigma_1$ is an appropriate Riemann surface codomain. In order to understand this function (and especially the inverse) we must specify the image sheet and the associated branch cut(s). Taking

$$\Omega_0 = \{z \in \mathbb{C} : z \in \mathbb{R}^+ \text{ or } \operatorname{Im} z > 0\}$$

we may specify

$$f(z) = \begin{cases} z^2 \in \Sigma_0, & z \in \Omega_0 \\ z^2 \in \Sigma_1, & z \in \mathbb{C}\backslash\Omega_0. \end{cases}$$

Similar considerations apply to each function $f : \mathbb{C} \to \mathbb{C}$ with $f(z) = z^n$ for some $n \in \mathbb{N}$, though the Riemann surface may be different. Similar considerations may be applied to more complicated complex functions usually given by or associated with power series.

Before we consider power series, let me present briefly the idea of the complex derivative in the context of single power functions. Given $f : \mathbb{C} \to \mathbb{C}$ and $z_0 \in \mathbb{C}$,

$$f'(z_0) = \lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0}$$

if the limit on the right exists. Applying this definition with $f(z) = z^2$, we have

$$f'(z_0) = \lim_{z \to z_0} \frac{z^2 - z_0^2}{z - z_0} = \lim_{z \to z_0} z + z_0 = 2z_0.$$

Thus, the derivative values are given in terms of one lower power:

$$f'(z) = 2z.$$

Recall $\operatorname{Re} z^2 = x^2 - y^2$ and $\operatorname{Im} z^2 = 2xy$ where we have written $z = x + iy$ with $x, y \in \mathbb{R}$. Writing also $u = u(x, y) = \operatorname{Re} z^2$ and $v = v(x, y) = \operatorname{Im} z^2$, we observe

$$\operatorname{Re} f' = \frac{\partial u}{\partial x} \qquad \text{and} \qquad \operatorname{Im} f' = \frac{\partial v}{\partial x}.$$

Similar identities always hold. We are using here an important identification between $\mathbb{R}^2$ and $\mathbb{C}$, specifically as suggested above there is a one-to-one correspondence between $\mathbb{R}^2$ and $\mathbb{C}$ with the number $x + iy$ associated to the point $(x, y) \in \mathbb{R}^2$. Real valued functions of two variables like $u : \mathbb{R}^2 \to \mathbb{R}$ by

$$u(x, y) = x^2 - y^2$$

have, and are analyzed using, partial derivatives

$$\frac{\partial u}{\partial x}(x,y) = \lim_{h\to 0} \frac{u(x+h,y)-u(x,y)}{h} = 2x$$

and

$$\frac{\partial u}{\partial y}(x,y) = \lim_{h\to 0} \frac{u(x,y+h)-u(x,y)}{h} = -2y.$$

The second quantity, i.e., u_y, does not immediately appear in the expression

$$f'(z) = \frac{\partial u}{\partial x} + i\ \frac{\partial v}{\partial x}.$$

Exercise 2.30. Show that whenever the limit

$$f'(z) = \lim_{w\to z} \frac{f(w)-f(z)}{w-z}$$

exists, then

$$\operatorname{Re} f' = \frac{\partial u}{\partial x} \qquad \text{and} \qquad \operatorname{Im} f' = \frac{\partial v}{\partial x}.$$

Looking back at our example involving $f(z) = z^2$ we note that

$$\operatorname{Im} f'(z) = 2y.$$

Also,

$$\frac{\partial u}{\partial y} = -2y.$$

This too is not a coincidence.

Theorem 1. (Cauchy-Riemann equations) If $f : \mathbb{C} \to \mathbb{C}$ is complex differentiable and we define $u : \mathbb{R}^2 \to \mathbb{R}$ and $v : \mathbb{R}^2 \to \mathbb{R}$ by

$$u(x,y) = \operatorname{Re} f(x+iy) \qquad \text{and} \qquad v(x,y) = \operatorname{Im} f(x+iy),$$

then

$$f'(z) = f'(x+iy) = \frac{\partial u}{\partial x} + i\ \frac{\partial v}{\partial x} = \frac{\partial v}{\partial y} - i\ \frac{\partial u}{\partial y}.$$

In particular,

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \qquad \text{and} \qquad \frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y}.$$

This system of partial differential equations is called the system of Cauchy-Riemann equations.

A function which is complex differentiable on all of $\mathbb{C}$ is called an **entire function**.

2.4 functions from series

The real exponential function $\exp : \mathbb{R} \to (0, \infty)$ may be defined in a variety of ways including as the unique solution of the initial value problem

$$\begin{cases} y' = y, & x \in \mathbb{R} \\ y(0) = 1 \end{cases}$$

for a differentiable function $y : \mathbb{R} \to \mathbb{R}$. Another possibility is to use a real power series

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

The complex exponential function is usually defined in terms of a power series:

$$e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}. \tag{2.12}$$

I will assume a few things about real series and say/derive a few things about complex power series in general. One of my first main objectives is to carefully justify[7] the formula

$$e^z \; e^w = e^{z+w}$$

for complex numbers z and w. From this we can mostly understand how to analyze the complex exponential (and its inverse the complex logarithm) in much the same way we can understand complex powers.

2.4.1 series of numbers

There are many things to remember (From Boas' Chapter 1 for example) about real series and power series. I will mention only one or two directly: First a series of positive or nonnegative terms is either unbounded or bounded in the sense that if $a_1, a_2, a_3, \ldots$ are positive then the partial sums

$$\sum_{n=1}^{K} a_n$$

[7] Boas gives this, or something like this, as her Problem 2.8.1.

form an increasing sequence which is either bounded (above) or unbounded. In the former case

$$\sum_{n=1}^{\infty} a_n \in \mathbb{R} \tag{2.13}$$

is the least upper bound of the sequence. Second, associated with each real series which we can denote formally by

$$\sum_{n=1}^{\infty} a_j$$

giving simply identity to the terms $a_1, a_2, a_3, \ldots$ and indication of the fact that it is wished to sum them but implying nothing analogous to (2.13) about the representation of a real number, there is a series of nonnegative terms

$$\sum_{n=1}^{\infty} |a_j|.$$

If this series is convergent in the sense of the boundedness of (2.13), then the original series is said to be **absolutely convergent**, and in this case, there is a real number represented by

$$\sum_{n=1}^{\infty} a_n \in \mathbb{R}$$

in the sense of convergence of partial sums (reviewed in the complex case below). In the case of a series of complex terms

$$\sum_{n=1}^{\infty} w_n$$

with $w_1, w_2, w_3, \ldots$ in $\mathbb{C}$ much of the same terminology applies. In particular, one may replace the simple absolute value with the modulus (function) and again declare the formal series of complex terms to be absolutely convergent[8] if

$$\sum_{n=1}^{\infty} |w_n| \in \mathbb{R}.$$

[8]It turns out absolute convergence, though still a relatively special case of convergence for a series, in the case of complex power series considered below is usually adequate in that context as we will explain later.

The analogous implication also applies meaning there is some complex number σ for which

$$\lim_{K\to\infty} \sigma_K = \sigma \tag{2.14}$$

where

$$\sigma_K = \sum_{n=1}^{K} w_n$$

denotes the K-th partial sum. In this case, we elaborate on the condition (2.14) which can also be expressed analogously as

$$\sum_{n=1}^{\infty} w_n = \sigma \in \mathbb{C}.$$

This means that for any $\epsilon > 0$, there is some $N > 0$ so that

$$K > N \qquad \Longrightarrow \qquad |\sigma_K - \sigma| < \epsilon \tag{2.15}$$

where

$$\sigma_K = \sum_{n=1}^{K} w_n.$$

Another important rephrasing of the convergence condition is the following: The series value σ has for each K a decomposition

$$\sigma = \sigma_K + (\sigma - \sigma_K)$$

into terms the first of which is a finite sum, i.e., a sum of finitely many complex numbers. The condition (2.15) may then be interpreted as saying something about each of these terms. First of all condition (2.15) says that for large K the partial sum σ_K is close to the limiting value σ so that

$$|\sigma_K - \sigma| < \epsilon.$$

Essentially the same quantity

$$\sigma - \sigma_K = \sum_{n=K+1}^{\infty} w_n$$

is an infinite series representing the "tail" of the convergent series which is of course also small when K is large. Thus, the complex number σ decomposes

into two terms one of which is a finite (partial) sum close to the limit σ and the other is an infinite "tail" series close to zero. One may call this perspective on convergence "series decomposition."

Another aspect of series decomposition which is often useful is the observation that certain bounds hold independent of the index K determining the decomposition. Say we consider the partial sum σ_K for any K. By the convergence condition, there is some $N > 0$ for which

$$|\sigma_{K_1} - \sigma| < 1 \qquad \text{whenever } K_1 > N.$$

There are finitely many terms $\sigma_1, \sigma_2, \ldots, \sigma_N$ with

$$M = \max\{|\sigma_1|, |\sigma_2|, \ldots, |\sigma_N|\} + |\sigma| + 1$$

a finite positive number independent of K. If $K \leq N$, then clearly

$$|\sigma_K| < M.$$

If $K > N$, then by the triangle inequality

$$|\sigma_K| \leq |\sigma_K - \sigma| + |\sigma| < |\sigma| + 1 \leq M.$$

Consequently we have a "universal bound"

$$|\sigma_K| < M \tag{2.16}$$

independent of K.

The second term in the series decomposition also satisfies a universal bound

$$|\sigma_K - \sigma| \leq |\sigma_K| + |\sigma| < 2M \tag{2.17}$$

independent of K, and as mentioned above this second "tail series" term $\sigma - \sigma_K$ is close to $0 \in \mathbb{C}$ according to (2.15).

Exercise 2.31. Notice the first universal bound above implies the sequence $\{|\sigma_K|\}_{K=1}^{\infty}$ is bounded above. A very important property of the real numbers is that a sequence of real numbers which is bounded above has a unique well-defined **least upper bound** which is a real number. Denote the least upper bound of $\{|\sigma_K|\}_{K=1}^{\infty}$ by

$$M = \sup\{|\sigma_K|\}_{K=1}^{\infty} = \sup_K |\sigma_K|. \tag{2.18}$$

Note carefully the number M is independent of any particular index in the sequence $\{|\sigma_K|\}_{K=0}^{\infty}$ though the symbol "K" does happen to appear in the definition. The situation is similar to the definite integral

$$\int_0^1 x^2\,dx = \frac{1}{3}$$

which is a value independent of any particular value for x in the interval $[0,1]$ though the symbol x happens to appear in the expression for the integral value. One can properly though perhaps somewhat humorously say "there is no x in the integral," and in the same way, there is no (particular) index "K" in the definition of the least upper bound M. To beat this poor dead horse, one can simply change the symbol to capture the same meaning:

$$\int_0^1 \xi^2\,d\xi = \frac{1}{3} \qquad \text{and} \qquad M = \sup\{|\sigma_J|\}_{J=1}^{\infty} = \sup_J |\sigma_J|.$$

(a) Show $|\sigma| \leq M$.

(b) Show $|\sigma_K| \leq M$ independent of K.

(c) Show $|\sigma_K - \sigma| \leq 2M$ independent of K.

(d) Show this value of the universal bound M given in (2.18) satisfies

$$M < \max\{|\sigma_1|, |\sigma_2|, \ldots, |\sigma_N|\} + |\sigma| + 1$$

for every $N \in \mathbb{N}$. Note the strict inequality in (2.16) and (2.17) compared to the weak inequalities of **(b)** and **(c)**.

Here is a simple application of the technical definition of convergence: A complex series

$$\sum_{n=1}^{\infty} w_n$$

is convergent if and only if the real and imaginary parts (of the partial sums) are convergent real series.

First assume $w_n = a_n + ib_n$ so that

$$\sigma_K = \sum_{n=1}^{K} w_n = \sum_{n=1}^{K} a_n + i \sum_{n=1}^{K} b_n$$

with

$$\sum_{n=1}^{\infty} a_n = A \in \mathbb{R} \qquad \text{and} \qquad \sum_{n=1}^{\infty} b_n = B \in \mathbb{R}.$$

Let $\epsilon > 0$ and set $\sigma = A + iB \in \mathbb{C}$. By the convergence of the real and imaginary parts, there is some $N > 0$ for which

$$K > N \qquad \Longrightarrow \qquad \begin{cases} \left| \sum_{n=1}^{K} a_n - A \right| < \frac{\epsilon}{2} \\ \left| \sum_{n=1}^{K} b_n - B \right| < \frac{\epsilon}{2} \end{cases}$$

For $K > N$ we have then

$$|\sigma_K - \sigma| = \left| \sum_{n=1}^{K} w_n - \sigma \right| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \tag{2.19}$$

In the last line we used the triangle inequality for complex numbers. I will now pause to prove that inequality and elaborate on how it has been used in (2.19). For any $z, w \in \mathbb{C}$,

$$\begin{aligned} |z+w|^2 &= (z+w)\overline{(z+w)} \\ &= (z+w)(\bar{z}+\bar{w}) \\ &= z\bar{z} + z\bar{w} + w\bar{z} + w\bar{w} \\ &= |z|^2 + z\bar{w} + w\bar{z} + |w|^2. \end{aligned}$$

Note that generally $\overline{(zw)} = \bar{z}\bar{w}$ because

$$\begin{cases} (x+iy)(\xi+i\eta) = x\xi - y\eta + i\,(x\eta + y\xi) \\ (x-iy)(\xi-i\eta) = x\xi - y\eta + i\,(-x\eta - y\xi). \end{cases} \qquad \text{and}$$

Also, $\overline{\bar{w}} = w$, so

$$\overline{z\bar{w}} = \bar{z}\overline{\bar{w}} = \bar{z}w.$$

This means

$$|z+w|^2 = |z|^2 + z\bar{w} + \overline{z\bar{w}} + |w|^2.$$

But also generally $z + \bar{z} = 2\,\mathrm{Re}\, z$, so

$$|z+w|^2 = |z|^2 + 2\,\mathrm{Re}(z\bar{w}) + |w|^2.$$

Here is a another useful inequality

$$\begin{aligned} \mathrm{Re}(z\bar{w}) &\leq |\,\mathrm{Re}(z\bar{w})| && (2.20)\\ &= \sqrt{[\mathrm{Re}(z\bar{w})]^2} && (2.21)\\ &\leq \sqrt{[\mathrm{Re}(z\bar{w})]^2 + [\mathrm{Im}(z\bar{w})]^2} && (2.22)\\ &= |z\bar{w}|. && (2.23) \end{aligned}$$

Exercise 2.32. Distill/generalize the useful inequality expressed in (2.20-2.23). Hint: Your formulation should involve only a single complex number.

Using (2.23) we conclude

$$|z+w|^2 \leq |z|^2 + 2|z\bar{w}| + |w|^2.$$

Finally, there generally holds $|zw| = |z|\ |w|$ because

$$|zw|^2 = zw\overline{(zw)} = zw(\bar{z}\bar{w}) = z\bar{z}w\bar{w} = |z|^2|w|^2$$

where it may be noted that we have used the associativity and commutativity of the field $\mathbb{C}$. This means in particular $|z\bar{w}| = |z||\bar{w}| = |z||w|$, and

$$|z+w|^2 \leq |z|^2 + 2|z||w| + |w|^2 = (|z| + |w|)^2.$$

Since $|z+w|$ and $|z| + |w|$ are both nonnegative, this gives the triangle inequality:

$$|z+w| \leq |z| + |w|$$

for any complex numbers $z, w \in \mathbb{C}$. Returning to the estimate (2.19) and

adding some omitted steps

$$\begin{aligned}
|\sigma_K - \sigma| &= \left| \sum_{n=1}^{K} w_n - \sigma \right| \\
&= \left| \sum_{n=1}^{K} a_n + i \sum_{n=1}^{K} b_n - (A + Bi) \right| \\
&= \left| \sum_{n=1}^{K} a_n - A + i \left(\sum_{n=1}^{K} b_n - B \right) \right| \\
&\leq \left| \sum_{n=1}^{K} a_n - A \right| + \left| i \left(\sum_{n=1}^{K} b_n - B \right) \right| \\
&= \left| \sum_{n=1}^{K} a_n - A \right| + |i| \left| \sum_{n=1}^{K} b_n - B \right| \\
&= \left| \sum_{n=1}^{K} a_n - A \right| + \left| \sum_{n=1}^{K} b_n - B \right| \\
&< \frac{\epsilon}{2} + \frac{\epsilon}{2} \\
&= \epsilon.
\end{aligned}$$

We have shown that if the real and imaginary parts of (the partial sums of) a complex series are convergent, then the complex series is convergent.

For the converse, again setting $\operatorname{Re} w_n = a_n$ and $\operatorname{Im} w_n = b_n$

$$\left| \sum_{n=1}^{K} a_n - \operatorname{Re} \sigma \right| \leq \left| \sum_{n=1}^{K} w_k - \sigma \right|. \tag{2.24}$$

This is because

$$\operatorname{Re} \left(\sum_{n=1}^{K} w_n - \sigma \right) = \sum_{n=1}^{K} a_n - \operatorname{Re} \sigma$$

and I have also used Exercise 2.32. Thus, for any $\epsilon > 0$, there is some $N > 0$ for which $K > N$ implies the right side of (2.24) is less than ϵ. Evidently, the same inequality applies to the left side:

$$\left| \sum_{n=1}^{K} a_n - \operatorname{Re} \sigma \right| < \epsilon,$$

and this means the real part of σ_K converges to the real part of

$$\sum_{n=0}^{\infty} w_n \in \mathbb{C}.$$

The imaginary part of σ_K converges similarly to $\operatorname{Im}\sigma$.

More involved applications using the universal bounds will be given below. Here is one more very simple application.[9]

Theorem 2. If the series

$$\sum_{n-1}^{\infty} w_n$$

of complex numbers is convergent in $\mathbb{C}$, then

$$\lim_{n\to\infty} w_n = 0. \tag{2.25}$$

That is, convergence of a series implies the individual terms must tend to zero.

Proof: If it is assumed (2.25) fails, then there is some fixed $\epsilon > 0$ and an increasing sequence of indices $n_1, n_2, n_3, \ldots$ with

$$\lim_{k\to\infty} n_k = \infty \qquad \text{and} \qquad |w_{n_k}| \geq \epsilon.$$

On the other hand setting

$$\sigma = \sum_{n=1}^{\infty} w_n \in \mathbb{C}$$

there is some N for which

$$\left| \sum_{n=1}^{K} w_n - \sigma \right| < \frac{\epsilon}{2} \qquad \text{whenever } K > N.$$

[9]Boas calls this the "preliminary test." It is also variously called the "necessary condition for convergence" or the "basic test."

Taking some index n_k with $n_k > N+1$, we see from the triangle inequality

$$\begin{aligned}|w_{n_k}| &= \left|\sum_{n=1}^{n_k-1} w_n - \sigma + w_{n_k} - \left(\sum_{n=1}^{n_k-1} w_n - \sigma\right)\right| \\ &\leq \left|\sum_{n=1}^{n_k} w_n - \sigma\right| + \left|\sum_{n=1}^{n_k-1} w_n - \sigma\right| \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} \\ &= \epsilon\end{aligned}$$

since $n_k > n_k - 1 > N$. This contradicts the assumption that $|w_{n_k}| \geq \epsilon$.

We conclude there must exist some $\epsilon > 0$ for which there is some $K > 0$ so that

$$|w_n| < \epsilon \qquad \text{whenever } n > K.$$

That is,

$$\lim_{n\to\infty} w_n = 0.$$

2.4.2 absolute convergence

Certainly not all complex series are absolutely convergent with the sum of the moduli of the terms

$$\sum_{n=1}^{\infty} |a_n| < \infty$$

giving a finite real value. In the considerations below, as will be explained later, when one is considering power series the notion of absolute convergence is mostly adequate. Here is a key result:

Theorem 3. If

$$\sum_{n=1}^{\infty} |w_n| < \infty, \tag{2.26}$$

then

$$\sigma = \sum_{n=1}^{\infty} w_n \in \mathbb{C}.$$

In other words, the real convergence of the associated series of moduli/absolute values implies the convergence of a complex series.

Proof: Let $\operatorname{Re} w_n = a_n$ and $\operatorname{Im} w_n = b_n$ as usual. Since

$$\sum_{n=1}^{K} |a_n| \leq \sum_{n-1}^{K} |w_n|$$

we know

$$\sum_{n=1}^{\infty} |a_n| < \infty.$$

This means

$$A = \sum_{n=1}^{\infty} a_n \in \mathbb{R}$$

is a well-defined real number because absolute convergence of a real series implies convergence of that series. Similarly,

$$B = \sum_{n=1}^{\infty} b_n \in \mathbb{R},$$

and we have a complex number $\sigma = A + iB$. Estimate (2.19) now applies so that

$$\sum_{n=1}^{\infty} w_n \in \mathbb{C}.$$

2.4.3 complex power series

I think we are very much ready to return to the exponential series given in (2.12) and make some observations and conclusions about it. First of all, for each $z \in \mathbb{C}$,

$$\sum_{n=0}^{\infty} \left| \frac{z^n}{n!} \right| = \sum_{n=0}^{\infty} \frac{|z|^n}{n!} = e^{|z|}, \tag{2.27}$$

so the series for e^z is absolutely convergent. Consequently, the series expression for e^z determines a unique complex number. I could leave the convergence of (2.27) as a fact for you to verify from what you know about real series, but perhaps it will be helpful to offer some comments on that real convergence by way of review.

The ratio of successive terms

$$\frac{|z|^{n+1}}{(n+1)!} \Big/ \frac{|z|^n}{n!}$$

is $|z|/(n+1)$ which certainly for fixed $z \in \mathbb{C}$ tends to zero as n tends to infinity. Perhaps you know the ratio test and how that test can be related to geometric series. Perhaps you would like to consider those details. With $z \in \mathbb{C}$ fixed as mentioned above, there is some N for which

$$n \geq N \qquad \Longrightarrow \qquad \frac{|z|}{n} < 1.$$

For $K > N$ we can write

$$\sigma_K = \sum_{n=1}^{N-1} \frac{|z|^n}{n!} + \sum_{n=N}^{K} \frac{|z|^n}{n!}.$$

The first N terms in this sum

$$\sum_{n=1}^{N-1} \frac{|z|^n}{n!}$$

now just constitute some fixed nonnegative real number. Let us also assume $z \in \mathbb{C}\backslash\{0\}$, since the analysis of the series is very easy in the case $z = 0$ with the resulting series clearly convergent to $e^0 = 1$. If $z \neq 0$, the last $K - N + 1$ terms then satisfy

$$\begin{aligned}
\sum_{n=N}^{K} \frac{|z|^n}{n!} &= \sum_{n=0}^{K-N} \frac{|z|^{N+n}}{(N+n)!} \\
&< \frac{|z|^N}{N!} \sum_{n=0}^{K-N} \frac{|z|^n}{N^n} \\
&\leq \frac{|z|^N}{N!} \sum_{n=0}^{\infty} \left(\frac{|z|}{N}\right)^n \\
&= \frac{|z|^N}{N!} \frac{1}{1 - |z|/N}.
\end{aligned}$$

In the last line, I have summed a geometric series with finite sum because the ratio $|z|/N < 1$. Thus, we see the exponential series is absolutely convergent for all $z \in \mathbb{C}$ defining the exponential function $\exp : \mathbb{C} \to \mathbb{C}$ by $\exp(z) = e^z$.

I would next like to justify the exponential formula $e^z e^w = e^{z+w}$ for complex numbers z and w. I'll start with a formal calculation which is

probably what Boas has in mind with her Problem 2.8.1.

$$e^z e^w = \left(\sum_{k=0}^{\infty} \frac{z^k}{k!}\right)\left(\sum_{\ell=0}^{\infty} \frac{w^\ell}{\ell!}\right) \tag{2.28}$$

$$= \sum_{k=0}^{\infty}\left[\frac{z^k}{k!}\left(\sum_{\ell=0}^{\infty} \frac{w^\ell}{\ell!}\right)\right] \tag{2.29}$$

$$= \sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} \frac{z^k w^\ell}{k!\ell!} \tag{2.30}$$

$$= \sum_{k,\ell=0}^{\infty} \frac{z^k w^\ell}{k!\ell!} \tag{2.31}$$

$$= \sum_{n=0}^{\infty}\sum_{k=0}^{n} \frac{z^k w^{n-k}}{k!(n-k)!} \tag{2.32}$$

$$= \sum_{n=0}^{\infty}\frac{1}{n!}\sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array}\right) z^k w^{n-k} \tag{2.33}$$

where

$$\left(\begin{array}{c} n \\ k \end{array}\right) = \frac{n!}{k!(n-k)!}$$

is the arithmetic expression for the "combination of n things taken k at a time" (where order doesn't matter). Continuing from here one can complete the assertion because

$$\sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array}\right) z^k w^{n-k} = (z+w)^n$$

and then

$$e^z e^w = \sum_{n=0}^{\infty} \frac{1}{n!}(z+w)^n = e^{z+w}.$$

Next let me consider each line carefully starting with (2.28). This first line is not objectionable. Here one is simply substituting the definitions of e^z and e^w. The series appearing here are well-defined complex numbers, and they can be multiplied. This is just what $e^z e^w$ means. The second and third lines are also essentially okay. In (2.29) the second factor is distributed to the terms of the first sum. Perhaps we have not gone through the details

of extending the distributive property to this context, but that extension is straightforward. In (2.30) again factors are simply distributed to the terms in the second sum. This is fine. One may start to be concerned about the order of the terms at this point with the double sum, but there is still a definite order: Each inner sum (infinite sum) is evaluated to obtain a specific complex number, then the outer sum of these complex numbers is calculated. If there is a justification for the distributive property, then this should all be fine.

Things start to go wrong in (2.31) where there is no longer any specified order for the summing, or at least one may assume no order is specified. Boas might suggest the (arbitrary) rearrangement of the terms here is justified by absolute convergence. On the one hand, we have not verified absolute convergence for either series

$$\sum_{k=0}^{\infty}\left|\sum_{\ell=0}^{\infty}\frac{z^k w^\ell}{k!\ell!}\right|$$

nor

$$\sum_{k,\ell=0}^{\infty}\frac{|z|^k|w|^\ell}{k!\ell!}$$

whatever the order of the terms may be in the second series. Furthermore, Boas' "fact" stated in section 9 of her Chapter 1 was stated only for real series and also given no justification even in that case. She seems to then take an analogous result for granted for complex series in Chapter 2 without even a statement. Maybe this is okay, but I think one can and should say something here. It is also somewhat interesting to see exactly how the rearrangement works which is, more or less, suggested by the expression in (2.32) though again there is some ambiguity. Essentially the final line (2.33) is just distribution again and the definition of the combination. From there the binomial expansion formula is used and the formula follows.

In order to clarify what happened in the "weak" steps in this calculation and give a better justification for them, I proceed using series decomposition as suggested above. There are several series to which the construction may be applied:

$$e^z,\ e^w,\ e^{|z|},\ e^{|w|},\ e^{z+w},\ e^{|z+w|},\ e^{|z|+|w|}$$

is a list of obvious ones; some are real valued, some are complex valued. All are absolutely convergent. One important thing to remember is that each

exponential expression represents a series with an associated decomposition. I will start with the real series

$$e^{|z|} = \sum_{k=0}^{\infty} \frac{|z|^k}{k!}.$$

Let $K \in \mathbb{N}$. Then

$$\sum_{k=0}^{K-1} \frac{|z|^k}{k!}$$

satisfies the universal estimate

$$\sum_{k=0}^{K-1} \frac{|z|^k}{k!} \leq M_1 \tag{2.34}$$

independent of K for some $M_1 > 0$. Also, given any $\epsilon > 0$, there is some N_1 for which the quantity

$$\left| \sum_{k=0}^{K-1} \frac{|z|^k}{k!} - e^z \right| = \left| \sum_{k=K}^{\infty} \frac{|z|^k}{k!} \right| < \epsilon \tag{2.35}$$

when $K > N_1$. In view of (2.34) and the triangle inequality, one has also the universal bound

$$\left| \sum_{k=0}^{K-1} \frac{|z|^k}{k!} - e^{|z|} \right| \leq 2M_1$$

independent of K.

Observe first that this real series yields estimates for

$$\sigma_K = \sum_{k=0}^{K-1} \frac{z^k}{k!} \qquad \text{and} \qquad \sigma_K - e^z = -\sum_{k=K}^{\infty} \frac{z^k}{k!}$$

as well. Specifically,

$$|\sigma_K| \leq \sum_{k=0}^{K-1} \frac{|z|^k}{k!} \leq M_1,$$

$$|\sigma_K - e^z| \leq \left| \sum_{k=0}^{K-1} \frac{|z|^k}{k!} - e^z \right| \leq 2M_1,$$

independent of K, and

$$|\sigma_K - e^z| \leq \sum_{k=K}^{\infty} \frac{|z|^k}{k!} < \epsilon \tag{2.36}$$

when $K > N_1$.

Similar assertions hold for

$$e^w = \sum_{\ell=0}^{\infty} \frac{w^\ell}{\ell!} - \mu_K \qquad \text{and} \qquad e^{|w|} = \sum_{\ell=0}^{\infty} \frac{|w|^\ell}{\ell!} - \sum_{\ell=0}^{K-1} \frac{|w|^\ell}{\ell!}$$

involving second bounds M_2 and N_2 where of course

$$\mu_K = \sum_{\ell=0}^{K-1} \frac{w^\ell}{\ell!}.$$

Notice that if we set $M_3 = \max\{M_1, M_2\}$ and $N_3 = \max\{N_1, N_2\}$, then we have all the estimates

$$|\sigma_K| \leq \sum_{k=0}^{K-1} \frac{|z|^k}{k!} \leq M_3, \tag{2.37}$$

$$|\mu_K| \leq \sum_{\ell=0}^{K-1} \frac{|w|^\ell}{\ell!} \leq M_3, \tag{2.38}$$

$$|\sigma_K - e^z| \leq \left| \sum_{k=0}^{K-1} \frac{|z|^k}{k!} - e^z \right| \leq 2M_3,$$

$$|\mu_K - e^w| \leq \left| \sum_{\ell=0}^{K-1} \frac{|w|^\ell}{k!} - e^w \right| \leq 2M_3,$$

independent of K, with

$$|\sigma_K - e^z| \leq \sum_{k=K}^{\infty} \frac{|z|^k}{k!} < \epsilon \tag{2.39}$$

and

$$|\mu_K - e^w| \leq \sum_{\ell=K}^{\infty} \frac{|w|^\ell}{\ell!} < \epsilon \tag{2.40}$$

when $K > N_3$.

We make one more application of series decomposition applied to $e^{|z|+|w|}$. From this we obtain M_4 for which

$$\left|\sum_{n=0}^{K-1}\frac{(z+w)^n}{n!}\right| \leq \sum_{n=0}^{K-1}\frac{(|z+w|)^n}{n!} \leq \sum_{n=0}^{K-1}\frac{(|z|+|w|)^n}{n!} \leq M_4,$$

and

$$\left|\sum_{n=K}^{\infty}\frac{(z+w)^n}{n!}\right| \leq \sum_{n=K}^{\infty}\frac{(|z+w|)^n}{n!} \leq \sum_{n=K}^{\infty}\frac{(|z|+|w|)^n}{n!} \leq 2M_4.$$

We also obtain N_4 for which $K > N_4$ implies

$$\left|\sum_{n=K}^{\infty}\frac{(z+w)^n}{n!}\right| \leq \sum_{n=K}^{\infty}\frac{(|z+w|)^n}{n!} \leq \sum_{n=K}^{\infty}\frac{(|z|+|w|)^n}{n!} \leq \epsilon. \tag{2.41}$$

Finally, we take $M = \max\{M_3, M_4, 1\}$ and $N = \max\{N_3, N_4\}$ so versions of all the above estimates hold with M and N appropriately in place of the various preliminary constants M_3, M_4, N_3 and N_4. The strictly positive quantity 1 is used to define M in order to avoid possible degeneracy associated with division by M later.

At this point, we make one last observation and adjustment to some of the "tail" estimates above. The universal bounds M_1, M_2, M_3, M_4, and consequently M are all independent of both any index K, any tolerance ϵ, and any cutoff N_1, N_2, N_3, or N_4 or N, appearing above. Because of this we may choose a different positive tolerance depending on these universal bounds. Accordingly, we start with $\epsilon > 0$ as we did before, but make the choice

$$A(\epsilon) = \min\left\{\frac{\epsilon}{5}, \frac{1}{2}, \frac{\epsilon}{5M}\right\} \tag{2.42}$$

so that (2.35) and (2.36) may be replaced with

$$|\sigma_K - e^z| \leq \sum_{k=K}^{\infty}\frac{|z|^k}{k!} < A(\epsilon) \tag{2.43}$$

for $K > N_1$. Similarly (2.40) is replaced with

$$|\mu_K - e^w| \leq \sum_{\ell=K}^{\infty}\frac{|w|^\ell}{\ell!} < A(\epsilon) \tag{2.44}$$

for $K > N_3 \geq N_1$. Next and last (2.41) is replaced with

$$\left| \sum_{n=K}^{\infty} \frac{(z+w)^n}{n!} \right| \leq \sum_{n=K}^{\infty} \frac{(|z+w|)^n}{n!} \leq \sum_{n=K}^{\infty} \frac{(|z|+|w|)^n}{n!} < A(\epsilon) \tag{2.45}$$

for $K > N \geq N_3$.

The choice of the curious and currently mysterious quantity $A(\epsilon)$ in (2.42) will be better motivated by the end of the argument, but one may simply note now that any positive quantity is allowed.

For any fixed K consider

$$e^z e^w = \left(\sum_{k=0}^{K-1} \frac{z^k}{k!} + \sum_{k=K}^{\infty} \frac{z^k}{k!} \right) \left(\sum_{\ell=0}^{K-1} \frac{w^\ell}{\ell!} + \sum_{\ell=K}^{\infty} \frac{w^\ell}{\ell!} \right)$$

and set

$$\nu_K = \sum_{n=0}^{K-1} \frac{(z+w)^n}{n!}.$$

Using the symbols σ_K, μ_K, and ν_K for the appropriate partial sums one finds

$$e^z e^w - e^{z+w} = [\sigma_K - (e^z - \sigma_K)][\mu_K - (e^w - \mu_K)] - [\nu_K - (e^{z+w} - \nu_K)].$$

The product

$$\begin{aligned} &[\sigma_K - (e^z - \sigma_K)][\mu_K - (e^w - \mu_K)] \\ &\quad = \sigma_K \mu_K - \sigma_K (e^w - \mu_K) - \mu_K (e^z - \sigma_K) + (e^z - \sigma_K)(e^w - \mu_K) \end{aligned} \tag{2.46}$$

contains the interesting term or (sub)product

$$\sigma_K \mu_K = \left(\sum_{k=0}^{K-1} \frac{z^k}{k!} \right) \left(\sum_{\ell=0}^{K-1} \frac{w^\ell}{\ell!} \right).$$

This is a product of finite sums, so rearrangements are completely unambiguous and relatively easy to understand. In fact, the product

$$\sum_{k,\ell=0}^{K-1} \frac{z^k w^\ell}{k!\ell!} \tag{2.47}$$

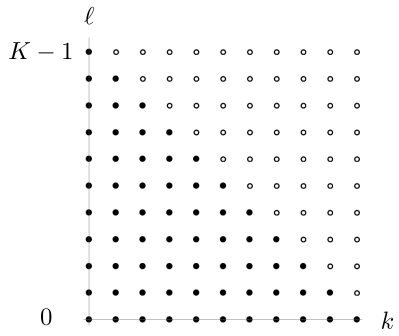

Figure 2.1: Finitely many lattice points $\{(0,0),(0,1),\ldots,(K-1,K-1)\}$ in a square corresponding to K^2 terms in a finite sum of complex numbers.

has exactly K^2 terms corresponding to the lattice points in $\{0,1,2,\ldots,K-1\}\times\{0,1,2,\ldots,K-1\}$ which may be visualized in the k,ℓ-coordinate plane as indicated in Figure 2.1.

If we consider the terms

$$\sum_{k=0}^{K-1}\sum_{\ell=0}^{K-1-k}\frac{z^k w^\ell}{k!\ell!}$$

corresponding to the lower triangle of lattice points, we see that since $0\le\ell\le K-1-k$ we have $k\le k+\ell\le K-1$. In particular, every pair of points (k,ℓ) in the lower triangle corresponds to some integer $n=k+\ell$ with $0\le n\le K-1$.

Furthermore, for each integer n with $0\le n\le K-1$ and each k with $0\le k\le n$ there is a unique $\ell=n-k$ for which $k+\ell=n$, and the pair (k,ℓ) satisfies $0\le\ell=n-k\le K-1-k$. This means (k,ℓ) is in the lower triangle. We conclude

$$\sum_{k=0}^{K-1}\sum_{\ell=0}^{K-1-k}\frac{z^k w^\ell}{k!\ell!}=\sum_{n=0}^{K-1}\sum_{k=1}^{K-1}\frac{z^k w^{n-k}}{k!(n-k)!}=\sum_{n=0}^{K-1}\frac{1}{n!}\sum_{k=1}^{K-1}\left(\begin{array}{c}n\\k\end{array}\right)z^k w^{n-k}=\nu_K.$$

This accounts for $K(K+1)/2$ of the K^2 terms in the sum (2.47) and captures an important part of what was happening in the formal calculation discussed at the beginning. Specifically, the terms corresponding to the lower triangle

give exactly ν_K. In a certain sense it is the remaining terms

$$\sum_{k=1}^{K-1}\sum_{\ell=K-k}^{K-1}\frac{z^k w^\ell}{k!\ell!}$$

corresponding to the (slightly smaller) upper triangle of lattice points marked by circles in Figure 2.1 the formal calculation ignored. We have shown then

$$\begin{aligned}
e^z e^w - e^{z+w} &= \sigma_k\mu_k - \nu_K \\
&\qquad - \sigma_K(e^w - \mu_K) - \mu_K(e^z - \sigma_K) + (e^z - \sigma_K)(e^w - \mu_K) \\
&\qquad\quad + (e^{z+w} - \nu_K) \\
&= \sum_{k=1}^{K-1}\sum_{\ell=K-k}^{K-1}\frac{z^k w^\ell}{k!\ell!} \\
&\qquad - \sigma_K(e^w - \mu_K) - \mu_K(e^z - \sigma_K) + (e^z - \sigma_K)(e^w - \mu_K) \\
&\qquad\quad + (e^{z+w} - \nu_K).
\end{aligned}$$

It follows that

$$\begin{aligned}
\left|e^z e^w - e^{z+w}\right| &\le \left|\sum_{k=1}^{K-1}\sum_{\ell=K-k}^{K-1}\frac{z^k w^\ell}{k!\ell!}\right| \\
&\qquad + |\sigma_K(e^w - \mu_K)| + |\mu_K(e^z - \sigma_K)| + |(e^z - \sigma_K)(e^w - \mu_K)| \\
&\qquad\quad + \left|e^{z+w} - \nu_K\right|.
\end{aligned}$$

If we can show each of the remaining terms

$$A_1 = \left|\sum_{k=1}^{K-1}\sum_{\ell=K-k}^{K-1}\frac{z^k w^\ell}{k!\ell!}\right|,$$

$$A_2 = |\sigma_K(e^w - \mu_K)|,$$

$$A_3 = |\mu_K(e^z - \sigma_K)|,$$

$$A_4 = |(e^z - \sigma_K)(e^w - \mu_K)|,$$

and

$$A_5 = \left|e^{z+w} - \nu_K\right|$$

is smaller than $\epsilon/5$ when K is large enough, then we obtain

$$\left|e^z e^w - e^{z+w}\right| < \epsilon$$

for every $\epsilon > 0$, and the identity $e^z e^w = e^{z+w}$ for complex exponentials must hold.

Taking $K > N$, the estimate for A_5 is given in (2.45) since $A(\epsilon) \leq \epsilon/5$.

The quantity A_4 involves the product of two "tails" each of which according to (2.43) and (2.44) is smaller in modulus than $A(\epsilon)$. Since $A(\epsilon) < 1$ we have

$$A_4 < [A(\epsilon)]^2 < A(\epsilon) \leq \frac{\epsilon}{5}.$$

For the term A_3, we may use the universal bound (2.38) along with (2.43) to conclude

$$A_3 = |\mu_K(\sigma_K - e^z)| \leq M\ A(\sigma) < \frac{\epsilon}{5}.$$

Similarly, (2.37) and (2.44) imply

$$A_2 = |\sigma_k(\mu_K - e^w)| \leq M\ A(\sigma) < \frac{\epsilon}{5}.$$

It remains to estimate the sum appearing in A_1.

$$A_1 = \left| \sum_{k=1}^{K-1} \sum_{\ell=K-k}^{K-1} \frac{z^k w^\ell}{k!\ell!} \right| \leq \sum_{k=1}^{K-1} \sum_{\ell=K-k}^{K-1} \frac{|z|^k |w|^\ell}{k!\ell!}.$$

Note this is still a finite sum and for each pair of indices (k, ℓ) there holds $k + \ell \geq K$ and $k + \ell \leq 2K - 2 = 2(K-1)$. This means each term appears in

$$\sum_{k=0}^{2(K-1)} \sum_{\ell=0}^{2(K-1)} \frac{|z|^k |w|^\ell}{k!\ell!} - \sum_{k=0}^{K-1} \sum_{\ell=0}^{K-1-k} \frac{|z|^k |w|^\ell}{k!\ell!}$$

where the second sum which has been extracted is the same sum corresponding to the lower triangle of lattice points discussed above. The remaining terms are all terms $|z|^k |w|^\ell / (k!\ell!)$ with $K \leq k + \ell \leq 2(K + \ell)$, and we have

$$\begin{aligned} A_1 &\leq \sum_{n=K}^{2(K+1)} \sum_{k=K}^{2(K+1)} \frac{|z|^k |w|^{n-k}}{k!(n-k)!} \qquad (2.48) \\ &= \sum_{n=K}^{2(K+1)} \frac{1}{n!} \left(\begin{array}{c} n \\ k \end{array} \right) |z|^k |w|^{n-k} \\ &= \sum_{n=K}^{2(K+1)} \frac{(|z| + |w|)^n}{n!}. \end{aligned}$$

The last finite sum of positive terms is clearly bounded above by the infinite sum

$$\sum_{n=K}^{\infty} \frac{(|z| + |w|)^n}{n!}$$

appearing in (2.45) from which we see

$$A_1 < A(\epsilon) \leq \frac{\epsilon}{5}$$

as required to complete the argument.

Exercise 2.33. Make a diagram of lattice points like that appearing in Figure 2.1 but representing the terms in the sum on the right in (2.48).

I will finish the immediate discussion by extending some of what was observed concerning the series for the complex exponential function to other complex valued function of a complex variable defined by series. This justifies the comment made earlier that the consideration of absolutely convergent complex series is often adequate for consideration of complex series.

Theorem 4. If a power series

$$\sum_{n=0}^{\infty} \alpha_n (z - z_0)^n$$

has

$$\sum_{n=0}^{\infty} \alpha_n (z_1 - z_0)^n$$

convergent (in any way) for some $z_1 \in \mathbb{C}$, then

$$\sum_{n=0}^{\infty} \alpha_n (z - z_0)^n$$

is absolutely convergent for every $z \in \mathbb{C}$ with $|z - z_0| < |z_1 - z_0|$.

Proof: We start by applying the basic necessary condition for convergence which says that if

$$\sum_{n=0}^{\infty} \alpha_n (z_1 - z_0)^n$$

converges then the individual terms must tend to zero:

$$\lim_{n\to\infty} \alpha_n (z_1 - z_0)^n = 0.$$

Thus, we can take N large enough so that $|\alpha_n||z_1 - z_0|^n < 1$ whenever $n \geq N$. For $K > N$ and $|z - z_0| < |z_1 - z_0|$ we then have

$$\sum_{n=0}^{K} |\alpha_n||z - z_0|^n = \sum_{n=0}^{N-1} |\alpha_n||z - z_0|^n + \sum_{n=N}^{K} |\alpha_n||z_1 - z_0|^n \frac{|z - z_0|^n}{|z_1 - z_0|^n}.$$

The first sum on the right is a fixed nonnegative real number. The second sum may be viewed as a partial sum for the series

$$\sum_{n=N}^{\infty} |\alpha_n||z - z_0|^n$$

of nonnegative terms which is bounded above by

$$\sum_{n=N}^{\infty} \left| \frac{z - z_0}{z_1 - z_0} \right|^n = \left| \frac{z - z_0}{z_1 - z_0} \right|^N \sum_{n=0}^{\infty} \left| \frac{z - z_0}{z_1 - z_0} \right|^n.$$

Note that

$$\sum_{n=0}^{\infty} \left| \frac{z - z_0}{z_1 - z_0} \right|^n$$

is a geometric series with ratio

$$\left| \frac{z - z_0}{z_1 - z_0} \right| < 1$$

and sum

$$\sum_{n=0}^{\infty} \left| \frac{z - z_0}{z_1 - z_0} \right|^n = \frac{1}{1 - \left| \frac{z - z_0}{z_1 - z_0} \right|} < \infty.$$

2.5 More (semiabstract) algebra

I may have mentioned above that the complex field $\mathbb{C}$ plays an important role in linear algebra and ordinary differential equations. Simply the fact

that $\mathbb{C}$ is a field containing $\mathbb{R}$ gives an initial indication of the importance. A more complete explanation arises through the consideration of polynomials.

Given any field F, the polynomial ring $F[x]$ and the isomorphic ring $\mathcal{P}(F)$ of polynomial functions with coefficients in F are linear spaces over F. In each case, it also makes sense to evaluate polynomials at points in F, i.e., to consider $p(a)$ with $a \in F$. It has also mentioned that a particular element $a \in F$ might be a zero for p or it might not.

Definition 4. A field F is said to be **algebraically complete** if every nonconstant polynomial $p \in \mathcal{P}(F)$ has a zero in F.

In these terms the special property of $\mathbb{C}$ can be made clear. The field $\mathbb{R}$ is not algebraically complete. The polynomial function with $p(x) = x^2 + 1$ never vanishes, and there are many other nonconstant polynomials with no real roots. Every nonconstant polynomial function $p \in \mathcal{P}(\mathbb{C})$ however does have a root. This fact has a name:

Theorem 5. *(fundamental theorem of algebra) Every nonconstant polynomial with complex coefficients has a zero in $\mathbb{C}$. The field $\mathbb{C}$ is algebraically complete.*

I would like to prove the fundamental theorem of algebra which is central to the study of linear algebra, for example, linear automorphisms

$$L : \mathbb{C}^n \to \mathbb{C}^n,$$

and complex valued functions $X : \mathbb{R} \to \mathbb{C}$ which are solutions of linear ordinary differential equations.

One can use **continuity** and a little understanding of complex functions to prove the fundamental theorem of algebra. Specifically, if we imagine there is a complex polynomial function $f : \mathbb{C} \to \mathbb{C}$ with values given by

$$f(z) = \sum_{n=0}^{k} \alpha_n z^n$$

for some $k > 0$ with $\alpha_k \neq 0$ and $f(z) \neq 0$ for every $z \in \mathbb{C}$, then first of all we may assume this polynomial satisfies $\alpha_k = 1$.

Exercise 2.34. Show that if $f : \mathbb{C} \to \mathbb{C}$ is a polynomial function with the properties above, namely

$$f(z) = \sum_{n=0}^{k} \alpha_n z^n, \qquad \alpha_k \neq 0, \qquad \text{and} \qquad f(z) \neq 0, \qquad \text{for} \qquad z \in \mathbb{C},$$

then the polynomial function $g : \mathbb{C} \to \mathbb{C}$ with

$$g(z) = \sum_{n=0}^{k} \beta_n z^n,$$

where $\beta_n = \alpha_n/\alpha_k$ satisfies $\beta_k = 1$ and $g(z) \neq 0$ for $z \in \mathbb{C}$.

Thus, we assume

$$f(z) = z^k + \sum_{n=0}^{k-1} \alpha_n z^n.$$

A circle with center $0 \in \mathbb{C}$ and radius $r > 0$ is given by

$$\Gamma_r = \{r(\cos t + i \sin t) = re^{it} : 0 \leq t \leq 2\pi\}.$$

We will consider the image of such a circle

$$f(\Gamma_r) = \{f(re^{it}) : 0 \leq t \leq 2\pi\}$$

for r on some interval $[\epsilon, R]$ where ϵ is some very small value and R is some very large value. Consider first the small value $r = \epsilon$. In this case,

$$|f(\epsilon e^{it})| = \left| \alpha_0 + \sum_{n=1}^{k} \alpha_n \alpha_n \epsilon^n e^{int} \right|, \tag{2.49}$$

and

$$\begin{aligned} |\alpha_0| &= \left| \alpha_0 + \sum_{n=1}^{k} \alpha_n \alpha_n \left(\epsilon e^{it}\right)^n - \sum_{n=1}^{k} \alpha_n \alpha_n \epsilon^n e^{int} \right| \\ &= \left| f(\epsilon e^{it}) - \sum_{n=1}^{k} \alpha_n \alpha_n \epsilon^n e^{int} \right| \\ &\leq |f(\epsilon e^{it})| + \left| \sum_{n=1}^{k} \alpha_n \alpha_n \epsilon^n e^{int} \right| \\ &\leq |f(\epsilon e^{it})| + \epsilon \sum \sum_{n=1}^{k} |\alpha_n| \epsilon^{n-1}. \end{aligned} \tag{2.50}$$

Note that $\alpha_0 = f(0) \neq 0$, so

$$|\alpha_0| > 0.$$

On the other hand, if we take $\epsilon < 1$ or even ϵ **very small**, which is sometimes indicated notationally by $\epsilon << 1$, then $\epsilon^{n-1} \leq 1$ for $n = 1, 2, \ldots, k$ so

$$\epsilon \sum \sum_{n=1}^{k} |\alpha_n| \epsilon^{n-1} \leq \epsilon \sum \sum_{n=1}^{k} |\alpha_n| = \epsilon M$$

where

$$M = \sum \sum_{n=0}^{k} |\alpha_n|$$

is some fixed nonnegative real number. In fact, since we know $\alpha_k = 1$ we know $M \geq 1$. In particular if

$$\epsilon < \frac{|\alpha_0|}{2M},$$

then using the inequality in (2.50) we have

$$|f(\epsilon e^{it}| > |\alpha_0| - \frac{|\alpha_0|}{2} = \frac{|\alpha_0|}{2} > 0.$$

2.6 Liouville's theorem

Theorem 6. *(Liouville's theorem) A bounded entire function, that is a function $f : \mathbb{C} \to \mathbb{C}$ which is complex differentiable and for which there exists some $M > 0$ with*

$$|f(z)| \leq M \qquad \text{for all } z \in \mathbb{C}$$

must be a constant function with $f(z) \equiv w_0$ for some $w_0 \in \mathbb{C}$.

Chapter 3

Linear Algebra

As mentioned above, linear algebra is the study of linear functions $f : V \to W$ where V and W are linear spaces. We will focus primarily on the special case $L : \mathbb{C}^n \to \mathbb{C}^n$ where $V = W$ is the complex vector space $\mathbb{C}^n$ and L is a linear operator.

3.1 Complex vector spaces

3.2 Polynomials