Appendix B: Mathematical Background

T.J. Bogdan

September 16, 2018

1. Introduction

Mathematics has frustrated generations of aspiring physical scientists for a variety of reasons. I believe that it does not have to be this way.

Perhaps at the heart of this issue is that physical scientists typically want to *use* mathematics to understand a real-universe problem, while mathematicians find intrinsic beauty in the subject and could care less whether one can do anything practical with it or not. G.H. Hardy's haunting memoir A Mathematician's Apology, is not in fact an apology for finding the subject beautiful in its own right, or having spent not one moment's thought on applying it to any practical benefit for humankind.

Ostensibly another reason for this state of affairs is that the mathematical landscape is very curiously connected with "underground tunnels", or "chutes and ladders", linking what would often appear to be very disparate intellectual locations. It is somewhat akin to opening the door of an old wooden rustic Biergarten in Bavaria and finding yourself surrounded by drying chilis at Georgia O'Keefe's ranch in Abiquiu, New Mexico.

And mathematicians who teach mathematics courses do not help physical scientists by prefacing their remarks with indications of whether what they will say next is *useful* in solving a problem, or merely *beautiful* in its meaning.

My goal in this Appendix is to provide you my own "Lonely Planet"—or for those old enough among you to remember such things, "Baedeker's"—guide to the mathematical landscape which underlies RMHD. It's nothing new or revelatory. But it is personal and comes from my many years of getting lost, over and over and over, again.

Here too, it is the journey, not the destination that counts.

2. Sets and Mappings

Almost all the mathematics that we will need can be thought of in terms of three basic concepts: (*i*) *sets, groups, collections, classes, organizations* of (*ii*) *elements, objects, things, entities* and (*iii*) the *mappings, associations, relations, connections, correspondences* between *them*—*them* being equivalently the organizations or the stuff that belongs in them. I've used a number of different words here to describe the individual "things", the "organizations" they belong to and the "connections" between them and/or the organizations of which they are members. I'll tend to use these various descriptors interchangeably. Rigorous and logical mathematical analysis usually has very precise definitions to distinguish between what precisely is meant by these different words. This is especially true of a *set*, which is a very basic mathematical construct.

A *set*, loosely speaking, is simply an assembly, or collection of objects, or things, or concepts, or whatever really, which we refer to as *elements*. I will try to consisently use caligraphic capital letters $\mathcal{A}, \mathcal{B}, \mathcal{C}$, etc, to denotes sets. If

an element, say x, belongs to a set $\mathcal{V}$ we write, $x \in \mathcal{V}$. Sets by themselves are very egalitarian, no element of a set is any better or any worse than any other element. One way to specify a set is to list all of its elements between curly brackets, such as

$$\mathcal{H} = \{\text{red}, \text{white}, \text{blue}\} \ ,$$

$$\mathcal{B} = \{\clubsuit, \diamondsuit, \spadesuit, \heartsuit\} \ ,$$

$$\mathcal{A} = \{a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z\} \ .$$

Order does not matter, only content; for example,

$$\mathcal{H} = \{\text{blue}, \text{white}, \text{red}\}$$

as well.

Listing is not the only way. Indeed there are some sets for which it would be impossible to list all their elements, for example, the set of all counting numbers, which has a special symbol reserved for it:

$$\mathbb{N} \equiv \{1, 2, 3, 4, 5, \cdots\} \ .$$

Notice we *can't* possibly list all of the elements of this set, although because of our familiarity with such things we know that 276 is a member of this set somewhere off in the dots.

Sometimes a set can be defined by what *property* or *attribute* its elements share, for example

$$\mathcal{V} = \{a, e, i, o, u\} \ ,$$

can also be written equivalently as

$$\mathcal{V} = \{\alpha \in \mathcal{A} \mid \alpha \text{ is a vowel}\}$$

which is read as "the set of all elements α in $\mathcal{A}$ such that α is a vowel". The vertical bar is read as "such that", the statement to the left of the bar indicates what larger set the elements are to be selected from, and the statement on the right of the bar provides the property the elements must have to be in the set $\mathcal{V}$. Notice that α stands for a *generic* element of $\mathcal{A}$. This works only if someone has provided us with a larger sample set from which to select α's based on an attribute.

This property-defines-set approach is fraught with several logical difficulties such as what after all is the bigger set from which we shall select objects that have a certain property in common. And even worse, how can we state a property (or properties) unambiguously so there is no possible source of confusion. This can lead to a variety of paradoxes or conundrums, such as, does the set of all sets contain itself as a member. There are ways to straighten such things out (by declaring that the set of all sets is not actually a set), formally at least, but we shall set these logical issues aside and move on.

A *mapping*, say φ, can be thought of as a relationship, association, correspondence or assignment that connects elements selected from one set $\mathcal{D}$—which we

refer to as the *domain*—to another set $\mathcal{R}$—which we call the *range*. In general, we express this idea by writing

$$\varphi : \mathcal{D} \to \mathcal{R} \ ;$$

φ maps $\mathcal{D}$ to $\mathcal{R}$. The only property that we insist a mapping must have is that a particular element of $\mathcal{D}$, say x, can be associated with *one and only one* element of $\mathcal{R}$. Call this element, y. This elemental association is expressed as

$$\varphi : x \mapsto y \ , \quad \text{or} \quad \varphi(x) = y \ .$$

Therefore, we admit the possibility that many different elements of $\mathcal{D}$ can be mapped to the same element of $\mathcal{R}$.

It may be that $\mathcal{D}$ and $\mathcal{R}$ are in fact the same set! We have to define what it means for two sets to be "equal", but the correct and sensible definition is that they must contain the same elements:

$$\mathcal{D} = \mathcal{R} \iff (\forall x, \ x \in \mathcal{D} \Rightarrow x \in \mathcal{R}) \wedge (\forall y, \ y \in \mathcal{R} \Rightarrow y \in \mathcal{D}) \ ,$$

which reads "$\mathcal{D}$ is identical to $\mathcal{R}$ if and only if, for every x in $\mathcal{D}$, x is also in $\mathcal{R}$, and, for every y in $\mathcal{R}$, y is also in $\mathcal{D}$".

Another way to express the same idea is to think in terms of *subsets*. The vowel set, $\mathcal{V}$, is obviously a subset of the entire alphabet $\mathcal{A}$, but the converse is not true. Symbolically, we write

$$\mathcal{V} \subset \mathcal{A} \ .$$

Another way to state this is that every element in $\mathcal{V}$ is also in $\mathcal{A}$. Obviously

$$\mathcal{A} \subset \mathcal{A} \ .$$

Two sets are therefore equal if each is a subset of the other

$$\mathcal{D} = \mathcal{R} \iff (\mathcal{D} \subset \mathcal{R}) \wedge (\mathcal{R} \subset \mathcal{D}) \ .$$

These last two equations give a glimpse of the syntax of mathematical logic. For example, we read $\iff$ as "if and only if", meaning the implication works in both directions. Whereas $\Rightarrow$ works only in one direction, and is read as "implies"—the converse implication need not (although it may) be true. The symbol $\wedge$ is the logical "and"—while $\vee$ is the logical "or". Finally, $\forall$ means "for every".

Although mappings from $\mathcal{D} \to \mathcal{R}$ have access to all of the elements of $\mathcal{R}$ they may not in fact use all of $\mathcal{R}$—the part, or *subset*, of $\mathcal{R}$ that they use is called the *image* of $\mathcal{D}$ in $\mathcal{R}$ under φ. Mappings that use all of $\mathcal{R}$ are said to be *surjective* or are a *surjection*. And mappings that associate one, and only one, element of $\mathcal{R}$ with an element of $\mathcal{D}$, are said to be *injective*, or an *injection*. Mappings that possess both of these properties are particularly important, and are called *bijections*. They are also said to be one-to-one and onto. A bijection φ has the

very imporant property that its inverse, say $\varphi^{-1} : \mathcal{R} \to \mathcal{D}$, or $\varphi^{-1}(y) = x$ is also a bijection. That is

$$\varphi : x \mapsto y \ , \qquad \Longleftrightarrow \qquad \varphi^{-1} : y \mapsto x \ .$$

Finally, notice, that any mapping $\varphi : \mathcal{D} \to \mathcal{R}$, also *defines* a set! It can be thought of as the set, say $\mathcal{P}$, of elements that are all the ordered pairs (x, y), where $x \in \mathcal{D}$ and $y \in \mathcal{R}$, and for which $\varphi(x) = y$ [and $\varphi^{-1}(y) = x$, if φ is a bijection].

Two sets are said to be *equivalent* if there is a bijection that connects them. Thus the set of colors $\mathcal{H}$ and the set

$$\mathbb{N}_3 = \{1, 2, 3\} \subset \mathbb{N}$$

are equivalent, although they certainly are not equal. An equivalence is an example of a mathematical construct called a *relationship*. The sets $\mathcal{H}$ and $\mathbb{N}_3$ are also equivalent to the set of three fruits

$$\mathcal{F} = \{\text{banana}, \text{apple}, \text{orange}\} \ ,$$

but they are not equivalent to $\mathcal{B}, \mathcal{V}$ or $\mathcal{A}$

Unlike mappings, relationships can connect one object, say the set of three colors, to more than one object, the set of three fruits *and* the set containing the first three counting integers. The collection of all objects that are equivalent form an *equivalence class*. Obviously the property that distinguishes this equivalence class is that each set contains just three distinct objects. So here is a first instance where mathematicians choose to make a distinction between "mapping" and "relation", and "set" and "class".

This notion of equivalence classes, allows us to define the size, or more precisely, the *cardinality* of a set as follows. For sets with a finite number of elements, say m, there is an obvious bijection from the set to $\mathbb{N}_m$, and so the cardinality of the set, or the cardinal number of the set, is just m. Sets that are not finite, like $\mathbb{N}$ for example, are said to be infinte. The cardinal number of a set is *infinite* (i.e., not finite, or *trans*finite) if there is a bijection from the set to a *proper* subset of itself. A proper subset is a subset that is not identical to the set itself—the vowels are a proper subset of the alphabet, but the alphabet is not a proper subset of itself.

To be concrete, consider again the set of counting numbers

$$\mathbb{N} = \{1, 2, 3, \cdots\}$$

and a proper subset that consists only of the even numbers

$$\mathbb{E} = \{2, 4, 6, \cdots\} \ .$$

Then the "doubling" mapping $\varphi : n \mapsto 2n$ is a bijection from $\mathbb{N}$ to $\mathbb{E}$, and obviously $\mathbb{E} \subset \mathbb{N}$. Therefore, both $\mathbb{N}$ and $\mathbb{E}$ are equivalent (but certainly not equal), and they belong to the same equivalence class, and have the same non-finite cardinality. We denote this by the first transfinite cardinal number $\aleph_0$ or

equvalently $\mathfrak{d}$ for $\mathfrak{d}$enumerable. In otherword, we can "count" or sequentially write out this first infinity, although we cannot actually ever get to the end of the list, we always know what comes next.

The cardinality (or size) of the set of all integers

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

and the set of all rational numbers $\mathbb{Q}$ is also $\aleph_0$. This is because one can find bijections between both of these sets and $\mathbb{N}$. (Try it!)

In a certain sense, $\aleph_0$ is the smallest "infinity"—sometimes called the countable infinity, because we can at least "list" a countable infinites worth of elements, as we have above. Are there "bigger" infinities? Yes, indeed, in fact, very many of them!

To explore this a little further we take a slight detour. From the set $\mathcal{H}$ of colors we can construct a bigger set which contains as its elements all the possible distinct subsets of $\mathcal{H}$. Specifically, this set is

$$\{\{\}, \{\text{red}\}, \{\text{white}\}, \{\text{blue}\}, \{\text{red}, \text{white}\}, \{\text{red}, \text{blue}\}, \{\text{blue}, \text{white}\}, \{\text{red}, \text{white}, \text{blue}\}\} \ .$$

The curious set

$$\emptyset \equiv \{\}$$

which contains no elements whatsoever is called the *empty set* and it has the property that it is a subset of every set you can think of (including itself). If you count them up, you will see that the cardinality of all the subsets of $\mathcal{H}$ is $8 = 2^3$, where the cardinality of $\mathcal{H}$ is 3. The cardinality of the set of all card suits $\mathcal{B}$ is 4, and if you make a list of all the subsets of $\mathcal{B}$ you will find there are $2^4 = 16$ of them. All the subsets of the vowels leads to $2^5 = 32$, and you definitely do not want to try to list all the $2^{26} = 67,108,864$ subsets of the alphabet.

In every case, we see that the set of all subsets of a given set has a greater cardinal number than the original set. Therefore the set of all subsets of $\mathbb{N}$ is undoubtedly bigger that $\mathbb{N}$ and its cardinality can be represented by $2^{\aleph_0}$ which is also written as $\mathfrak{c}$ for $\mathfrak{c}$ontinuum. The set of all real numbers, which we denote by $\mathbb{R}$, and which we cannot possibly list between curly brackets, is equivalent to the set of all subsets of $\mathbb{N}$ and so it is a bigger infinity than the size of $\mathbb{N}$. The set of all subsets of $\mathbb{R}$, if you can wrap your mind around that, must be a bigger infinity still, and so there is an endless parade of bigger and bigger transfinite cardinal numbers.

Besides determining how big a set is, the next very useful thing you can do is to *order* a set and remove the egalitarianism. Because there is an obvious bijection for all sets of finite cardinality to the set $\mathbb{N}_m$, all finite sets can be ordered in any number of different ways. And, of course, we know from practical experience that the sets $\mathbb{Z}$ and $\mathbb{R}$ can be ordered because we know things like $2.78654... > -56.67877...$ are true. It must be said that there is a lot of mathematical rigor that underlies a statement like this which we commonly take for granted. It rests upon the concept of $\mathbb{Z}$ having a proper subset of *positive* integers, say $\mathbb{Z}^+$, a proper subset of *negative* integers, say $\mathbb{Z}^-$, and one lonely

identity element, 0, in a proper subset by itself. Then we can be precise and say things like

$$n > m \iff n - m \in \mathbb{Z}^+ \ .$$

Set's, like $\mathbb{Z}$ can not only be ordered, but some can be *well-ordered*—which means that any subset of $\mathbb{Z}^+$ has a least or smallest element.

Some additional useful set-theoretic concepts involve binary operations on sets themselves. The *intersection* of two sets, denoted $\bigcap$, takes two sets, $\mathcal{A}$ and $\mathcal{B}$ and makes a third set, say $\mathcal{C}$ from them according to

$$\mathcal{C} = \mathcal{A} \bigcap \mathcal{B} = \{x | x \in \mathcal{A} \ \wedge \ x \in \mathcal{B}\} \ ,$$

or in otherwords, it is the set of all elements x that are in $\mathcal{A}$ *and* in $\mathcal{B}$ as well. Clearly $\mathcal{C}$ must be a subset of both $\mathcal{A}$ *and* $\mathcal{B}$,

$$\mathcal{C} = \mathcal{A} \bigcap \mathcal{B} \implies \mathcal{C} \subset \mathcal{B} \ \wedge \ \mathcal{C} \subset \mathcal{A} \ .$$

Notice that the arrow of implication goes only in one direction. That is because a mutual subset, $\mathcal{C}$ could be smaller than the full intersection of $\mathcal{A}$ and $\mathcal{B}$. If there is no element x that is in both $\mathcal{A}$ and $\mathcal{B}$ we write

$$\emptyset = \mathcal{A} \bigcap \mathcal{B},$$

and say that $\mathcal{A}$ and $\mathcal{B}$ are *disjoint*. Thus, for logical consistency, we require that the empty set is a subset of every set, including $\mathcal{A}$ and $\mathcal{B}$.

If intersections are in some sense "exclusive", *unions*, denoted $\bigcup$, are their "inclusive" counterpart. We say

$$\mathcal{D} = \mathcal{A} \bigcup \mathcal{B} = \{x | x \in \mathcal{A} \ \vee \ x \in \mathcal{B}\} \ .$$

Notice that all we have done relative to intersection is flipped the "and" for an "or". To be in the union of two sets, all you need is to be in one or the other of the sets (you can also be in both). Thus

$$\mathcal{D} = \mathcal{A} \bigcup \mathcal{B} \implies \mathcal{A} \subset \mathcal{D} \ \wedge \ \mathcal{B} \subset \mathcal{D} \ .$$

Again, the implication goes one way because we could have a $\mathcal{D}$ that is bigger than the union of $\mathcal{A}$ and $\mathcal{B}$. The intersection of two sets is also a subset of the union of two sets.

Order does not matter in intersections and unions of sets, and indeed, we can form the intersections and unions of many sets sequentially, where again the order in which we pair them up does not matter.

The final concept is the idea of the *complement* of a subset. Unfortunately there is no standard notation for this. To make matters clear, consider the set of consonants,

$$\mathcal{C} = \{b, c, d, f, g, h, j, k, l, m, n, p, q, r, s, t, v, w, x, y, z\} \ ,$$

$$\mathcal{C} = \{\alpha \in \mathcal{A} \mid \alpha \text{ is a consonant}\} \ ,$$
$$\mathcal{C} = \{\alpha \in \mathcal{A} \mid \alpha \text{ is not a vowel}\} \ .$$

We say $\mathcal{C}$, the consonants, are the *complement* of the vowels $\mathcal{V}$ in the bigger set of the entire alphabet $\mathcal{A}$, and we write

$$\mathcal{C} = \mathcal{A} \backslash \mathcal{V} \ .$$

and it must also be true that

$$\mathcal{V} = \mathcal{A} \backslash \mathcal{C} \ .$$

The intersection of a proper subset (of some bigger set) and its complement is the empty set. The union of a proper subset (of some bigger set) and its complement is the bigger set itself. A collection of disjoint proper subsets of some bigger set whose union is the bigger set itself, is called a *partition* of the bigger set. For example, the set (of two sets)

$$\{\mathcal{C}, \mathcal{V}\}$$

is a partition of the alphabet $\mathcal{A}$. And all of the following sets (of sets) are different partitions of the colors $\mathcal{H}$

$$\{\{\text{red}\}, \{\text{white}\}, \{\text{blue}\}\} \ ,$$
$$\{\{\text{red}\}, \{\text{blue}, \text{white}\}\} \ ,$$
$$\{\{\text{white}\}, \{\text{red}, \text{blue}\}\} \ .$$

(Are there any others?) Therefore, any given set, like $\mathcal{A}$, can have many different partitions, but fewer partitions than individual subsets.

3. Algebraic Structures

Another direction you can go is to begin to introduce *structure* on a set. Typical of algebraic structure is a *binary operation* that acts upon ordered pairs of elements of a set and associates with each orderd pair a third element, which is also in the set. We've used the word "operation" here to distinguish it from mappings which we typically think of connecting a *single* element from a set with another element in that, or perhaps some other, set.

There are various properties, or laws, that one would like such a binary operation—denoted here generically by $\circ$—to possess. In order of increasing sophistication, we can list the following familiar laws:

$$I.\ \textit{Closure Law}: \qquad \forall\ x, y \in \mathcal{V}, \qquad x \circ y \in \mathcal{V} \ .$$
$$II.\ \textit{Associative Law}: \quad \forall\ x, y, z \in \mathcal{V}, \qquad x \circ (y \circ z) = (x \circ y) \circ z \ .$$
$$III.\ \textit{Identity}: \quad \exists\ e_\circ \in \mathcal{V} \mid \forall x \in \mathcal{V}, \qquad e_\circ \circ x = x \circ e_\circ = x \ .$$
$$IV.\ \textit{Inverse}: \quad \forall\ x \in \mathcal{V},\ \exists\ x_\circ \in \mathcal{V} \mid \qquad x \circ x_\circ = x_\circ \circ x = e_\circ \ .$$
$$V.\ \textit{Commutative Law}: \qquad \forall\ x, y \in \mathcal{V}, \qquad x \circ y = y \circ x \ .$$

Read the symbol $\exists$ as "there exists", and as before $|$ as "such that". A set which is endowed with a single binary operation satisfying the first four of these laws is called a *group*. And if the fifth law is also valid, it is called a *commutative*, or an *Abelian* group. You might take some scrap paper and convince yourself that $\mathbb{R}$ is a group under both addition and multiplication separately, and that $\mathbb{Z}$ is a group under addition only but not multiplication.

A set that is equipped with a single binary operation that is only closed is called a *magma*. There is not too much you can do, or say, about magma's to be honest—you have to introduce a little more sturcture before you get to something that is rich in properties and behaviors. An associative magma is called a *semi-group*. A semi-group that comes equipped with an identity element is a *monoid*. A monoid in which each element has an (unique) inverse is a *group*.

The additional stucture imposed upon a set by the group properties (Laws I.-IV.) proves to be very powerful. There is one and only one group which contains one, two and three elements, respectively. There are two distinct groups with four elements, and again, only one group with five elements. All of these groups have the additional, but by no means necessary property that the order of multiplication is irrelevant, i.e., Law V.: $x \circ y = y \circ x$ holds true. When we look at groups with six elements we find our first so-called *non-Abelian* group, where $x \circ y \neq y \circ x$.

Groups with finite cardinality are often associated with geometric symmetries. For example the dihedral group D_4 which has $2 \times 4 = 8$ elements is equivalent to the 4 rotations by 90 degrees and the 4 reflections that are all the possible transformations which map a square to a square in two dimensions. Groups with transfinite cardinality are also very important and can be related, as we shall see, to the continuous symmetries of space-time. These groups are also called *Lie Groups*.

Although every group, regarded as a set, has many subsets—in fact 2^n of them if n is the cardinal number of the group—not every subset is in fact a *subgroup*. A subgroup, say $\mathcal{U}$ of a group $\mathcal{V}$ is a proper subset of $\mathcal{V}$ which is a group in its own right—that is, it satisfies the Laws I.-IV. on its own with no help necessary from the rest of $\mathcal{V}$. This means the identity element must be in every subgroup of a larger group. Many groups have no subgroups at all, except the trivial subgroup: the set consisting only of the identity element by itself. For example, groups whose cardinal number is a prime number can have no subgroups. Therefore, if we found some way to endow the color hues $\mathcal{H}$, or the vowels with a group structure, they could have no subgroups. The card suits, the consonants and the alphabet are not prevented from haveng subgroups by this theorem.

This raises the interesting question of precisely how one might endow these sets with a group structure, and how many different ways there might be to do so. Both questions have interesting answers. For every set of finite cardinality we can find a way to construct a group from it. And for some values of the cardinality (but, interestingly, not all) there can be more than one way to impose this algebraic structure!

The *cyclic groups of order* n, sometimes denoted by $\mathbb{Z}_n$, provide a constructive answer to the first question. One useful representation for $\mathbb{Z}_n$ is

$$\mathbb{Z}_n = \{0, 1, 2, \cdots, n-3, n-2, n-1\} \ .$$

The usual addition of integers provides our binary operation, but with one interesting twist—whenever addition gives us a number that is bigger than $n-1$, i.e., that appears not to be in our group, we subtract off n to return it to the set. To be concrete consider the case $n = 4$, and

$$\mathbb{Z}_4 = \{0, 1, 2, 3\} \ .$$

To ensure this is a group we require, 1+3=2+2=0, and 2+3=1, instead of 4 and 5, which are not in our group. This is an Abelian group. The identity is 0, and the inverse of 1 is 3, and 2 is its own inverse. Indeed, $\mathbb{Z}_4$ contains a non trivial subgroup,

$$\{0, 2\} \subset \mathbb{Z}_4 \ .$$

Notice that this subgroup "behaves" in exactly the same fashion as the cyclic group

$$\mathbb{Z}_2 = \{0, 1\}$$

which uses the rule 1+1=0 in place of 1+1=2.

The term "behaves" can be given a precise mathematical meaning—*isomorphic*—which we shall get to presently. Another set of 4 elements that "behaves" in exactly the same fashon as $\mathbb{Z}_4$ is

$$\{1, a, a^2, a^3\}$$

where our binary operation is now multiplication. Here $aa = a^2$ and $aaa = a^3$ and to make this a group we require $aaaa = a^4 = 1$. You can verify that $a = \sqrt{-1}$, will do the trick, so $a^2 = -1$ and $a^3 = -\sqrt{-1}$. Note the structure preserving correspondence, or *isomorphism*,

$$0 \leftrightarrow 1, \qquad 1 \leftrightarrow a, \qquad 2 \leftrightarrow a^2, \qquad 3 \leftrightarrow a^3 \ ,$$

between our two representations of the cyclic group $\mathbb{Z}_4$. So, for all intents and purposes

$$\{1, a, a^2, a^3\}$$

is just another representation of $\mathbb{Z}_4$.

However, the so-called *Klein 4-group*,

$$\{1, a, b, ab\}$$

with the property $aa = a^2 = bb = b^2 = 1$ is also a group, but it does not behave like $\mathbb{Z}_4$ at all. The group's fourth element is the product of a and b, which can be written either as ab or ba, since this too is an Abelian group. Therefore $abab = (ab)^2 = abba = aa = 1$. Each element of the group that is not the identity is its own inverse, and the product of any two non-identity elements is

the third non-identity element. The Klein group can also be throught of as the dihedral group of order 2, D_2. The general dihedral group is

$$D_n = \{1, a, a^2, a^{n-1}, b, ba, ba^2, ba^{n-1}\} \ ,$$

where $a^n = b^2 = 1$ and $aba = b$.

There are no other distinct groups of cardinal number 4, and the cyclic group of order 5 is the only group with cardinal number 5. At cardinal number of 6, we again have one more group in addition to the cyclic group of order 6, and this is the first non-Abelian group, the dihedral group of order 3:

$$D_3 = \{1, a, a^2, b, ab, ba\} \ ,$$

where $a^3 = b^2 = 1$, and $aba = b$. One might be tempted to guess that when n is not a prime, there are two distinct groups, but this is not true. For $n = 8$ and 12, there are 5 distinct groups, and only the cyclic group for $n = 15$. Thereafter all rhyme or reason seems to go out the window as there are 14 distinct groups for $n = 16$.

Much has been written about, and indeed much can be deduced about the properties of groups, which we will not bother to repeat here. From this fundamental concept of a set-as-an-algebraic-group there are basically two directions one can go. One can add additional sets into the picture—which we do in the next section—or one can introduce additional binary operations, which we do below.

If we now introduce a *second* independent binary operation—say $\boxdot$—that maps pairs of elements from the group to another element in the group we have more complicated alegbraic stuctures called *rings*, *integral domains*, and (*algebraic*, as opposed to gravitational, electromagnetic, or radiation) *fields*. A *field* is a set $\mathcal{V}$ equipped with *two* binary operations—usually called addition ($\circ$) and multiplication ($\boxdot$)—such that Laws I.-V. are satisfied for each operation separately (the identity element for multiplication $e_\boxdot \neq e_\circ$), as well as the two additional laws

$$\textit{VI. Distributive Law}: \quad \forall\, x, y, z \in \mathcal{V}, \qquad x \boxdot (y \circ z) = (x \boxdot y) \circ (x \boxdot z) \ ,$$

$$\textit{VII. Cancellation Law}: \quad \forall\, x, y \in \mathcal{V}, \quad x \boxdot y = e_\circ \iff x = e_\circ \ \wedge \ y = e_\circ .$$

Usually, we simplify the notation by taking $\circ = +$ and dropping $\boxdot$ entirely in favor of juxtaposition. The identities are then usually abbreviated as $e_\circ =$ "0", and $e_\boxdot =$ "1", so we have the rational and real numbers in mind as the paradigm of an algebraic field.

Bijections that preserve the algebraic structure of two different sets are incredibly important and are given a special name: *isomorphisms*. Let $\varphi : \mathcal{D} \to \mathcal{R}$ be an isomorphism between two algebraic fields, then

$$a + b = c \ \text{ in } \mathcal{D} \iff \varphi(a) + \varphi(b) = \varphi(c) \ \text{ in } \mathcal{R} \ ,$$

$$ab = c \ \text{ in } \mathcal{D} \iff \varphi(a)\varphi(b) = \varphi(c) \ \text{ in } \mathcal{R} \ ,$$

and so forth. An isomorphism from an alegbraic structure to itself is called an *automorphism*. The identity mapping $\varphi(a) = a$, is the *trivial* automorphism. Nontrivial automorphisms, when they exist, are related to the symmetries of the algebraic structure and play a very important role.

When an isomorphism exists between two algebraic structures they can be regarded as equivalent. Take, for example our cyclic group,

$$\mathbb{Z}_3 = \{0, 1, 2\} \ .$$

This is an algebraic (number) field if we treat addition and multiplication in the usual fashion, *except* whenever we end up with a "3" as the result of a calculation, we replace it with "0", and a "4" is replaced by a "1". Because there are bijections from this algebraic number field to the set of three colors and three fruits, we can use the algebra of $\mathbb{Z}_3$ to endow an algebraic structure to the set of colors or fruits and render the bijection an isomorphism. In other words, we can build a logically consistent structure (indeed more than one of them) where it makes sense to say what the additive inverse of a banana is! Some authors will choose to write this algebraic field as

$$\mathbb{Z}_3 = \{\overline{0}, \overline{1}, \overline{2}\} \ ,$$

to alert the reader that the usual laws of addition and multiplication have to be modified by casting out multiples of "3". This is also known as *modular arithmetic*, thus

$$2 + 2 = 1 \text{ mod } 3$$

is equivalent to

$$\overline{2} + \overline{2} = \overline{1} \ .$$

Lest you think that all of this is fairly empty in the sense that anything can be made into anything, consider the next bigger cyclic group

$$\mathbb{Z}_4 = \{0, 1, 2, 3\} \ .$$

Given our success with its smaller friend, we might try to make this into a field by the usual replacement trick of mod 4 arithmetic:

$$4 \to 0 \ , 5 \to 1, \text{ and } 6 \to 2 \ .$$

This does *not* a field make, however, because $2 \cdot 2 = 0$ which violates the cancellation law (however, it is still a group with respect to modular addition and multiplication separately). One of two alternatives must be true: either (i) there is no way to make $\mathbb{Z}_4$ into a field by any means, or (ii) a different arithmetic than the one we selected above is required to make a field with just 4 elements. There is a very beautiful general result here that tells us that $\mathbb{Z}_n$ can be made into an algebraic field if and only if n is any prime number raised to an integer power. If that integer power is 1, like $n = 3$ for example, the standard trick we used with $\mathbb{Z}_3$ will work for both addition and multiplication. If that power is not

1, as in $n = 4 = 2^2$, then alternative (ii) is true, but we must find a different arithmetic than the standard trick. You might amuse yourself by selecting 0 and 1 as the (additive and multiplicative) identity elements and replacing 2 by a generic symbol a. Now, convince yourself that 3 has to be the multiplicative inverse of a, call it $a^{-1} \neq a$. Then use all the laws that govern a field to determine what $1+a$, $1+a^{-1}$, and $a+a^{-1}$ equal. Finally, notice that there was only one way to endow $\mathbb{Z}_4$ with a field structure, and because $6 = 2 \cdot 3$, no matter how hard you try you will not succeed in making $\mathbb{Z}_6$ into a field! Therefore, the alphabet $\mathcal{A}$ cannot be made into an algebraic field, and it is pointless to enquire as to the multiplicative inverse of a consonant (but not a vowel, because $\mathbb{Z}_5$ can be made into a algebraic field).

4. Vector Spaces

We can build even more complicated and useful mathematical structures by combining *two* sets with *four* binary operations. A *Vector space*, $\mathcal{V}(\mathbb{F})$, *over* an algrebraic field $\mathbb{F}$ is just such a object.

The vector space itself, is the set $\mathcal{V}$, whose elements are called vectors, which we will denote in boldface— e.g., $\mathbf{x} \in \mathcal{V}$. There is a single binary operation defined on $\mathcal{V}$, called *vector addition*, $+$, which satisfies the five laws of a group. This means there is an identity element for addition, which we denote by $\mathbf{0}$, and each vector $\mathbf{x}$ has an additive inverse, which we will write as $-\mathbf{x}$.

In addition to $\mathcal{V}$, we have an auxiliary set of scalars $\mathbb{F}$ which is usually the field of real numbers $\mathbb{R}$, but which can in fact be any (algebraic) field. It comes equipped with *two* more binary operations, being the usual *scalar addition* and *scalar multiplication*. In so far as scalar addition is a different operation than vector addition, if one were being very meticulous, it might be useful to use something different than the "+" sign to denote this operation. To distinguish the vectors from the scalars, we'll use lower case Greek letters for the scalars. The fourth and final binary operation is different from what we have seen before, it involves multiplying a scalar times a vector resulting in another vector. Again, if we were being careful, we would avoid using simple juxtaposition for this binary operation since this also stands for multiplication of scalars by scalars. However, these distinctions become moot if we require

$$\forall\ \alpha \in \mathbb{F}, \forall\ \mathbf{x} \in \mathcal{V}, \qquad \alpha\mathbf{x} = \mathbf{x}\alpha \in \mathcal{V}\ ,$$

$$\forall\ \alpha, \beta \in \mathbb{F}, \forall\ \mathbf{x}, \mathbf{y} \in \mathcal{V}, \qquad \alpha(\mathbf{x} + \beta\mathbf{y}) = \alpha\mathbf{x} + \alpha(\beta\mathbf{y}) = \alpha\mathbf{x} + (\alpha\beta)\mathbf{y}\ ,$$

$$\forall\ \alpha, \beta \in \mathbb{F}, \forall\ \mathbf{x}, \mathbf{y} \in \mathcal{V}, \qquad (\alpha + \beta)(\mathbf{x} + \mathbf{y}) = \alpha\mathbf{x} + \beta\mathbf{x} + \alpha\mathbf{y} + \beta\mathbf{y}\ .$$

Notice that all four binary operations are at play in these equations but it is sufficient to get by with just one symbol $(+)$ and juxtaposition in practice without any confusion or ambiguity.

Vector spaces that possess yet an additional (*fifth*) binary operation, akin to vector multiplication, are called *algebras*, and are in some sense the pinnacle of algebraic structure. The very name *vector* gives the geometrical sense of the elements being associated with points in some space of a given dimension. And this is certainly one incarnation of a vector space. But thanks to isomorphisms,

we can also consider an element $\mathbf{x}$ in a more abstract sense, as a banana, say, or the color red, or the club suit. Multiplication by a real number allows the banana to take on a variety of guises in the vector space, i.e., $2\mathbf{x}$ might be a banana that is twice as long as $\mathbf{x}$, and the inverse banana of unit length is $-\mathbf{x}$, and so on.

Precisely because of the intrinsic ability to multiply a vector by a scalar from the field $\mathbb{F}$ and get another vector, regarded as sets, vectors fields can be quite large. Indeed, minimally, a vector space $\mathcal{V}$ would need to contain the zero vector $\mathbf{0}$, and say at least one other vector, let's call it $\mathbf{x} \neq \mathbf{0}$. But since every $\alpha\mathbf{x}$ is also in $\mathcal{V}$, this $\mathcal{V}$ must be the same size as $\mathbb{F}$ (note: $\mathbf{0} = 0\mathbf{x}$ where "0" is the additive identity of $\mathbb{F}$). The cardinality of the set $\mathcal{V}$ is therefore greater or equal the cardinality of the set $\mathbb{F}$.

This simple example helps to introduce the idea of a *basis* and the *dimension* of a vector space. The set $\mathcal{B} = \{\mathbf{x}\}$, provided $\mathbf{x} \neq \mathbf{0}$, with one lonely element, is a basis for for our minimal vector space $\mathcal{V}$ (which contains more than one element) since any element of $\mathcal{V}$ can be expressed as $\alpha\mathbf{x}$ for some unique $\alpha \in \mathbb{F}$. Notice that $\mathcal{B}_\alpha = \{\alpha\mathbf{x}\}$ for any $\alpha \in \mathbb{F}, \alpha \neq 0$, is an equally good basis for $\mathcal{V}$. What all these bases have in common is that they contain just one element. And so we say that the *dimension* of this particular vector space is one, although the cardinality or size of this vector space can be much larger as it is equal to that of $\mathbb{F}$. The smallest finite field that one can build is $\mathbb{Z}_2 = \{0, 1\}$, with cardinal number 2. And therefore a one-dimensional vector space over the field $\mathbb{Z}_2$ also has a cardinal number of 2.

Vector spaces of dimension greater than one are easy to envision. A basis $\mathcal{B} = \{\mathbf{x}_i\}$ is any subset of $\mathcal{V}$ with the largest cardinal number one can find such that the only possible solution of

$$\sum_{i=1}^{n} \alpha_i \mathbf{x}_i = \mathbf{0}$$

is $\alpha_1 = \alpha_2 = \cdots = \alpha_{n-1} = \alpha_n = 0$. The integer n is the dimension of the vector space, and we say the set $\mathcal{B} = \{\mathbf{x}_i\}$ of vectors is *linearly independent*, because it is not possible to express any one of them as some linear combination of the others. Again, there can be many different bases of $\mathcal{V}$, but they all have to have the same number of elements or cardinality. Sometimes we adjust our notation $\mathcal{V} \to \mathcal{V}_n$ to indicate the dimensionality of our vectors space is n. Vector spaces can have infinite (better, transfinite) dimensionality as well as transfinite size and then we tend to omit the subscript! These are very large vector spaces.

It follows that given some basis $\mathcal{B}$, *any* element, $\mathbf{a} \in \mathcal{V}_n(\mathbb{F})$ can be expressed uniquely as

$$\mathbf{a} = \sum_{i=1}^{n} \alpha_i \mathbf{x}_i \ ,$$

for some set of $\{\alpha_i\} \in \mathbb{F}$. Viewed as an ordered n-tuple, $(\alpha_1, \alpha_2, \cdots, \alpha_n)$ can also be regarded as the *coordinates* of $\mathbf{a}$ with respect to the basis $\mathcal{B} = \{\mathbf{x}_i\}$. Hence,

every basis generates a coordinate system on the vector space, and, since

$$\mathbf{a} \leftrightarrow (\alpha_1, \alpha_2, \cdots, \alpha_n)$$

is an isomorphism, every vector space of $\mathcal{V}_n(\mathbb{F})$ of dimension n is isomorphic to the Cartesian product space

$$\mathbb{F} \otimes \mathbb{F} \otimes \cdots \otimes \mathbb{F} \equiv \mathbb{F}^n \ .$$

It should be pointed out here that this isomorphism has some peculiar implications. For example, since $\mathcal{V}_1(\mathbb{F})$ is isomorphic to $\mathbb{F}$ itself, it follows that any algebraic field $\mathbb{F}$ can also be regarded as a one-dimensional vector space over itself! (Think: Bavaria meets Georgia O'Keefe!)

There are as many coordinate systems as there are bases, and some coordinate systems turn out to be more useful than others, but there are a vast number of bases for any vector space and no one basis is better or worse than any other.

Another useful concept surrounding vector spaces is the of a *dual* vector space. Let $\mathcal{V}_n(\mathbb{F})$ be our vector space over a field $\mathbb{F}$, whose elements we will denote by boldface, $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$, etc. (We belabor the point here because notation becomes all important in getting a handle on this concept.) Next we define a *linear functional* on $\mathcal{V}_n(\mathbb{F})$ as a mapping $y : \mathcal{V}_n \to \mathbb{F}$, such that

$$y(\alpha\mathbf{x} + \beta\mathbf{z}) = \alpha y(\mathbf{x}) + \beta y(\mathbf{z}) \in \mathbb{F} \ .$$

We'll use italic lowercase Latin letters for linear functionals, keeping in mind that there is absolutely no particular relationship between the element $\mathbf{y} \in \mathcal{V}$ and the linear functional y. Notice that for *any* linear functional y, it must be the case that

$$y(\mathbf{0}) = y(0\mathbf{0}) = 0y(\mathbf{0}) = 0$$

where, remember, 0 is the additive identity in the field of scalars $\mathbb{F}$ and $\mathbf{0}$ is the additive identity vector in the vector field $\mathcal{V}$. A particularly important linear functional is

$$o(\mathbf{x}) = 0$$

for all $\mathbf{x}$ in $\mathcal{V}$. With this linear functional in mind, it is now fairly easy to show that the set of all linear functionals on $\mathcal{V}$ is itself a vector space over the field $\mathbb{F}$. We call it the *dual space* and denote it as $\mathcal{V}^\dagger$ to indicate that it derives its existence so to speak from $\mathcal{V}$.

A more useful (in many ways) notation for the same concept is the square bracket:

$$y(\alpha\mathbf{x} + \beta\mathbf{z}) \equiv [\alpha\mathbf{x} + \beta\mathbf{z}, y] = \alpha[\mathbf{x}, y] + \beta[\mathbf{z}, y]$$

which can be regarded as a *bilinear functional* from $\mathcal{V} \times \mathcal{V}^\dagger \to \mathbb{F}$, owing to the fact that $\mathcal{V}^\dagger$ is also a vector space over $\mathbb{F}$:

$$[\mathbf{x}, \alpha y + \beta z] = \alpha[\mathbf{x}, y] + \beta[\mathbf{x}, z] \ .$$

Now for the mind boggler. Since $\mathcal{V}^\dagger(\mathbb{F})$ is a vector space, we can dream up a new object which is a linear functional on elements of $\mathcal{V}^\dagger(\mathbb{F})$ [which are themselves linear functionals on $\mathcal{V}(\mathbb{F})$], the collection of all of which must be another vector space $[\mathcal{V}^\dagger]^\dagger(\mathbb{F})$. Thankfully, some thought and a stiff drink, convinces one that the dual of the dual space is in fact, for all intents and purposes the same object that we started out with, in other words, there is an isomorphism from $[\mathcal{V}^\dagger]^\dagger(\mathbb{F})$ to $\mathcal{V}(\mathbb{F})$ and so the two spaces are essentially self-dual. The dual space turns out to be useful when we consider transformations or operators that act on the elements of a vector space.

Vector spaces $\mathcal{V}_n(\mathbb{F})$ with dimension $n \geq 2$ possess many proper subsets which themselves are vector spaces on their own. Such subsets are called, *subspaces*, or *manifolds*. Notice that this behavior is quite different from that of groups, many of which do not prossess *any* proper subgroups. Perhaps this is a reflection of the more sophisticated algebraic structure of vector spaces. For example, consider the two-dimensional Cartesian plane $\mathbb{R}^2 = \mathcal{V}_2(\mathbb{R})$. *Any* straight line that passes through the origin, $\mathbf{0}$, is a manifold. Notice that each one of these manifolds is equivalent to $\mathbb{R} = \mathcal{U}_1(\mathbb{R}) \subset \mathbb{R}^2 = \mathcal{V}_2(\mathbb{R})$. Thus the dimension of a manifold is always less than the dimension of the entire vector space which it is a part of. Two other points are worth noting and are easily seen from this example. First, the complement of a manifold is *not* a manifold. Second, a straight line that does not pass through the origin, $\mathbf{0}$, is *not* a manifold. This is because every manifold must contain the identity element, or origin, of the vector space, $\mathbf{0}$. Therefore the intersection of any two manifolds is not empty, but must contain at the very least, the origin.

These aspects of mainfolds permit one to construct larger vector spaces out of smaller vector spaces. For example, let $\mathcal{V}_n(\mathbb{F})$ and $\mathcal{U}_m(\mathbb{F})$ be two distinct vector spaces over the same algebraic field $\mathbb{F}$. If we take pains to ensure that the identity $\mathbf{0}$ in $\mathcal{V}_n(\mathbb{F})$ and its counterpart in $\mathcal{U}_m(\mathbb{F})$, are both taken individually to be precisely the same identity element in the bigger set,

$$\mathcal{W}_{m+n}(\mathbb{F}) = \mathcal{U}_m(\mathbb{F}) \bigcup \mathcal{V}_n(\mathbb{F}) \equiv \mathcal{U}_m(\mathbb{F}) \oplus \mathcal{V}_n(\mathbb{F}) \ ,$$

then, as our notation already gives away, then union of these two vector spaces creates another bigger vector space with dimension $m + n$. This process for building bigger vector spaces should be contrasted with

$$\mathcal{V}_n(\mathbb{F}) = \mathbb{F}^n = \mathbb{F} \otimes \mathbb{F} \otimes \cdots \otimes \mathbb{F} \ ,$$

$$\mathcal{V}_n(\mathbb{F}) = \mathcal{V}_1(\mathbb{F}) \otimes \mathcal{V}_1(\mathbb{F}) \otimes \cdots \otimes \mathcal{V}_1(\mathbb{F}) \ .$$

Given *any* subset $\mathcal{M} \subset \mathcal{V}_n(\mathbb{F})$ we can always minimally enlarge (if necessary) this subset to make it into a vector space in its own right by simply forming a (bigger) set from all the possible linear combinations of elements of $\mathcal{M}$ with coefficients from the field $\mathbb{F}$. We call this enlarged set the *span* of $\mathcal{M}$ and write it as $\text{span}(\mathcal{M})$. If $\text{span}(\mathcal{M})$ is not all of $\mathcal{V}_n(\mathbb{F})$, then it is a proper subset and a manifold of dimension less than n. Finally, the set containing only the origin/identity,

$$\mathcal{O}_0(\mathbb{F}) = \{\mathbf{0}\} \ ,$$

is a manifold of dimension zero of every vector space.

Next we turn to some additional structure that may be imposed on a vector space, the *norm*, *metric* and *inner product*. Vector spaces need not possess this additional structure, but when they do, many more powerful results can be obtained. Generally speaking we wish to distinguish here between *normed vector spaces*, and *inner product vector spaces*. Both spaces are *metric vector spaces*, implying that we can define distances between vectors for both. Inner product vector spaces contain normed vector spaces as a proper subset. An inner product can be used to define a norm, but, a norm cannot be used to define an inner product.

A *norm* defined on a vector space $\mathcal{V}$ over a field $\mathbb{F}$ is a mapping from $\mathcal{V}$ to the nonnegative subset of an ordered field $\mathbb{F}$. For example, if $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$ we mean the positive real numbers and zero. We'll use the notation $|\mathbf{x}|$ to denote the norm of some element $\mathbf{x} \in \mathcal{V}$. The properties a norm must satisfy are:

$$\forall\ \mathbf{x} \in \mathcal{V},\ |\mathbf{x}| \geq 0,\ \text{and}\ |\mathbf{x}| = 0 \iff \mathbf{x} = \mathbf{0}\ ,$$

$$\forall\ \alpha \in \mathbb{F}\ \text{and}\ \mathbf{x} \in \mathcal{V},\ |\alpha\mathbf{x}| = |\alpha||\mathbf{x}|\ ,$$

$$\forall\ \mathbf{x}, \mathbf{y} \in \mathcal{V},\ |\mathbf{x} + \mathbf{y}| \geq |\mathbf{x}| + |\mathbf{y}|\ .$$

Many authors prefer to use $||\mathbf{x}||$ especially if $\mathbb{F} = \mathbb{C}$, and they wish to distinguish between the *modulus* of a complex number $\alpha \in \mathbb{C}$, $|\alpha| = \sqrt{\alpha\alpha^\star}$, where $\alpha^\star$ is the complex conjugate of α, and the norm of a vector.

Like bases, there can be many different norms which satisfy these requirements. A vector space that is equipped with a norm is called a *normed* vector space. When you have norm, you also have a *metric* and vice-versa, because we can define the *distance* between two element of a vector space, $\mathbf{x}$ and $\mathbf{y}$ as the norm of their difference $|\mathbf{x} - \mathbf{y}|$ (better stated as the sum of $\mathbf{x}$ and the inverse of $\mathbf{y}$). A metric, $d(\mathbf{x}, \mathbf{y})$, is a bilinear mapping from $\mathcal{V} \times \mathcal{V}$ to the nonnegative subset of $\mathbb{F}$. Its properties are

$$\forall\ \mathbf{x}, \mathbf{y} \in \mathcal{V},\ d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})\ ,$$

$$d(\mathbf{x}, \mathbf{y}) = 0 \iff \mathbf{x} = \mathbf{y}\ ,$$

$$\forall\ \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{V},\ d(\mathbf{x}, \mathbf{z}) \leq d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z})\ .$$

It is natural to use the norm defined by the metric or the metric defined by the norm, but, again, many different metrics are possible for the same vector space. Vector spaces with a metric are called *metric* spaces. A normed vector space that is *complete*, which loosely means that the limit of every convergent sequence of vectors also lies within the vector space, is called a *Banach Space*.

The epitome of auxiliary structure on a vector space is an *inner product*, which is a bilinear mapping from $\mathcal{V} \times \mathcal{V}$ to all of $\mathbb{F}$. If you have an inner product then you necessarily have *both* a metric and a norm. The converse, however, is not true. We'll denote the inner product by $\langle \mathbf{x}, \mathbf{y} \rangle$. The properties an inner product must satisfy are

$$\langle \mathbf{x}, \mathbf{x} \rangle \in \text{the nonnegative subset of } \mathbb{F}\ , \mathbb{F}^+ \bigcup \{0\}$$

$$\langle \mathbf{x}, \mathbf{x} \rangle = 0 \iff \mathbf{x} = \mathbf{0} \ .$$

Hence we can define the natural norm associated with the inner product as

$$|\mathbf{x}| \equiv \langle \mathbf{x}, \mathbf{x} \rangle^{1/2} \ ,$$

and the metric as

$$d(\mathbf{x}, \mathbf{y}) \equiv \langle \mathbf{x} - \mathbf{y}, \mathbf{x} - \mathbf{y} \rangle^{1/2} \ .$$

An inner product must satisfy four additional properties:

$$\forall \ \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{V}, \ \langle \mathbf{x} + \mathbf{y}, \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle \ ,$$

$$\forall \ \alpha \in \mathbb{F} \text{ and } \mathbf{x}, \mathbf{y} \in \mathcal{V}, \ \langle \alpha \mathbf{x}, \mathbf{y} \rangle = \alpha \langle \mathbf{x}, \mathbf{y} \rangle \ ,$$

$$\forall \ \mathbf{x}, \mathbf{y} \in \mathcal{V}, \ \langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle^{\star} \ ,$$

$$\forall \ \mathbf{x}, \mathbf{y} \in \mathcal{V}, \ |\langle \mathbf{x}, \mathbf{y} \rangle|^2 \leq \langle \mathbf{x}, \mathbf{x} \rangle \langle \mathbf{y}, \mathbf{y} \rangle \ .$$

If $\mathbb{F} = \mathbb{C}$ the star "$\star$" denotes complex conjugation, for $\mathbb{F} = \mathbb{R}$ it can be omitted.

Inner product spaces have lots of structure and are therefore rich in properties and powerful results. For these reasons we typically like to work with inner product spaces. In Euclidean geometry on $\mathbb{R}^n$, the inner product is simply the dot product of vectors in the usual sense. With an inner product we can, in a very real sense, do geometry with distances and angles and so forth. The properties listed above allow us to extend these notions to more abstract settings where $\mathbb{F}$ can be some other ordered field.

Two elements of a vector space are *orthogonal* if their inner product is precisely zero. Hence of all the bases $\mathcal{B}$ we can employ for a vector space $\mathcal{V}$, there is a certain affinity for those whose elements are mutually orthogonal, i.e.,

$$\mathcal{B} = \{\mathbf{x}_1, \mathbf{x}_2, \cdots \mathbf{x}_n\} \text{ , and } \langle \mathbf{x}_i, \mathbf{x}_j \rangle = 0 \text{ for } i \neq j$$

and of unit norm, $\langle \mathbf{x}_i, \mathbf{x}_i \rangle = 1$. If $\mathcal{B}$ is *any* orthogonal set of vectors (not necessarily a basis, and not necessarily of unit norm) then

$$\big| \sum_i \mathbf{x}_i \big|^2 = \sum_i |\mathbf{x}_i|^2$$

generalizes the Pythagoran Theorem, and further, if $\mathcal{B}$ is an ortho*normal* set, then Bessel's Inequality holds

$$\sum_i |\langle \mathbf{y}, \mathbf{x}_i \rangle|^2 \leq |\mathbf{y}|^2$$

for any $\mathbf{y}$. An inner product vector space that is *complete*, which loosely means that the limit of every convergent sequence of vectors also lies within the vector space, is called a *Hilbert Space*.

5. Mappings Between Vector Spaces

Mappings between different (or the same) vector spaces, which must share the same field of scalars, particularly if they preserve alegraic structure, are very

important. *Linear maps*, $A : \mathcal{V}_n(\mathbb{F}) \to \mathcal{U}_m(\mathbb{F})$, between two vector spaces have the property

$$A(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha A\mathbf{x} + \beta A\mathbf{y},\ \forall\ \mathbf{x}, \mathbf{y} \in \mathcal{V}_n(\mathbb{F}), \forall\ \alpha, \beta \in \mathbb{F}\ .$$

Since $A\mathbf{x}, A\mathbf{y} \in \mathcal{U}_m(\mathbb{F})$ so are $\alpha A\mathbf{x}, \beta A\mathbf{y}$. Notation again! Italic capital Latin letters will be reserved for linear maps. (Remember: vector spaces are caligraphics $\mathcal{V}$, fields are $\mathbb{F}$, elements of a vector space are generally boldface $\mathbf{x}$, linear functionals of the dual space are italic lower case Latin y, and linear maps, or operators, between vector spaces are upper case italic Latin A—got it?)

Because of the isomorphism $\mathcal{U}_1(\mathbb{F}) \leftrightarrow \mathbb{F}$, it follows that the linear functionals we introduced in the previous section, which live in $\mathcal{V}^\dagger$, can be regarded as linear maps from their (dual) vector space $\mathcal{V}$ to the one-dimensional vector space $\mathcal{U}(\mathbb{F}) = \mathbb{F}$. (Think: Gerogia O'Keefe in Bavaria!)

Let $\mathcal{B} = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n\}$ be a basis for $\mathcal{V}_n(\mathbb{F})$ and let $\mathcal{C} = \{\mathbf{y}_1.\mathbf{y}_2, \cdots \mathbf{y}_m\}$ be a basis for $\mathcal{U}_m(\mathbb{F})$. Suppose a linear mapping A maps $\mathbf{b} \in \mathcal{V}_n(\mathbb{F})$ to $\mathbf{c} \in \mathcal{U}_m(\mathbb{F})$, i.e., $A\mathbf{b} = \mathbf{c}$. Then because $\mathbf{b}$ and $\mathbf{c}$ have unique coordinates in each of their respectives bases, $\mathcal{B}$ and $\mathcal{C}$, $\mathbf{b} = \beta_i\mathbf{x}_i$ and $\mathbf{c} = \gamma_j\mathbf{y}_j$, it follow that the action of A can be represented by a rectangular array of entries selected from $\mathbb{F}$, say A_{ij}, which satisfy

$$\gamma_j = A_{ij}\beta_i\ .$$

The array A_{ij} is the unique representation of the linear operator A for the bases $\mathcal{B}$ and $\mathcal{C}$. For different bases, the entries in the array are different of course.

It is true that the set of all linear mappings form $\mathcal{V}_n(\mathbb{F})$ to $\mathcal{U}_m(\mathbb{F})$ form a nm dimensional vector space over $\mathbb{F}$ with respect to the operation of adding entries in a rectangular array representation, i.e., if A, B, are any two linear maps we can form a new linear map by simply adding the respective ij-entries in their rectangular arrays. But, interestingly enough, this aspect of linear maps turns out not to be of much importance. Instead, it is the successive application of linear maps, something more akin to multiplication than addition, which is more fruitful. And this cannot be done here because our linear maps act only upon $\mathcal{V}_m(\mathbb{F})$ and they have no meaning applied to $\mathcal{U}_m(\mathbb{F})$, for example.

Linear maps from a vector space to itself are particularly important and are often called *operators*. And operators can be applied successively, which is called *composition*. And with respect to addition, defined in the last paragraph, operators must *also* form a vector space over the field $\mathbb{F}$. To show this, we need of course the linear operator

$$O\mathbf{x} = \mathbf{0}\ , \forall \mathbf{x} \in \mathcal{V}_n\ ,$$

which plays the role analogous to that of the origin $\mathbf{0} \in \mathcal{V}_n(\mathbb{F})$. The identity operator

$$I\mathbf{x} = \mathbf{x}\ , \forall \mathbf{x} \in \mathcal{V}_n\ ,$$

also proves useful in what follows. Finally notice that, as with linear functionals,

$$A\mathbf{0} = \mathbf{0}\ , \forall A\ .$$

Composition of operators is defined by:

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

where B is first applied to $\mathbf{x}$ followed by A. However, it need *not* be the case that $AB = BA$, and it is certainly possible for $AB = O$ for $A \neq O$ and $B \neq O$. However, $IA = AI$, and if $AB = BA$ one says that the two opeators *commute* with one another.

The two special operators, O and I are useful in distinguishing between what one might regard as good, or sensible, operators and ones that are ill-behaved. The identity operator enjoys (in a trivial sense) the two properties that, its *range* is all of $\mathcal{V}_n$, and, it is one-to-one, in that $I\mathbf{x} = I\mathbf{y} \implies \mathbf{x} = \mathbf{y}$. Hence there exists an inverse operator I^{-1} for composition, such that

$$I\mathbf{x} = \mathbf{y} \iff \mathbf{y} = I^{-1}\mathbf{x} \ .$$

In this trivial case, of course, $I^{-1} = I$ and $\mathbf{y} = \mathbf{x}$. But a nontrivial linear operator A for which these two properties obtain is said to be *invertible* with a unique inverse A^{-1} (generally not equal to A) that commutes with A

$$AA^{-1} = A^{-1}A \ .$$

The operator O on the other hand violates both of these properties and is therefore not invertible. O maps all of $\mathcal{V}_n$ to $\mathcal{O}_0 = \{\mathbf{0}\} \subset \mathcal{V}_n$.

Of *all* the possible linear maps from a vector space $\mathcal{V}_n(\mathbb{F})$, those which *are* invertible with respect to composition form a *group* with respect to the binary operation compostion. The identity for the group is I. This group is called the *General Linear Group*, and is denoted by $\mathrm{GL}_n(\mathbb{F})$. The operator O is not a member of this group.

In so far as *every* operator, when referred to a basis $\mathcal{B}$ for $\mathcal{V}_n(\mathbb{F})$ is isomorphic to an n^2 square array of scalars selected from $\mathbb{F}$, what distinguishes the membership A of $\mathrm{GL}_n(\mathbb{F})$ is their matrix determinants do not vanish:

$$\det[A] \neq 0 \ .$$

Because *all* of the entries in the matrix representation of O are zero, $\det[O]$=0.

It turns out that O is not alone in having this property, there are many other linear maps which have zero determinant as well. For example, *projections*, are operators that isolate the component of a given element $\mathbf{x}$ which lives entirely within a certain manifold.

The *kernel* of a linear operator A is the manifold of $\mathcal{V}_n(\mathbb{F})$ which under the action of A maps to $\mathbf{0}$. For example, the identity mapping, I (which is in $\mathrm{GL}_n(\mathbb{F})$), maps only $\mathbf{0}$ [equivalently $\mathcal{O}_0(\mathbb{F})$] to $\mathbf{0}$, while the O operator maps all of $\mathcal{V}_n(\mathbb{F})$ to $\mathbf{0}$. This sort of brackets the possibilities. Indeed all the operators in $\mathrm{GL}_n(\mathbb{F})$ have kernels of dimension 0, while all the operators excluded from membership have kernels with dimensions somewhere between 1 and n, inclusive. It follows that over the complement of the kernel of A in $\mathcal{V}_n(\mathbb{F})$, A is in fact invertible! This is consistent with the members of $\mathrm{GL}_n(\mathbb{F})$ being invertible over

the entire vector space, while O is invertible nowhere. The complement of the kernel of A with the union of the set $\{\mathbf{0}\}$, is itself a manifold, and therefore a vector space. So by paring down our original vector space $\mathcal{V}_n(\mathbb{F})$ into lower dimensional manifolds, it is possible to "rehabilitate" some of our ill-behaved operators that did not make the initial cut into $\mathrm{GL}_n(\mathbb{F})$, as they can now be placed in a $\mathrm{GL}_m(\mathbb{F})$, albeit with $1 \leq m < n$.

A closely allied concept to the kernel of a linear operator is its spectrum of eigenvalues. For any linear operator A we can determine the n-th degree polynomial $\det[A - \lambda I] = 0$. The eigenvalues λ associated with the members of $\mathrm{GL}_n(\mathbb{F})$ are nonzero. Those which do not belong to this group have at least one zero eigenvalue.

Much more can be said about eigenvalues and what they imply about the operators they are derived from. Limitations of space and time prevent us from venturing further into this important area, but the two books by Halmos [**H4,5**] will satisfy your curiosity about what we have left out.

6. The Real, the Complex And the Truly Bizarre Numbers

The counting numbers $\mathbb{N} = \{1, 2, 3, ...\}$ do not form a group under addition (or multiplication for that matter). Equivalently, an equation of the form

$$x + a = b$$

need not (although it may) have a solution $x \in \mathbb{N}$ for arbitrary $a, b \in \mathbb{N}$. In the larger set of integers, $\mathbb{Z}$, however, this equation *always* has an unique solution given by

$$x = -a + b$$

since every a has an (additive) inverse, $-a$. However, the equation

$$ax = b$$

need not (although it may) have a solution $x \in \mathbb{Z}$ for arbitrary $a, b \in \mathbb{Z}$. It will have a solution if and only if a is an *integral divisor* of b. For this reason, $\mathbb{Z}$ is called an *integral domain*.

In the larger set of rational numbers $\mathbb{Q}$, however, this equation *always* has an unique solution given by

$$x = a^{-1}b$$

since every $a \neq 0$ has an (multiplicative) inverse, a^{-1}. And for this reason, $\mathbb{Q}$ is called a *field*.

On the other hand, the equation

$$ax^2 = b$$

need not (although it may) have a solution $x \in \mathbb{Q}$ for arbitrary $a, b \in \mathbb{Q}$. In the larger set of real numbers $\mathbb{R}$ this equation will have a solution, but *only* if $b/a \geq 0$. When it does have a nonzero solution, in fact, it has two solutions

because $-x$ is also a solution. In the *still* larger set of complex numbers $\mathbb{C}$, this equation *always* has solutions, period. Indeed, in $\mathbb{C}$ the equation

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x = a_0$$

always has solutions for $x \in \mathbb{C}$—in general n solutions—for any positive integer n and arbitrary $a_i \in \mathbb{C}$.

There are many ways to enlarge $\mathbb{R}$ to make $\mathbb{C}$. An enlightening one is the Calyley-Dickson method. Begin with an algebraic field, in this case, say the real numbers $\mathbb{R}$. We now form a new set of objects being ordered pairs of elements drawn from the field $\mathbb{R}$: (a, b). We *define* the binary operations of addition, and multiplication as follows

$$(a, b) + (c, d) = (a + c, b + d)$$

$$(a, b)(c, d) = (ac - bd, bc + ad)$$

where the addition and multiplications *inside* the parentheses are those appropriate to the field $\mathbb{R}$. This field, is, in fact isomorphic to $\mathbb{C}$. It contains $\mathbb{R}$ as a subfield consisting of all ordered pairs $(a, 0)$. It also solves the quadratic equation

$$x^2 + 1 = 0 \ ,$$

the solutions obviously being $x = (0, \pm 1)$. But the set of all orderd pairs (a, b) with $a, b \in \mathbb{R}$ is also isomorphic to $\mathbb{R}^2$, and therefore, the two dimensional vector space over the field of real numbers $\mathcal{V}_2[\mathbb{R}]$. But because we can also multiply vectors from the vector space, not just add them and multiply them by scalars from $\mathbb{R}$, we have something richer in structure than merely a vector space—we have a *division algebra*—in $\mathbb{C}$

By exploiting the isomosphisms between the complex numbers and $\mathcal{V}_2[\mathbb{R}]$, we can define an inner product as

$$(a, b) \cdot (c, d) = \frac{1}{2}[(a, b)(c, -d) + (a, -b)(c, d)] \equiv \frac{1}{2}[(a, b)(c, d)^\star + (a, b)^\star (c, d)] \ ,$$

where the *conjugate* of an element (a, b), written $(a, b)^\star \equiv (a, -b)$. With this inner product we have a norm

$$|(a, b)| = \sqrt{(a, b) \cdot (a, b)} = \sqrt{a^2 + b^2} \ ,$$

and therefore a metric as well! In a very real sense, $\mathbb{C}$ embodies all the mathematical structures we have discussed in this Appendix and might rightly be considered the epitome of the long sequences of sets that we began back with the set of three colors.

To simplify notation, one usually refers to an element of $\mathbb{C}$ simply as a complex variable $z = x + iy$, where it is understood that $i = \sqrt{-1}$, and x and y are drawn from $\mathbb{R}$. And in this fashion, the corresponding basis for the vector space is conveniently $\mathcal{B} = \{1, i\}$, as expected. So $\mathbb{C}$ has two dimensions as a vector space.

One might, as William Rowan Hamilton did for 20 years, wonder if there is a three-dimensional analogue to $\mathbb{C}$. Alas, there is not. There is however, a four-dimensional structure, called the *quaternions*, denoted by $\mathbb{H}$ in honor of their discoverer, Hamilton, which can be generated from $\mathbb{C}$ by the same Cayley-Dickson process, i.e., we consider a set of ordered pairs of elements drawn now from $\mathbb{C}$ (as opposed to $\mathbb{R}$). As this is a four-dimensional division algebra, a quaternion can be expressed as, $\zeta = x + iy + ju + kv$, where x, y, u, v are drawn from $\mathbb{R}$,

$$i^2 = j^2 = k^2 = ijk = -1 \ , \qquad ij = -ji = k \ ,$$

and i gets two additional distinct friends, j and k. The basis for this division algebra viewed as a vector space is $\mathcal{B} = \{1, i, j, k\}$. (Note the isomorphism to the Klein 4-group!) The quaternions, $\mathbb{H}$, however, just fail at being an algebraic field, because multiplication of quaternions is not commutative! This in a very real sense limits their usefulness for doing things like calculus and analysis. They do come with a norm, however. If $\zeta = x + iy + ju + kv$, then we can define its conjugate as $\zeta^\star \equiv x - iy - ju - kv$, and the product $\zeta\zeta^\star = \zeta^\star\zeta = x^2 + y^2 + u^2 + z^2 \geq 0$. And because we have a norm, we also have a metric. Can you find an inner product?

In fact, any attempt to enlarge $\mathbb{C}$ requires that we give up some attribute which the complex numbers enjoyed. Some algebraic extensions of the complex numbers choose to give up the cancellation law, meaning that two nonzero vectors can be multiplied together to give the **0** vector—so we no longer have a *division* algebra. Examples of these extensions are *Grassmann* and *Clifford Algebras*.

If we apply Cayley-Dickson again to the quaternions, things get even worse. The elements of the ensuing eight-dimensional division algebra are called the *octonions*, and their collection is denoted by $\mathbb{O}$. For the octonions, alas, not only is multiplication non-commutative, it is also non-associative. And after that, there are no more division algebras, period. Each higher dimensional division algebra contains the previous one from which it was generated, symbolically,

$$\mathbb{R} \subset \mathbb{C} \subset \mathbb{H} \subset \mathbb{O} \ ,$$

and of course,

$$\mathbb{E} \subset \mathbb{N} \subset \mathbb{Z} \subset \mathbb{R} \ .$$

7. Complex Analysis & Fourier Transforms

The discussion of the previous section suggests that it may be more fruitful to go back to $\mathbb{C}$ and concentrate our efforts on this set of numbers, since, afterall, it has the most robust sense of algebraic structure we can endow a set with, and at the same time, it provides solutions to every algebraic equation we can dream up. In fact, not just one solution, but all of them.

What makes the complex numbers useful is our ability to do calculus with them. Calculus, or more generally analysis, requires as its logical basis certain topological concepts regarding the proximity, density and orientation of the elements of our vector space. For example, how closely are they spaced? Is

there an element that is—in terms of the metric—arbitrarily close to another element? Can I separate two distinct elements? Can I actually get to where I am going? Do limits exist?

The essential building blocks here, are *open* and *closed* sets and operations carried out on those sets, such as intersections, unions, and closures. Topology is often called the study of all the "c"-concepts: *continuous*, *compact*, *connected*, *complete*, *convergent*, and *covered*. Vector spaces and the mappings defined on those vector spaces must satisfy some of these "c"-concepts for us to be able to differentiate and integrate them.

This brings us to the concept of *analytic (holomorphic) functions*—or a special subset of all the mappings $\varphi : \mathbb{C} \to \mathbb{C}$. Let $\zeta(z) = \xi(x, y) + i\eta(x, y)$ be a function of the complex variable $z = x + iy$—that is $\varphi : z \mapsto \zeta$. In what follows, we'll try to reserve z, ζ and when necessary w, f, and F for complex variables from $\mathbb{C}$, and everything else will be a real number from $\mathbb{R}$.

It follow that $\xi(x, y)$ and $\eta(x, y)$ are real functions of two real variables x, y—that is $\xi, \eta : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$. The function ζ is said to be *analytic* at the point z if, and only if

$$\frac{\partial \xi}{\partial x} = \frac{\partial \eta}{\partial y} \quad \text{and} \quad \frac{\partial \xi}{\partial y} = -\frac{\partial \eta}{\partial x} \ .$$

These are called the *Cauchy-Riemann Equations*.

If we restrict our attention to functions which are analytic *almost everywhere*, then we can do calculus with these functions, which means computing derivatives and calculating integrals. Two remarks are worth adding here. First, the Cauchy-Riemann Equations imply that

$$\frac{\partial^2 \xi}{\partial x^2} + \frac{\partial^2 \xi}{\partial y^2} = \frac{\partial^2 \eta}{\partial x^2} + \frac{\partial^2 \eta}{\partial y^2} = 0 \ ,$$

for any analytic $\zeta(z)$. That is, the real and imaginary parts of an analytic function regarded as real functions of the two-dimensional Cartesian coordinates (x, y), are solutions of Lapace's equation! Second, they are orthogonal,

$$\nabla \xi \cdot \nabla \eta = 0 \ .$$

The function

$$\zeta(z) = z^\star$$

is analytic nowhere! However, as long as one sets the complex conjugation function aside, virtually any other function that you can care to think of turns our to be analytic almost everywhere! Any polynomial is analytic everywhere. A function that is analytic everywhere is called an *entire* function. The exponential function, defined by the convergent power series

$$\exp z \equiv \sum_{k=0}^{\infty} \frac{z^k}{k!} \equiv \cos z + i \sin z$$

is another entire function, from which we derive so to speak the two trigonometric functions. It is centrally important, because often we use it to express the complex variable

$$z = x + iy = \sqrt{x^2 + y^2}\ \exp[i \arctan(y/x)]$$

in terms of a *modulus*, $\sqrt{x^2 + y^2}$, and a *phase*, $\arctan(y/x)$. The phase is ill-defined up to an integer multiple of 2π, which as we shall see, can lead to some fascinating consequences.

The fact that one must forget about the complex conjugation function entirely in dealing with analytic functions and calculus suggests that complex variables—although they are a two-dimensional vector space—are essentially one-dimensional in character, since the only way we can extract the real or imaginary parts separately is through complex conjugation. Hence it is appropriate to think of z as a complex variable (singular!), that just happens to come with two components, x, y—Cayley-Dickson's ordered pair (x, y).

Any rational function—a ratio of polynomials—is analytic almost everywhere, *except* at a countable handful of isolated points. These isolated points are the (no greater than n) zeros of the denominator, which is a polynomial of degree n. The Cauchy-Riemann equations break down at these isolated points where the derivatives become infinite. These points are called *poles*. A function which is analytic everywhere except at a finite number of isolated poles is called a *meromorphic* function.

The n-th root function is analytic almost everywhere, with the exception of the origin, $z = 0$, and (more or less) *any* line that connects the origin to infinity. This line—which one has significant latitude in its placement— is called a *branch cut*. To see how this works, let us consider the square-root function:

$$\zeta(z) = \sqrt{z}\ , \qquad \xi + i\eta = \sqrt{x + iy}\ ,$$

or

$$x = \xi^2 - \eta^2\ , \qquad y = 2\xi\eta\ .$$

Use the equation on the left to eliminate η from the equation on the right. This gives a quadratic equation for ξ^2, which, since ξ^2 cannot be negative, has the *unique* solution

$$\xi^2 = \frac{1}{2}\Big(x + \sqrt{x^2 + y^2}\Big) \geq 0\ .$$

For ξ itself, we have two choices:

$$\xi = \pm\frac{1}{\sqrt{2}}\sqrt{x + \sqrt{x^2 + y^2}}\ .$$

And therefore,

$$\eta = \pm\text{sign}(y)\frac{1}{\sqrt{2}}\sqrt{-x + \sqrt{x^2 + y^2}}\ .$$

No matter how we select between the the two choices, a branch cut, or a curve across which either ξ or η changes discontinuously is unavoidable. Pick the plus

sign everywhere. Then $\xi(x,y)$ is nonnegative, and is zero for $y=0$ and $x\leq 0$. On the other hand $\eta(x,y)$ is zero for $y=0$ and $x\geq 0$ and so the flip of sign caused by the sign of y causes no problem here. However, for $y=0$ and $x\leq 0$, η is double-valued. This phenomenon is characteristic of a branch cut. If you try to get around this by changing the sign of the $\pm$, you can eliminate the double-valuedness of η, but now it shows up in ξ.

Also worth mentioning is the logarithm, defined to be the inverse of the exponential,

$$\log z = \log(x+iy) \equiv \frac{1}{2}\log(x^2+y^2) + i[\arctan(y/x) + 2n\pi] \ ,$$

which, owing to the indeterminate multiples of 2π, also comes equipped with a branch cut if we want a single-value version, like we did for our square-root.

Having restricted from here on our attention to functions which are analytic almost everywhere (except at isolated poles and across branch cuts), and which satsify the Cauchy-Riemann conditions, we can rely on a number of extremely sweeping and powerful properties of analytic functions. The derivative of an analytic function

$$\frac{d\zeta}{dz} = \lim_{|w|\to 0} \frac{\zeta(z+w)-\zeta(z)}{w}$$

is itself an analytic function of the complex variable z and is independent of the specific choice of $w=u+iv$. A direct corollary of this is that any analytic function $\zeta(z)$ can be differentiated as many times as one wishes, and the result is always an analytic function! So analytic functions are infinitely differentiable.

Next, all the rules that apply to differentiation of functions of a real variable apply directly to the differentiation of analytic functions of a complex variable, e.g.,

$$\frac{d}{dz}(z^\alpha + w)^\beta = \alpha\beta z^{\alpha-1}(z^\alpha+w)^{\beta-1} \ .$$

Also, functions defined in terms of convergent infinite series are still defined by these same series now of a complex variable.

Complex integration is even more remarkable. In computing an integral, such as

$$\int dz\ \zeta(z) = \int (dx+idy)\ [\xi(x,y)+i\eta(x,y)] \ ,$$

one must in general specify the precise path, or *contour*, the integration follows in the complex plane—*unless* $\zeta(z)$ is analytic in the domain of interest! Then, because analytic functions are infinitely differentiable, for some analytic function $f(z)$, $\zeta = df/dz$, and the integral depends only upon the starting and stopping endpoints. *Any* integration path that connects these two endpoints gives the same result. Therefore, if $\zeta(z)$ is analytic inside and along any closed path in the complex plane, the integral along that path must vanish,

$$\oint dz\ \zeta(z) = 0 \ .$$

This is an exceedingly powerful result.

If a function is meromorphic—that is, it is not analytic only at isolated poles—then Cauchy found that this result can be generalized as follows. Suppose $\zeta(z)$ is analytic inside and along a closed contour in the complex plane. Then consider the integral

$$\frac{n!}{2\pi i}\oint dz \frac{\zeta(z)}{(z-w)^{n+1}} \ ,$$

where $n = 0, 1, 2, ...$, and so forth. If the complex variable w lies *outside* of the contour, then the integrand is analytic along and inside of the contour, and so the integral is exactly zero. If, on the other hand, w lies *inside* the closed integration contour, then the integral is simply $\pm d^n\zeta(w)/dw^n$. *We choose the plus sign if the contour circles the pole in a counterclockwise fashion, and the minus sign if it circles the pole in a clockwise fashion.* Putting $\zeta(w) = 1$ shows that for *any* w

$$\frac{n!}{2\pi i}\oint dz \frac{1}{(z-w)^{n+1}} = 0 \ , \quad \text{for } n = 1, 2, 3, ...$$

Many more absolutely fascinating results are possible! An example is the following, due again, to Cauchy. If $f(z)$ is a meromorphic function—it's only singularities are isolated poles—inside a closed contour, then

$$\frac{1}{2\pi i}\oint dz \frac{f'(z)}{f(z)} = N - P \ ,$$

where N is the number of zeros and P is the number of poles that $f(z)$ has inside of this contour. If a pole or a zero has a double, a triple, or an n-tuple multiplicity, then it contributes n to N or P.

It only remains to consider what happens if w—or more generally a pole of the integrand—should lie exactly *on* the contour. Strictly speaking, one should never try to integrate through a pole (or across a branch cut). One can get as close to the pole as one likes, but at the last moment it is necessary to divert above or below the pole. In one case, the pole lies within the contour and in the other case it lies outside of the contour, and it can be the case that the value of the integral changes by a finite definite amount whether this tiny deviation goes slightly above or below the pole. And that is simply how complex integration works. However, one can choose to *define* something called the *principal value integral* by taking the average of the result obtained by diverting above or below and associating this, so to speak, with integrating directly through the singularity or pole. In other words, for counterclockwise circling,

$$\oint dz \frac{1}{z-w} = 2\pi i \ , \quad \text{if } w \text{ is inside the contour} \ ,$$

$$\oint dz \frac{1}{z-w} = 0 \ , \quad \text{if } w \text{ is outside the contour} \ ,$$

$$P.V. \oint dz \frac{1}{z-w} \equiv \pi i \ , \text{if } w \text{ is "on" the contour} \ .$$

Using these results we can also define many new analytic functions in terms of integrals. For example, the Gamma Function

$$\Gamma(w) \equiv \int_0^\infty dt\ t^{w-1}e^{-t}\ ,$$

where the integration contour is along the positive real axis, interpolates the factorial function:

$$\Gamma(n+1) = n!\ ,$$

where $n = 0, 1, 2, ...$, and so on. Provided we do not cross any branch cuts, the contour can be displaced off the real axis by adding any amount of

$$\oint dz\ z^{w-1}e^{-z} = 0\ ,$$

to the above expression, for which the integrand is analytic inside and along the closed contour. Another functions that can be defined in terms of an integral is the inverse function of the exponential,

$$\log w \equiv \int_1^w dz\ \frac{1}{z}\ .$$

Notice, in particular, that like the square root function, the logarithm requires a branch cut somewhere. By design, as $w \to 1$ the integral goes to zero. But suppose the contour takes w around the origin in a counterclockwise fashion *before* it returns to 1. Then we know that the value of the integral is $2\pi i$ and not zero. Thus, to avoid the double-valued nature of this definition, it is necessary to connect the origin to infinity by a branch cut where there is a discontinous jump in the value of the logarithm. Another way to see this is to note that the above definition implies that

$$\log(u+iv) = \frac{1}{2}\log(u^2+v^2) + i\theta$$

where $\tan\theta = v/u$, where θ can be advanced or retarded by any integer multiple of 2π with no consquence. For example, restricting $-\pi \le \theta \le \pi$ places the branch cut along the negative real axis, and the imaginary part of $\log w$ jumps by 2π in crossing the branch cut.

Another means to generate an analytic function F is by an integral of another analytic function:

$$F(w) = \int_{C(z)} dz\ e^{-iwz}f(z)$$

with $w = u + iv$ and $z = x + iy$, and where f is analytic along the integration contour C, which is not necessarily closed but may have end points. Notice that we have to specify this contour to get a unique definition for F because of Cauchy's amazing results about integrating analytic functions around closed curves. The important point here is that the exponential function is an entire

function with no singularities anywhere in the complex plane. Therefore, if we choose to employ a closed contour, F will be nonzero only if f has poles within that contour. Otherwise it is more practical, like in our definition of the Gamma Function, to choose a countour which is not closed.

Selecting the entire real axis ($y = 0$)for the C contour leads to the definition of the *Fourier Transform* of $f(x)$:

$$F(w) = \int_{-\infty}^{\infty} dx\ e^{-iwx} f(x)\ .$$

A remarkable property of $F(w)$ is that we can retrieve $f(z)$ from

$$f(z) = \frac{1}{2\pi} \int_{\gamma(w)} dw\ e^{iwz} F(w)\ ,$$

for some suitable choice of contour $\gamma(w)$ (which may not be the real axis). Be aware that the factor of 2π gets split up all over the place in various definitions which people employ, and some people replace $i \to -i$ to get forward and backward transforms exchanged. Like units, the best advice I can give here is to pick some definition and stick with it.

When the contour γ can be taken to run along the real axis ($v = 0$), we get the usual definition of the Fourier Transform pairs in terms of functions of the two real variables, x and u (i.e., $z = x + iy$ and $w = u + iv$)

$$F(u) = \int_{-\infty}^{\infty} dx\ e^{-iux} f(x)\ \equiv \mathfrak{F}_u[f(x)],$$

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} du\ e^{iux} F(u)\ \equiv \mathfrak{F}_x{}^{-1}[F(u)].$$

It is worth mentioning that not every function $f(x)$ has a well-defined Fourier transform according to these formulae. The function $f(z) = 1$ does not, for example. In some instances by allowing a non-zero v back into the picture, we can find areas off the real axis ($v = 0$) were $F(w)$ is analytic, and then we can make "sense" of otherwise "senseless" integrals by mean of *analytic continuation* of $F(w)$ back to the real axis! More about this later.

Two important theorems for Fourier Transform pairs are the following

$$\mathfrak{F}_u \left[\int_{-\infty}^{\infty} dx\ f(x')g(x - x') \right] = \mathfrak{F}_u[f(x)] \cdot \mathfrak{F}_u[g(x)]\ ,$$

and

$$\int_{-\infty}^{\infty} dx\ |f(x)|^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} du\ |\mathfrak{F}_u[f(x)]|^2\ .$$

And finally, the following two formulas are particuarly useful in applications,

$$\mathfrak{F}_u \left[\frac{d^n f}{dx^n} \right] = (iu)^n \mathfrak{F}_u[f(x)]\ ,$$

$$\mathfrak{F}_u\left[x^n f(x)\right]] = \left(i\frac{d}{du}\right)^n \mathfrak{F}_u[f(x)] \ .$$

A closely allied transform pair is named after Laplace. Here, we take the contour $C(w)$ to run only along the positive real axis according to

$$\int_0^\infty dx \ e^{-iwx} f(x) \ ,$$

and repace $w = -i\zeta = -i(\xi + i\eta) = \eta - i\xi = u + iv$, (which amounts to a 90° rotation in the complex plane) giving

$$\hat{f}(\zeta) = \int_0^\infty dx \ e^{-\zeta x} f(x) \ \equiv \mathfrak{L}_\zeta[f(x)]$$

the *Laplace Transform* of $f(x)$. I use a hat "$\hat{}$" over a lower case f to distinguish this from the Fourier Transform of $f(x)$, which I called $F(w)$.

Now it gets fun! Define the Heaviside Step Function, $\theta(x)$ by

$$\theta(x) = 0 \ , \ \text{for } x < 0, \ \ \theta(x) = 1 \ , \ \text{for } x > 0 \ ,$$

so that

$$\hat{f}(\zeta) = \int_{-\infty}^\infty dx \ e^{-\zeta x} f(x)\theta(x)$$

$$\hat{f}(\xi + i\eta) = \int_{-\infty}^\infty dx \ e^{-i\eta x} \left[e^{-\xi x} f(x)\theta(x)\right]$$

$$\hat{f}(\xi + i\eta) = \mathfrak{F}_\eta \left[e^{-\xi x} f(x)\theta(x)\right] \ ,$$

where, recall

$$F(\eta) = \mathfrak{F}_\eta[f(x)] \ !$$

This allows us to write

$$e^{-\xi x} f(x)\theta(x) = \frac{1}{2\pi} \int_{-\infty}^\infty d\eta e^{i\eta x} \hat{f}(\xi + i\eta)$$

or

$$f(x)\theta(x) = \frac{1}{2\pi} \int_{-\infty}^\infty d\eta \ e^{(\xi + i\eta)x} \hat{f}(\xi + i\eta)$$

$$f(x)\theta(x) = \frac{1}{2\pi i} \int_{\xi - i\infty}^{\xi + i\infty} d\zeta \ e^{\zeta x} \hat{f}(\zeta) \equiv \mathfrak{L}_x^{-1}[\hat{f}(\zeta)] \ ,$$

where

$$\hat{f}(\zeta) = \int_0^\infty dx \ e^{-\zeta x} f(x) \ \equiv \mathfrak{L}_\zeta[f(x)] \ .$$

gives our Laplace Transform pair.

Notice the interesting role played by the Heaviside Step Function in all of this, which requires that

$$\mathfrak{L}_x^{-1}[\hat{f}(\zeta)] = 0 \ , \ \text{for } x < 0.$$

In turn, this assures us that $\hat{f}(\zeta)$ is analytic for all ξ for which the integral

$$\hat{f}(\xi + i\eta) = \int_0^\infty dx\ e^{-(\xi+i\eta)x} f(x)$$

exists. For example, if

$$\int_0^\infty dx\ e^{-i\eta x} f(x)$$

is finite, then $\hat{f}(\zeta)$ is analytic for all $\xi \geq 0$.

We conclude with saying a little bit about *analytic continuation*. The Cauchy-Riemann conditions are so prescriptive that they can be used to "uniquely" extend a function that is analytic in some "patch" of the complex plane to the entire complex plane. We have to be a little careful in what we mean by uniqueness and what constitutes a patch of sufficient size to enable this construction. Also, this extension may often result in isolated singularities, and, as we have seen previously, branch cuts.

Take for example, the analytic function represented by the infinite series

$$1 + z + z^2 + z^3 + \cdots\ ,$$

which converges to an analytic function provided $|z| < 1$. This is a sufficiently big patch of $\mathbb{C}$ that we can analytically continue it everywhere. To do so, we notice that when $|z| < 1$ the series coincides with

$$\frac{1}{1-z}\ ,$$

which, *without the restriction* $|z| < 1$, is in fact the analytic continuation of the function defined by this infinite series. It's defined everywhere in the complex plane except at $z = 1$, where it has an isolated singularity (a simple pole). The underlying concept here is that if two different expressions agree over some patch of the complex plane, then they are both representations of the *same* analytic function over whatever their individual domains of validity might be. In this case, the infinite series representation is valid only in the disk $|z| < 1$, while the fractional representation is valid everywhere except at a single point, so there is no particular advantage necessarily to using the infinite series in the disk and the fraction elsewhere. The next example provides a situation where neither of the two representations is valid everywhere.

As a second example, consider the Gamma Function,

$$\Gamma(w) = \int_0^\infty dt\ t^{w-1} e^{-t}$$

which defines $\Gamma(w)$, $w = u + iv$, only for $u > 0$. We need to find a different representation to figure out what it is doing in the other half of the complex plane. To analytically continue it into $u \leq 0$ we form the product

$$\Gamma(w)\Gamma(1-w) = \int_0^\infty ds \int_0^\infty dt\ t^{w-1} s^{-w} e^{-(s+t)}\ ,$$

which defines an analytic function only in the strip $0 < u < 1$. We do the t integral first by making a change of variable $t = xs$, so $dt = sdx$, giving

$$\Gamma(w)\Gamma(1-w) = \int_0^\infty ds \int_0^\infty du\ u^{w-1}e^{-s(1+u)}\ .$$

We now do the s integral to obtain

$$\Gamma(w)\Gamma(1-w) = \int_0^\infty du\ \frac{u^{w-1}}{1+u} = \frac{\pi}{\sin \pi w}\ .$$

In the $0 < u < 1$ patch of the complex plane, the product $\Gamma(w)\Gamma(1-w)$ agrees with $\pi \csc \pi w$, so, in fact, it must be $\pi \csc \pi w$ everywhere else! Therefore,

$$\Gamma(1-w) = \pi \csc \pi w\ \frac{1}{\Gamma(w)}\ ,$$

valid for $u > 0$ provides a valid (analytic continuation) representation of the Gamma Function for $z = 1 - w = x + iy$ with $x < 1$. It tells us incidentally, that the Gamma Function has a series of simple poles at all the non-positive integers. The original defining representation

$$\Gamma(w) = \int_0^\infty dt\ t^{w-1}e^{-t}$$

and

$$\Gamma(1-w) = \pi \csc \pi w\ \frac{1}{\Gamma(w)}\ ,$$

agree everywhere on the strip $0 < u < 1$ or equivalently $0 < x < 1$, so they represent the same function. While each individually fails to cover the entire complex plane, taken together they define the Gamma Function everywhere!

As a third example, suppose

$$f(z) = \sum_{k=0}^{\infty} f_k z^k$$

defines an analytic function only within a disk $|z| < a$ for some nonzero real constant a. To analytically continue $f(z)$ outside of this disk, we consider a slighlty different function

$$F(z) = \sum_{k=0}^{\infty} \frac{f_k}{k!} z^k\ ,$$

which will converge to an analytic function for all z. It then follows that

$$f(z) = \int_0^\infty dt\ F(zt)e^{-t}$$

will agree with $f(z)$ within the disk of radius a and will extend beyond this into a polygonal region bounded by the nearest singularities of $f(z)$. For example,

if $f(z)$ has a single pole located at $z = a + i0$, then the integral is well defined for $x < a$. If the pole is at $z = -a + i0$, the integral is well defined for $x > -a$, and if the pole is at $z = 0 + ia$, then the integral is well defined for $y < a$. This trick is called *Borel Summation*.

Our final, example of analytic continuation is due to the famous Indian mathematician Ramanujan, and it is remarkable in the sense that if we only know a function $f(z)$ at every non-negative integer, $z = 0, 1, 2, 3, \cdots$ then

$$f(-z) = \frac{\sin \pi z}{\pi} \int_0^\infty dt\ t^{z-1} \sum_{n=0}^{\infty} (-t)^n f(n)\ ,$$

provides f everywhere the integral converges! This really stretches the imagination of what a "patch" constitutes. One can do many beautiful things with this result, for example, if you wondered how we obtained

$$\int_0^\infty du\ \frac{u^{w-1}}{1+u} = \frac{\pi}{\sin \pi w}\ ,$$

consider the implication of Ramanujan's formula for the given entire function $f(z) = 1$. You can also use this formula to derive the analytic continuation of the Gamma Function, directly. [Hint: Let $f(n) = 1/n!$ and follow through.] Finally, Ramanujan's formula can be recast as

$$f(-z) = \frac{1}{\Gamma(z)} \int_0^\infty dt\ t^{z-1} \sum_{n=0}^{\infty} \frac{(-t)^n}{n!} f(n)\ ,$$

with better opportunities for the integral to be defined over more of the complex plane. Notice that setting $f(z) = 1$ in *this* formula tells us nothing we did not already know!

8. Lie Groups and Lie Algebras

For groups of size $\mathfrak{c} = 2^{\aleph_0}$, we can no longer just list the elements, but in fact, must index them with a continuous parameter (or a countable number of parameters, if appropriate), say $\alpha \in \mathbb{R}$ or $\mathbb{C}$. And because of the nice topological properties of $\mathbb{R}$, like continuity, completeness, and so on, these objects, known as *Lie Groups*, are in fact manifolds. We have already encountered a Lie Group in the general linear group $\mathrm{GL}_n(\mathbb{F})$ of invertible transformations from an n-dimensional vector space to itself. To ensure the necessary topologies, we shall restrict $\mathbb{F}$ to be either $\mathbb{R}$ or $\mathbb{C}$ in what follows.

Perhaps the simplest Lie Group to consider first is called $\mathrm{U}_1(\mathbb{C})$, the *Unitary Group*. As a subgroup of the General Linear Group of 1x1 matrices with complex coefficients, $\mathrm{GL}_1(\mathbb{C})$, there is admittedly not much of "matrices" actually involved here. Both groups just consist of the complex numbers $\mathbb{C}$, viewed as a group (remember it is actually field) under multiplication only—we forget about addition here. What sets $\mathrm{U}_1(\mathbb{C})$ apart is that we restrict its membership to complex numbers with unit modulus, or, which can be found on the unit

circle $|z| = 1$. Therefore, it requires a single real parameter, let's call it θ to parameterize the elements of this group which can be written as

$$e^{i\theta} \ , \ \theta \in \mathbb{R} \ .$$

It is easy to demonstrate that $\mathrm{U}_1(\mathbb{C})$ is in fact a group under multiplication.

If we apply an element of $\mathrm{U}_1(\mathbb{C})$ to any complex number z we effectively rotate it counterclockwise about the origin through an angle θ. So $\mathrm{U}_1(\mathbb{C})$ looks for all intents and purposes like the symmetry group of rotations on the two-dimensional space $\mathbb{C}$.

But, because there is an isomorphism between $\mathbb{C}$ and $\mathbb{R}^2$ regarded as vector spaces, it seems there must be a subgroup of $\mathrm{GL}_2(\mathbb{R})$ that effects rotations just like this in $\mathbb{R}^2$. In fact, there is, of course. It is $\mathrm{SO}_2(\mathbb{R})$, the *Special Orthogonal Group* also known as the *circle group*. The elements of $\mathrm{SO}_2(\mathbb{R})$ can be represented by 2x2 orthogonal matrices

$$Ro(\theta) \equiv \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \ ,$$

indexed by a continuous parameter $\theta \in \mathbb{R}$. This is a commutative group with respect to multiplication. Multiplication of a vector in $\mathbb{R}^2$, by $Ro(\theta)$ corresponds to a counterclockwise rotation through an angle θ, so this Lie Group describes one-dimensional rotations about the origin in a two-dimensional space, also. In other words, it is the *symmetry group* of rotations in a two-dimensional space as well. So, $\mathrm{U}_1(\mathbb{C})$ and $\mathrm{SO}_2(\mathbb{R})$ are basically the same Lie Group.

What distinguishes $\mathrm{SO}_2(\mathbb{R})$ from $\mathrm{GL}_2(\mathbb{R})$ is that $\det[Ro(\theta)]$=1, and that $Ro(-\theta) = Ro(\theta)^T = Ro(\theta)^{-1}$.

The rotation group in two dimensions $\mathrm{SO}_2(\mathbb{R})$, comes equipped with one *group generator*:

$$S \equiv \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \ .$$

It follows that any element of the group, $Ro(\theta)$, can be expressed both as a matrix exponential of the parameter times the group generator, or as a linear combination of the group generator and the identity matrix:

$$Ro(\theta) = \exp(\theta S) = I\cos\theta + S\sin\theta = \sum_{k=0}^{\infty} \frac{\theta^k}{k!} S^k \ ,$$

where

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

Notice that $S^2 = -I$, $S^3 = -S$, and $S^4 = I$, so that S in fact behaves just like $i = \sqrt{-1}$. And it is this aspect of S which allows us to state the isomorphism between $\mathrm{SO}_2(\mathbb{R})$ and $\mathrm{U}_1(\mathbb{C})$ as

$$\exp(\theta S) = \sum_{k=0}^{\infty} \frac{\theta^k}{k!} S^k \ \leftrightarrow e^{i\theta}$$

Next in order of complication must come $\mathrm{SU}_2(\mathbb{C})$ and $\mathrm{SO}_3(\mathbb{R})$, which, our intuition suggests, ought to be the same object masquerading in different guises, as we just discovered for $\mathrm{SO}_2(\mathbb{R})$ and $\mathrm{U}_1(\mathbb{C})$. This turns out to be "mostly" true.

The *Special Unitary Group* $\mathrm{SU}_2(\mathbb{C})$ consists of all 2x2 matrices of the form

$$\begin{pmatrix} z & w \\ -w^\star & z^\star \end{pmatrix}$$

where $|z|^2 + |w|^2 = 1$, and $z = x + iy, w = u + iv \in \mathbb{C}$. It requires three real parameters to index each element of the group. This is reassuring because rotations in three-dimensions also require three parameters, typically, the three Euler angles. There are now three group generators,

$$S_1 \equiv \frac{1}{2}\begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix} , \qquad S_2 \equiv \frac{1}{2}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} , \qquad S_3 \equiv \frac{1}{2}\begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix} .$$

These are simply the three Pauli Spin Matrices divided by $2i$. The group generators satisfy the *commutation relations*:

$$[S_i, S_j] \equiv S_i S_j - S_j S_i = \epsilon_{ijk} S_k ,$$

and we also have

$$S_1^2 = S_2^2 = S_3^2 = -\frac{1}{4} I ,$$

$$S_1^2 + S_2^2 + S_3^2 = -\frac{3}{4} I = -\frac{1}{2}\left(\frac{1}{2} + 1\right) I .$$

This reminds us of the behavior of the total angular momentum operator for a spin-1/2 particle in quantum mechanics. The correspondence becomes complete if we rescale each $S_j \to i\hbar S_j$.

Should all of this strike you as strangely familiar, feel obliged to award yourself some bonus points! In fact, twice the three group generators behave in exactly the same fashion as the nontrivial basis vectors for the quaternions, $\mathbb{H}$:

$$2S_1 \leftrightarrow i , \qquad 2S_2 \leftrightarrow j , \qquad 2S_3 \leftrightarrow k .$$

As a quaternion ζ can be expressed as

$$\zeta = x + iy + ju + kv , \quad x, y, u, v \in \mathbb{R}$$

so can a general element of $\mathrm{SU}_2(\mathbb{C})$ be expressed as a linear superposition of the unit matrix and the three group generators. The four real coefficients that appear in such a superposition are not independent owing to the fact that the element belongs to $\mathrm{SU}_2(\mathbb{C})$. And to finish the isomorphism, the corresponding quaterions must have unit norm, so $x^2 + y^2 + u^2 + z^2 = |z|^2 + |w|^2 = 1$ and lie on the surface of a sphere of unit radius in a four-dimensional space!

The Special Orthogonal Group $\mathrm{SO}_3(\mathbb{R})$ consists of all 3x3 matrices with unit determinant, and with the property that the transpose of a matrix is also its inverse. The three group generators are

$$S_1 \equiv \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \ , \qquad S_2 \equiv \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix} \ , \qquad S_3 \equiv \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \ .$$

They satisfy the same commutator relation that we derived for the $\mathrm{SU}_2(\mathbb{C})$ generators:

$$[S_i, S_j] = S_i S_j - S_j S_i = \epsilon_{ijk} S_k \ ,$$

only now,

$$S_1^2 + S_2^2 + S_3^2 = -2I = -1\,(1+1)\,I \ ,$$

which appears more appropriate for a spin-1 particle under the rescaling $S_j \to i\hbar S_j$. We now have

$$S_i^3 = -S_i \ , i = 1, 2, 3 \ .$$

The generator S_3 corresponds to rotations about the x_3-axis, which creates elements of the form

$$Ro(0, 0, \theta_3) = \begin{pmatrix} \cos\theta_3 & -\sin\theta_3 & 0 \\ \sin\theta_3 & \cos\theta_3 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

with similar statements for S_2 and S_1, i.e.,

$$Ro(\theta_1, 0, 0) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_1 & -\sin\theta_1 \\ 0 & \sin\theta_1 & \cos\theta_1 \end{pmatrix}$$

$$Ro(0, \theta_2, 0) = \begin{pmatrix} \cos\theta_2 & 0 & \sin\theta_2 \\ 0 & 1 & 0 \\ -\sin\theta_2 & 0 & \cos\theta_2 \end{pmatrix} \ .$$

Notice that the 2x2 subblocks of $Ro(0, 0, \theta_3)$ and $Ro(0, 0, \theta_1)$ look just like $Ro(\theta)$ of $\mathrm{SO}_2(\mathbb{R})$. There is a sign flip for rotations about the x_2-axis. If you are a glutton for algebra, you can of course multiply these three matrices out to get a general rotation through all three angles. However, order matters here because $\mathrm{SO}_3(\mathbb{R})$ is not a commutative group! Therefore it is meaningless to write an element as $Ro(\alpha, \beta, \gamma)$ without indicating in what order we carried out these rotations, precisely because

$$Ro(\theta_1, 0, 0) Ro(0, \theta_2, 0) \neq Ro(0, \theta_2, 0) Ro(\theta_1, 0, 0)$$

or equivalently

$$e^{\theta_1 S_1} e^{\theta_2 S_2} \neq e^{\theta_2 S_2} e^{\theta_1 S_1} \ .$$

The anticipated isomorphism between $\mathrm{SU}_2(\mathbb{C})$ (and the quaternions of unit norm!) and $\mathrm{SO}_3(\mathbb{R})$ takes the form

$$z = \cos\frac{\beta}{2}\cos\frac{\alpha+\gamma}{2} + i\cos\frac{\beta}{2}\sin\frac{\alpha+\gamma}{2} = x + iy \ ,$$

$$w = \sin\frac{\beta}{2}\cos\frac{\alpha-\gamma}{2} - i\sin\frac{\beta}{2}\sin\frac{\alpha-\gamma}{2} = u + iv,$$

in terms of *half angles*! If you've been following this all carefully, then you should be able to figure out how to identify α, β, γ with $\theta_1, \theta_2, \theta_3$. Good luck!

Of course, $\mathrm{SO}_3(\mathbb{R})$ is an essential element of our space-time symmetries and forms a subgroup of the Galilean and Poincaré Groups. In quantum mechanics, $\mathrm{U}_1(\mathbb{C})$ is the gauge symmetry group of electromagnetism, $\mathrm{SU}_2(\mathbb{C})$ is the gauge symmetry group of the electro-weak force, and $\mathrm{SU}_3(\mathbb{C})$ is the symmetry group of the strong force. $\mathrm{SU}_3(\mathbb{C})$ consists of all 3x3 complex matrices such that the conjugated transpose of an element is its inverse, and the determinant of each element is one. This group requires 8 group generators, which are called the Gell-Mann matrices, and 8 continuous real parameters.

Both $\mathrm{SO}_3(\mathbb{R})$ and $\mathrm{SU}_2(\mathbb{C})$ had a (single) suggestive quantum mechanical identity involving the sums of squares of the group generators. Such identities are called *Casimirs*. $\mathrm{SU}_3(\mathbb{C})$ has both a quadratic *and* a cubic Casimir.

To conclude, I'd like to leave you with a few ponderables.

First, if

$$\mathrm{U}_1(\mathbb{C}) \mapsto \text{ QED} \ , \qquad \mathrm{SU}_2(\mathbb{C}) \mapsto \text{ electroweak} \ , \qquad \mathrm{SU}_3(\mathbb{C}) \mapsto \text{ QCD} \ ,$$

what of $\mathrm{SU}_4(\mathbb{C})$? Its elements are 4x4 complex matrices, our space-time comes with 4 dimensions. Some of the elements of the Poincaré Group can be represented as 4x4 matrices. There are 15 group generators and 3 Casimirs. Is this a road to quantum gravity? If it was, don't you think we would hardly have been the first to figure out that 4 comes after 1, 2, and 3? So why doesn't it work? Is it because there is no division algrebra left at 16 dimensions (remember the quaternions were 4 dimensional and the octonions are 8 dimensional).

Second, what sort of Lie Groups analogous to the SO and the SU series could we build if we filled our matrices with quaternions drawn from $\mathbb{H}$ instead of using $\mathbb{R}$ and $\mathbb{C}$? I'd start, of course, with the 1x1 matrices of quaternions which have unit norm and see where that leads [to $\mathrm{SU}_2(\mathbb{C})$ of course, as we just discovered]. Then look up *symplectic Lie Groups*!

Third, linear systems of differential equations for a vector of dependent variables $\mathbf{x} = \{x_t(t), x_2(t), \cdots x_n(t)\}$ can be written in matrix form

$$\frac{d}{dt}\mathbf{x} = A(t)\mathbf{x} + \mathbf{y}(t) \ ,$$

so it is tempting to think about developing a solution as a matrix exponential along the general approach we used for solving the equation of radiative transfer:

$$\mathbf{x}(t) \text{“} = \text{”} \int_0^t d\tau \exp\left(-\int_t^\tau dsA(s)\right) \ \mathbf{y}(\tau) + \exp\left(\int_0^t dsA(s)\right) \ \mathbf{x}(0) \ .$$

Indeed, if A is restricted to 4x4 matrices, this is exactly the problem we face to solve the transfer equation for polarized radiation! Numerically, of course, we would approach this by breaking ds into some finite number intervals and adding together the sequence of A matrices. But what if $A(s)$ and $A(s + \Delta S)$ do not commute with one another? Do we really mean

$$\exp\left(\int_0^t dsA(s)\right) = \lim_{n\to\infty} \exp\left[\frac{t}{n}\sum_{i=1}^{n} A\left(\frac{2i-1}{2n}\right)\right] = \lim_{n\to\infty}\prod_{i=1}^{n} \exp\left[\frac{t}{n}A\left(\frac{2i-1}{2n}\right)\right]$$

when we are dealing with matrices that do not commute? How do we assess the order? An important clue to sorting such things out is the Baker-Campbel-Hausdorff formula:

$$e^A e^B = e^{C(A,B)}$$

where

$$C = A + B + \frac{1}{2}[A, B] + \frac{1}{12}\left([A, [A, B] + [B, [B, A]]\right) + \cdots \ ,$$

a series which does not terminate, and where all the subsequent terms always carry a commutator $[A, B]$. So if A and B commute, so do their matrix exponentials, and order does not matter.

9. Further Reading

Perhaps the all around best single book for much of the material in this Appendix is the classic

⋆[**BM 1**] Garrett Birkhoff & Saunders MacLane, A Survey of Modern Algebra, (New York, NY: The Macmillan Company; 1950), xi+450,

which is a very readable and accessible book even for the would be astrophysicist.

For groups and rudiments of group theory, supplement this tome with

⋆[**J 2**] David Joyner, Adventures in Group Theory. Rubik's Cube, Merlin's Machine & Other Mathematical Toys, (Baltimore, MD: Johns Hopkins University Press; 2002), xv+262 ,

a much more whimsical, but no less serious, foray into lots of things one can actually do with finite groups. Pages 168-172 provide a fascinating table of all the finite groups with cardinal number of 25 or less. I can spend hours looking at this and never get bored.

For vector spaces, I have relied heavily on the two superb contributions by Paul Halmos,

⋆[**H 4**] Paul R. Halmos, Finite-dimensional Vector Spaces, 2nd Edn., (Mineola, NY: Dover Publications; 2017), viii+200,

[**H 5**] Paul R. Halmos, Introduction to Hilbert Space and the Theory of Spectral Multiplicity, (Mansfield Center, CT: Martino Publishing; 2013), -+114.

Halmos builds things so logically and consistently that it is hard to get confused, although, I have found it necessary in the above to try to improve upon his notation. To go beyond Halmos, try

[**AG 1**] N.I. Akhiezer & I.M. Glazman, Theory of Linear Operators in Hilbert Space. Two Volumes Bound as One, (New York, NY: Dover Publications; 1993), xi+147/ii+218.

I have given short shrift to topology and (real) analysis, which is required to bridge the gap between vector spaces over $\mathbb{R}$ or $\mathbb{C}$ and calculus on manifolds. In some sense, once you have a metric space, then much of the fascinating and intriguing bits about topology become, well, largely unnecessary. *Had* I tried to fill this gap, I would have made use of two wonderful little volumes by Michael Gemignani. Start with
[**G 6**] Michael C. Gemignani, Elementary Topology, 2nd Edn., (Reading, MA: Addison-Wesley Publishing Company; 1972), xi+270,
and then finish up with,
[**G 7**] Michael Gemignani, Introduction to Real Analysis, (Philadelphia, PA: W.B. Saunders; 1971), viii+160.
Finally, for metric spaces in general, try
[**B 8**] Victor Bryant, Metric Spaces. Iteration and Application, (Cambridge, UK: Cambridge University Press; 1996), vi+104,
[**C 8**] E.T. Copson, Metric Spaces, (Cambridge, UK: Cambridge University Press; 1968), vii+143.

Of the vast number of books that have been written on or about complex analysis, I am completely smitten by a relatively new offering by Bengt Fornberg and Cécile Piret,
⋆[**FP 1**] Bengt Fornberg & Cécile Piret, An Illustrated Introduction to Analytic Functions, in press.
Look for it!

For Lie Groups and associated Lie Algebras, Penrose [**P 8**], Moriyasu [**M 2**], Cantwell [**C 1**], and Gelfand et al [**GMS 1**] will all reward your study with different aspects of a complicated subject. In addition, a really beautiful synthesis is provided by
[**S 9**] Stephanie Frank Singer, Linearity, Symmetry, and Prediction In the Hydrogen Atom, (New York, NY: Springer; 2005), xiv+396.
For more tradiitional mathematically-oriented offerings, see
[**J 3**] Nathan Jacobson, Lie Algebras, (New York, NY: Dover Publications; 1979), ix+331,
[**L 5**] Harry J. Lipkin, Lie Groups for Pedestrians, (Mineola, NY: Dover Publications; 2002), ix+182,
[**H 6**] Robert Hermann, Lie Groups for Physicists, (New York, NY: W.A. Benjamin; 1968), ix+193.

For the wild and whacky world of transfinite cardinal arithmetic, I like
⋆[**K 4**] E. Kamke, Theory of Sets, (New York, NY: Dover Publications; 1950), vii+144.

Lastly, the following two volumes in the Princeton Companion Series are a *must* for looking beyond what little I could share with you here.

⋆[**G 8**] Timothy Gowers, ed., The Princeton Companion to Mathematics, (Princeton, NJ: Princeton University Press; 2008). xxi+1034,
[**H 7**] Nicholas J. Higham, ed., The Princeton Companion to Applied Mathematics, (Princeton, NJ: Princeton University Press; 2015), xvii+994.

10. Appendix A: Transfinite Cardinal Arithmetic

For no other reason that it is fun and so bizarre, we present here Georg Cantor's "arithmetic" of the transfinite cardinal numbers. From §2, recall that the smallest "infinity", is the countable infinity of the integers, which Cantor denoted by $\aleph_0$. It is not known whether the size of the set of real numbers, designated $2^{\aleph_0}$ is the next largest transfinite cardinal number, or if there is one (or more) transfinite cardinal numbers between these two. It's an open problem, and if you find yourself with some spare time on your hands you might try solving it. At any rate, we do know that for any finite integer $n \in \mathbb{N}$,

$$0 < n < \aleph_0 < 2^{\aleph_0} \ .$$

The last designation is consonant with the fact that there is a one-to-one mapping between the real numbers and the set of all subsets of the integers.

It's easy to make bigger things still. For example, the set of all subsets of the real numbers has to be, by definition, bigger than the set of real numbers (because there is no one-to-one correspondence between them), this set can be put in one-to-one correspondence with the set of all functions $f(x)$ defined on the interval $x \in [0, 1]$ for example. As the notation gets unwieldy here, this even larger transfinite cardinal number is sometimes just designated $\mathfrak{f}$ (for $\mathfrak{f}$unctions) and $2^{\aleph_0} = \mathfrak{c}$ (for $\mathfrak{c}$ontinuum) and $\aleph_0 = \mathfrak{d}$ (for $\mathfrak{d}$enumerable).

Now for the fun part, because the even and the odd numbers both have size $\mathfrak{d}$ and we can combine them to give all the integers $\mathbb{Z}$ which also has size $\mathfrak{d}$, so it must therefore be the case that in this sense

$$\mathfrak{d} + \mathfrak{d} = \mathfrak{d} \implies 2\mathfrak{d} = \mathfrak{d} \ .$$

By dividing the integers $\mathbb{Z}$ into any finite number n number of disjoint subsets that have size $\mathfrak{d}$, which can be combined to create $\mathbb{Z}$ or simply using induction on the previous equation, you can hopefully convince yourself that

$$n\mathfrak{d} = \mathfrak{d} \ ,$$

for any finite n. Perhaps even more bizarre is (hint: think about disjoint subsets of $\mathbb{Z}$ that might be generated by the prime numbers)

$$\mathfrak{d}\mathfrak{d} = \mathfrak{d} \implies \mathfrak{d}^2 = \mathfrak{d} \ .$$

Notice an uncanny resemblance of $\mathfrak{d}$ to zero! But we can carry out induction on this equation to obtain the even more unsettling result

$$\mathfrak{d} = \mathfrak{d}^n \ ,$$

for any finite integer n.

However, we will not be able to carry the induction on indefinitely, because, as we noted previously

$$2^{\mathfrak{d}} = \mathfrak{c} > \mathfrak{d}$$

is bigger than $\mathfrak{d}$. The logic beyond this point becomes much more intricate. We'll simply quote the analogous arithmetic for $\mathfrak{c}$:

$$\mathfrak{c} = 2^{\mathfrak{d}} = n^{\mathfrak{d}} = \mathfrak{d}^{\mathfrak{d}} = \mathfrak{c}^{n} = \mathfrak{c}^{\mathfrak{d}} \ ,$$

where you will notice, that the only thing missing is anything raised to the $\mathfrak{c}$ power. Which, of course, makes sense because we know

$$2^{\mathfrak{c}} = \mathfrak{f}$$

is bigger than $\mathfrak{c}$.

Now, without knowing anything except how to unravel puzzles, and look for symmetries, can you write down the analogous arithmetic for $\mathfrak{f}$ based solely on the arithemtic of $\mathfrak{d}$ and $\mathfrak{c}$?