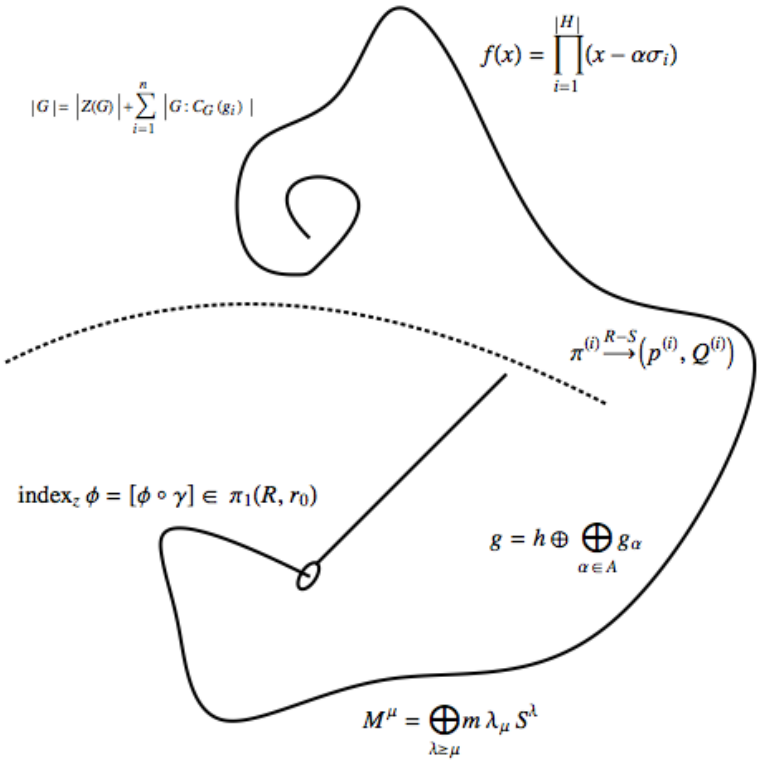
Capstone in Mathematics

The Critical Thread:



The Symmetric Group in Galois Theory, Representation Theory, Representations of Lie Algebras, Combinatorics, and Topology

Author:
Caleb G. McWhorter
Ithaca College

Supervisors:
Dr. Emilie Wiesner
Dr. David Brown

August 31, 2015

Contents

Dear Sir or Madam, will you read my book, it took me years to write, will you take a look?

John Lennon and Paul McCartney, *Paperback Writer*, single

1. Thanks

I would like to thank my advisor Dr. David Brown for listening while I talked through the big ideas and pointing out when I missed the small ones. I would like to thank my thesis advisor Dr. Emilie Wiesner for all her support and help in the comprehension of the Theorems and showing that indeed drawing boxes can be useful! I would also like to thank Dr. James West for his thorough covering of Universal Covering Spaces, without which, the Topology section would not be possible. Also, Professor Keith Dennis without whom the remarks about Tensor Products and Canonical forms would not be possible. Moreover, I'd like to give thanks to Dr. Teresa Moore, who helped me finally see proofs for what they are. To my friends and family, especially my friends Stephen Gorgone, Cole Lechleiter, and Ethan Wennberg for endlessly having to hear about what I was doing. It is always most helpful to have someone to listen as I talk out the ideas. Finally, I'd like to thank the rest of my professors at Ithaca College and Cornell University who have forever shaped my Mathematical thinking.

2. Forward to the Reader

The purpose of this text is not only to demonstrate the knowledge I have accumulated during my time as an undergraduate at Ithaca College and time spent at Cornell University, but as a demonstration of my ability to apply that Mathematics and learn new concepts independently. Moreover, it is to demonstrate my ability to communicate that knowledge to others. Accordingly, this text was written for the advanced undergraduate Mathematics major. Many of the proofs belong to the cited sources. However, when those texts are absent or "lacking" they have been filled in by myself. The purpose is not to replicate or replace those texts but to provide the shortest route to the concept being discussed, i.e. to eliminate "extra" or "extraneous" concepts. That is, to connect fields and concepts in Mathematics sufficiently enough to see the details while removing enough details to clearly present the big picture - the sinews binding the fields of Mathematics. The interested reader is of course suggested to read the cited texts. Finally, although this text will provide the necessary knowledge to follow the argumentation, the text in no way explains all possible concepts and much is assumed to the reader. Though much of this knowledge is unnecessary to understand the text, there are key concepts or terminology the reader should be familiar with to make full use of this text. For simplicity, especially if the reader desires only to read a portion of the text, the assumed knowledge is broken down here:

- Galois Theory: The division algorithm, group properties and definition, subgroups, special groups (such as $Z(G)$, $C(g)$, cyclic groups, permutation groups to some extent, $\mathrm{Aut}(G)$, $\mathrm{Inn}(G)$, $U(n)$, $\mathrm{stab}_G(i)$, $\mathrm{orb}_G(i)$, $\ker(\varphi)$, $\mathrm{End}(G)$), the Isomorphism and Homomorphism theorems, cosets, Lagrange's Theorem, external and internal direct products, normal groups, factor groups, Cauchy's Theorem for abelian groups, Sylow's Theorems, Fundamental Theorem of Abelian Groups, ring properties and definitions, ideals, integral domains (ID), fields, ring homomorphisms, polynomial rings, reducibility, factorizations of polynomials, Eisenstein Criterion, Cyclotomic polynomials, unique factorization domains (UFD), principal ideal domains (PID), ascending/descending chains, vector spaces, vector subspaces, linear independence/dependence, bases, and some basic knowledge of tensor products.
- Representation Theory: Some of the above group theory knowledge, basic knowledge of the permutation group, cosets, isomorphisms, homomorphisms, special groups (as above), vector spaces, vector subspaces, linear independence/dependence, bases, and some knowledge about matrices.
- Lie Representations: Much if not all of the knowledge from Galois Theory, vector spaces, vector subspaces, linear independence/dependence, bases, eigenvalues, bilinear forms, euclidean spaces, some tensor products, algebras. Though not necessary, it is also useful to know the special matrix groups and their algebras (such as $O(n)$, $U(n)$, $SU(n)$, $Sp(n)$, $\mathfrak{sl}(n,\mathbb{C})$, $\mathfrak{so}(n)$, $\mathfrak{sp}(n)$, et cetera), and commutative diagrams.
- Topology: The definition of a topological space, basis, closed sets, boundary, closure, interior, dense, continuous functions, homeomorphisms, open/closed maps, limit points, "special topologies" (such as the standard, order, product, subspace,

metric, and quotient topologies), countability, metrics, Hausdorfness, connected spaces, compactness, components, separation axioms, countability axioms, normality, Urysohn lemma, Tietze Extension Theorem, Tychonoff Theorem, and some complex analysis (such as winding numbers and various aspects of contour integration).

- Combinatorics: Not much is assumed here. All that is required is some knowledge of basic counting principals and a good understanding of group theory.

Moreover, there things that are not easily expressed in a list which the reader is expected to have, such as the ability to follow proofs and understand typical Mathematical jargon, such as "if and only if", "...that makes the following diagram commute", "for all" and its negation "there exists at least one...", "agrees with φ on ...", et cetera. Though these concepts are assumed, a few of the above are briefly stated or reviewed for the reader for issues of clarity. Finally, this text in no way is exhaustive or even sufficient in its description in the topics it contains. However, the text should contain the knowledge that is necessary to understand the basic principals. However, the focus will be on the bigger picture and the connections to simpler mathematics or to other fields of Mathematics. The interested reader is of course suggested to continue their reading to more in-depth texts - the works cited page is a good start.

Finally, I have worked strenuously to create a coherent system of language, definition, and notation in this paper - especially working from texts with very different notation for the same topic. While it is standard for an algebraist to write $a\phi$, whereas an analyst would write $\phi(a)$ and the topologist would write either, I believe it is clearest when written in the form $\phi(a)$ to avoid confusion when many symbols are present. Moreover, some texts will have commutative diagrams written inverted from what is seen here. Again, I believe the way I have done it is clearest within the context in which I am working, especially when a diagram would not often be written. The order of the chapters have been presented such that some knowledge that would otherwise need to have been assumed has already been presented in a previous chapter, helping to create a somewhat smooth flow between concepts. Lastly and importantly, any errors are my own and do not reflect on the sources or my instructors. Any errors the reader wishes to report or suggestions for revision can be sent to cbgmcwhorter@gmail.com and would gladly be taken with my deepest gratitude.

Introduction

Mathematics is not a linear object. Though it may start off in such a fashion, soon it is a interweaving fabric of logic and definitions which create a coherent system of solving problems. Indeed, at the heart of Mathematics is logic. Once a definition in Mathematics is made, all the results that follow were always true, we just did not see them. Mathematics is then not a journey of construction or proof but of quest for understanding. The most powerful Mathematics comes when the various fields work together to solve a problem. Often, the first proof is not the simplest. It is the profession of the mathematician to look for a better way or a shorter way. Indeed, many of the results one reads in textbooks seem rather trivial. Why where these results so slow to arrive? The proof one often sees is decades of careful dissection and understanding which often comes when more discoveries had been made and a coherent theory created. Moreover, the proofs, especially of the more difficult questions or those which are most general, come only when one combines the fields of Mathematics into a powerful theorem proving machine. For example, the classification of all finite simple groups took over 100 years and all the tools of Analysis, Lie Theory, Group and Field Theory, and Topology to finally put to rest in a total proof length of over 10,000 mathematical journal pages. Today, after great work in careful reduction this number has been reduced substantially to a 8 volume set of books.

More often than not, the thread which binds the fields together and patches holes is not a singular theorem but rather a broader concept that bridges the gap. The concept of looking at the possible ways to count arrangements of a collection of objects - permutations - is a simple idea but is seen in nearly every field of Mathematics in some form. Its power to count and view the structure of an object is enhanced by a theory that came much later than the idea of a permutation itself - the concept of a group. As Fraleigh said, "never underestimate a theorem which counts something." Indeed, I would contest that there is not a field of Mathematics that does not use the concept of a group in one way or another. The concept of a group forces a coherent structure on an object which only further enhances the probing power of the symmetry of an object. In fact, embedded in the concepts of a group is the idea of a permutation. Any binary operation in a group takes an element and "permutes" that element through (left/right) multiplication to another element in the group. This viewpoint yields is exactly Cayley's Theorem:

Theorem 2.1. *(Cayley's Theorem) Every group is isomorphic to a group of permutations.*

Proof: Let G be a group. We prove the theorem in 3 steps:

1. We construct a mapping from G to itself and show that it is a permutation.
2. Using the mapping in (1), we construct a group of permutations S_G.
3. We show that G is isomorphic to S_G.

We now will follow the roadmap we have just established.

1. Let $P_g : G \to G$ be a map defined by $x \mapsto gx$ for a fixed $g \in G$. To show that P_g is a permutation of G, we need to show that it is a bijection of G to itself. Suppose

that $P_g(x) = P_g(x')$, then $gx = gx'$ and left cancelation yields $x = x'$, showing P_g is injective. For surjectivity, suppose we wish to produce $x \in G$. Because $g, x \in G$ and G is a group, $g^{-1} \in G$ and so is $g^{-1}x$. But then $P_g(g^{-1}x) = g(g^{-1}x) = (gg^{-1})x = ex = x$, demonstrating surjectivity. Therefore, P_g is a permutation.

2. The construction of the group should be obvious. Take S_G to be the set defined by $S_G = \{P_g \,|\, g \in G\}$. Since each P_g is a permutation, S_G is a set of permutations. However, we need to show that S_G is a permutation group. Hence, we need to give it a binary operation, show it is nonempty, contains identity, and is closed under inverses and its operation. Let the operation on S_G be function composition. We know that there is an identity in S_G since $e \in G$ then P_e is the identity in S_G because if $P_g \in S_G$ and $x \in G$. This shows that S_G is nonempty, then
$$(P_eP_g)(x) = P_e(P_g(x)) = P_e(gx) = e(gx) = gx = P_g(x)$$
and
$$(P_gP_e)(x) = P_g(P_e(x)) = P_g(ex) = g(ex) = gx = P_g(x)$$
Similarly, it is simple to show that S_G is closed under its binary operation. Suppose $P_g, P_h \in S_G$ and $x \in G$, then
$$(P_gP_h)(x) = P_g(P_h(x)) = P_g(hx) = g(hx) = (gh)x = P_{gh}(x)$$
Since $g, h \in G$, we know $gh \in G$ and this shows that $(P_gP_h)(x) = P_{gh}(x)$. We now need only show that each element in S_G has an inverse. Since G is a group, if $g \in G$ then $g^{-1} \in G$ and the inverse of P_g has the inverse $P_{g^{-1}}$ as
$$(P_gP_{g^{-1}})(x) = P_{gg^{-1}}(x) = P_e(x)$$
and
$$(P_{g^{-1}}P_g)(x) = P_{g^{-1}g}(x) = P_e(x)$$
Therefore, S_G is a group.

3. Finally, we show that $G \cong S_G$, which means we need to find an isomorphism between them. But this isomorphism has practically been made for us! Let $\varphi : G \to S_G$ be the mapping defined by $g \mapsto P_g$. We need to show that φ is surjective, injective, and a homomorphism. First, to see that it is surjective is simple since each $P_g \in S_G$ must come from at least one $g \in G$, namely g itself. To see injectivity, let $g, h \in G$ and suppose $\varphi(g) = \varphi(h)$, then $P_g = P_h$ for all $x \in G$. Letting $x = e \in G$, we see that
$$\begin{aligned} P_g(e) &= P_h(e) \\ ge &= he \\ g &= h \end{aligned}$$
Finally, we need to show that φ is a homomorphism. But this follows directly from the properties of P_g. Suppose that $x, y \in G$ then
$$\varphi(xy) = P_{xy} = P_xP_y = \varphi(x)\varphi(y)$$
Therefore, φ is an isomorphism and $G \cong S_G$, completing the theorem. □

First, notice we constructed $P_g(x) = gx$ through left multiplication, this is called the left regular representation. There is nothing wrong with having chosen $P_g(x) = xg$, called the right regular representation. Cayley's Theorem would still have followed in the same way. However, we may obtain a different subgroup of S_n (as the group action in any subgroup of S_n is almost never commutative it would follow that left multiplication by an element would unlikely correspond to the same permutation induced by applying the permutation on the right). But this group would still have to be isomorphic to the one constructed by the left regular representation. Second, notice that Cayley's Theorem holds for any group, even an infinite one!

Groups are the study of objects through a binary operation. Cayley's Theorem then tells us that this concept is exactly analogous to examining the symmetry in the way the group can be arranged or permuted. We should expect then that the symmetric group should have a critical role in the study of groups. Indeed, the symmetric group appears in some way in almost every field of Algebra. However, its presence is often not obvious. Our goal is to motivate and develop some of the bigger theorems of Mathematics, with an emphasis on illustrating where the concept of a permutation is applied.

3. Combinatorics

Introduction. We start by showing an application of the symmetric group in perhaps its most natural element - combinatorics. However, one should not make the mistake of thinking that the symmetric group as a counting object. Yes, the symmetric group can be used to count things. The order of the symmetric group may count the possible number of arrangements of an object but the elements of the symmetric group, as the reader should well be aware, are maps of the set to another possible arrangement of its elements. Therefore, as stated before, the symmetric group in some way informs us about the *symmetry* of a collection of objects (hence the name the symmetric group). The number of these symmetries is what we tend to count.

Here, we will first present how symmetries can be considered as groups using very visual examples. Then we extend this idea to group actions and show how these are again the same idea as a permutation. Finally, we give very special groups resulting from the concept of group actions and conclude with their culmination, Burnside's Theorem, to show how this can be applied in very real world examples.

Example 3.1. We will show that the group of rigid motions of the cube (that is motions you could obtain by simply picking the cube and turning it in some way) has order 24. This is the same as saying that there are 24 ways to rigidly move the cube onto itself. Let H be the set of vertices of the cube, i.e. $H = \{1, 2, \cdots, 8\}$. Label the vertices of the cube H such that 1 is adjacent to 2 and so forth. Consider the rigid motions of vertex 1. There are 7 different vertices to which 1 can be sent, not including the motion which fixes the cube. For each of these 8 total rigid motions, there are 3 possibilities for the placement of vertex 2 as it must be adjacent to vertex 1. This yields a total of $3 \cdot (7 + 1) = 24$ total possible rigid motions of the cube. Since a rigid motion must fix the edge connecting vertex 1 and vertex 2, this is sufficient to determine the cube. Hence, $|G| = 24$.

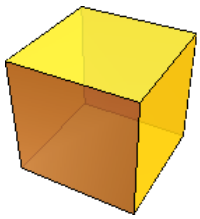


Figure 1. A standard cube in 3-space.

Notice in this example that the group of rigid motions are maps of the object onto itself and its order is the number of possible rigid motions. The group of rigid motions can be thought of as a permutation group. Indeed, the rigid motions must be a subgroup of the symmetric group, not just from an a fortiori argument resulting from Cayley's Theorem. We illustrate this with the following example:

Example 3.2. Let G be the group of rigid motions of the cube. We will show that $G \cong S_4$. We have already shown in Example 1 that $|G| = 24$. So we already know that G has the

same order of S_4. Observe that the cube has 4 diagonals and that any rigid motion induces a permutation of the diagonals. To show that the groups are isomorphic, it is sufficient to show that all 24 permutations of these diagonals come as a result of a rigid motion. Label the diagonals arbitrarily as 1, 2, 3, and 4. Consider a rigid rotation that would take a cube sitting on a flat surface and push it onto one of the sides who has an edge against the flat surface, this corresponds to the permutation (1234). A rotation of a right angle that is also right to the one corresponding to (1234) is (1423). But by closure of a group, the image of our group of rotations must at least contain the subgroup generated by $(1234), (1423)$: $\{(), (1234), (1234)(1234), (1234)(1234)(1234), (1423)(1423), (1423)(1423)(1234),$ $(1423)(1423)(1234)(1234), (1423)(1423)(1234)(1234)(1234)\}$. Moreover, it must contain the element (1234)(1423), which has order 3. Recall that the order group elements and subgroups must divide the order of a group. Our permutation group already contains a subgroup of order 8 and an element of order 3, the smallest possibility for the order of a group is 24. But then since this is the total order of our group, the rotations must yield all 24 possible rotations.

So how do we get the symmetric group to count for us? In the examples of above, we create a group which somehow contains the object of our attention and count the group elements which we are interested in. For examples, above we were interested in the number of rigid motions and the order of the group itself was the number of such motions. In general, the order of the whole group counts more than the total number of whatever it is we are interested in. Instead, we will often have to look at subgroups of G or perhaps create a quotient group. Typically, this will be done using subgroups such as the orbit or stabilizer. However, these are subgroups not under the typical group operation but through the concept of a group action.

3.1. **Group Actions.**

Definition 3.1. (Group Action) Given a group G and a set S, then a group action of G on S is a map $G \times S \to S$, denoted $g \cdot s$ and satisfies

1. $1 \cdot s = s$
2. $g_1 \cdot (g_2 \cdot s) = (g_1 g_2) \cdot s$

for all $g_1, g_2 \in G$ and $s \in S$. This is called the left action of G on S. When the set and group are clear, the operation is omitted and is simply written gs. The right action is defined analogously.

One should note that in the proof of Cayley's Theorem, we have already used the concept of a group action. We created an action where S was the group G, i.e. $G \times G \to G$ defined by $g \cdot s = gs$. In fact, this is a well known and used group action. Let G be a group and S be a set then define a map $\sigma_g : S \to S$ given by $\sigma_g(s) = g \cdot s$. Then for every fixed $g \in G$, σ_g is a permutation of S and the $\varphi : G \to S_S$ from the group G to the set of permutations on S, defined by $g \mapsto \sigma_g$ (the permutation representation), is a homomorphism. But this should have already all been clear from our proof of Cayley's Theorem.

Remark 3.1. We can also, similar to the idea of Cayley's Theorem, say G acts a nonempty set S if there is a homomorphism $\varphi : G \to S_S$. These definitions are equivalent.

Proof: Assume that G is a group and S is a nonempty set. We need to show the two group action axioms are satisfied. First, assume that there is a homomorphism $\varphi : G \to S_S$. Then for all $g, h \in G$ and $s \in S$,

$$(gh) \cdot s = \varphi(gh)(s) = (\varphi(g)\varphi(h))(s) = \varphi(g)(\varphi(h)(s)) = g \cdot (h \cdot s)$$

and because φ is a homomorphism, $e \cdot s = \varphi(e)(s) = s$.

Now assume that G acts on S. Let $\varphi : G \to S_S$ be a map defined by $\varphi(g)(x) = g \cdot x$. To show that $\varphi(g)$ is injective, suppose $\varphi(g)(x) = \varphi(g)(y)$, then $g \cdot x = g \cdot y$ and

$$\begin{aligned} g \cdot x &= g \cdot y \\ g^{-1} \cdot (g \cdot x) &= g^{-1} \cdot (g \cdot y) \\ (g^{-1}g) \cdot x &= (g^{-1}g) \cdot y \\ e \cdot x &= e \cdot y \\ x &= y \end{aligned}$$

Surjectivity is simple. Choose $s \in S$, then consider $g^{-1} \cdot s \in S$

$$\varphi(g)(g^{-1} \cdot s) = g \cdot (g^{-1} \cdot s) = (gg^{-1}) \cdot s = e \cdot s = s$$

Therefore, $\varphi(g)$ is a bijection and it only remains to show that φ is a homomorphism.

$$\varphi(gh)(x) = (gh) \cdot x = g \cdot (h \cdot x) = \varphi(g)(\varphi(h)(x)) = (\varphi(g)\varphi(h))(x)$$

But then the two definitions are equivalent. So a group action definition is equivalent of thinking of the group action as some permutation of S.

But why bother to consider groups actions? It is very similar to the inherent binary operation in a group G, so how do we get more from it? Indeed, the binary operation of a group is a group action of a group on itself. Obviously, we have already seen the use in proving Cayley's Theorem. However, we proved it earlier without ever having to consider our mappings as a group action. So where is the use? Using the idea of group action, one can study a set (or perhaps let the set be a group!) by forcing more structure on the set using another group. One can then use the group to learn more about the set, or use the set to learn about the group. Moreover, group actions lead to a partial converse of Lagrange's Theorem - Sylow's Theorem, which we will not examine here. The concept of permutations and group actions are powerful in their own right but together they prove extremely useful. We focus on three particularly useful sets created through group actions.

Definition 3.2. (Kernel) Given a group G acting on a nonempty set S, the kernel of the action is the set of all elements of G with act trivially on S:

$$\ker = \{g \in G \,|\, g \cdot s = s \quad \text{for all } s \in S\}$$

Note this kernel is not the same as a kernel of a homomorphism but follows the same idea, those elements which act trivially.

Definition 3.3. (Orbit) Let G be a group acting on a nonempty set S. Then the equivalence class $\mathcal{O}_G(s) = \{\, g \cdot s \,|\, g \in G\} = \text{orb}_G(s)$ is called the orbit of s under the action of G. Moreover, if there is only one orbit under the action of G on S, the action is called transitive.

Definition 3.4. (Stabilizer) Let G be a group acting on a nonempty set S. Then the stabilizer of $s \in S$ is $\text{stab}_G(s) = \{g \in G \,|\, g \cdot s = s\}$. In other words, the stabilizer of s is the set of $g \in G$ which act trivially on s.

Remark 3.2. It is easily checked that the stabilizer, orbit, and kernel are subgroups of G. In fact, the kernel is a normal subgroup of G.

Of course, we could have defined the kernel in terms of the stabilizer as $\ker = \bigcap_{s \in S} \text{stab}_G(s)$. This also shows that $\ker \leq \text{stab}_G(s)$. Of course, one would think there should be an intimate relation between actions and homomorphisms given they both have a kernel. This is indeed correct, suppose a group G acts on a nonempty set S and let $\varphi : G \to S_S$. Then one can obtain a group action through

$$g \cdot s = \varphi(g)(s)$$

This defined action is exactly the same as $\ker \varphi$. Moreover, there is an interesting relation between the orbit and stabilizer of a group G acting on a set S.

Theorem 3.1. *(Orbit-Stabilizer Theorem) Let G be a finite group of permutations acting on a set S. Then for any $s \in S$, $|G : stab_G(s)| = |orb_G(s)|$. Equivalently,*

$$|G| = |orb_G(s)|\,|stab_G(s)|$$

Proof: By Lagrange's Theorem, $|G : \text{orb}_G(s)|$ is the number of left cosets of $\text{orb}_G(s)$ in G. We create a bijection between these left cosets and $\text{stab}_G(s)$. Let $\mathcal{C}$ denote the set of left cosets of $\text{orb}_G(s)$ in G and let $\varphi : \mathcal{C} \to \text{stab}_G(s)$ be a map defined by $g \cdot S \mapsto g\,\text{stab}_G(s)$. We need to show

1. φ is well-defined.
2. φ is injective.
3. φ is surjective.

We can do (1.) and (2.) quite easily.

1 & 2. Fix a $s \in S$ and $g_1, g_2 \in G$ such that $g_1 \cdot s = g_2 \cdot s \in \text{orb}_G(s)$. Then

$$\begin{aligned} g_1 \cdot s &= g_2 \cdot s \\ g_2^{-1}(g_1 \cdot s) &= s \\ (g_2^{-1} g_1) \cdot s &= s \end{aligned}$$

This occurs of course if and only if $g_2^{-1} g_1 \in \text{stab}_G(s)$. But then $\varphi(g_1 \cdot s) = \varphi(g_2 \cdot s)$ and the mapping is well-defined. Since at each step we had an "if and only if", φ is injective. If there were no such g_1, g_2, then the result follows immediately as every $g \cdot s$ is unique. Hence, the action is well-defined and trivially injective.

3. Surjectivity is trivial for any $g\,\mathrm{stab}_G(s)$ came from at least one $g \in G$ as $g \cdot s \mapsto g\,\mathrm{stab}_G(s)$.

It follows that φ is a bijection. Therefore, we must have

$$\frac{|G|}{|\mathrm{orb}_G(s)|} = |\mathrm{stab}_G(s)|$$

or equivalently, $|G| = |\mathrm{orb}_G(s)|\,|\mathrm{stab}_G(s)|$. □

3.2. **Orbits and Fixed Points.** With sufficient review and definitions at hand, we are prepared to address the issue of putting groups to combinatorial use. It is clear that given a set of n elements, say X_n, that the number of possible orderings of the elements of n is $n!$. However, these types of counting questions are uninteresting. Often, we are interested in distinct arrangements. For example, a 3×3 checkered board is full of blank tiles are we are given the colors black, red, and blue with which to color the board. Not all possible colors of this checkered board using 3 colors would be unique, see Figure 2. It will be our goal to use groups to determine equivalent arrangements and count them. We follow Grove's "Groups and Characters" closely for its brevity in proofs[1].

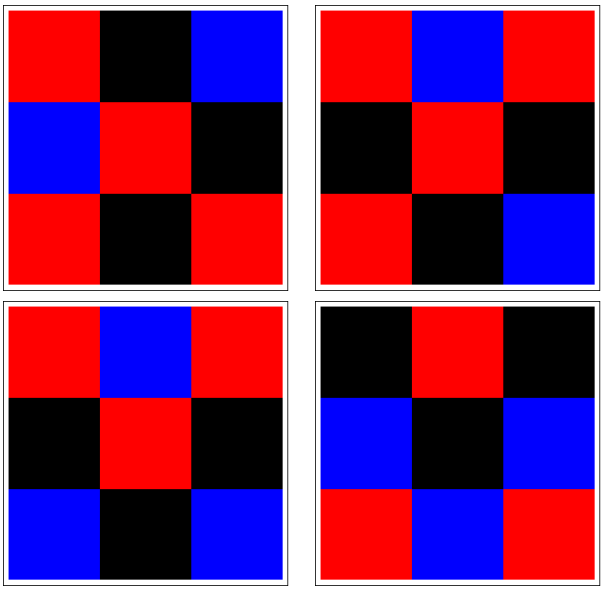


FIGURE 2. Here we see 4 possible 3×3 checkerboard tilings using 3 colors. Notice the top two are actually what we would consider to be equivalent as a 90° rotation counter-clockwise of the upper right board results in the upper left board. We then consider these arrangements of color to be equivalent. However, notice that the bottom two boards are not equivalent.

Definition 3.5. (Fix) Let G be a group acting on a set S. Given a $x \in G$, the $\text{fix}(x) = \{s \in S \,|\, x \cdot s = s\}$ or the set of elements $s \in S$ that are fixed by x.

Definition 3.6. (Character) Let G be a group acting on a set S. Then the character θ of the action is defined by $\theta(x) = |\text{fix}(x)|$. Therefore, $\theta : G \to \mathbb{N}$ is a function.

Lemma 3.1. *(Characters are Class Functions) The character θ of a group action is constant on the conjugacy classes of G, i.e. θ is a class function on G.*

Proof: We show that $\text{fix}(yxy^{-1}) = y\,\text{fix}(x)y^{-1}$. Suppose that $x, y \in G$. Let $r \in \text{fix}(x)$, then clearly $r \in \text{fix}(yxy^{-1})$ and $y\,\text{fix}(x)y^{-1} \subseteq \text{fix}(yxy^{-1})$. Next, assume that $r \in \text{fix}(yxy^{-1})$. Then $yxy^{-1} \cdot r = r$ and it follows that $yry^{-1} = r$. But this implies that $r \in y\,\text{fix}(x)y^{-1}$ and then $\text{fix}(yxy^{-1}) \subseteq y\,\text{fix}(x)y^{-1}$. Therefore, $\text{fix}(yxy^{-1}) = y\,\text{fix}(x)y^{-1}$. Finally, it then follows

$$\theta(yxy^{-1}) = |\text{fix}(yxy^{-1})| = |y\,\text{fix}(x)y^{-1}| = |\text{fix}(x) = \theta(x)$$

□

We then get the following very useful formula:

Theorem 3.2. *(Burnside's Theorem) If G is a finite group and acts on a set S with character θ then the number, k, of distinct G-orbits in S is*

$$k = \frac{1}{|G|} \sum_{x \in G} |\mathit{fix}(x)| = \frac{1}{|G|} \sum_{x \in G} \theta(x)$$

Proof: Let $\mathcal{M} = \{(x, s) \in G \times S \,|\, xsx^{-1} = s\}$. Given a $x \in G$, there are $\theta(x)$ such ordered pairs in S. But given any $s \in S$, there must be $|\text{stab}_G(s)|$ ordered pairs in S. Therefore,

$$|S| = \sum_{s \in S} |\text{stab}_G(s)| = \sum_{x \in G} \theta(x)$$

Observing the first sum on the left, we can evaluate the sum by letting $\mathcal{O}_1, \mathcal{O}_2, \cdots, \mathcal{O}_k$ be the distinct G-orbits in S. Choose a $s_i \in \mathcal{O}_i$ for some $1 \leq i \leq k$. Then by applying the Orbit-Stabilizer Theorem, we have

$$\sum_{s \in S} |\text{stab}_G(s)| = \sum_{i=1}^{k} |\mathcal{O}_i|\,|\text{stab}_G(s_i)| = k|G|$$

But then we have $k|G| = \sum_{x \in G} \theta(x)$. □

Theorem 3.2 is also known as the Burnside's Orbit Formula or simply the Orbit Formula. Although, this formula was proven earlier by Frobenius and appeared even earlier in the works of Cauchy. But what do fixes have to do with permutations? We already know that group actions are exactly parallel to the idea that there is a homomorphism $\varphi : G \to S_S$, i.e. a group into a permutation of a set S. Then $\text{fix}(x)$ is simply the set of elements that are left unchanged by the action of x and this is all we need focus on. Thinking back to the

cube in Example 1, the set S was the set of vertices, which told use the cube configuration. Then G was acting on this set of vertices. We exemplify this and its counting uses in the following two examples.

Example 3.3. Consider making a necklace out of a 6 beads using blue and red beads. Suppose for each necklace we are going to use 3 blue and 3 red beads. Basic combinatorics tells us there are $\binom{6}{3} = 20$ possible necklaces. However, as in our example in Figure 2, simply rotating some of these necklaces on ones neck yields one of the other necklace arrangements. Hence, a retailer would not be happy paying for a box of 20 "unique" necklaces when they are really getting less. More combinatorics could be applied to find the number of unique necklaces. However, we can do it here simply using Burnside's Theorem. Arrange the 6 beads into a hexagon and consider the rigid motions of the hexagon. Let G be the group of rigid motions, then $G = \{\text{rot}_{0^\circ}, \text{rot}_{60^\circ}, \text{rot}_{120^\circ}, \text{rot}_{180^\circ}, \text{rot}_{240^\circ}, \text{rot}_{300^\circ}\}$. We simple need to count how many of the 20 possible necklaces each of these elements fix. That is, if S is the set of vertices of the hexagon and G is the set of rigid motions of the hexagon, then we can let G act on S. Two designs are the same if they are in the same G-orbit.

$g \in G$	Number of Designs Fixed
rot_{0°	20
rot_{60°	0
rot_{120°	2
rot_{180°	0
rot_{240°	2
rot_{300°	0

Since $|G| = 6$, the total number of possible unique necklaces is

$$\frac{1}{6}(20 + 0 + 2 + 0 + 2 + 0) = \frac{1}{6}(24) = 4$$

We can see the possible unique necklaces in Figure 3

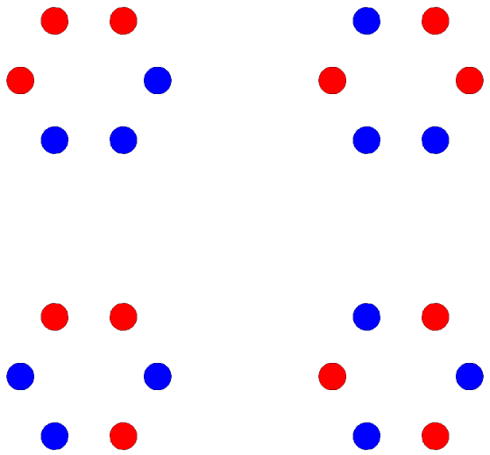


FIGURE 3. The four possible necklaces using 3 red and 3 blue beads. All other 16 possible necklaces are rigid motions of one of these 4 necklaces.

Example 3.4. Suppose we wish to color the edges of a regular tetrahedron using only the colors black, red, and blue. Since there are 6 edges and 3 colors, simply counting gives a total number of $3^6 = 729$ possible colorings. However, obviously some of these colorings are equivalent through a simple rigid motion. We wish to count the unique ways of coloring the tetrahedron. As before, let $S = \{1, 2, 3, 4\}$ be the set of vertices of the tetrahedron and G be the group of rigid motions of the tetrahedron. We let G act on S and two colorings are equivalent only if they are in the same G-orbit. One can show that the group of rigid motions of the tetrahedron is A_4 and we know that $|A_4| = 12$. Moreover, one can show that $A_4 = \langle (234), (12)(34) \rangle$ and that there is one element of order 1 (the identity), 3 elements of order 2 (the disjoint product of 2-cycles), and 8 elements of order 3 (the 3 cycles).

We take these types of elements one-by-one. The identity must trivially fix all possible colorings of the tetrahedron, so $|\text{fix}(e)| = 729$. It is sufficient to consider one of the generators since we have one of each of the two remaining element types. Consider (234) in A_4. Then (234) must take edge 12 to edge 13, edge 13 to edge 14, and edge 14 to edge 12. This means these edges are the same color. Moreover, (234) takes edge 23 to edge 34, edge 24 to edge 32, and edge 34 to edge 42 so they must have the same coloring. For each of these sets of colorings, we have 3 possible colors giving a total of $3^2 = 9$ possible colorings of the tetrahedron fixed by (234). Therefore, $|\text{fix}(\,(234)\,)| = 9$. Finally, consider (12)(34). Edge 12 and edge 34 are fixed by (12)(34) and each can be colored in any of the three ways giving $3(3) = 9$ colorings for these edges. Moreover, edge 14 and edge 23 are exchanged using the rotation (12)(34) and are therefore have the same color (for 3 possible colorings). Finally, edge 13 and edge 24 are exchanged using the rotation (12)(34) so they must have

the same color (with 3 possible choices). There are $3(3) = 9$ possible way to combine the colors of edges 14,24 and edges 13,24. Then the total number of ways to combine this with the colorings of edges 12 and 34 is $9(9) = 81$. Therefore, $|\text{fix}(\,(12)(34)\,)| = 81$. Now using Burnside's Theorem, the number of possible colorings, N, is

$$\begin{aligned} N =& \frac{1}{|G|}\sum_{a\in A_4}|\text{fix}(a)| \\ =& \frac{1}{12}\Big(729 + 3(81) + 8(9)\Big) \\ =& 87 \end{aligned}$$

Burnside's Theorem also tells a bit more about the permutations induced by the action of G under special circumstances.

Corollary 3.1. *If G is transitive on a nonempty set S then G has an element x having no fixed points.*

Proof: The number of distinct G-orbits is $k = \frac{1}{|G|}\sum_{x\in G}\theta(x)$. But since G is transitive, $1 = \frac{1}{|G|}\sum_{x\in G}\theta(x)$. But this is equivalent to saying that the average number of fixed points for an element in g is one. But then not all elements can have more than one fixed point. Since the identity in G has more than one fixed point. Therefore, there must be an element with less than one fixed point, i.e. zero fixed points. □

4. Galois Theory

Recall from high school Algebra that the quadratic formula solves for the roots of the polynomial. However, these roots are not always in the real numbers. But the roots are always in the complex numbers. Recall also that a first degree polynomial always has a root in the reals that is easily obtained. Do such formulas exist for all polynomials? This is the general question we seek to answer. Though we may think this is an Analysis question as we are working with continuous and differentiable functions. Indeed, introductory Calculus almost programs you to think of max, mins, slopes, concavity, and zeros when you hear the word polynomial. However, this is not the only approach. Instead, we look at polynomials as rings and investigate their properties. Indeed, the reader should already have some questions what it means that some polynomials do/do not have roots in the reals but always have roots in the complex plane. In answering this question we look at Galois Theory, which links together into a coherent framework the theory of fields and that of groups into a powerful computational tool. However, we need to develop the necessary theory before discussing a theorem such as the Fundamental Theorem of Galois Theory. First, we shall look at the simplest rings and ways to partition them into useful ways. Our initial approach will closely follow that of Gallian's *Contemporary Abstract Algebra* for its brevity and simplicity.[2]

4.1. **Factor Rings and Ideals.**

Normal groups and their corresponding factor/quotient groups are a powerful tool of Group Theory. Furthermore, they are used in the proof of Cauchy's Theorem and in Algebraic Topology in mappings of the fundamental group. It is then desirable to construct an analogous concept for rings. The question at hand is given an arbitrary ring R, is it possible to partition the ring in such a way so that the partitions are rings themselves when inheriting the operations from R. The partitions we seek are called ideals and are treated similar to quotient groups.

Definition 4.1. (Ideals)
A subring S of a ring R is called a (two-sided) ideal of R if for every $r \in R$ and every $s \in S$, $sr, rs \in S$. An ideal which is a proper subring of R is a proper ideal, otherwise it is called an improper ideal. This ideal is called a two-sided ideal. Right and left ideals are defined similarly.

A subring S of a ring R is an ideal if it is somehow "fixed" when acted on by elements of R. This is not to say that operation under elements of R is equivalent to an identity mapping but rather that R keeps S the same and merely permutes the elements of S amongst themselves. In fact, every ring comes equipped with two ideals: $\{0\}$ (the trivial ideal) and the ring R itself, also trivial.

Example 4.1. The set $n\mathbb{Z}$ is an ideal of $\mathbb{Z}$ for any integer n.

Example 4.2. $\langle x \rangle$ is an ideal of $\mathbb{R}[x]$.

Example 4.3. The set of all $n \times n$ matrices with bottom row zero is a right ideal in M_n, the set of all $n \times n$ matrices. Similarly, the set of all $n \times n$ matches with last column zero is a left ideal of M_n.

Example 4.4. The ring of all continuous functions from $\mathbb{R}$ to $\mathbb{R}$, $C(\mathbb{R})$, where the operation is pointwise multiplication, then $C_{f(1)=0}$, or all continuous functions f such that $f(1) = 0$, is an ideal.

However, one needs a way of telling if a subset of R is an ideal. Luckily, there is a simple way of telling whether or not a given subring of R is an ideal.

Theorem 4.1. *(Ideal Test)*
A nonempty subset S of a ring R is an ideal of R if the following hold
1. $s, t \in S$ then $s - t \in S$.
2. $s \in S$ and $r \in R$ then $rs, sr \in S$.

Condition 1 says that the subset is a subring and condition 2 is what forces it to be an ideal. Ideals are then used to construct our factor rings. As with groups, given an ideal I of R, the factor ring $r + I$ is a ring under the operations inherited from R. We first show given any ring R, we can always construct a factor ring, as every ring always has the trivial ideals.

Theorem 4.2. *(Existence of Factor Rings)*
Let R be a ring and S be a subring of R. The set of cosets $\{r + S \,|\, r \in R\}$ is a ring under the operations

$$(s + S) + (t + S) = (s + t) + S$$
$$(s + S)(t + S) = st + S$$

if and only if S is an ideal of R.

The proof is simple and left to other texts or the interested reader. One only need check that the operations only maintain closure when the quotient factor is an ideal (many of the operations come trivially and the proof really reduces down to showing that the ring is well defined only if we are working with an ideal).

Example 4.5. The subset $\{0, 3\}$ in $\mathbb{Z}_6$ is an ideal. Moreover, the factor ring $\mathbb{Z}_6/\{0, 3\}$ has three elements: $0 + \{0, 3\}, 1 + \{0, 3\}, 2 + \{0, 3\}$ and $\mathbb{Z}_6/\{0, 3\} \cong \mathbb{Z}_3$

Example 4.6. Recall from example 4.1 that $n\mathbb{Z}$ is an ideal of $\mathbb{Z}$. The factor ring $\mathbb{Z}/n\mathbb{Z} = \{0 + n\mathbb{Z}, 1 + n\mathbb{Z}, \cdots, (n - 1) + n\mathbb{Z}\}$ is equivalent to mod n arithmetic and therefore is isomorphic to $\mathbb{Z}_n$.

Looking at the examples above, we notice that many of the factor rings are integral domains or fields. This fact is dependent on the ideal used to construct the factor ring. But then that begs the question, "what conditions on the ideal force the factor ring to be an integral domain or a field?". To answer this, special types of ideals need to be considered.

Definition 4.2. (Principal Ideal)
Let R be a commutative ring with unity and let $a \in R$. The set $\langle a \rangle = \{ra \,|\, r \in R\}$ is an ideal of R called the (left) principal ideal generated by a, the right and two-sided principal ideal are defined analogously. That is, a principal ideal is an ideal generated by a single element.

Notice that a principal ideal is an ideal generated by a single element $a \in R$. Moreover, notice in a commutative ring there is no need to specify which type of principal ideal domain is meant as they are all equivalent. Principal ideals are essential in the study of rings, especially polynomial rings. Eventually they are generalized to very important types of rings called Noetherian rings (said simply, a ring where every ideal is finitely generated).

Remark 4.1. Notice all principal ideals are ideals but not all ideals are principal. Consider the ring $\mathbb{C}[x, y]$, polynomials in two variables with complex coefficients. The ideal formed by $\langle x, y\rangle$, or all polynomials in x, y with constant term 0, is not a principal ideal. Why? If p is a generator for the ideal $\langle x, y\rangle$, then $p \mid x$ and $p \mid y$. But this can't happen unless p is a nonzero constant, the only one of which in $\langle x, y\rangle$ is 0, a contradiction.

However, there are more types of useful ideals.

Definition 4.3. (Prime Ideal)
A prime ideal I of a commutative ring R is a proper ideal of R such that $a, b \in R$ and $ab \in I$ imply that $a \in I$ or $b \in I$.

Example 4.7. The ideal generated by $y^2 - x^3 - x - 1$ is a prime ideal in $\mathbb{C}[x, y]$ (in fact, this is an important fact in the theory of elliptic curves). It is a good exercise to show that this is an ideal.

Example 4.8. $p\mathbb{Z}$, where p is prime, is a prime ideal of $\mathbb{Z}$.

Example 4.9. The ideal $\langle x^3\rangle$ is not prime in $\mathbb{Z}_9[x]$ as it the ideal contains $(x + 3)^3 = x^3 + 9x^2 + 27x + 27 = x^3$ but does not contain $x + 3$.

Notice that prime ideals are the ring analogue of the fact that if $p \mid (ab)$ then $p \mid a$ or $p \mid b$, where p is prime and $a, b \in \mathbb{Z}$.

Definition 4.4. (Maximal Ideal)
A maximal ideal of a commutative ring R is a proper ideal of R such that, whenever I is an ideal of R and $I' \subseteq I \subset R$, then $I' = I$ or $I = R$.

Example 4.10. All principal ideals $\langle p\rangle$, where p is prime, are maximal ideals in the ring $\mathbb{Z}$.

Example 4.11. The ideal $4\mathbb{Z}$ is a maximal ideal in $2\mathbb{Z}$ but is not a maximal ideal in the ring $\mathbb{Z}$.

The definition of a maximal ideal of course implies that the only ideal which can properly contain a maximal idea is the ring R itself. Maximal ideals are what eventually we will call simple rings. They have very important roles. However, one must wonder whether every ring must have a maximal ideal. In fact, every ring does not have to have a maximal ideal. However, using an argument with Zorn's Lemma, we can show that every ring with identity has a maximal ideal. Without this condition a ring need not have a maximal ideal. Moreover, commutative rings with identity have unique maximal ideals, as is easy to show. It is too complicated for our scope here to give an example of a ring without maximal ideal. In any case, we are now equipped to answer the question of when a factor ring is an integral domain or a field.

Theorem 4.3. *(R/I is an I.D. iff I Prime)*
Let R be a commutative ring with unity and let I be an ideal of R. Then R/I is an integral domain if and only if I is prime.

Proof:
$\Rightarrow$: Suppose that R/I is an integral domain with $ab \in I$. Then we have

$$(a+I)(b+I) = ab + I = I$$

So either $a + I = I$ or $b + I = I$, i.e. $a \in I$ or $b \in I$. Therefore, I is a prime ideal.

$\Leftarrow$: First, note that R/I is a commutative ring with unity for any proper ideal I. Suppose that I is prime and

$$(a+I)(b+I) = ab + I = 0 + I = I$$

Then $ab \in I$ and then $a \in I$ or $b \in I$. So either $a + I$ or $b + I$ is the zero factor in R/I. $\square$

This then implies that if I is a maximal ideal in a commutative ring R then R/I is also an integral domain since I is necessarily prime. Moreover, the other direction is true also giving us the following theorem.

Theorem 4.4. *(R/I Field iff I Maximal)*
Let R be a commutative ring with unity and let I be an ideal of R. Then R/I is a field if and only if I is maximal (has no proper ideals).

Proof:
$\Rightarrow$: Suppose that R/I is a field and that A is an ideal of R that properly contains I. Let $a \in A$ but $a \notin I$. Then $a + I$ is a nonzero element of R/I. There is then an element $b + I$ such that

$$(a+I)(b+I) = 1 + I$$

Since $a \in A$, $ab \in I$. Now because

$$1 + I = (a+I)(b+I) = ab + I$$

we have $1 - ab \in I \subset A$. Notice that $(1 - an) + ab = 1 \in A$ and then $A = R$.

$\Leftarrow$: Suppose that I is maximal and $a \in R$ and $a \notin I$. We need to show that $a + I$ has a multiplicative inverse (the rest of the field properties then follow trivially). Take $A = \{ar + i \,|\, r \in R, i \in I\}$. A is then an ideal of R that properly contains I. But since I is maximal, we must have $A = R$. Therefore, $1 \in A$. Suppose that $1 = ab + i'$, where $i' \in I$. Then we must have

$$1 + I = ab + i' + I = ab + I = (a+I)(b+I)$$

$\square$

We then get the next result trivially.

Corollary 4.1. *(Maximal then Prime)*
Every maximal ideal in a commutative ring R with unity is necessarily a prime ideal.

Proof: If I is a maximal ideal in a commutative ring R with unity, then the factor ring R/I is a field and hence an integral domain. But then I is a prime ideal by Theorem 4.3. □

Of course, the converse is not true. Consider $\overline{x} = \{f(x) \in \mathbb{Z}[x] \,|\, f(0) = 0\}$. This is a prime ideal as if $f(x)g(x) \in \overline{x}$, then as f, g are integers, $f(0)g(0) = 0$ implies that $f(0) = 0$ or $g(0) = 0$. So one of them is in $\overline{x}$. However, this ideal is not maximal as $\overline{x} \subset \langle x, 2\rangle \subset \mathbb{Z}[x]$.

4.2. **Extension Fields.** A question in early Mathematics was how does one find the zeros of a polynomial? Although geometric constructions and arguments were originally preferred, analytic solutions were known. Indeed, even the Babylonians were aware of the quadratic formula. But not all polynomials were believed to have a zero. For example, the polynomial $x^2 + 1$ does not appear to have any zeros in $\mathbb{R}$. Though not initially given much consideration, the complex number $i = \sqrt{-1}$ is a zero for $x^2 + 1$. However, notice that $i \in \mathbb{C}$ but $i \notin \mathbb{R}$, making the following observation by Cauchy in 1847, surprising.

Remark 4.2. $\mathbb{R}[x]/\langle x^2 + 1\rangle$ contains a zero of $x^2 + 1$ Consider the principal ideal generated by $x^2 + 1$.

$$\langle x^2 + 1\rangle = \{f(x)(x^2 + 1) \,|\, f(x) \in \mathbb{R}[x]\}$$

Then we have

$$\begin{aligned}\mathbb{R}[x]/\langle x^2 + 1\rangle =& \{g(x) + \langle x^2 + 1\rangle \,|\, g(x) \in \mathbb{R}[x]\}\\ =& \{(ax + b) + \langle x^2 + 1\rangle \,|\, a, b \in \mathbb{R}\}\end{aligned}$$

But then using the division algorithm

$$\begin{aligned}g(x) + \langle x^2 + 1\rangle =& \big(q(x)(x^2 + 1) + r(x)\big) + \langle x^2 + 1\rangle\\ =& r(x) + \langle x^2 + 1\rangle\end{aligned}$$

Notice this is because $\langle x^2+1\rangle$ is the zero in this quotient ring. Moreover, we have $\mathbb{R}[x]/\langle x^2+1\rangle \cong \mathbb{C}$ (which we will show much later).

Therefore, $x^2 + 1$ has no zeros in the field $\mathbb{R}$ but does indeed have zeros in the field $\mathbb{R}[x]/\langle x^2+1\rangle \cong \mathbb{C}$. We have constructed a field with real coefficients such that x^2+1 has a zero without using complex numbers! In any case, this shows that a field must be specified when a polynomial $f(x)$ is said to have no zeros . But does every polynomial have a zero in some field? Indeed, it is true that given any polynomial in a field F that does not contain a zero for the polynomial, we can extend the field into a larger field E that contains F and the zeros of $f(x)$.

Definition 4.5. A field E is an extension field F if $F \subseteq E$ and the operations of F are those of E restricted to F.

Theorem 4.5. *Kronecker's Theorem (Fundamental Theorem of Field Theory)*
Let F be a field and let $f(x)$ be a nonconstant polynomial in $F[x]$. Then there is an extension field E of F in which $f(x)$ has a zero.

Proof: Since F is a field, $F[x]$ is a unique factorization domain. Then for any $f(x) \in F[x]$, $f(x)$ has an irreducible factor, even if it is $f(x)$ itself. Call this factor $p(x)$. Since $p(x)$ is irreducible, $\langle p(x)\rangle$ is maximal and $E = F[x]/\langle p(x)\rangle$ is a field. Let ϕ be a mapping $\phi : F \to E$ defined by

$$\phi(a) = a + \langle p(x)\rangle$$

Note that ϕ is injective and operation preserving. E then has a subfield that is isomorphic to F. Now we show that $p(x)$ has a zero in E. Write $p(x)$ as

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

But from the following, it is then so that in E, $x + \langle p(x)\rangle$ is a zero of $p(x)$.

$$\begin{aligned} p(x + \langle p(x)\rangle) =& a_n(x + \langle p(x)\rangle)^n + a_{n-1}(x + \langle p(x)\rangle)^{n-1} + \cdots + a_1(x + \langle p(x)\rangle) + a_0 \\ =& a_n(x^n + \langle p(x)\rangle) + a_{n-1}(x^{n-1} + \langle p(x)\rangle) + \cdots + a_1(x + \langle p(x)\rangle) + a_0 \\ =& a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 + \langle p(x)\rangle \\ =& p(x) + \langle p(x)\rangle \\ =& \langle p(x)\rangle \end{aligned}$$

□

Notice that extension fields are not necessarily unique. For example, $x^2 - 2 \in \mathbb{Q}[x]$ does not have a zero in $\mathbb{Q}$ but does have zeros in $\mathbb{R}, \mathbb{C}$, and $\mathbb{Q}[\sqrt{2}]$. We would like to find a field that contains the zero of a given polynomial but is not so large that it loses the structure of the original field. Moreover, it would be helpful if we could identify this field; that is, have the field be unique. This smallest extension field is called the splitting field. Moreover, it has the property we want - that it is unique up to isomorphism, as we will show.

4.3. **Splitting Fields.**

Definition 4.6. (Splitting Field)
Let E be an extension field of F and let $f(x) \in F[x]$. We say that $f(x)$ splits in E if $f(x)$ can be factored as a product of linear factors in $E[x]$. We call E a splitting field for $f(x)$ over F if $f(x)$ splits in E but in no proper subfield of E.

Similarly to how every polynomial in some polynomial ring has an extension field in which it has a zero, it also has a splitting field. Indeed, this must be the case because by Kronecker's Theorem, it must have a extension field.

Theorem 4.6. *(Existence of Splitting Fields)*
Let F be a field and let $f(x)$ be a nonconstant element of $F[x]$. Then there exists a splitting field E for $f(x)$ over F.

Proof: We will prove this by induction. If $\deg f(x) = 1$, then F itself is a splitting field for $f(x)$. Now assume that the theorem is true for all fields with degree less than $n = \deg f(x)$. By Kronecker's Theorem, there is an extension E of F for which $f(x)$ has a zero. Call this zero a, where $a \in E$. We can then factor out this zero and write $f(x) = (x - a)g(x)$, where $g(x) \in E[x]$. But then we have $\deg g(x) < \deg f(x)$, so by the induction hypothesis, there is a smallest field K that contains E and also all the zeros of $g(x)$. Suppose the zeros of $g(x)$ are $z_1, z_2, \cdots, z_n \in K \subset E$. Then the splitting field for $f(x)$ is then $F(a, z_1, \cdots, z_n)$. □

Example 4.12. Consider the polynomial $f(x) = (x^2 - 2)(x^2 + 1)$, where $f(x) \in \mathbb{Q}[x]$. Clearly, $f(x)$ has no zeros in $\mathbb{Q}[x]$ as the zeros of $f(x)$ are $\pm\sqrt{2}$ and $\pm i$. But then using the concept from Theorem 4.6, the splitting field for $f(x)$ over $\mathbb{Q}$ is

$$\mathbb{Q}(\sqrt{2}, i) = \mathbb{Q}(\sqrt{2})(i) = \{(a + b\sqrt{2}) + (c + d\sqrt{2})i \mid a, b, c, d \in \mathbb{Q}\}$$

Notice that similarly to extension fields, splitting fields are primarily constructed using factor rings. Though this should not be surprising as a splitting field is simply a smallest extension field. Now we show a useful relation between a splitting field for a irreducible polynomial and its corresponding factor ring.

Theorem 4.7. *($F(a) \cong F[x]/\langle p(x)\rangle$)*
Let F be a field and let $p(x) \in F[x]$ be irreducible over F. If a is a zero of $p(x)$ in some extension E of F, then $F(a)$ is isomorphic to $F[x]/\langle p(x)\rangle$. Furthermore, if $\deg p(x) = n$ then every member of $F(a)$ can be uniquely expressed in the form

$$c_{n-1}a^{n-1} + c_{n-2}a^{n-2} + \cdots + c_1 a + c_0$$

where $c_0, c_1, \cdots, c_{n-1} \in F$.

Proof: Consider the mapping $\phi_a : F[x] \to F(a)$, called the evaluation homomorphism, given by $\phi_a(f(x)) = f(a)$. It is clear that ϕ_a is a ring homomorphism (since $(f + g)(a) = f(a) + g(a)$ and $(f \cdot g)(a) = f(a)g(a)$); moreover, $p(x) \in \ker \phi$. Because $p(x)$ is irreducible, $\langle p(x)\rangle$ is maximal ideal. Since $p(x) \neq 1$, we have $\ker \phi_a \neq F[x]$ as $f(x) = 1 \in F[x]$, we must have $\ker \phi_a = \langle p(x)\rangle$. Specifically, if $f(x) = 1$, then $f(a) \neq 0$ and $f(x) \notin \ker \phi$, then $\ker \phi \neq F[x]$. By the First Isomorphism Theorem for rings and the fact that $F[x]/\langle p(x)\rangle$ is a field, $\phi_a(F[x])$ is a subfield of $F(a)$. Note that $F(a)$ is a splitting field and since $F \subset \phi_a(F[x])$ and $a \in \phi_a(F[x])$, we have

$$F[x]/\langle p(x)\rangle \cong \phi_a(F[x]) = F(a)$$

Finally, it can be easily shown with the division algorithm that every element in $F[x]/\langle p(x)\rangle$ can be uniquely expressed in the form

$$c_{n-1}x^{n-1} + \cdots + c_0 + \langle p(x)\rangle$$

with $c_0, c_1, \cdots, c_{n-1} \in F$. Then the natural isomorphism from $F[x]/\langle p(x)\rangle$ to $F(a)$ carries $c_k x^k + \langle p(x)\rangle$ to $c_k a^k$. □

Corollary 4.2. $(F(a) \cong F(b))$
Let F be a field and let $p(x) \in F[x]$ be irreducible over F. If a is a zero of $p(x)$ in some extension E of F and b is zero of $p(x)$ in some extension E' of F, then $F(a) \cong F(b)$.

Proof: By Theorem 4.7, $F(a) \cong F[x]/\langle p(x)\rangle \cong F(b)$ □

When working with mappings of vector spaces, it is most fruitful to observe how the mapping acts on the basis for the vector space. Knowing the behavior of the basis under the mapping completely determines the behavior of the new space formed. In looking at mappings between fields and their extensions and between extensions, it would be useful to have a basis for the extension field. The power of Theorem 4.7 is that we now can write a basis for $F[x]/\langle p(x)\rangle \cong F(a)$ by considering the obvious basis for $F(a)$ - that is, $\{1, a, a^2, \cdots, a^{n-1}\}$.

Example 4.13. Consider the irreducible polynomial $f(x) = x^6 - 2 \in \mathbb{Q}[x]$. Because $f(x)$ has a $\sqrt[6]{2}$ as a zero, by Theorem 4.7, the set $\{1, 2^{\frac{1}{6}}, 2^{\frac{2}{6}}, 2^{\frac{3}{6}}, 2^{\frac{4}{6}}, 2^{\frac{5}{6}}\}$ serves as a basis for $\mathbb{Q}(\sqrt{2})$ over $\mathbb{Q}$. Therefore,

$$\mathbb{Q}(\sqrt[6]{2}) = \{a_0 + a_1 2^{\frac{1}{6}} + a_2 2^{\frac{2}{6}} + a_3 2^{\frac{3}{6}} + a_4 2^{\frac{4}{6}} + a_5 2^{\frac{5}{6}} \mid a_i \in \mathbb{Q}\}$$

is a field isomorphic to $\mathbb{Q}[x]/]\langle x^6 - 2\rangle$.

Recall that extensions fields are not unique. That is, given a polynomial in a field, we can find many fields containing the zeros of the polynomial. In the above example, we carefully constructed a splitting field (the smallest extension field) for $x^6 - 2$. However, there is no reason we cannot find several smallest extension fields. Indeed, in the previous theorems we see that both $\mathbb{Q}[x]/\langle x^6 - 2\rangle$ and $\mathbb{Q}(\sqrt[6]{2})$ are both splitting fields for $x^6 - 2 \in \mathbb{Q}[x]$ over $\mathbb{Q}$. They may have different structures, but they are isomorphic. The previous corollary suggests that if one used a different root one produces an isomorphic splitting field. Are all splitting fields then unique up to isomorphism?

Lemma 4.1. *Let F be a field and $p(x) \in F[x]$ be irreducible over F. Let a be a zero of $p(x)$ in some extension of F. If ϕ is a field isomorphism from F to F' and b is a zero of $\phi(p(x))$ in some extension of F'. Then there is an isomorphism from $F(a)$ to $F'(b)$ that agrees with ϕ on F and carries a to b.*

$$F(a) \xrightarrow{\alpha} F[x]/\langle p(x)\rangle \xrightarrow{\overline{\phi}} F'[x]/\langle \phi(p(x))\rangle \xrightarrow{\beta} F'(b)$$

Proof: Since $p(x)$ is irreducible over F, $\phi(p(x))$ is irreducible over F'. Simple calculations show that the mapping $\overline{\phi} : F[x]/\langle p(x)\rangle \to F'[x]/\langle \phi(p(x))\rangle$ given by

$$f(x) + \langle p(x)\rangle \mapsto \phi(f(x)) + \langle \phi(p(x))\rangle$$

is a field isomorphism. By Theorem 4.7, there is an isomorphism α from $F(a)$ to $F[x]/\langle p(x)\rangle$ that is identity on F and carries a to $x + \langle p(x)\rangle$. Similarly, there is an isomorphism β from $F'[x]/\langle \phi(p(x))\rangle$ to $F'(b)$ that is the identity on F' and carries $x + \langle \phi(p(x))\rangle$ to b. Thus, the composition $\beta \circ \overline{\phi} \circ \alpha : F(a) \to F'(b)$ is the desired mapping. □

Theorem 4.8. *(Extending $\phi : F \to F'$)*
Let ϕ be an isomorphism from a field F to a field F' and $f(x) \in F[x]$. If E is a splitting field for $f(x)$ over F and E' is a splitting field for $\phi(f(x))$ over F', then there is an isomorphism from E to E' that agrees with ϕ on F.

$$\begin{array}{ccc} E & \dashrightarrow & E' \\ \cup & & \cup \\ F & \longrightarrow & F' \end{array}$$

Proof: We prove this by induction on deg $f(x)$. If $\deg f(x) = 1$, then we trivially have $E = F$ and $E' = F'$ so the mapping ϕ is all that is required. Now suppose that this is true for polynomials $f(x)$ up to degree n. Let $p(x)$ be an irreducible factor of $f(x)$ and let a be a zero of $p(x)$ in E and let b be a zero of $\phi(p(x))$ in E'. Using Lemma 4.1, there is an isomorphism α from $F(a)$ to $F'(b)$ that agrees with ϕ on F and carries a to b. Now rewrite $f(x)$ as $f(x) = (x - a)g(x)$, where $g(x) \in F(a)[x]$. Then E is a splitting field for $g(x)$ over $F(a)$ and E' is a splitting field for $\alpha(g(x))$ over $F'(b)$. Since $\deg g(x) < \deg f(x)$, by the induction hypothesis there is an isomorphism from E to E' that agrees with α on $F(a)$ and therefore with ϕ on F. □

The uniqueness of splitting fields then follows trivially:

Corollary 4.3. *(Splitting Fields are Unique)*
Let F be a field and let $f(x) \in F[x]$. Then any two splitting fields of $f(x)$ over F are isomorphic.

Proof: Suppose that E and E' are splitting fields for $f(x)$ over F. Then using Theorem 4.8 we simply allow our ϕ to be the identity map from F to F. □

Taking this theorem into account, one needn't use the language "a splitting field" for a polynomial $f(x)$ over F. Instead, one simply says "the splitting field" for $f(x)$ over F, as they must all be isomorphic by the preceding corollary. What have we accomplished thus far? We have shown that for every nonconstant polynomial, that any polynomial must split in either the field in which it sits or in some extension. Moreover, there is a smallest extension in which it splits that is unique up to isomorphism. However, we have not taken into consideration how these polynomials must split. Indeed, is this splitting unique and do these splittings have any other unique properties? Upon further investigation we can find special criterion and properties for zeros of multiplicity (multiple zeros). The first question we answer is when will an extension have multiple zeros. Is it the case that if $f(x)$ in F has a multiple zero that $f'(x)$ in E will have a multiple zero? To answer these questions, we take into account the derivative (though this is not necessary, I believe it is a cleaner argument).

Definition 4.7. (Derivative)
Let $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ be a polynomial in $F[x]$. The derivative of $f(x)$, denoted $f'(x)$, is the polynomial $f'(x) = na_n x^n + (n-1)a_{n-1}x^{n-1} + \cdots + a_1$ in $F[x]$.

Notice we have defined the derivative without using any of the notions of Calculus. Indeed, it turns out we only need look at the particular polynomial to answer the questions we have at hand. We do not need to use any Calculus but merely Ring Theory. But of course we know this to be the derivative of $f(x)$ from Calculus and thus give it the same name. This shows the intimate relation between polynomials in Calculus and polynomials when considered an element in a ring.

Theorem 4.9. *(Criterion for Multiple Zeros)*
A polynomial $f(x)$ over a field F has a multiple zero in some extension E if and only if $f(x)$ and $f'(x)$ have a common factor of positive degree in $F[x]$.

Proof: $\Rightarrow$: If a is a multiple zero of $f(x)$ in some extension E then there is a $g(x)$ in $E[x]$ such that $f(x) = (x-a)^2 g(x)$. Since $f'(x) = (x-a)^2 g'(x) + 2(x-a)g(x)$ (as one can check algebraically), then $f'(a) = 0$. Then $x - a$ is a factor of both $f(x)$ and $f'(x)$ in the extension E of F. Now if $f(x)$ and $f'(x)$ have no common divisor of positive degree in $F[x]$, there are polynomials $h(x), k(x) \in F[x]$ such that $f(x)h(x) + f'(x)k(x) = 1$. If we think of $f(x)h(x) + f'(x)k(x)$ as being an element of $E[x]$ (NOT $F[x]$), we see that $x - a$ is a factor of 1. But this cannot be so; therefore, $f(x)$ and $f'(x)$ must have a common divisor of positive degree in $F[x]$.
$\Leftarrow$: Conversely, suppose that $f(x)$ and $f'(x)$ have a common factor of positive degree. Let a be a zero of the common factor. Then a is a zero of $f(x)$ and $f'(x)$. Since a is a zero of $f(x)$, there is a polynomial $q(x)$ such that $f(x) = (x-a)q(x)$. Then $f'(x) = (x-a)q'(x) + q(x)$ and $0 = f'(a) = q(a)$. Therefore, $x - a$ is a factor of $q(x)$ and a is a multiple zero of $f(x)$. □

We then use this theorem to prove a stronger condition for the existence of multiple zeros of irreducible polynomials in a field.

Theorem 4.10. *(Zeros of Irreducibles)*
Let $f(x)$ be an irreducible polynomial over a field F. If F has characteristic 0, then $f(x)$ has no multiple zeros. If F has prime characteristic, then $f(x)$ has a multiple zero if it is of the form $f(x) = g(x^p)$ for some $g(x)$ in $F[x]$.

Proof: Let $f(x)$ be an irreducible polynomial over F. If $f(x)$ has a multiple zero, then by Theorem 4.9, $f(x)$ and $f'(x)$ have a common divisor of positive degree in $F[x]$. Since up to associates the only divisor of positive degree of $f(x)$ in $F[x]$ is itself, $f(x) \mid f'(x)$. But a polynomial over a field cannot divide a polynomial of smaller degree, then it must be the case that $f'(x) = 0$. Writing out $f'(x)$, we have

$$f'(x) = na_n x^{n-1} + (n-1)a_{n-1}x^{n-2} + \cdots + a_1$$

Then $f'(x) = 0$ only when $a_i = 0$ for $i = 1, 2, 3, \cdots, n$. Then $f'(x) = a_0$. But when char $F = 0$, but $f'(x) = a_0$ is not an irreducible polynomial. This contradicts the hypothesis that $f(x)$ is irreducible over F. Therefore, $f(x)$ has no multiple zeros.

If the character of F is prime, then we have $a_i = 0$ when p does not divide i. Thus, the only powers of x that appear in the sum $a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ are those of the form $x^{pj} = (x^p)^j$. It follows that $f(x) = g(x^p)$ for some $g(x) \in F[x]$. □

This theorem tells that an irreducible polynomial in any field with characteristic 0 cannot have multiple zeros. It would be useful if we could extend this to a larger class of fields than just those with char = 0.

Definition 4.8. (Perfect Fields)
A field F is called perfect if F has characteristic 0 or if F has characteristic p and $F^p = \{a^p \mid a \in F\} = F$.

The most famous of these families being the finite fields, which turn out to all be perfect; this will come into play later. For now, we now have yet a stronger criterion for multiple zeros.

Theorem 4.11. *(Criterion for Multiple Zeros)*
If $f(x)$ is an irreducible polynomial over a perfect field F, then $f(x)$ has no multiple zeros.

Proof: When F has characteristic 0, we have already shown that any irreducible polynomial in such a field has no multiple zeros. Now assume that $f(x) \in F[x]$ is irreducible over a perfect field F of characteristic p and that $f(x)$ has multiple zeros. By Theorem 4.10, $f(x) = g(x^p)$ for some $g(x) \in F[x]$. Let $g(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$. Since $F^p = F$, each a_i in F can be written in the form b_i^p in F for some b_i in F. However,

$$\begin{aligned} f(x) = g(x^p) =& b_n^p x^{pm} + b_{n-1}^p x^{p(n-1)} + \cdots + b_1^p x^p + b_0^p \\ =& (b_n x^n + b_{n-1}x^{n-1} + \cdots + b_1 x + b_0)^p \\ =& (h(x))^p \end{aligned}$$

where $h(x) \in F[x]$ and the middle equality follows from the fact that in a field of prime characteristic p $(x_1 + \cdots + x_n)^p = x_1^p + \cdots + x_n^p$. But then $f(x)$ is reducible. □

Furthermore, the zeros of an irreducible polynomial have a surprising property in their extension field.

Theorem 4.12. *(Zeros of Irreducibles over Splitting Fields)*
Let $f(x)$ be an irreducible polynomial over a field F and let E be the splitting field of $f(x)$ over F. Then all the zeros of $f(x)$ in E have the same multiplicity.

Proof: Suppose that a, b are distinct zeros of $f(x)$ in E. If a has multiplicity m, then in $E[x]$ we may write $f(x) = (x - a)^m g(x)$, with $g(x) \in E[x]$. By Lemma 4.1 and Theorem 4.8, there is a field isomorphism ϕ from E to itself that carries a to b and acts as the identity on F. Therefore,

$$f(x) = \phi(f(x)) = (x - b)^m \phi(g(x))$$

and the multiplicity of b is greater than or equal to the multiplicity of a. If we interchange the roles of a and b, the multiplicity of a must be greater than or equal to the multiplicity of b. Therefore, a and b must have the same multiplicity. □

But then it immediately follows how an irreducible polynomial with multiple zeros must factor in its extension field.

Theorem 4.13. *(Factorization of Irreducibles over Splitting Fields)*
Let $f(x)$ be an irreducible polynomial over a field F and let E be a splitting field of $f(x)$. Then $f(x)$ has the form

$$a(x-a_1)^n(x-a_2)^n\cdots(x-a_t)^n$$

where $a_1, a_2, \cdots, a_t$ are distinct elements of E and $a \in F$, with $nin\mathbb{Z}_+$.

4.4. **Algebraic Extensions.** We have already shown our first important result: every polynomial has a zero in some extension and hence is always factorable in some extension. Furthermore, we described this factorization in the extension field, including for polynomials with zeros of multiplicity, and given criterion for multiple zeros. However, our focus has been on the zeros and the polynomials in the field and its extension. We have not looked at other elements in the field to consider what implications this might have for our field if these element have certain properties.

Definition 4.9. (Extensions)
Let E be an extension field of a field F and let $a \in F$. We call a algebraic over F if a is the zero of some nonzero polynomial in $F[x]$. If a is not algebraic over F, it is called transcendental over F. An extension E of F is called an algebraic extension of F if every element of E is algebraic over F. If E is not an algebraic extension of F, it is called a transcendental extension of F. An extension of F of the form $F(a)$ is called a simple extension of F.

What we have done is taken each element in a field and put them into one of two possible classes: algebraic elements or transcendental elements. But why make the distinction between elements in a field?

Theorem 4.14. *(Extension Characterization)*
Let E be an extension field of the field F and let $a \in E$. If a is transcendental over F, then $F(a) \approx F(x)$. If a is algebraic over F, then $F(a) \approx F[x]/\langle p(x)\rangle$, where $p(x)$ is a polynomial in $F[x]$ of minimum degree such that $p(a) = 0$. Moreover, $p(x)$ is irreducible over F.

Proof: Consider the homomorphism $\phi_\alpha : F[x] \to F(a)$ given by $f(x) \to f(a)$. If a is transcendental over F, then $\ker\phi = \{0\}$ because a is the zero of no polynomial and $f(a) = 0$ if and only if $f(x) = 0$. So we may extend ϕ to an isomorphism $\overline{\phi} : F(x) \to F(a)$ by defining $\phi\big(\frac{f(x)}{g(x)}\big) = \frac{f(x)}{g(x)}$.

If a is algebraic over F, then $\ker\phi \neq 0$. But recalling properties of ideals, there is a polynomial $p(x)$ in $F[x]$ such that $\ker\phi = \langle p(x)\rangle$ and $p(x)$ has minimum degree among all

nonzero elements of $\ker \phi$ (as $F[x]$ is a principal ideal domain). Thus, $p(a) = 0$ and because $p(x)$ is a polynomial of minimum degree with this property, it is irreducible over F. $\square$

Moreover, the previous theorem can easily adapted to show that the monic irreducible polynomial is unique and that $p(x) \mid f(x)$.

Theorem 4.15. *(Uniqueness)*
If a is algebraic over a field F, then there is a unique monic irreducible polynomial $p(x)$ in $F[x]$ such that $p(a) = 0$.

Theorem 4.16. *(Divisibility)*
Let a algebraic over F and let $p(x)$ be the minimal polynomial for a over F. If $f(x) \in F[x]$ and $f(a) = 0$, then $p(x)$ divides $f(x)$ in $F[x]$.

The previous theorems hint that at this point a change of viewpoint would be fruitful. We now will view the field F and its extension E in a different way. We can think of E as a vector space over F, that is the elements of E are the basis vectors where the elements of F are scalars. Then we can discuss extensions using the language of basis and dimension.

Definition 4.10. (Extension Degree)
Let E be an extension field of a field F. We say that E has degree n over F and write $[E : F] = n$ if E has dimension n as a vector space over F. If $[E : F]$ is finite, E is called a finite extension of F; otherwise, we say that E is an infinite extension of F.

Example 4.14. A simple example familiar to the reader is the complex numbers, $\mathbb{C}$. The complex numbers $\mathbb{C}$ have dimension 2 over the reals with a possible basis of $\{1, i\}$.

Example 4.15. If a is algebraic over F and its minimal polynomial over F has degree n, so by Theorem 4.7, $\{1, a, a^2, \cdots, a^{n-1}\}$ is a basis for $F(a)$ over F. Therefore, $[F(a) : F] = n$. We say that a has degree n over F.

All finite extensions have sufficient criterion to be algebraic extensions.

Theorem 4.17. *(Finite then Algebraic)*
If E is a finite extension of F, then E is an algebraic extension of F.

Proof: Suppose that $[E : F] = n$ and $a \in E$. Then the set $\{1, a, a^2, \cdots, a^n\}$ is linearly dependent over F. There are elements $c_0, c_1, \cdots, c_n$ in F not all zero so that

$$c_n a^n + c_{n-1} a^{n-1} + \cdots + c_1 a + c_0 = 0$$

Clearly, a is a zero of the nonzero polynomial

$$f(x) = c_n x^n + c_{n-1} x^{n-1} + \cdots + c_1 x + c_0$$

$\square$

However, it is important to note that the converse of this theorem is not true. If the converse were to be true then the degrees of every algebraic extension E over G would be bounded. However, $\mathbb{Q}(\sqrt{2}, \sqrt[3]{2}, \sqrt[4]{2}, \cdots)$ is an algebraic extension of $\mathbb{Q}$ that contains

elements of every degree over $\mathbb{Q}$. Now we show that the analogue of Cauchy's Theorem holds for fields as well. We will use this theorem often in later results.

Theorem 4.18. *(Algebraic Orders)*
Let K be a finite extension field of the field E and let E be a finite extension field of the field F. Then K is a finite extension field of F and $[K : F] = [K : E][E : F]$.

Proof: Suppose that $X = \{x_1, x_2, \cdots, x_n\}$ be a basis for K over E and let $Y = \{y_1, y_2, \cdots, y_m\}$ be a basis for E over F. It suffices to show that

$$YX = \{y_j x_i \,|\, 1 \leq j \leq m, 1 \leq i \leq n\}$$

is a basis for K over F. Let $a \in K$, then there are elements $b_1, b_2, \cdots, b_n \in E$ such that

$$a = b_1 x_1 + b_2 x_2 + \cdots + b_n x_n$$

For each $i = 1, 2, \cdots, n$, there are elements $c_{i1}, c_{i2}, \cdots, c_{im} \in F$ such that

$$b_i = c_{i1} y_1 + c_{i2} y_2 + \cdots + c_{im} y_m$$

Therefore,

$$a = \sum_{i=1}^{n} b_i x_i = \sum_{i=1}^{n} \left(\sum_{j=1}^{m} c_{ij} y_j \right) x_i = \sum_{i,j} c_{ij} (y_j x_i)$$

so YX spans K over F. Suppose that there are elements c_{ij} in F such that

$$0 = \sum_{ij} c_{ij} (y_j x_i) = \sum_{i} \left(\sum_{j} \left(c_{ij} y_j \right) \right) x_i$$

Since each $\sum_j c_{ij} y_j \in E$ and X is a basis for K over E, then

$$\sum_{j} c_{ij} y_j = 0$$

for each i. But each $c_{ij} \in F$ and Y is a basis for E over F, so each $c_{ij} = 0$. Therefore, YX is a linearly independent set over F. □

But given that the extension E of a field F is also a field, $[E : F] = n$ if and only if E is isomorphic to F^n as vector spaces.

Example 4.16. A basis for $\mathbb{Q}(\sqrt{3}, \sqrt{5})$ has a basis $\{1, \sqrt{3}\}$ over $\mathbb{Q}(\sqrt{5})$ and $\{1, \sqrt{5}\}$ is a basis for $\mathbb{Q}(\sqrt{5})$ for $\mathbb{Q}(\sqrt{5})$ over $\mathbb{Q}$. Then $\{1, \sqrt{3}, \sqrt{5}, \sqrt{3}\sqrt{5}\}$ is a basis for $\mathbb{Q}(\sqrt{3}, \sqrt{5})$ over $\mathbb{Q}$.

Theorem 4.19. *(Primitive Element Theorem)*
If F is a field of characteristic 0 and a and b are algebraic over F, then there is an element c in $F(a, b)$ such that $F(a, b) = F(c)$.

Proof: Let $p(x)$ and $q(x)$ be minimal polynomials over F for a and b, respectively. In some extension K of F, let $a_1, a_2, \cdots, a_m$ and $b_1, b_2, \cdots, b_n$ be the distinct zeros of $p(x)$ and $q(x)$, respectively, where $a = a_1$ and $b = b_1$. Choose an element $d \in F$ such that

$$d \neq \frac{a_i - a}{b - b_j}$$

for all $i \geq 1$ and $j > 1$, In particular, $a_i \neq a + d(b - b_j)$ for $j > 1$.

Now we need show that $c = a + db$ is such that $F(a, b) = F(c)$. It is obvious that $F(c) \subseteq F(a, b)$. To show that $F(a, b) \subseteq F(c)$, we show that $b \in F(c)$, then $b, c, d \in F(c)$ and $a = c - b\,d$. Consider the polynomials $q(x)$ and $r(x) = p(c - d\,x)$ over $F(c)$. Because $q(b) = 0$ and $r(b) = p(c - d\,b) = p(a) = 0$, both $q(x)$ and $r(x)$ are divisible by the minimal polynomial $s(x)$ for b over $F(c)$. Because $s(x) \in F(c)[x]$. We prove $s(x) = x - b$, that is $b \in F(c)$. Since $s(x)$ is a common divisor of $q(x)$ and $r(x)$, the only possible zeros of $s(x)$ in K are the zeros of $q(x)$ that are also the zeros of $r(x)$. But $r(b_j) = p(c - d\,b_j) = p(a + d\,b - d\,b_j) = p(a + d(b - b_j))$ and d was chosen such that $a + d(b - b_j) \neq a_i$ for $j > 1$. If follows that b is the only zero of $s(x)$ in $K[x]$. Therefore, $s(x) = (x - b)^u$. Since $s(x)$ is irreducible and F has characteristic 0. But then by Theorem 4.12, $u = 1$. □

It should come as no surprise that extending an algebraic extension yields an algebraic extension.

Theorem 4.20. *(Algebraic is Algebraic)*
If K is an algebraic extension of E and E is an algebraic extension of F, then K is an algebraic extension of F.

Proof: Let $a \in K$. We need to show that a belongs to some finite extension of F, by Theorem 4.17. Since a is algebraic over E, we know that a is the zero of some irreducible polynomial in $E[x]$, say $p(x) = b_n x^n + \cdots + b_0$. We construct a tower of field extensions of F as follows,

$$\begin{aligned} F_0 =& F(b_0) \\ F_1 =& F_0(b_1), \cdots, F_n = F_{n-1}(b_n) \end{aligned}$$

In particular, $F_n = F(b_0, b_1, \cdots, b_n)$. So $p(x) \in F_n[x]$. Thus, $[F_n(a) : F_n] = n$. Because each b_i is algebraic over F; each $[F_{i+1} : F_i]$ is finite. Therefore,

$$[F_n(a) : F] = [F_n(a) : F_n][F_n : F_{n-1}] \cdots [F_1 : F_0][F_0 : F]$$

is finite. □

It would be beneficial to us further down the road if we could group all algebraic elements together in some way. Yes, we have already grouped elements in a field into algebraic and transcendental classes. However, do they form a concrete structure? In fact, they do. All algebraic elements in an extension E form a subfield of E with which we can work.

Corollary 4.4. *(Algebraic Subfields)*
Let E be an extension field of the field F. Then the set of all elements of E that are algebraic over F is a subfield of E.

Proof: Suppose that $a, b \in E$ are algebraic over F and $b \neq 0$. We need to show that we have a subfield, we need show that $a+b, a-b, ab$, and a/b are all algebraic over F. However, it suffices to show that $[F(a,b) : F]$ is finite (since each of these four elements belong to $F(a,b)$). However, note that

$$[F(a,b) : F] = [F(a,b) : F(b)][F(b) : F]$$

Because a is algebraic over F, it is algebraic over $F(b)$. Therefore, both $[F(a,b) : F(b)]$ and $[F(b) : F]$ are finite. □

If E is an extension of F then the subfield over all elements that are algebraic over F is called the algebraic closure of F in E. We considered extension fields and found the smallest possible extension field possible, i.e. the splitting field. Similarly, when we worked with ideals, we know the smallest possible ideal is the trivial ideal and we also discussed the largest proper ideal, the maximal ideal. It is clear that we can find a smallest algebraic extension. However, is there a largest-smallest algebraic extension? That is, an algebraic extension E of F that has no proper algebraic extension is an extension which is its own closure (similar to the idea of a closed set in topology). For such a field to exist, it is necessary that every polynomial in $E[x]$ splits in E. Otherwise, by the Fundamental Theorem of Field Theory, E would have a proper algebraic extension. If every member of $E[x]$ happen to split in E and K were an algebraic extension of E, then every member of K is a zero of some element in $E[x]$. But all the elements in $E[x]$ have their zeros in E. A field for which there is no algebraic extension is called algebraically closed.

It was proved in 1910, by Ernst Steinitz, that every field F has, up to isomorphism, a unique algebraic extension that is algebraically closed (to show that every field has an algebraic extension that is closed one needs to use Zorn's lemma which is equivalent to the Axiom of Choice!). In fact, $\mathbb{C}$ is algebraically closed. This was shown by Gauss and is known as the Fundamental Theorem of Algebra (that is every polynomial in $\mathbb{R}[x]$ has a root in $\mathbb{C}[x]$, see the Appendix for two different proofs of this). If one is familiar with the quaternions, $\mathbb{H}$, one may misbelieve that $\mathbb{H}$ could be an algebraic extension of $\mathbb{C}$ as it properly contains $\mathbb{C}$. However, $\mathbb{H}$ is not a field but rather a skew-field. If any polynomial $h(x)$ in $\mathbb{H}[x]$ has a zero, then $h(x)$ must have infinitely many. For example, if i, j, k are zeros of a polynomial $h(x) \in \mathbb{H}[x]$, then every scalar multiple of i, j, k is a zero for $h(x)$. In fact, $\mathbb{H}$ is the smallest and up to isomorphism unique object which properly contains $\mathbb{C}$ and is not itself a field but a skew field.

4.5. **Algebraic and Simple Extensions.** We will now discuss further properties of algebraic fields and simple extensions. For the remainder of the discussion of groups and fields, we will turn to Fraleigh's text *An Introduction to Abstract Algebra* for it's precise language and its readability.[3] First, we will introduce some notation and remind the reader of previous results.

Theorem 4.21. *(Unique Irreducible Polynomial)*
Let E be an extension field of F and let $\alpha \in F$, where $\alpha \neq 0$ and α is algebraic over F. Then there is an irreducible polynomial $p(x) \in F[x]$ such that $p(\alpha) = 0$. This irreducible polynomial $p(x)$ is uniquely determined up to a constant factor in F and is a polynomial of minimal degree greater than 1 in $F[x]$ having α as a zero. If $f(\alpha) = 0$ for $f(x) \in F[x]$, with $f(x) \neq 0$, then $p(x) \mid f(x)$.

Proof: Let ϕ_α be the evaluation homomorphism $\phi_\alpha : F[x] \to E$ defined by

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

The $\ker \phi_\alpha$ is an ideal and because F is a field, $F[x]$ is a principal ideal domain so $\ker \phi_\alpha$ is a principal ideal generated by some element $p(x) \in F[x]$. Clearly, $\langle p(x) \rangle$ consists precisely of those elements of $F[x]$ having α as a zero. Thus, if $p(x)$ is a polynomial of minimal degree greater than 1 having α as a zero and any other such polynomial of the same degree as $p(x)$ must be of the form $a \cdot p(x)$ for some $a \in F$.

It only remains to show that $p(x)$ is irreducible. If $p(x) = r(x)s(x)$ were a possible factorization of $p(x)$ into polynomials of lower degree, then $p(\alpha) = 0$ would imply that $r(\alpha)s(\alpha) = 0$, since E is a field, either $r(\alpha) = 0$ or $s(\alpha) = 0$. This contradicts the fact that $p(x)$ is of minimal degree greater than 1 such that $p(\alpha) = 0$. Hence, $p(x)$ is irreducible. □

So after multiplying by a constant in F, we can force the coefficient of the higher power of x in $p(x)$ to be 1. This is called a monic polynomial (though for a monic polynomial we do not require anything about reducibility).

Definition 4.11. (Irreducible Polynomials Continued)
Let E be an extension field of a field F and let $\alpha \in F$ be algebraic over F. The unique monic polynomial $p(x)$ of the previous theorem is the irreducible polynomial for α over F and is denoted $\text{irr}(\alpha, F)$. The degree of $\text{irr}(\alpha, F)$ is the degree of α over F, denoted $\deg(\alpha, F)$.

Example 4.17. Clearly, $\text{irr}(\sqrt{2}, \mathbb{Q}) = x^2 - 2$. However, $\alpha = \sqrt{1 + \sqrt{3}} \in \mathbb{R}$ is a zero for $x^4 - 2x^2 - 2 \in \mathbb{Q}[x]$. By the Eisenstein criterion, $x^4 - 2x^2 - 2$ is irreducible over $\mathbb{Q}$. So $\text{irr}(\sqrt{1 + \sqrt{3}}, \mathbb{Q}) = x^4 - 2x^2 - 2$. Therefore, $\sqrt{1 + \sqrt{3}}$ is algebraic element of degree 4 over $\mathbb{Q}$.

Remark 4.3. We never say that α is algebraic but rather that α is algebraic over F. Similarly, we never say the degree of α but rather the degree of α over F. For example, $\sqrt{2} \in \mathbb{R}$ is algebraic of degree 1 over $\mathbb{R}$ but is algebraic of degree 2 in $\mathbb{Q}$.

Then consider a special type of extension field E.

Definition 4.12. (Simple Extension)
An extension field E of a field F is a simple extension of F if $E = F(\alpha)$ for some $\alpha \in E$.

Specifically, we want to consider the case where α is algebraic over F.

Theorem 4.22. *(Algebraic Forms)*
Let E be a simple extension $F(\alpha)$ of a field F and let α be algebraic over F. Let the degree of $irr(\alpha, F)$ be $n \geq 1$. Then every element β of $E = F(\alpha)$ can be uniquely expressed in the form

$$\beta = b_0 + b_1\alpha + \cdots + b_{n-1}\alpha^{n-1}$$

where the b_i are in F.

Proof: This follows from Theorem 4.7. □

Example 4.18. The polynomial $p(x) = x^2 + x + 1 \in \mathbb{Z}_2[x]$ is irreducible over $\mathbb{Z}_2$ since neither 0 or 1 are zeros of $p(x)$. We know by the Fundamental Theorem for Fields that there is an extension field E of $\mathbb{Z}_2[x]$ containing a zero α of $x^2 + x + 1$. $\mathbb{Z}_2(\alpha)$ has elements $0 + 0\alpha = 0, 1 + 0\alpha = 1, 0 + 1\alpha = \alpha$, and $1 + 1\alpha = 1 + \alpha$. This is a new finite field of four elements.

Example 4.19. Recall earlier that we had stated $\mathbb{R}[x]/\langle x^2 + 1\rangle$ is isomorphic to $\mathbb{C}$. We are finally prepared to show this. Let $\alpha = x + \langle x^2 + 1\rangle$. Then $\mathbb{R}(\alpha) = \mathbb{R}[x]/\langle x^2 + 1\rangle$ and consists of all elements of the form $a + b\alpha$ for $a, b \in \mathbb{R}$. But since $\alpha^2 + 1 = 0$, we can clearly see that α plays the role of $i \in \mathbb{C}$ and that $a + b\alpha$ plays the role of $x + yi \in \mathbb{C}$. Therefore, $\mathbb{R}(\alpha) \cong \mathbb{C}$. So we have found an elegant algebraic way to construct $\mathbb{C}$ from $\mathbb{R}$.

We remind the reader of the following results:

Theorem 4.23. *Let E be an extension field of F. Then*

$$\overline{F}_E = \{\alpha \in E \,|\, \alpha \text{ is algebraic over } F\}$$

is a subfield of E, the algebraic closure of F in E.

which have already shown this in Theorem 4.4. However, we did not state the following corollary:

Corollary 4.5. *The set of all algebraic numbers forms a field.*

Moreover, we have previously discussed what it means for a field to be algebraically closed, that every polynomial in $F[x]$ has a zero in F. But then this means we can express every polynomial in $F[x]$ as a product of linear factors! Moreover, this gives us a necessary criterion for a field to be algebraically closed.

Theorem 4.24. *(Polynomials are Linear Products) A field F is algebraically closed if and only if every nonconstant polynomial in $F[x]$ factors in $F[x]$ into linear factors.*

Proof: $\Rightarrow$: Let F be algebraically closed and let $f(x)$ be a nonconstant polynomial in $F[x]$. Then $f(x)$ has a zero $a \in F$. But then $f(x)$ can be written as $f(x) = (x - a)g(x)$. Then if $g(x)$ is nonconstant, it has a zero $b \in F$ and $f(x) = (x - a)(x - b)h(x)$. Since our polynomial is of finite degree, it can only have finitely many zeros. Hence, this process is finite and gives a factorization of $f(x) \in F[x]$ into linear factors.

$\Leftarrow$: Suppose that every nonconstant polynomial of $F[x]$ has a factorization into linear factors. If $ax - b$,$a, b \in \mathbb{Z}, a \neq 0$, is a linear factor of $f(x)$, then $b/a \in F$ is a zero of $f(x)$. Thus, F is algebraically closed. □

Corollary 4.6. *An algebraically closed field F has no proper algebraic extensions.*

Proof: Let E be an algebraic extension of F, so F is a subfield of E. Then if $\alpha \in E$, $\text{irr}(\alpha, F) = x - \alpha$. Then by Theorem 4.24, since F is algebraically closed. Therefore, $\alpha \in F$ and we have $F = E$. □

4.6. **Field Automorphisms.** The reader may have already forgotten our ultimate goal, to investigate the solvability of polynomials. Of course this means we focus on the zeros of polynomials and ultimately their irreducibility. Thus far, we have shown that every irreducible has a field in which it has a zero. We then looked at the two classes of numbers in such fields: algebraic and transcendental. However, we shift our focus now to looking at the isomorphisms of an extension E to itself, that is the automorphisms of E. Our goal for the moment is to show that for an extension field E of a field F with $\alpha, \beta \in E$, that $\alpha, \beta \in E$ have the same algebraic properties if and only if $\text{irr}(\alpha, F) = \text{irr}(\beta, F)$. To do this, we show the existence of an isomorphism $\psi_{\alpha,\beta}$ of $F(\alpha)$ onto $F(\beta)$ which maps each elements which acts as the identity on F and maps α onto β, in the case where $\text{irr}(\alpha, F) = \text{irr}(\beta, F)$. This will become our main tool in studying different algebraic extensions. The function $\psi_{\alpha,\beta}$ can be thought of as the "basic isomorphism of algebraic field theory." First, we introduce some new concepts.

Definition 4.13. (Conjugate)
Let E be an algebraic extension of a field F. Two elements $\alpha, \beta \in E$ are conjugate over F if $\text{irr}(\alpha, F) = \text{irr}(\beta, F)$. That is, if α and β are zeros of the same irreducible polynomial over F.

This is precisely the idea of complex conjugation.

Example 4.20. Consider the polynomial $x^2 + 1 \in \mathbb{R}[x]$ with extension field $\mathbb{C}$. Then the elements i and $-i$ are conjugates in $\mathbb{C}$ since they are both roots of the polynomial $x^2 + 1$ in $\mathbb{R}[x]$.

Theorem 4.25. *(Basic Isomorphisms of Algebraic Field Theory)*
Let F be a field and let α and β be algebraic over F with $\deg(\alpha, F) = n$. *The map $\psi_{\alpha,\beta} : F(\alpha) \to F(\beta)$, defined by*

$$\psi_{\alpha,\beta}(c_0 + c_1\alpha + \cdots + c_{n-1}\alpha^{n-1}) = c_0 + c_1\beta + \cdots + c_{n-1}\beta^{n-1}$$

for $c_i \in F$ is an isomorphism of $F(\alpha)$ onto $F(\beta)$ if and only if α and β are conjugate over F.

Proof: $\Rightarrow$: Suppose that $\psi_{\alpha,\beta} : F(\alpha) \to F(\beta)$ is an isomorphism as stated in the theorem. Let $\text{irr}(\alpha, F) = a_0 + a_1x + \cdots + a_nx^n$. Then $a_0 + a_1\alpha + \cdots + a_n\alpha^n = 0$, so

$$0 = \psi_{\alpha,\beta}(a_0 + a_1\alpha + \cdots + a_n\alpha^n) = a_0 + a_1\beta + \cdots + a_n\beta^n$$

But this implies that $\text{irr}(\beta, F) \mid \text{irr}(\alpha, F)$. A similar argument shows that $(\psi_{\alpha,\beta})^{-1} = \psi_{\beta,\alpha}$ shows that $\text{irr}(\alpha, F) \mod \text{irr}(\beta, F)$. Therefore, since both of the polynomials are monic, $\text{irr}(\alpha, F) = \text{irr}(\beta, F)$, and so α and β are conjugate over F.

$\Leftarrow$: Suppose that $\text{irr}(\alpha, F) = \text{irr}(\beta, F) = p(x)$. Then the evaluation homomorphisms $\phi_\alpha : F[x] \to F(\alpha)$ and $\phi_\beta : F[x] \to F(\alpha)$ both have the same kernel $\langle p(x) \rangle$. There is a natural isomorphism ψ_α, corresponding to $\phi_\alpha : F[x] \to F(\alpha)$, mapping $F[x]/\langle p(x) \rangle$ onto $\phi_\alpha(F[x]) = F(\alpha)$. Similarly, ϕ_β gives rise to an isomorphism ψ_β mapping $F[x]/\langle p(x) \rangle$ onto $F(\beta)$. Let $\psi_{\alpha,\beta} = \psi_\beta((\psi_\alpha)^{-1})$. As the composition of two isomorphisms $\psi_{\alpha,\beta}$ is also an isomorphism and maps $F(\alpha)$ onto $F(\beta)$. Also, for $(c_0 + c_1\alpha + \cdots + c_{n-1}\alpha^{n-1}) \in F(\alpha)$, we have

$$\begin{aligned}
\psi_{\alpha,\beta}(c_0 + c_1\alpha + \cdots + c_{n-1}\alpha^{n-1}) =& \psi_\alpha^{-1}(c_0 + c_1\alpha + \cdots + c_{n-1}\alpha^{n-1}) \\
=& \psi_\beta\Big((c_0 + c_1\alpha + \cdots + c_{n-1}\alpha^{n-1}) + \langle p(x) \rangle \Big) \\
=& c_0 + c_1\beta + \cdots + c_{n-1}\beta^{n-1}
\end{aligned}$$

Thus, the mapping $\psi_{\alpha,\beta}$ is the map in the theorem statement. □

Corollary 4.7. *(Conjugates and the Unique Isomorphism)*
Let α be algebraic over a field F. Every isomorphism ψ mapping $F(\alpha)$ into $\overline{F}$ such that $a\psi = a$ for $a \in F$ maps α onto a conjugate β of α over F. Conversely, for each conjugate β of α over F, there exists exactly one isomorphism $\psi_{\alpha,\beta}$ of $F(\alpha)$ into $\overline{F}$ mapping α onto β and mapping each $a \in F$ onto itself.

Proof: $\Rightarrow$: Let ψ be an isomorphism mapping $F(\alpha)$ into $\overline{F}$ such that $a\psi = a$ for $a \in F$. Let $\text{irr}(\alpha, F) = a_0 + a_1x + \cdots + a_nx^n = 0$. Then

$$a_0 + a_1\alpha + \cdots + a_n\alpha^n = 0$$

Therefore,

$$0 = \psi(a_0 + a_1\alpha + \cdots + a_n\alpha^n) = a_0 + a_1\psi(\alpha) + \cdots + a_n\psi(\alpha)^n$$

and $\beta = \alpha\psi$ is a conjugate of α.

$\Leftarrow$: Suppose for each conjugate β of α over F, the isomorphism $\psi_{\alpha,\beta}$ is an isomorphism with the desired properties. That $\psi_{\alpha,\beta}$ is the only such isomorphism follows from the fact that an isomorphism of $F(\alpha)$ is completely determined by its values on the elements of F and its value on α. □

The next result one should be familiar with from high school algebra, that is complex roots occur in conjugate pairs for $f(x) \in \mathbb{R}[x]$.

Corollary 4.8. *(Conjugate Pairs)*
Let $f(x) \in \mathbb{R}[x]$. If $f(a + bi) = 0$ for $a + bi \in \mathbb{C}$, where $a, b \in \mathbb{R}$, then $f(a - bi) = 0$ also.

Proof: We have seen that $\mathbb{C} = \mathbb{R}(i)$ and that $\mathbb{C} = \mathbb{R}(-i)$. Now

$$\text{irr}(i, \mathbb{R}) = \text{irr}(-i, \mathbb{R}) = x^2 + 1$$

so that i and $-i$ are conjugate over $\mathbb{R}$. Then the map $\psi_{i,-i} : \mathbb{C} \to \mathbb{C}$ given by $\psi_{i,-i}(a+bi) = a - bi$ is an isomorphism. Thus, if for $a_i \in \mathbb{R}$

$$f(a + bi) = a_0 + a_1(a + bi) + \cdots + a_n(a + bi)^n = 0$$

then

$$\begin{aligned} 0 = \psi_{i,-i}\big(\psi_{i,-i}(f(a + bi))\big) =& a_0 + a_1(a - bi) + \cdots + a_n(a - bi)^n \\ =& f(a - bi) \end{aligned}$$

that is, $f(a - bi) = 0$ also. □

Notice from Corollary 4.8, an isomorphism from a field to itself need not be trivial. The set of all mappings of a field to itself forms a group called the automorphism group.

Definition 4.14. (Field Automorphisms)
An isomorphism of a field onto itself is an automorphism of the field.

Next, we consider this automorphism as a simple permutation of the field.

Definition 4.15. (Automorphism as Permutations)
If σ is an isomorphism of a field E into some field, then an element a of E is left fixed by σ if $a\sigma = a$. A collection S of isomorphisms of E leaves a subfield F of E fixed if each $a \in F$ is left fixed by every $\sigma \in S$. If $\{\sigma\}$ leaves F fixed, then σ leaves F fixed.

Notice that we have started to make a shift from a focus of fields to a discussion about groups. We shall show that the set of field automorphisms form a group.

Theorem 4.26. *(Automorphism form a Group)*
The set of all automorphisms of a field E is a group under function composition.

Proof: Multiplication of automorphisms of E is defined by function composition and is thus associative (as it is permutation multiplication). The identity permutation $i : E \to E$ given by $i(\alpha) = \alpha$ for all $\alpha \in E$ is obviously an automorphism of E. If σ is an automorphism, then the permutation σ^{-1} is also obviously an automorphism. Thus, all automorphisms of E form a subgroup of S_E, the group of all permutations of E. □

The idea of Galois theory is to relate the properties of groups and fields. The following theorems describes essentially all properties of fixed elements of E.

Theorem 4.27. *Let $\{\sigma_i \mid i \in I\}$ be a collection of automorphisms of a field E. Then the set $E_{\{\sigma_i\}}$ of all $a \in E$ left fixed by every σ for $i \in I$ forms a subfield of E.*

Proof: If $\sigma(a) = a$ and $\sigma_i(b) = b$ for all $i \in I$, then

$$\sigma_i(a \pm b_i) = \sigma_i(a) + \sigma_i(b) = a \pm b$$

and

$$\sigma_i(ab) = \sigma_i(a)\sigma_i(b) = ab$$

for all $i \in I$. Also, if $b \neq 0$, then

$$\sigma_i(a/b) = \sigma_i(a)/\sigma_i(b) = a/b$$

for all $i \in I$. Since the σ_i are automorphisms, we have

$$\sigma_i(0) = 0 \qquad \sigma_i(1) = 1$$

for all $i \in I$. Hence, $0, 1 \in E_{\{\sigma_i\}}$. Thus, $E_{\{\sigma_i\}}$ is a subfield of E. □

Definition 4.16. (Fixed Field)
The field $E_{\{\sigma_i\}}$ of Theorem 4.27 is the fixed field of $\{\sigma \mid i \in I\}$. For a single automorphism σ, we shall refer to $E_{\{\sigma\}}$ as the fixed field of σ.

Example 4.21. Consider the automorphism $\psi_{\sqrt{2},-\sqrt{2}}$ of $\mathbb{Q}(\sqrt{2})$ given by

$$\psi_{\sqrt{2},-\sqrt{2}}(a + b\sqrt{2}) = a - b\sqrt{2}$$

for $a, b \in \mathbb{Q}$. However, $a - b\sqrt{2} = a + b\sqrt{2}$ if and only if $b = 0$. Thus, the fixed field of $\psi_{\sqrt{2},-\sqrt{2}}$ is $\mathbb{Q}$.

Theorem 4.28. *(Automorphisms Fixing F)*
Let E be a field and let F be a subfield of E. Then the set of all $G(E/F)$ of all automorphisms of E leaving F fixed forms a subgroup of the group of all automorphisms of E. Furthermore, $F \leq E_{G(E/F)}$

Proof: For $\sigma, \tau \in G(E/F)$ and $a \in F$, we have

$$(\tau\sigma)(a) = \tau(\sigma(a)) = \tau(a) = a$$

so $\sigma\tau \in G(E/F)$. Clearly, the identity automorphism i is in $G(E/F)$. Also, if $\sigma(a) = a$ for $a \in F$, then $a = \sigma^{-1}(a)$, so $\sigma \in G(E/F)$ implies that $\sigma^{-1} \in G(E/F)$. Thus, $G(E/F)$ is a subgroup of the group of all automorphisms of E. □

Definition 4.17. (Fixed Automorphisms)
The group $G(E/F)$ of Theorem 4.28 is the group of automorphisms of E leaving F fixed, or the group of E over F.

It is important to note that the notation E/F does not refer to a quotient space. Rather, it reminds one that E is the extension of a field F.

Example 4.22. Consider the field $\mathbb{Q}(\sqrt{2}, \sqrt{3})$. Viewing $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ as $(\mathbb{Q}(\sqrt{3}))(\sqrt{2})$, the basic isomorphism $\psi_{\sqrt{2},-\sqrt{2}}$ defined by

$$\psi_{\sqrt{2},-\sqrt{2}}(a + b\sqrt{2}) = a - b\sqrt{2}$$

for $a, b \in \mathbb{Q}(\sqrt{3})$ is an automorphism of $\mathbb{Q}(\sqrt{2}, \sqrt{3})$, having $\mathbb{Q}(\sqrt{3})$ as a fixed field. Similarly, we can repeat the same process with $\psi_{\sqrt{3},-\sqrt{3}}$ as an automorphism of $\mathbb{Q}(\sqrt{2}, \sqrt{3})$

fixing $\mathbb{Q}(\sqrt{3})$. Since the product of automorphisms is again an automorphism, consider $\psi_{\sqrt{3},-\sqrt{3}}(\psi_{\sqrt{2},-\sqrt{2}})$. Then let

$$\begin{aligned} i =& \text{ identity automorphism} \\ \sigma_1 =& \psi_{\sqrt{2},-\sqrt{2}} \\ \sigma_2 =& \psi_{\sqrt{3},-\sqrt{3}} \\ \sigma_3 =& \psi_{\sqrt{2},-\sqrt{2}}\ \psi_{\sqrt{3},-\sqrt{3}} \end{aligned}$$

The group of all automorphisms of $\mathbb{Q}(\sqrt{2},\sqrt{3})$ must have a fixed field, which must contain $\mathbb{Q}$ since every automorphism of a field leaves 1 fixed and hence the prime subfield fixed. A basis for $\mathbb{Q}(\sqrt{2},\sqrt{3})$ over $\mathbb{Q}$ is $\{1,\sqrt{2},\sqrt{3},\sqrt{2}\sqrt{3}\}$. It is easy to see that $\mathbb{Q}$ is exactly the fixed field of $\{i,\sigma_1,\sigma_2,\sigma_3\}$. Then the group $G=\{i,\sigma_1,\sigma_2,\sigma_3\}$ is a group under automorphism under function composition. Moreover, G is isomorphic to the Klein 4-group.

Suppose that F is a finite field. It turns out, as we will later show, that the automorphism group of F is cyclic. Though typically one doesn't care about the choice of generator for a cyclic group (supposing there is more than one). However, for the automorphism group of a finite field there is a natural choice. This natural choice is the canonical (natural) generator, called the Frobenius automorphism (also known as the Frobenius substitution).

Theorem 4.29. *Let F be a finite field of characteristic p. Then the map $\sigma_p : F \to F$ defined by $a\sigma_p = a^p$ for $a \in F$ is an automorphism called the Frobenius automorphism. Morevoer, $F_{\{\sigma_p\}} \cong \mathbb{Z}_p$.*

Proof: Let $a,b \in F$. Applying the binomial theorem to $(a+b)^p$, we have

$$\begin{aligned} (a+b)^p =& a^p + (p\cdot 1)a^{p-1}b + \left(\frac{p(p-1)}{2}\cdot 1\right)a^{p-2}b^2 + \cdots + (p\cdot 1)ab^{p-1} + b^p \\ =& a^p + 0a^{p-1}b + 0a^{p-2}b^2 + \cdots + 0ab^{p-1} + b^p \\ =& a^p + b^p \end{aligned}$$

Thus, we have

$$\sigma_p(a+b) = (a+b)^p = a^p + b^p = \sigma_p(a) + \sigma_p(b)$$

and of course

$$\sigma_p(ab) = (ab)^p = a^p b^p = \sigma_p(a)\sigma_p(b)$$

so σ_p is a homomorphism. If $\sigma_p(a) = 0$, then $a^p = 0$ and $a = 0$. So the kernel of σ_p is $\{0\}$ and σ_p is an isomorphic mapping. Finally, since F is finite, σ_p is onto by basic counting principles. Thus, σ_p is an automorphism of F.

The prime field $\mathbb{Z}_p$ must be contained, up to isomorphism, in F since F is of characteristic p. For $c \in \mathbb{Z}_p$, we have $\sigma_p(c) = c^p = c$, by Fermat's Theorem. Thus the polynomial $x^p - x$ has p zeros in F, namely the elements of $\mathbb{Z}_p$. But a polynomial of degree n over a field can have at most n zeros in the field. Since the elements fixed under σ_p are precisely the zeros in F of $x^p - x$, we see that

$$\mathbb{Z}_p = F_{\{\sigma_p\}}$$

□

This is ironic in the sense that high school students tend to want to "distribute" powers, meaning they believe

$$(a+b)^n = a^n + b^n$$

and are told that this is not true. However, we see that in a field of prime characteristic, this is indeed the case!

4.7. **Splitting Fields, Separable Extensions, and Totally Inseparable Extensions.** We will now revisit splitting fields but revisit them using our new terminology and in a different light. The splitting field of a field F turns out to be a field such that every isomorphic mapping of E to $\overline{F}$ leaving $\overline{F}$ fixed is an automorphism of E, the extension field. So for every $\alpha \in E$, all conjugates of α over F must be in E as well. We can then reformulate the definition of a splitting field as follows.

Definition 4.18. (Splitting Field) Let F be a field with algebraic closure $\overline{F}$. Let $\{f_i(x) \mid i \in I\}$ be a collection of polynomials in $F[x]$. A field $E \leq \overline{F}$ is the splitting field of $\{f_i(x) \mid i \in I\}$ over F if E is the smallest subfield of $\overline{F}$ containing F and all the zeros in $\overline{F}$ of each of the $f_i(x)$ for each $i \in I$. A field $K \leq \overline{F}$ is a splitting field over F if it is the splitting field of some set of polynomials in $F[x]$.

Then we can given a criterion for being a splitting field in terms of mappings.

Theorem 4.30. *(Splitting Field Criterion)*
A field E, where $F \leq E \leq \overline{F}$, is a splitting field over F if and only if every automorphism of $\overline{F}$ leaving F fixed maps E onto itself and thus induces an automorphism of E leaving F fixed.

Proof:
$\Rightarrow$: Let E be a splitting field over F in $\overline{F}$ of $\{f_i(x) \mid i \in I\}$ and let σ be an automorphism of $\overline{F}$ leaving F fixed. Let $\{\alpha_j \mid j \in J\}$ be the collection of all zeros in $\overline{F}$ of all of the $f_i(x)$ for $i \in I$. For a fixed α_j, $F(\alpha_j)$ has elements all expressions of the form

$$g(\alpha_j) = a_0 + a_1\alpha_j + \cdots + a_{n_{j-1}}\alpha_j^{n_j-1}$$

where n_j is the degree of $\text{irr}(\alpha, F)$ and $a_k \in F$. Consider the set S of all finite sums of finite products of elements of the form $g(\alpha_j)$ for all $j \in J$. The set S is a subset of E closed under addition and multiplication and containing 0,1 and the additive inverses in each element in the set. Each element of S is in some $F(\alpha_1, \cdots, \alpha_{j_r}) \subseteq S$, S also contains the multiples of α_j for $j \in J$. By the definition of a plotting field E of $\{f_i(x) \mid i \in I\}$, $S = E$. But then $\{\alpha_j \mid j \in J\}$ generates E over F. Then the value of σ on any element of E is completely determined by the values of $\sigma(\alpha_j)$. But $\sigma(\alpha_j)$ must also be a zero of $\text{irr}(\alpha, F)$. But $\text{irr}(\alpha, F)$ divides $f_i(x)$ when $f_i(\alpha_j) = 0$, so $\sigma(\alpha_j) \in E$. Thus, σ maps E into E isomorphically. However, the same is true of the automorphism σ^{-1} of $\overline{F}$, since for $\beta \in E$,

$$\beta = \sigma(\beta\sigma^{-1})$$

Then we see that σ maps E onto E and induces an automorphism of E.

$\Leftarrow$: Suppose that every automorphism of $\overline{F}$ leaving F fixed induces an automorphism of E. Let $g(x)$ be an irreducible polynomial in $F[x]$ having a zero α in E. If β is any zero of $g(x)$ in $\overline{F}$, then there is a basic isomorphism $\psi_{\alpha,\beta}$ of $F(\alpha)$ leaving F fixed. Then $\psi_{\alpha,\beta}$ can be extended to an isomorphism τ of $\overline{F}$ into $\overline{F}$. Then

$$\tau^{-1} : \tau(\overline{F}) \to \overline{F}$$

can be extended to an isomorphism mapping $\overline{F}$ into $\overline{F}$. Since the image of τ^{-1} is already all of $\overline{F}$, τ must have been onto $\overline{F}$, so τ is an automorphism of $\overline{F}$ leaving F fixed. By assumption, τ induces an automorphism of E, so $\tau(\alpha) = \beta$ is in E. So if $g(x)$ is an irreducible polynomial in $F[x]$ having one zero in E, then all zeros of $g(x)$ in $\overline{F}$ are in E. Hence, if $\{g_k(x)\}$ is the set of all irreducible polynomials in $F[x]$ having a zero in E, then E is the splitting field of $\{g_k(x)\}$. □

Definition 4.19. (Splits)
Let E be an extension field of a field F. A polynomial $f(x) \in F[x]$ splits in E if it factors into a product of linear factors in $E[x]$.

Corollary 4.9. *If $E \leq \overline{F}$ is a splitting field over F, then every irreducible polynomial in $F[x]$ having a zero in E splits in E.*

Proof: If E is a splitting field over F in $\overline{F}$, then every automorphism of $\overline{F}$ induces an automorphism of E. The second half of Theorem 5.5 then showed precisely that E is also the splitting field over F of the set $\{g_k(x)\}$ of all irreducible polynomials in $F[x]$ having a zero in E. Thus an irreducible polynomial $f(x)$ of $F[x]$ having a zero in E has all its zeros in $\overline{F}$ in E. Therefore, its factorization into linear factors in $\overline{F}[x]$, actually taking place in $E[x]$. Therefore, $f(x)$ splits in E. □

Definition 4.20. (Index)
Let E be a finite extension of a field F. The number of isomorphism of E into $\overline{F}$ leaving F fixed is the index $\{E : F\}$ of E over F.

Corollary 4.10. *If $E \leq \overline{F}$ is a splitting field over F, then every isomorphic mapping of E into $\overline{F}$ leaving F fixed is actually an automorphism of E. In particular, if E is a splitting field of finite degree over F, then*

$$\{E : F\} = |G(E/F)|$$

Proof: Every isomorphism σ mapping E into $\overline{F}$ leaving F fixed can be extended to an automorphism τ of $\overline{F}$. If E is a splitting field over F, then by Theorem 5.5, τ restricted to E, that is σ in Theorem 5.5, is an automorphism of E. Thus for a splitting field E over F, every isomorphic mapping of E into $\overline{F}$ leaving F fixed is an automorphism of E. The equation $\{E : F\} = |G(E/F)|$ then follows immediately for a splitting field E of finite degree over F, since $\{E : F\}$ was defined as the number of different isomorphic mappings

of E into $\overline{F}$ leaving F fixed. □

Recall that splitting fields are unique. Then there is only one fixed algebraic closure $\overline{F}$ of F. When are the $[E : F]$ and $\{E : F\}$ equivalent? We show that for a simple algebraic extension $F(\alpha)$ of F, there is only one identity isomorphism from F into F for each distinct zero of $\text{irr}(\alpha, F)$; moreover, these are the only extensions of the isomorphism. Therefore, $\{F(\alpha) : F\}$ is the number of distinct zeros of $\text{irr}(\alpha, F)$.

Theorem 4.31. *Let E be an algebraic extension of a field F. Then then there exists a finite number of elements $\alpha_1, \alpha_2, \cdots, \alpha_n$ in E such that $E = F(\alpha_1, \cdots, \alpha_n)$ if and only if E is a finite-dimensional vector space over F, i.e., if and only if E is a finite extension of F.*

Proof: $\Rightarrow$: Suppose that $E = F(\alpha_1, \cdots, \alpha_n)$. Since E is an algebraic extension of F, each α_i is algebraic over F, so clearly each α_i is algebraic over every extension field of F in E. Thus, $F(\alpha_1)$ is algebraic over F and in general, $F(\alpha_1, \cdots, \alpha_j)$ is algebraic over $F(\alpha_1, \cdots, \alpha_{j-1})$ for $j = 2, \cdots, n$. By Theorem 4.20 applied to the sequence of finite extensions

$$F, F(\alpha_1), F(\alpha_1, \alpha_2), \cdots, F(\alpha_1, \cdots, \alpha_n) = E$$

then shows that E is a finite extension of F.

$\Leftarrow$: Suppose that E is a finite algebraic extension of F. If $[E : F] = 1$, then $E = F(1) = F$ and we are done. If $E \neq F$, let $\alpha_1 \in E$, where $\alpha_1 \notin F$. Then $[F(\alpha_1) : F] > 1$. If $F(\alpha_1) = E$, we are done; if not, let $\alpha_2 \in E$, where $\alpha_2 \notin F(\alpha_1)$. Continuing this process, we see that since $[E : F]$ is finite, we must arrive at α_n, such that

$$F(\alpha_1, \alpha_2, \cdots, \alpha_n) = E$$

□

Theorem 4.32. *($\{E : F\} \mid [E : F]$) If E is a finite extension of F, then $\{E : F\}$ divides $[E : F]$.*

Proof: By Theorem 4.31, if E is finite over F, then $E = F(\alpha_1, \cdots, \alpha_n)$, where $\alpha_i \in \overline{F}$. Let $\text{irr}(\alpha_i, F(\alpha_1, \cdots, \alpha_{i-1})$ have α_i as one of n_i distinct zeros which are all of a common multiplicity v_i, then because all the zeros of $f(x)$ in $\overline{F}$ have the same multiplicity and has a linear factorization in $\overline{F}[x]$,

$$[E : F] = \prod_i n_i v_i$$

and

$$\{E : F\} = \prod_i n_i$$

Therefore, $\{E : F\}$ divides $[E : F]$. □

Definition 4.21. (Separable Extension)
A finite extension E of F is a separable extension of F is $\{E : F\} = [E : F]$. An element α of $\overline{F}$ is separable over F is $F(\alpha)$ is a separable extension of F. An irreducible polynomial $f(x) \in F[x]$ is separable over F if every zero of $f(x)$ in $\overline{F}$ is separable over F.

However, we have just shown that $\{E : F\}$ is the number of distinct zeros of $\text{irr}(\alpha, F)$. Moreover, the multiplicity of α in $\text{irr}(\alpha, F)$ must be the same as the multiplicity of each conjugate of α over F. Therefore, α is separable over F if and only if $\text{irr}(\alpha, F)$ has all zeros of multiplicity 1. It then immediately follows that an irreducible polynomial $f(x) \in F[x]$ is separable over F if and only if $f(x)$ has all no repeated roots. The notion of totally inseparable extensions is derived similarly.

Definition 4.22. (Totally Inseparable Extensions)
A finite extension E of field F is a totally inseparable extension of F if $\{E : F\} = 1 < [E : F]$. An element α of $\overline{F}$ is totally inseparable over F if $F(\alpha)$ is totally inseparable over F.

As we had before, α is totally inseparable over F if and only if $\text{irr}(\alpha, F)$ has only one zero which is of multiplicity greater than 1.

Example 4.23. $\mathbb{Z}_p(y)$ is totally inseparable over $\mathbb{Z}_p(y^p)$, where y is an indeterminate.

The following two results exactly parallel the notion we had for separable extensions.

Theorem 4.33. *If K is a finite extension of E, E is a finite extension of F and $F < E < K$, then K is totally inseparable over F if and only if K is totally inseparable over E and E is totally inseparable over F.*

Proof:
$\Rightarrow$: Since $F < E < K$, we have $[K : E] > 1$ and $[E : F] > 1$. Suppose that K is totally inseparable over F. Then$\{K : F\} = 1$ and

$$\{K : F\} = \{K : E\}\{E : F\}$$

so we must have

$$\{K : E\} = 1 < [K : E] \quad \text{and} \quad \{E : F\} = 1 < [E : F]$$

Thus K is totally inseparable over E and E is totally inseparable over F.
$\Leftarrow$: If K is totally inseparable over E and E is totally inseparable over F, then

$$\{K : F\} = \{K : E\}\{E : F\} = (1)(1) = 1$$

and $[K : F] < 1$. Thus, K is totally inseparable over F. □

The previous theorem can then be extended by induction to any finite proper tower of finite extensions. The final field is totally inseparable extension of the initial one only if each field is totally inseparable extension of the preceding one.

Corollary 4.11. *If E is a finite extension of F, then E is totally inseparable over F if and only if each α in E, $\alpha \notin F$, is totally inseparable over F.*

Proof: $\Rightarrow$: Suppose that E is totally inseparable over F and let $\alpha \in E$ with $\alpha \notin F$. Then

$$F < F(\alpha) \leq E$$

If $F(\alpha) = E$, we are done. If $F < F(\alpha) < E$, then by Theorem 4.33 shows that since E is totally inseparable over F, $F(\alpha)$ is totally inseparable over F.

$\Leftarrow$: Suppose that for every $\alpha \in E$, with $\alpha \notin F$, α is totally inseparable over F. Since E is finite over F, there exist $\alpha_1, \cdots, \alpha_n$ such that

$$F < F(\alpha_1) < F(\alpha_1, \alpha_2) < \cdots < F(\alpha_1, \alpha_2, \cdots, \alpha_n) = E$$

Now since α_i is totally inseparable over F, α_i is totally inseparable over $F(\alpha_1, \cdots, \alpha_{i-1})$, for $q(x) = \text{irr}(\alpha_i, F(\alpha_1, \cdots, \alpha_{i-1}))$ divides $\text{irr}(\alpha_i, F)$ so that α_i is the only zero of $q(x)$ and is of multiplicity greater than 1. Thus, $F(\alpha_1, \cdots, \alpha_i)$ is totally inseparable over $F(\alpha_1, \cdots, \alpha_{i-1})$ and E is totally inseparable over F by Theorem 4.33, extended by induction. □

Theorem 4.34. *Let F have characteristic $p \neq 0$ and let E be a finite extension of F. Then $\alpha \in E$, $\alpha \notin F$, is totally inseparable over F if and only if there is some integer $t \geq 1$ such that $\alpha^{p^t} \in F$. Furthermore, there is a simple extension K of F, with $F \leq K \leq E$ such that K is separable over F and either $E = K$ or E is totally inseparable over K.*

Proof:

$\Rightarrow$: Let $\alpha \in E$ and $\alpha \notin F$ be totally inseparable over F. Then $\text{irr}(\alpha, F)$ has just one zero α of multiplicity greater than 1 and must be of the form

$$x^{p^t} - \alpha^{p^t}$$

Hence, $\alpha^{p^t} \in F$ for some $t \geq 1$.

$\Leftarrow$: If $\alpha^{p^t} \in F$ for some $t \geq 1$, where $\alpha \in E$ and $\alpha \notin F$, then

$$x^{p^t} - \alpha^{p^t} = (x - \alpha)^{p^t}$$

and $(x^{p^t} - \alpha^{p^t}) \in F[x]$, showing that $\text{irr}(\alpha, F)$ divides $(x - \alpha)^{p^t}$. Therefore, $\text{irr}(\alpha, F)$ has α as its only zero and this zero is of multiplicity greater than 1. So α is totally inseparable over F. □

Definition 4.23. (Separable Closure)
The unique filed K of theorem 4.34 is the separable closure of F in E.

The purpose of Theorem 4.34 gives the exact structure of totally inseparable extensions of a field with characteristic p. That is to say, we can create such an extension simply by adding the elements to obtain larger fields, that is the pth roots of unity.

4.8. **Finite Fields.** We have already made mention that all finite fields are perfect. We are ready to show that for every prime p and positive integer n, that there is only one, up to isomorphism, finite field with order p^n. This field is referred to as the Galois field of order p^n. Here, the properties of cyclic groups come into play heavily and will play the

most critical role in our proof of the insolvability of the quintic. First, we show that every finite field must have an order that is a prime power.

Theorem 4.35. *(Order of Finite Fields)*
Let E be a finite extension of degree n over a finite field F. If F has q elements, then E has q^n elements.

Proof: Let $\{\alpha_1, \cdots, \alpha_n\}$ be a basis for E as a vector space over F. Then every $\beta \in E$ can be uniquely written in the form

$$\beta = b_1\alpha_1 + \cdots + b_n\alpha_n$$

for each $b_i \in F$. Since each b_i may be any of the q elements of F, the total number of such distinct linear combinations of the α_i is q^n. □

Corollary 4.12. *If E is a finite field of characteristic p, then E contains exactly p^n elements for some positive integer n.*

Proof: Every finite field E is a finite extension of a prime field isomorphic to the field $\mathbb{Z}_p$, where p is the characteristic of E. The result then follows immediately from Theorem 4.35. □

Now let's look at the structure of a finite field under its operation.

Theorem 4.36. *A finite field E of p^n elements is the splitting field of $x^{p^n} - x$ over its prime subfield $\mathbb{Z}_p$ (up to isomorphism).*

Proof: Let E be a finite field with p^n elements, where p is the characteristic of F. The set E^* of nonzero elements of E forms a multiplicative group of order $p^n - 1$ under field multiplication. For $\alpha \in E^*$, the order of α in this group divides the order $p^n - 1$ of the group. Thus for $\alpha \in E^*$, we have $\alpha^{p^n-1} = 1$, so $\alpha^{p^n} = \alpha$. Therefore, every element in E is a zero of $x^{p^n} - x$. Since $x^{p^n} - x$ can have at most p^n roots, we see that E is the splitting field of $x^{p^n} - x$ over $\mathbb{Z}_p$. □

This theorem tells us that we can form any finite field from a prime subfield of the field, much in the way we can always take any positive integer and obtain it through a multiple of a prime number.

Definition 4.24. (Root of Unity)
An element α of a field is an nth root of unity if $\alpha^n = 1$. It is a primitive nth root of unity if $\alpha^n = 1$ and $\alpha^m \neq 1$ for $0 < m < n$.

Therefore, using the language of this definition, the nonzero elements of a finite field with characteristic p^n are the $(p^n - 1)$th roots of unity. Typically, the nth roots of unity are denoted by U_n an form a group under multiplication. Moreover, the group U_n is cyclic. However, we prove a more general result.

Theorem 4.37. *If G is a finite multiplicative subgroup of the multiplicative group $\langle F^*, \cdot \rangle$ of nonzero elements of a field F, then G is cyclic.*

Proof: Because G is a finite abelian group, G is isomorphic to a direct product $\mathbb{Z}_{m_1} \oplus \mathbb{Z}_{m_2} \oplus \cdots \mathbb{Z}_{m_n}$ of cyclic groups by the Fundamental Theorem for Abelian Groups, where m_i divides m_{i+1}. Think of each of the $\mathbb{Z}_{m_i}$ as a cyclic group of order m_i in multiplicative notation. Then for $a_i \in \mathbb{Z}_{m_i}$, $a_i^{m_i} = 1$, so $a_i^{m_n} = 1$, since m_i divides m_n. Thus, for all $\alpha \in G$, we have $\alpha^{m_n} = 1$, so every element of G is a zero of $x^{m_n} - 1$. But G has $\prod_{i=1}^{n} m_i$ elements. However, $x^{m_n} - 1$ can have at most m_n zeros in a field. Therefore, $n = 1$ and G is cyclic. □

Corollary 4.13. *The multiplicative group of all nonzero elements of a finite field under field multiplication is cyclic.*

Proof: This result follows trivially from the proof of the Theorem 4.37. □

Corollary 4.14. *A finite extension E of a finite field F is a simple extension of F.*

Proof: Let α be a generator for the cyclic group E^* of nonzero elements of E. Then it is trivial that $E = F(\alpha)$. □

Example 4.24. Consider the finite field $\mathbb{Z}_{11}$. By Corollary 1 to Theorem 4.37, $\langle \mathbb{Z}_{11}, \cdot \rangle$ is cyclic. The generator of $\mathbb{Z}_{11}^*$ is easy to finite as by Fermat's Little Theorem, $a^{p-1} \equiv 1 \mod p$, where p is prime. Therefore, 2 is a generator of $\mathbb{Z}_{11}^*$ as $2^{10} \equiv 1 \mod 11$. So 2 is a primitive 10th root of unity in $\mathbb{Z}_{11}$. But then by the Fundamental Theorem of Abelian Groups, all the primitive roots of unity in $\mathbb{Z}_{11}$ are of the form 2^n, where n is relatively prime to 10.

4.9. **Galois Theory.** "Galois theory is a showpiece of mathematical unification, brining together several different branches of the subject and creating a powerful machine for the study of problems of considerable historical and mathematical importance."
Ian Stewart, *Galois Theory*

We have built the basic parts to our "machine". We will now put all the basic pieces together to form a power machine for looking at the relation between groups and fields. We will then later use this tool to complete our goal of proving the insolvability of polynomials of degree $n \geq 5$.

The power of Galois Theory is the realization that there is an intimate connection between the lattice of subgroups and the lattice of subfields. Specifically, Galois theory examines the intimate connection between the lattice of subfields of an algebraic extension E of a field F to the subgroup structure on a particular group, the set of all automorphisms from E to F while acting as the identity on F. Ultimately, this discovery came while trying to solve polynomials by radicals. In fact, we will show that there is a injective correspondence between the set of all automorphisms from E to F acting as the identity on F and all subfields of an extension E containing F. If one looks at the proof of the

insolvability of the quintic given by Abel, though certainly not exactly as it appears today and does not use modern language, Abel's idea is exactly that of Galois Theory.

First, we will remind the reader of all the necessary accomplishments we have made in the preceding sections which will come together to allow us to prove the Fundamental Theorem of Galois Theory.

1. Let $F \leq E \leq \overline{F}$, $\alpha \in E$ and let β be conjugate of α over F. Then there is an isomorphism $\psi_{\alpha,\beta}$ mapping $F(\alpha)$ onto $F(\beta)$ which leaves F fixed and maps α onto β.
2. If $F \leq E \leq \overline{F}$ and $\alpha \in E$, then an automorphism σ of $\overline{F}$ which leaves F fixed must map α onto some conjugate of α over F.
3. If $F \leq E$, the collection of all automorphisms of E leaving F fixed forms a group $G(E/F)$. For any subset S of $G(E/F)$, the set of all elements of E left fixed by all elements of S is a field E_S. Moreover, $F \leq E_{G(E/F)}$.
4. A field E, $F \leq E \leq \overline{F}$, is a splitting field over F if and only if every isomorphism of E into $\overline{F}$ leaving F fixed is an automorphism of E. If E is a finite extension and a splitting field over F, then $|G(E/F)| = \{E : F\}$.
5. If E is a finite extension of F, then $\{E : F\}$ divides $[E : F]$. If E is also separable over F, then $\{E : F\} = [E : F]$. Also, E is separable over F if and only if $\text{irr}(\alpha, F)$ has all zeros of multiplicity 1 for every $\alpha \in E$.
6. If E is a finite extension of F and is a separable splitting field over F, then $|(G(E/F)$

Our focus is on automorphisms of a finite extension K of F that leaves F fixed and

$$[K : F] = \{K : F\}$$

But these are exactly the separable splitting fields over F!

Definition 4.25. (Finite Normal Extension)
A finite extension K of F is a finite normal extension of F is K is a separable splitting field over F.

We need but one more theorem before we can state our main result.

Theorem 4.38. *(Induced Automorphism)*
Let K be a finite normal extension of F and let E be an extension of F, where $F \leq E \leq K \leq \overline{F}$. Then K is a finite normal extension of E and $G(K/E)$ is precisely the subgroup of $G(K/F)$ consisting of all those automorphisms which leave E fixed. Moreover, two automorphisms σ and τ in $G(K/F)$ induce the same isomorphism of E and into $\overline{F}$ if and only if they are in the same right coset of $G(K/E)$ in $G(K/F)$.

Proof: If K is the splitting field of a set $\{f_i(x) \,|\, i \in I\}$ of polynomials in $F[x]$, then clearly K is the splitting field over E of this same set of polynomials viewed as elements of $E[x]$. Theorem 4.33 shows that K is separable over E, since K is separable over F. Thus, K is a normal extension of E. This proves the first part.

Clearly, every element $G(K/E)$ is an automorphism of K leaving F fixed, since it even leaves the possibly larger field E fixed. Thus, $G(K/E)$ can be viewed as a subset of $G(K/F)$. Since $G(K/E)$ is a group under function composition also, we see that $G(K/E) \leq G(K/F)$.

Finally, for σ and τ in $G(K/F)$, σ and τ are in the same right coset of $G(K/E)$ if and only if $\tau^{-1}(\sigma) \in G(K/E)$ or if and only if $\sigma = \tau(\mu)$ for $\mu \in G(K/E)$. But if $\sigma = \tau(\mu)$ for $\mu \in G(K/E)$, then for $\alpha \in E$, we have

$$\sigma(\alpha) = (\tau\mu)(\alpha) = \tau(\mu(\alpha)) = \tau(\alpha)$$

since $\mu(\alpha) = \alpha$ for $\alpha \in E$. Conversely, if $\sigma(\alpha) = \tau(\sigma)$ for all $\sigma \in E$, then

$$\tau^{-1}(\sigma(\alpha)) = \alpha$$

for all $\sigma \in E$, so $\sigma\tau^{-1}$ leaves E fixed and $\mu = \sigma\tau^{-1}$ is thus in $G(K/E)$. □

The heart of Galois Theory says that for a finite normal extension K of a field F, there must be an injective correspondence between the subgroups of $G(K/F)$ and all the intermediate fields E, where $F \leq E \leq K$. But this correspondence associates with each intermediate field E the subgroup $G(K/E)$! We can reverse this process. We illustrate this with an example.

Example 4.25. Let $K = \mathbb{Q}(\sqrt{2}, \sqrt{3})$. Now K is a normal extension of $\mathbb{Q}$ and there are four automorphisms of K leaving $\mathbb{Q}$ fixed (which we found in an earlier example).

$$\begin{aligned} i &: \text{identity map} \\ \sigma_1 &: \sqrt{2} \mapsto -\sqrt{2}, \sqrt{6} \mapsto -\sqrt{6} \\ \sigma_2 &: \sqrt{3} \mapsto -\sqrt{3}, \sqrt{6} \mapsto -\sqrt{6} \\ \sigma_3 &: \sqrt{2} \mapsto -\sqrt{2}, \sqrt{3} \mapsto -\sqrt{3} \end{aligned}$$

We already stated that $G = \{i, \sigma_1, \sigma_2, \sigma_3\}$ is isomorphic to the Klein 4-group. Below we see all possible subgroups of G paired with its corresponding intermediate field which it fixes.

$$\begin{aligned} \{i, \sigma_1, \sigma_2, \sigma_3\} &\leftrightarrow \mathbb{Q} \\ \{i, \sigma_1\} &\leftrightarrow \mathbb{Q}(\sqrt{3}) \\ \{i, \sigma_2\} &\leftrightarrow \mathbb{Q}(\sqrt{2}) \\ \{i, \sigma_3\} &\leftrightarrow \mathbb{Q}(\sqrt{2}\sqrt{3}) \\ \{i\} &\leftrightarrow \mathbb{Q}(\sqrt{2}, \sqrt{3}) \end{aligned}$$

Notice all the subgroups of the Abelian group $\{i, \sigma_1, \sigma_2, \sigma_3\}$ are normal subgroups of G. Moreover, all the intermediate fields are normal extensions of $\mathbb{Q}$.

Notice from the example that if one subgroup is contained in another, then the larger corresponds to the smaller of the two corresponding fixed fields, why? The larger the subgroup the more possible automorphisms that are possible. Vise versa, the smaller the fixed field, the fewer the elements that could be left fixed. Furthermore, notice in the example that groups at the top of the chain correspond to the fields on the bottom of the chain. But then the lattice of subgroups is corresponds to the lattice of subfields

turned upside-down! This is known as the lattice inversion principle and this was the key observation made by Galois.

Definition 4.26. (Galois Group) If K is a finite normal extension of a field F, then $G(K/F)$ is called the Galois group of K over F.

We finally can state the fruits of all our labor.

Theorem 4.39. *(The Fundamental Theorem of Galois Theory) Let K be a finite normal extension of a field F with Galois group $G(K/F)$. For a field E, where $F \leq E \leq K$, let $E\lambda$ be the subgroup of $G(K/F)$ leaving E fixed. Then λ is a injective map of the set of all such intermediate fields E onto the set of all subgroups of $G(K/F)$. The following properties hold for λ:*

1. $\lambda(E) = G(K/E)$
2. $E = K_{G(K/E)} = K_{\lambda(E)}$
3. *For* $H \leq G(K/F), \lambda(K_H) = H$
4. $[K : E] = |\lambda(E)|$; $[E : F] = \{G(K/F) : \lambda(E)\}$, *the number of cosets of* $\lambda(E)$ *in* $G(K/F)$.
5. *E is a normal extension of F if and only if $E\lambda$ is a normal subgroup of $G(K/F)$. When $E\lambda$ is a normal subgroup of $G(K/F)$, then*
$$G(E/F) \cong G(K/F)/G(K/E)$$
6. *The lattice of subgroups of $G(K/F)$ is the inverted lattice of intermediate fields of K over F.*

Proof:

1. Property 1 follows straight from the definition of λ.
2. From Theorem 4.28, we know that
$$E \leq K_{G(K/E)}$$
Let $\alpha \in K$, where $\alpha \notin E$. Since K is a normal extension of E, we can find an automorphism an automorphism of K leaving E fixed and mapping α onto a different zero of $\text{irr}(\alpha, F)$. This implies that
$$K_{G(K/E)} \leq E$$
so $E = K_{G(K/E)}$. Then we have shown Property 2. Moreover, we have shown that λ is injective because if $\lambda(E_1) = \lambda(E_2)$, then we have
$$E_1 = K_{\lambda(E_1)} = K_{\lambda(E_2)} = E_2$$
3. Property 3 is equivalent to showing that λ is surjective. For $H \leq G(K/F)$, we have $H \leq \lambda(K_H)$, for H surely is included in the set of all automorphisms leaving K_H fixed. Here we use the property $[K : E] = \{K : E\}$.. We know that $H \leq \lambda(K_H) \leq G(K/F)$. Thus, what we really must show is that it is impossible to have H a proper subgroup of $\lambda(K_H)$. We assume that $H < \lambda(K_H)$ and derive a contradiction. If K_H is infinite, then as a finite separable extension of an infinite field, $K = K_H(\alpha)$ for some $\alpha \in K$, by

the Primitive Element Theorem. On the other hand, if K_H is finite, then we still have $K = K_H(\alpha)$ for some $\alpha \in K$, by the second Corollary to Theorem 4.37. Let

$$n = [K : K_H] = \{K : K_H\} = |G(K/K_H)|$$

Then $H < G(K/K_H)$ implies that $|H| < |G(K/K_H)| = n$. Thus, we would have to have $|H| < [K : K_H] = n$. Let the elements of H be $\sigma_1, \cdots, \sigma_{|H|}$ and consider the polynomial

$$f(x) = \prod_{i=1}^{|H|} \big(x - \sigma_i(\alpha)\big)$$

Then $f(x)$ is of degree $|H| < n$. Now the coefficients of each power of x in $f(x)$ are symmetric expressions in the $\sigma_i(\alpha)$. Thus, the coefficients are invariant under each isomorphism $\sigma_i \in H$, since if $\sigma \in H$, then $\sigma(\sigma_1), \cdots, \sigma(\sigma_{|H|})$ is again the sequence $\sigma_1, \cdots, \sigma_{|H|}$, up to ordering, H being a group. Hence, $f(x)$ has coefficients in K_H and since some σ_i is the identity, we see that some $\sigma_i(\alpha) = \alpha$, so $f(\alpha) = 0$. Therefore, we would have

$$\deg(\alpha, K_H) \leq |H| < n = [K : K_H] = [K_H(\alpha) : K_H]$$

which is impossible.

4. This comes from $[K : E] = \{K : E\}$,$[E : F] = \{E : F\}$, and the final statement of Theorem 4.38
5. Every extension E of F, $F \leq E \leq K$, is separable over F by Theorem 4.37. Thus, E is normal over F if and only if E is a splitting field over F. By the Isomorphism Extension Theorem, every isomorphism of E into $\overline{F}$ leaving F fixed can be extended to an automorphism of K, since K is normal over F. Thus, the automorphisms of $G(K/F)$ induce all possible isomorphisms of E into $\overline{F}$ leaving F fixed. This shows that E is a splitting field over F and hence normal over F if and only if for all $\sigma \in G(K/F)$ and $\alpha \in E$

$$\sigma(\alpha) \in E$$

By Property 2, E is the fixed field of $G(K/E)$, so $(\alpha\sigma) \in E$ if and only if for all $\tau \in G(K/E)$

$$\tau(\sigma(\alpha)) = \sigma(\alpha)$$

However, this holds if and only if

$$(\sigma\tau\sigma^{-1})(\alpha) = \alpha$$

for all $\alpha \in E$, $\sigma \in G(K/F)$, and $\tau \in G(K/E)$. But this means that for all $\sigma \in G(K/F)$ and $\tau \in G(K/E)$, $\sigma\tau\sigma^{-1}$ leaves every element of E fixed, that is

$$(\sigma\tau\sigma^{-1}) \in G(K/E)$$

This is precisely the condition that $G(K/E)$ be a normal subgroup of $G(K/F)$. Finally, we only need show that when E is a normal extension of F, $G(E/F) \cong G(K/F)/G(K/E)$. For $\sigma \in G(K/F)$, let σ_E be the automorphism of E induced by σ (assuming that E is a

normal extension of F). Thus, $\sigma_E \in G(E/F)$. The map $\phi : G(K/F) \to G(E/F)$ given by

$$\sigma_\phi = \sigma_E$$

for $\sigma \in G(K/F)$ is obviously a homomorphism. By the Isomorphism Extension Theorem, every automorphism of E leaving F fixed can be extended to some automorphism of K, i.e. it is τ_E for some $\tau \in G(K/F)$. Thus, ϕ is onto $G(E/F)$. The kernel of ϕ is clearly $G(K/E)$. Therefore, by the Fundamental Isomorphism Theorem, $G(E/F) \cong G(K/F)/G(K/E)$. Furthermore, this isomorphism is the natural one. □

The power of the Fundamental Theorem of Galois Theory comes from its ability to relate the lattice of subfields to the lattice of subfields and their corresponding subgroups/subfields. In general, it is difficult to find all possible subfields of a given field. However, it is a simpler exercise to try to find all possible subgroups of a given group. Hence, the Fundamental Theorem gives us a powerful computation saving device.

4.10. **Solvability and Galois Groups of Polynomials.** We have reached the final stretch. In this section, we will reach our ultimate goal of showing the insolvability of the quintic. However, first we must define what we mean when we say solvable. Certainly there are quintics which are "solvable". For example, take the quintic $f(x) = x^5 - 15x^4 + 85x^3 - 225x^2 + 274x - 120 \in \mathbb{R}[x]$. Then $f(x)$ has zeros 1, 2, 3, 4, and 5 in $\mathbb{R}$. When we say solvable, we mean solvability by radicals.

Definition 4.27. (Extension by Radicals)
An extension K of a field F is an extension of F by radicals if there are elements $\alpha_1, \cdots, \alpha_r$ and positive integers $n_1, \cdots, n_r$ such that $K = F(\alpha_1, \cdots, \alpha_r)$, $\alpha_1^{n_1} \in F$ and $\alpha_i^{n_i} \in F(\alpha_1, \cdots, \alpha_{i-1})$ for $1 < i \leq r$.

Definition 4.28. (Solvable by Radicals)
A polynomial $f(x) \in F[x]$ is solvable by radicals over F if the splitting field K of $f(x)$over F is an extension of F by radicals.

This simply means that a polynomial $f(x) \in F[x]$ is solvable if we can obtain all possible zeros of $f(x)$ by adjoining the nth roots (n dependent) to F. Hence, a polynomial $f(x) \in F[x]$ is solvable by radicals over F if we can obtain every zero of $f(x)$ by using a finite sequence of the operations of addition, subtraction, multiplication, division, and taking roots, starting with the elements of F. We take this question about solvability and turn it into a question about field extensions. Here we will follow the proof given in *Abstract Algebra* by Dummit and Foote[4], see the Appendix for the derivations of the formulas for finding roots of the linear, quadratic, cubic, and quartic equations as well as proofs of the Fundamental Theorem of Algebra. We first discuss the Galois groups of a polynomial.

If K is a Galois extension of F then K is the splitting field for some separable polynomial $f(x)$ over F. Furthermore, any automorphism $\sigma \in G(K/F)$ cannot change the roots of $f(x)$. Hence, σ is a permutation of the roots and is uniquely determined by the permutation

of the roots (since the permutations generate K over F). But then each $\sigma \in G(K/F)$. This then gives the injection

$$G(K/F) \to S_n$$

Therefore, as with all groups, we can think of the Galois groups as a subgroup of some permutation group. By the Fundamental Theorem of Galois Theory, the degree of the splitting field is the same as the order of the Galois group. But recall that we can write a polynomial $f(x)$ as a product of irreducibles in the form $f_1(x)f_2(x)\cdots f_k(x)$, where each of the $f_i(x)$ has some degree n_i. Since the Galois group merely permutes the roots of $f(x)$, i.e. it permutes the roots of the irreducibles amongst themselves, it must be the case that

$$G(K/F) \leq S_{n_1} \oplus \cdots \oplus S_{n_k}$$

Similarly with Cayley's Theorem, it is an open mathematical question whether every finite group embeds in some Galois group. Next, we introduce some terminology to show that a general polynomial of degree n has S_n as its Galois group.

Definition 4.29. (Symmetric Functions)
Let $x_1, x_2, \cdots, x_n$ be indeterminates. The elementary symmetric functions $s_1, s_2, \cdots, s_n$ are defined by

$$\begin{aligned} s_0 =& 1 \\ s_1 =& x_1 + x_2 + \cdots + x_n \\ s_2 =& x_1x_2 + x_1x_3 + \cdots + x_2x_3 + x_2x_4 + \cdots + x_{n-1}x_n \\ &\vdots \\ s_n =& x_1x_2\cdots x_n \end{aligned}$$

Now we will rigorously define what we mean when we say the general polynomial.

Definition 4.30. (General Polynomial)
The general polynomial of degree n is the polynomial $f(x)$ given by

$$(x - x_1)(x - x_2)\cdots(x - x_n)$$

Where the roots of $f(x)$ are $x_1, x_2, \cdots, x_n$.

Theorem 4.40. *(General Polynomial Coefficients)*
The coefficient of the ith power of the general polynomial $f(x)$ of degree n is the elementary symmetric function s_{n-i}.

Proof: We prove this by induction. The proof is trivial for the $n = 0, 1$ cases. We look at the $n = 2$ case explicitly here. Let $n = 2$, then the general polynomial has the form

$$f_2(x) = (x - x_1)(x - x_2) = x^2 - (x_1 + x_2)x + x_1x_2$$

But notice the coefficients are precisely s_0, s_1, s_2. Now assume this is true for $n = 1, 2, 3, 4, \cdots, k$. Consider the general polynomial

$$f_{k+1}(x) = (x - x_1)(x - x_2)(x - x_3)\cdots(x - x_k)(x - x_{k+1})$$

By the induction hypothesis, we have

$$f_{k+1}(x) = (x^k s_{0,k} + x^{k-1} s_{1,k} + \cdots + s_{k,k})(x - x_{k+1})$$

Applying simple algebra, we find that

$$\begin{aligned} f_{k+1}(x) =& (x^k s_{0,k} + x^{k-1} s_{1,k} + \cdots + s_{k,k})(x - x_{k+1}) \\ =& (x^{k+1} s_{0,k} + x^k s_{1,k} + \cdots + x s_{k,k}) + (-x^k s_{0,k} x_{k+1} - x^{k-1} s_{1,k} x_{x+1} - \cdots + s_{k,k} x_{k+1}) \\ =& s_{0,k} x^{k+1} + (s_{1,k} - s_{0,k} x_{k+1}) x^k + \cdots + s_{k,k} x_{k+1} \end{aligned}$$

However, notice that

$$\begin{aligned} s_{0,k+1} =& s_{0,k} \\ s_{1,k+1} =& s_{1,k} - s_{0,k} x_{k+1} \\ & \vdots \\ s_{k+1,k+1} =& s_{k,k} x_{k+1} \end{aligned}$$

Then the coefficients of the nth power term of $f_{k+1}(x)$ are precisely $s_{n-(k+1)}$. □

It is important to notice that for any field F, the extension $F(\alpha_1, \alpha_2, \cdots, \alpha_n)$ is a Galois extension of the field $F(s_1, s_2, \cdots, s_n)$ since it is a splitting field of the general polynomial of degree n. If given a rational function $F(\alpha_1, \alpha_2, \cdots, \alpha_n)$, $\sigma \in S_n$ acts on it by permuting the roots. This automorphism of $F(\alpha_1, \alpha_2, \cdots, \alpha_n)$ identifies S_n as a subgroup of $\mathrm{Aut}(F(\alpha_1, \alpha_2, \cdots, \alpha_n))$. However, the symmetric functions remainder fixed, up to order, under this permutation, hence the name symmetric. Then the (sub)field $F(\alpha_1, \alpha_2, \cdots, \alpha_n)$ is contained in the fixed field of S_n. Using the Fundamental Theorem of Galois Theory, the fixed field of S_n must have index $n!$ in $F(\alpha_1, \alpha_2, \cdots, \alpha_n)$. Then we have

$$[F(\alpha_1, \alpha_2, \cdots, \alpha_n : F(s_1, s_2, \cdots, s_n] \leq n!$$

because $F(\alpha_1, \alpha_2, \cdots, \alpha_n)$ is the splitting field over $F(s_1, s_2, \cdots, s_n)$. But then the two must be equivalent. Therefore, $F(s_1, s_2, \cdots, s_m)$ is the fixed field of S_n.

Definition 4.31. (Symmetric)
A rational function $f(x_1, x_2, \cdots, x_n)$ is called symmetric if it is not changed by any permutation of the variables $x_1, x_2, \cdots, x_n$.

Theorem 4.41. *(Fundamental Theorem on Symmetric Functions)*
Any symmetric function in the variables $x_1, x_2, \cdots, x_n$ is a rational function in the elementary symmetric functions $s_1, s_2, \cdots, s_n$.

Proof: A symmetric function lies in the fixed field of S_n. Hence, it is a rational function in $s_1, s_2, \cdots, s_n$. □

Moreover, one can show that there is no rational function "between" the symmetric polynomials. That is to say, there are no polynomial relations between them. With this fact, it makes no difference whether one considers a polynomial $f(x)$ to have indeterminate roots or indeterminate coefficients.

Definition 4.32. (Discriminant)
THe discriminant D of $x_1, x_2, \cdots, x_n$ by the formula

$$D = \prod_{i<j}(x_i - x_j)^2$$

Define the discriminant of a polynomial to be the discriminate of the roots of the polynomial. Moreover, the discriminant is a symmetric function in $x_1, \cdots, x_n$ and therefore is an element of $K = F(s_1, s_2, \cdots, s_n)$.

Recall that in a typical high school course, one learns that the discriminant of a quadratic $ax^2 + bx + c$ is $b^2 - 4ac$. Notice that this is a symmetric function and uses the indeterminate coefficients of the polynomial. Recall that in high school one used this to learn the types of roots the quadratic had. This is because of the intimate relationship between the coefficients of the quadratic and the roots themselves.

Theorem 4.42. *If char $F \neq 2$, then the permutation $\sigma \in S_n$ is an element of A_n if and only if it fixed the square root of the discriminant D (i.e., the discriminate $D \in F$ is the square of some element $f \in F$).*

Proof: A permutation $\sigma \in S_n$ is a member of the alternation group A_n if and only if it fixes the product

$$\sqrt{D} = \prod_{i<j}(x_i - x_j) \in \mathbb{Z}[x_1, x_2, \cdots, x_n]$$

Then by the Fundamental Theorem of Galois Theory, if F has characteristic different than 2 then $\sqrt{D}$ generates the fixed field of A_n and generates a quadratic extension of K. □

We now make two important remakes about the discriminant.

Remark 4.4. The discriminant D is symmetric in the roots of $f(x)$ and is fixed by all the automorphisms of the Galois group of $f(x)$. Then by the Fundamental Theorem of Galois Theory, $D \in F$.

Remark 4.5. The discriminant can be written as a polynomial in the coefficients of $f(x)$. Since

$$\sqrt{D} = \prod_{i<j}(\alpha_i - \alpha_j)$$

$\sqrt{D}$ is always contained in the splitting field for $f(x)$. So if the roots of $f(x)$ are distinct, we can fix an order of the roots and consider the Galois group of $f(x)$ as merely a permutation of the roots and hence a subgroup of S_n.

This is exactly why one looked at the discriminant in high school. For example, if $D = 0$, then we have detected the presence of roots of multiplicity.

We have already discussed what it means to be solvable by radicals. We will use simple radical extensions, that is extensions obtained simply by adjoining the nth root of an element $a \in F$ to F. Notice that all the roots of $x^n - a$ differ only by the factors of the nth roots of unity for $a \in F$. Therefore, if we adjoin a root we obtain a Galois extension

if and only if the new field contains the nth root of unity. This new field "behaves well" over the original field F when the field F already contains the necessary roots of unity.

Remark 4.6. The symbol $\sqrt[n]{a}$, for $a \in F$, is used to denote the root of the polynomial $x^n - a \in F[x]$.

Definition 4.33. (Cyclic Extension)
The extension K/F is said to be cyclic if it is Galois with a cyclic with a cyclic Galois group.

We will need something called the Lagrange resolvent for the proof of the next theorem.

Definition 4.34. (Lagrange Resolvent)
For $\alpha \in K$ and any nth root of unity ζ, define the Lagrange resolvent $(\alpha, \zeta) \in K$ by

$$(\alpha, \zeta) = \alpha\zeta\sigma(\alpha) + \zeta^2\sigma^2(\alpha) + \cdots + \zeta^{n-1}\sigma^{n-1}(\alpha)$$

Theorem 4.43. *(Characteristics of Unity)*
Let F be a field of characteristic relatively prime to n which contains the nth roots of unity if and only if the extension $F(\sqrt{a})$ for $a \in F$ is cyclic over F of degree dividing n.

Proof:
$\Rightarrow$: The extension $K = F(\sqrt[n]{a})$ is Galois over F if F contains the nth roots of unity since it is the splitting field for $x^n - a$. For any $\sigma \in G(K/F)$, $\sigma(\sqrt[n]{a})$ is another root of this polynomial, hence $\sigma(\sqrt[n]{a}) = \zeta_\sigma \sqrt[n]{a}$ for some nth root of unity ζ_σ. This gives the mapping

$$\begin{aligned} G(K/K) &\to \mu_n \\ \sigma &\mapsto \zeta_\sigma \end{aligned}$$

where μ_n denotes the group of nth roots of unity. Since F contains μ_n, every nth root of unity is fixed by every element of $G(K/F)$. Hence

$$\begin{aligned} \sigma\tau(\sqrt[n]{a}) &= \sigma(\zeta_\tau \sqrt[n]{a}) \\ &= \zeta_\tau\sigma(\sqrt[n]{a}) \\ &= \zeta_\tau\zeta_\sigma \sqrt[n]{a} = \zeta_\sigma\zeta_\tau \sqrt[n]{a} \end{aligned}$$

this shows that $\zeta_{\sigma\tau} = \zeta_\sigma\zeta_\tau$, therefore the map is a homomorphism whose kernel consists precisely of those automorphisms which fix $\sqrt[n]{a}$, namely the identity. This gives an injection of $G(K/F)$ into the cyclic group μ_n of order n, which completes the theorem.

$\Leftarrow$: Let K be any cyclic extension of degree n over a field F of characteristic relatively prime to n which contains the nth roots of unity. Let σ be a generator for the cyclic group $G(K/F)$. If we apply the automorphism σ to the Lagrange resolvent, (α, ζ), we obtain

$$\sigma(\alpha, \zeta) = \sigma\alpha + \zeta\sigma^2(\alpha) + \zeta^2\sigma^3(\alpha) + \cdots + \zeta^{n-1}\sigma^n(\alpha)$$

since ζ is an element of the base field F and is fixed by σ, we have $\zeta^n = 1$ in μ_n and $\sigma^n = 1$ in $G(K/F)$, so this can be written

$$\begin{aligned}\sigma(\alpha,\zeta) =& \sigma\alpha + \zeta\sigma^2(\alpha) + \zeta^2\sigma^3(\alpha) + \cdots + \zeta^{-1}\alpha \\ =& \zeta^{-1}(\alpha + \zeta\sigma(\alpha) + \zeta^2\sigma^2(\alpha) + \cdots + \zeta^{n-1}\sigma^{n-1}(\alpha)) \\ =& \zeta^{-1}(\alpha,\zeta)\end{aligned}$$

It then follows that

$$\sigma(\alpha,\zeta)^n = (\zeta^{-1})^n(\alpha,\zeta)^n = (\alpha,\zeta)^n$$

so that $(\alpha,\zeta)^n$ is fixed by $G(K/F)$. Therefore, $(\alpha,\zeta)^n \in F$ for any $\alpha \in K$. Let ζ be a primitive nth root of unity. By the linear independence of the automorphisms $1, \sigma, \cdots, \sigma^{n-1}$ (our basis), there is an element $\alpha \in K$ with $(\alpha,\zeta) \neq 0$. Iterating the process above, by induction, we find that

$$\sigma^i(\alpha,\zeta) = \zeta^{-i}(\alpha,\zeta), \qquad i = 0, 1, 2, \cdots$$

and it follows that σ^i does not fix α, ζ for any $i < n$. Therefore, this element cannot lie in any proper subfield of K, so $K = F((\alpha,\zeta))$. Since we have proved that $(\alpha,\zeta)^n = a \in F$, we have $F(\sqrt[n]{a}) = F((\alpha,\zeta)) = K$. □

Moreover, we see that every such field must have this form. This is known as Kummer theory. We give a basic proof outline below.

Theorem 4.44. *Any cyclic extension of degree n over a field F of characteristic relatively prime to n contains the nth roots of unity if and only if it is of the form $F(\sqrt[n]{a})$ for some $a \in F$.*

"Proof": Let F be a field of characteristic relatively prime to n which contains the nth roots of unity. If we take elements $a_1, a_2, \cdots, a_k \in F^*$, then the extension

$$F(\sqrt[n]{a_1}, \sqrt[n]{a_2}, \cdots, \sqrt[n]{a_k})$$

is an abelian extension of F whose Galois group is of exponent n. Conversely, any abelian extension of exponent n is of this form. Denote by $(F^*)^n$ the subgroup of the multiplicative group F^* consisting of the nth powers of the nonzero elements of F. The factor group $F^*/(F^*)^n$ is an abelian group of exponent n. The Galois group of the extension above is isomorphic to the group generated in $F^*/(F^*)^n$ by the elements $a_1, \cdots, a_k$ and two extensions as above are equal if and only if their associated groups in $F^*/(F^*)^n$ are equal. Hence, the finitely generate subgroups of $F^*/(F^*)^n$ classify the abelian extensions of exponent n over fields containing the nth roots of unity (and characteristic relatively prime to n). These extensions are called Kummer extensions and can by generalized to larger cases by induction. □

Recall that we say something is solvable by radicals if we can create a chain or tower of simple radical expressions, each a root extension of the previous. When we consider radical extensions, we may always adjoin roots of unity as they they are radicals. Then cyclic extensions are radical extensions and the converse also holds.

Theorem 4.45. *If α is contained in a root extension K, then α is contained in a root extension which is Galois over F, where each extension K_{i+1}/K_i is cyclic.*

Proof: Let L be the Galois closure of K over F. For any $\sigma \in G(L/F)$, we have the chain of subfields

$$F = \sigma K_0 \subset \sigma K_1 \subset \cdots \subset \sigma K_i \subset K_{i+1} \subset \cdots \subset \sigma K_s = \sigma K$$

where $\sigma K_{i+1}/\sigma K_i$ is again a simple radical extension as it is generated by the element $\sigma(\sqrt[n_i]{a_i})$, which is a root of $x^{n_i} - \sigma(a_i)$ over $\sigma(K_i)$. Then the composition of two root extensions is again a root extension. It follows that the composition of all the conjugate fields $\sigma(K)$ for $\sigma \in G(L/F)$ is again a root extension. This field is precisely L, then α is contained in a Galois root extension.

Adjoin the n_ith roots of unity to F for all the roots $\sqrt[n_i]{a}$ for the simple radical extensions in the Galois root extension K/F to obtain a field, say F', and form the composition of F' with the root extension

$$F \subseteq F' = F'K_0 \subseteq F'K_1 \subseteq \cdots \subseteq F'K_i \subseteq F'K_{i+1} \subseteq \cdots \subseteq F'K_s = F'K$$

The field $F'K$ is a Galois extension of F since it is the composition of two Galpois extensions. The extension from F to $F' = F'K_0$ can be represented as a chain of subfields with each individual extension being cyclic (as is true with any Abelian extension). Each extension $F'K_{i+1}/F'K_i$ is a simple radical extension and since we now have the appropriate roots of unity in the base fields, each of these individual extensions from F' to $F'K$ is a cyclic extension by Theorem 4.43. Therefore, $F'K/F$ is a root extension which is Galois over F with cyclic intermediate extensions. □

4.11. **Insolvability of the Quintic.** We now have reached the end. Recall our original question was if there was a general formula for the zeros of a polynomial of degree 5 or higher. We then looked at the properties of zeros by finding fields which contained the zeros of a given irreducible polynomial, i.e. extension fields and splitting fields. Then we classified elements in a field as algebraic or transcendental based on whether they were a zero of a polynomial in the field. Then we looked at the properties of automorphic mappings of the extension field which fixed the base field. We called this the Galois group and showed that it must permute the roots of a polynomial in the base field amongst themselves. However, we then showed there was no difference between examining the roots in terms of the indeterminates or by observing the undetermined coefficients. Then we created root extensions and found that the roots must be contained in a Galois group over a field F and the quotient groups of the tower of root extensions must be cyclic. We complete our proof of the insolvability of the general polynomial of degree greater than or equal to 5 by showing that a group is solvable if and only if its Galois group is solvable. But we showed that such a group for polynomials is a subgroup of S_n. Then showing that S_n contains only the single normal nonabelian simple subgroup A_n for $n \geq 5$ will be sufficient.

Theorem 4.46. *(Radical Solvable iff Galois Solvable) The polynomial $f(x)$ can be solved by radicals if and only if its Galois group is a solvable group.*

Proof:

$\Rightarrow$: Assume that $f(x)$ can be solved by radicals. Then each root of $f(x)$ is contained in an extension as in Theorem 4.45. Let G_i be the subgroups corresponding to the subfields K_i, $i = 0, 1, \cdots, s-1$. Since

$$G(K_{i+1}/K_i) = G_i/G_{i+1}, \qquad i = 0, 1, \cdots, s-1$$

it follows that the Galois group $G = G(L/F)$ is a solvable group. The field L contains the splitting field for $f(x)$ so the Galois group of $f(x)$ is a quotient group of the solvable group G and therefore solvable.

$\Leftarrow$: Suppose that the Galois group G of $f(x)$ is a solvable group and let K be the splitting field of $f(x)$.Taking the fixed fields of the subgroups in a chain for G, gives the chain

$$F = K_0 \subset K_1 \subset \cdots \subset K_i \subset K_{i+1} \subset \cdots \subset K_s = K$$

where K_{i+1}/K_i, $i = 0, 1, \cdots, s-1$ is a cyclic extension of degree n_i. Let F' be the cyclotomic field over F of all roots of unity of order n_i, $i = 0, 1, \cdots, s-1$ and form the composite fields $K_i' = F'K_i$. Then we have the sequence of extensions

$$F \subseteq F' = F'K_0 \subseteq F'K_1 \subseteq \cdots \subseteq F'K_i \subseteq F'K_{i+1} \subseteq \cdots \subseteq F'K_s = F'K$$

Then extension $F'K_{i+1}/F'K_i$ is cyclic of degree dividing n_i, $i = 0, 1, \cdots, s-1$. Since we now have the appropriate roots of unity in the base fields, each of these cyclic extensions is a simple radical extension. Then each of the roots of $f(x)$ is contained in the root extension $F'K$ so that $f(x)$ can be solved by radicals. □

Now we show that the only normal subgroup of S_n for $n \geq 5$ is A_n and then we show that A_n is simple for $n \geq 5$.

Theorem 4.47. *The only normal subgroup of S_n is A_n for $n \geq 5$ and the only subgroup of S_n with index 2 is A_n.*

Proof: Suppose that N is a normal subgroup of S_n with N nontrivial. We show that $A_n \subseteq N$ and therefore $N = A_n$ or $N = S_n$. Choose a permutation $\sigma \in N$ with σ not the identity permutation. So there is an index i such that $\sigma(i) \neq i$. Choose an element $j \in \{1, 2, 3, \cdots, n\}$ such that $j \neq i$ and $j \neq \sigma(i)$. They define $\alpha = (ij)$. Then we have

$$\sigma\ \alpha\ \sigma^{-1}\ \alpha^{-1} = (\sigma(i)\ \sigma(j))(i\ j)$$

But neither $\sigma(i) = i$ nor $j = \sigma(j)$ and $\sigma(i) = \sigma(j)$, then the 2-cycles $(\sigma(i)\ \sigma(j))$ and $(i\ j)$ are not equal. Therefore, their product cannot be the identity. Then we have $\sigma\ \alpha \neq \alpha\ \sigma$.

Since N is a normal subgroup, $\sigma\ \alpha\ \sigma^{-1}\ \alpha^{-1} \in N$. From our construction, we have $\sigma(i) \neq i, j$. It follows then that $(\sigma(i)\ \sigma(j))(i\ j)$ must have the same cycle type, which is $(2, 2)$. If $\sigma(j)$ one of i, j, then $(\sigma(i)\ \sigma(j))(i\ j)$ must be a 3-cycle. Therefore, N must contain a permutation of type $(2, 2)$ or be a 3-cycle. However, since N is a normal subgroup of S_n, it must contain all $(2, 2)$ cycles or all 3-cycles.

However, we know that A_n is generated by 3-cycles. Moreover, for $n > 5$, A_n is generated by permutations of type $(2, 2)$ as we can write any 3-cycle $(x\ y\ z)$ as

$$(x\ y\ z) = (x\ y)(w\ v)(w\ v)(y\ z)$$

so long as neither w, v are one of the x, y, z. Since for $n \geq 5$, all 3-cycles are conjugate in A_n. Why? Suppose that σ is a 3-cycle in A_n with $n \geq 5$. Then we can conjugate to (1 2 3) for some permutation $\tau \in S_n$. That is,

$$(1\ 2\ 3) = \tau\ \sigma\ \tau^{-1}$$

If $\tau \in A_n$, then the proof is complete. If $\tau \notin A_n$, then let $\tau = (4\ 5)$, notice that $\tau \in A_n$. Now when we conjugate, we find

$$\tau\ \sigma\ \tau^{-1} = (4\ 5)\overline{\tau}\ \sigma\ \overline{\tau}^{-1} = (4\ 5)(1\ 2\ 3)(4\ 5) = (1\ 2\ 3)$$

for some $\overline{\tau} \in S_n$. But then all this shows that $A_n \subset N$ and N must contain a generator for A_n; therefore, either $N = A_n$ or $N = S_n$. □

Theorem 4.48. *A_n is simple for $n \geq 5$.*

Proof: We apply induction on n. Let $n = 5$, then we first look at the conjugacy classes of A_5 and their orders. The representatives of the cycle types of even permutations can be taken to be

$$1, \quad (1\ 2\ 3), \quad (1\ 2\ 3\ 4\ 5), \quad (1\ 2)(3\ 4)$$

The only centralizers of the 3-cycles and 5-cycles which are contained in A_5 are

$$C_{A_5}\big((1\ 2\ 3)\big) = \langle (1\ 2\ 3)\rangle, \quad C_{A_5}\big((1\ 2\ 3\ 4\ 5)\big) = \langle (1\ 2\ 3\ 4\ 5)\rangle$$

These groups have order 3 and 5 with index 20 and 12, respectively. Therefore, there are 20 distinct conjugates of (1 2 3) and 12 distinct conjugates of (1 2 3 4 5) in A_5. Since there are twenty 3-cycles in S_5 and all of these must lie in A_5, we have that all 20 3-cycles are conjugate in A_5. There are 24 5-cycles in A_5. However, only 12 of these distinct conjugates of the 5-cycles (1 2 3 4 5). Thus, some 5-cycle, say σ, is not conjugate to (1 2 3 4 5) in A_5 (for example, $\sigma = (1\ 3\ 5\ 2\ 4)$. Any element of S_5 conjugating (1 2 3 4 5) to (1 3 5 2 4) must be an odd permutation. So σ also has 12 distinct conjugates in A_5, accounting for all 24. So the conjugacy classes of A_5 have order 1,15,20,12, and 12. Now suppose that H is a normal subgroup of A_5. Then H would be the union of the conjugacy classes of A_5. But then the order of H would be a divisor of 60 (the order of A_5) and be the sum of some collection of the integers $\{1, 12, 12, 15, 20\}$. But then the only possibilities are $|H| = 1$ or $|H| = 60$. Hence, the only normal subgroups of A_5 are trivial.

Now we proceed with the induction. Assume that $n \geq 6$ and let $G = A_n$. Assume that there exists a normal subgroup H of G with H nontrivial ($H \neq 1$ and $H \neq G$). For each $i \in \{1, 2, 3, \cdots, n\}$, let G_i be the stabilizer of i in the natural action of G on $i \in \{1, 2, 3, \cdots, n\}$. Therefore, $G_i \leq G$ and $G_i \cong A_{n-1}$. By induction, G_i is simple for $1 \leq i \leq n$.

Now suppose first that there is some $\tau \in H$ with $\tau \neq 1$ but $\tau(i) = i$ for some $i \in \{1, 2, 3, \cdots, n\}$. Since $\tau \in H \cap G_i$ and $H \cap G_i$ being normal in G_i, by the simplicity of G_i we must have $H \cap G_i = G_i$, that is

$$G_i \leq H$$

However, $\sigma G_i \sigma^{-1} = G_{\sigma(i)}$, so for all $i, \sigma G_i \sigma^{-1} \leq \sigma H \sigma^{-1} = H$. Thus,

$$G_k \leq H$$

for all $j \in \{1, 2, \cdots, n\}$. Any $\lambda \in A_n$ may be written as the product of an even number of $2t$ transpositions, so

$$\lambda = \lambda_1 \lambda_2 \cdots \lambda_t$$

where λ_k is a product of two transpositions. Since $n > 4$, each $\lambda_k \in G_j$, for some j. Hence,

$$G = \langle G_1, G_2, \cdots, G_n \rangle \leq H$$

which is a contradiction. Therefore, if $\tau \notin 1$ is an element of H, then $\tau(i) \neq i$ for an $i \in \{1, 2, \cdots, n\}$. That is, no nonidentity element of H fixes any element of $\{1, 2, 3, \cdots, n\}$.

It then follows that if τ_1, τ_2 are elements of H with

$$\tau_1(i) = \tau_2(i)$$

for some i, then

$$\tau_1 = \tau_2$$

since then we would have $\tau_2^{-1}\tau_1(i) = i$. Now suppose that there exists a $\tau \in H$ such that the cycles decomposition of τ contains a cycle of length greater than 3, say

$$\tau = (a_1 a_2 a_3 \cdots)(b_1 b_2 \cdots) \cdots$$

Let $\sigma \in G$ be an element with $\sigma(a_1) = a_1, \sigma(a_2) = a_2$ but $\sigma(a_3) \neq a_3$ (such an element exists since in A_n since $n \geq 5$). Then we have

$$\tau_1 = \sigma\tau\sigma^{-1} = (a_1 a_2 \sigma(a_3) \cdots)(\sigma(b_1)\sigma(b_2) \cdots) \cdots$$

so τ and τ_1 are distinct elements of H with $\tau(a_1) = \tau_1(a_1) = a_2$, contrary to the fact that if $\tau_1(i) = \tau_2(i)$ for some i, then $\tau_1 = \tau_2$. So the only 2-cycle that can appear in the cycle decomposition of nonidentity elements of H.

Now let $\tau \in H$, with $\tau \neq 1$, so that

$$\tau = (a_1 a_2)(a_3 a_4)(a_5 a_6) \cdots$$

Let $\sigma = (a_1 a_2)(a_3 a_4)(a_5 a_6) \in G$. Then

$$\tau_1 = \sigma\tau\sigma^{-1} = (a_1 a_2)(a_3 a_4)(a_5 a_6) \cdots$$

hence τ and τ_1 are distinct elements of H with $\tau(a_1) = \tau_1(a_1) = a_2$, again contrary to the fact that if $\tau_1(i) = \tau_2(i)$ for some i, then $\tau_1 = \tau_2$. □

Finally, the proof we have been waiting for:

Theorem 4.49. *The general equation of degree n cannot be solved by radicals for $n \geq 5$.*

Proof: For $n \geq 5$, the group S_n has only A_n as the nontrivial proper normal subgroup, which is simple. Because A_n is not abelian for $n \geq 5$, it cannot be cyclic and then S_n cannot be solvable and therefore the Galois group is not solvable. $\square$

Notice how short the proof is! So short in fact, that it is almost lost in the sea of theorems we developed to answer it. This shows the power of Galois theory - that a difficult question as the solvability of all polynomials of certain degrees can be answered so efficiently. However, the path to this power is certainly nontrivial.

5. Representation Theory

Introduction. Here we look at how the one can study the symmetric groups with what are called representations, which will be defined shortly. By Cayley's Theorem, every finite group can be thought of as a subgroup of the permutation group S_n for some n. Our goal is to understand groups in a very different way. Every group of course comes equipped with a binary operation. However, how the group behaves becomes a difficult problem to answer. The goal of representation theory is to take a elements of a group and represent them as matrices. Operations in the group then are linear transformations on a vector space and the group operations become matrix operations. That is we study the symmetries of a group G acting on a set X with a few extra structure restrictions. In this way, questions about the group structure are reduced to linear algebra questions, which are more concrete and well understood. Here then we make a change from the language of groups to the language of modules and algebras. In group theory, we look at factor groups to understand the group structure by looking at smaller pieces of the group. Similarly, we will create a representation for a group and then break it up into irreducible pieces. If all the irreducible pieces can be understood the symmetries of the space are understood and hence the original group. Finally, it turns out that these representations for the symmetric group can be quickly found and understood using diagrams called Young Tableaux. For its clarity and brevity, we closely follow the language and proofs found in Sagan's "The Symmetric Group." Throughout this section we assume that the group G under discussion is finite.

5.1. **Permutations.** Though we have been discussing the properties of permutations throughout this paper, we have not had occasion to work with many of its properties and representation. Therefore, we shall briefly review these here to be sure of our language and definitions throughout our discussion of representations. We call $\pi \in S_n$ a permutation. We can represent a permutation by specifying each action of π on $\{1, 2, \cdots, n\}$. For example, take $\pi \in S_4$ given by

$$\pi(1) = 2 \quad \pi(2) = 4 \quad \pi(3) = 3 \quad \pi(4) = 1$$

But this is not only very time consuming but repetitive. It is simpler to represent π using a matrix. Taking π as before, we can represent it using the matrix

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 3 & 1 \end{pmatrix}$$

where the top line is the element under consideration and the element directly below it is where it is sent under π. Elements missing from the matrix notation are assumed to be mapped to themselves under π. Now since the top line is fixed, we can drop it to obtain the one line cycle notation.

$$\pi = (1\ 2\ 4)$$

where again elements missing from $(\cdots)$ are assumed to be fixed. This is one of the most common way of presenting a permutation π. Looking at the cycle notation for π, it is clear that π breaks the elements into cycles of possibly varying lengths. The permutation

π creates a partition on the set $\{1, 2, \cdots, n\}$. It would then be useful to develop a language so that we can talk about permutations in terms of partition on a set.

Definition 5.1. (k-Cycle/Cycle Type) A k-cycle or cycle of length k is a cycle containing k elements. The cycle type type or the type of a permutation π is an expression in the form $(1^{n_1}, 2^{n_2}, \cdots, m^{n_j})$, where n_j is the is the number n of k-cycles of length j in π.

Example 5.1. Consider the permutations $\pi = (123)(45)$ and $\rho = (12)(45)$ in S_5. Then π has cycle type $(1^0, 2^1, 3^1)$ and ρ has cycle type $(1^0, 2^2)$. Notice their product $\pi\rho = (13)$ has cycle type $(1^0, 2^1)$ and $\rho\pi = (23)$ has cycle type $(1^0, 2^1)$.

Example 5.2. Take the permutation (partition) of $\pi \in S_{14}$ given by

$$\pi = (1\ 2)(3\ 4)(5)(6\ 7\ 8)(9\ 10\ 11\ 12\ 13\ 14)$$

contains the 2-cycles (1 2) and (3 4) and the 1-cycle (or fixed point, 5). The cycle type of the permutation π is $(1^1, 2^1, 3^1, 4^0, 5^0, 6^1)$ or $(1^1, 2^1, 3^1, 6^1)$.

With this we can talk directly about partitions of a set as permutations on the set.

Definition 5.2. (Partitions) A partition on a set of n elements is a sequence

$$\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_l)$$

where all the λ_i are weakly decreasing and $\sum_{i=1}^{l} \lambda_i = n$

Notice in the following examples, the partition is essentially the cycle type.

Example 5.3. Consider again the permutations $\pi = (123)(45)$ and $\rho = (12)(45)$ in S_5. Then π corresponds to the partition $(3, 2)$ and ρ to the partition $(2, 2)$.

Example 5.4. Take the permutation from Example 5.2,

$$\pi = (1\ 2)(3\ 4)(5)(6\ 7\ 8)(9\ 10\ 11\ 12\ 13\ 14)$$

this corresponds to the partition $\lambda = (6, 3, 2, 2, 1)$.

It should be clear that we are going to often be interested in partitions of a set, in this case a group, because these partitions correspond to permutations of the set. Recall that we say two elements g, h of a group are conjugate if there is a $k \in G$ such that

$$g = khk^{-1}$$

Lemma 5.1. *Conjugacy classes form an equivalence relation.*

Proof: Let $g \sim h$ if $h = khk^{-1}$ for some $k \in G$.

1. Reflexive: $g \sim g$ as $g = ege^{-1} = g$.
2. Symmetric: If $g \sim h$ then there is a $k \in G$ such that $g = khk^{-1}$. Then $h = k^{-1}gk$. So let $k' = k^{-1}$ and then $h = k'gk'^{-1}$ and h is conjugate to g.
3. Transitive: Suppose $g \sim h$ and $h \sim k$. Then there is a $m, n \in G$ such that $g = mhm^{-1}$ and $h = nkn^{-1}$. Then $g = mhm^{-1} = m(nkn^{-1})m^{-1} = (mn)k(mn)^{-1}$ and mn conjugates k to g so $g \sim k$. □

The set of all elements that are conjugate to an element g is called the equivalence class, K_g of g. Notice that as before we partitioned the set with numbers, conjugacy forms a set partition of the set/group into conjugacy classes. The question is, how does one calculate the number of conjugacy classes in a group? Since a permutation only permutes those elements which are in the same cycle type, two permutations are in the same conjugacy class if they have the same cycle type. In fact, the other direction holds as well. But then there is a natural bijection between the cosets of Z_g and K_g, giving the nice relation on the cardinality of G

$$|G| = |K_g||Z_g| \qquad \text{or} \qquad |K_g| = \frac{|G|}{|Z_g|}$$

Theorem 5.1. *($|G| = |K_g||Z_g|$) Given a group G and an element $g \in G$, there is a bijection between the cosets of Z_g and elements of K_g so that $|G| = |K_g||Z_g|$.*

Proof: Suppose a group G acts on a set S. Let $s \in S$. Recall that the stabilizer of s in G is $G_s = \{g \in G \,|\, gs = s\}$ and the orbit of s is $O_s = \{gs : g \in G\}$. First, we show that G_s is a subgroup of G. Notice that $e \in G_s$ because it fixes s. Therefore, G_s is nonempty. Assume that $a, b \in G_s$. Then $as = s$ and $bs = s$. Furthermore, $b^{-1} \in G_s$ because

$$\begin{aligned} s =& es \\ =& (b^{-1}b)s \\ =& b^{-1}(bs) \\ =& b^{-1}s \end{aligned}$$

But since $b^{-1}s = s$, then $s = as = ab^{-1}$ and G_s is a subgroup of G. The natural mapping from G to O_s is surjective by definition since s is acted on by all of G. Define a mapping $\phi : G/G_s \to O_s$ by

$$gG_s \to g \cdot s$$

We need only show that this is injective to show that ϕ is a bijection. But this quickly follows from the fact that if $g, g' \in G$ and $gs = g's$ then $s = (g^{-1}g')s$. Then g and g' are in the same coset. It is clear that the number of elements in O_s is the number of left cosets of G_s in G. This is given by $|G|/|G_s|$. Then we have

$$O_s = \frac{|G|}{|G_s|}$$

The $g \in G$ act on s if and only if they lie in the same left coset of Z_g. But then this is exactly the idea that

$$|K_s| = \frac{|G|}{|Z_g|}$$

□

If we make the situation less general by letting $G = S_n$. If $g \in G$, $g \in S_n$, g is a permutation π with some cycle type λ. We can then write K_λ and obtain the following for permutations of cycle type λ

Theorem 5.2. *(Centralizer Order) If $\lambda = (1^{n_1}, 2^{n_2}, \cdots, m^{n_j})$ and $g \in S_n$ has cycle-type λ, then $|Z_g|$ depends only on λ and*

$$z_\lambda \stackrel{def}{=} |Z_g| = 1^{n_1} n_1! 2^{n_2} n_2! \cdots m^{n_j} n_j!$$

Proof: Given a $h \in Z_g$, h will either permute the cycles of length i amongst themselves, permute elements within the cycles, or both. Since there are m_i cycles of length i, there are $m_i!$ ways to permute them. Moreover, there are i^{m_i} ways to rotate the elements with a cycle of length i. To give the total number, we need only multiply and then we obtain the relation in the theorem statement. □

Applying Theorem 5.1 to the symmetric group, we have

$$k_\lambda = \frac{n!}{z_\lambda}$$

where $k_\lambda = |K_\lambda|$ and of course z_λ can be calculated as in Theorem 5.2. One particular aspect of permutations we will make use of is the sign of a permutation. Recall that we can always decompose a permutation into transpositions (permutations of the form $\pi = (i, j)$). Furthermore, the symmetric group S_n is generated by adjacent transpositions $(1\ 2), (3\ 4), \cdots, (n-1\ n)$. When we write a permutation π in terms of two-cycles, then the sign of the permutation is

$$\text{sgn}(\pi) = (-1)^k$$

where k is the number of 2-cycles in π when it is written as the product of two-cycles exclusively. Moreover, the sgn of a permutation is independent of the decomposition of π into two-cycles, i.e it is well-defined, even though the decomposition of π into two cycles need not be unique. It easily follows then that

$$\text{sgn}(\pi\sigma) = \text{sgn}(\pi)\text{sgn}(\sigma)$$

5.2. **Representations.** We now begin our goal of representing a group using matrices. We attempt to imbed our group into a matrix space and study the group operation there where it is equivalent to linear transformations on group elements. These questions are more well understood and will allow us some insight on the more complicated structure of the group and perhaps allow us to break the group into smaller parts. To do this we take a group and associate each element with a matrix in $GL_n(\mathbb{C})$, the set of $n \times n$ invertible matrices with entries in $\mathbb{C}$. Using this we are able to create an algebra. An algebra is a vector space with an associative multiplication of vectors and hence imposes a ring structure on the space.

Definition 5.3. (Representation)
A matrix representation of a group G is a group homomorphism

$$X : G \to GL_d$$

Equivalently, to each $g \in G$ is assigned a matrix $X(g) \in GL_d$ such that

1. $X(e) = I$ the identity matrix
2. $X(gh) = X(g)X(h)$ for all $g, h \in G$

The parameter d is called the degree, or dimension, of the representation, denoted $\deg X$.

One might wonder why we only require a homomorphism and not an isomorphism between the space. First, finding an isomorphism between a group G and a subspace of GL_n for some n is generally more difficult than finding a homomorphism or even finding subgroups of G! Second, we only need insight into the group structure. Requiring an isomorphism would be equivalent trying to study the group structure as a whole, the only difference being it is then in GL_n. Only then it would require more work to translate the whole group into GL_n. Requiring only the homomorphism allows us to eventually break the group into pieces and see a bit of its structure without having to understand it in its entirety. Furthermore, since we only see a homomorphism we can vary the dimension of the representation, d, and obtain different matrix representations which each may tells us something different about the group.

Example 5.5. All groups have a matrix representation as we can place, as usual, the trivial representation on the group. That is we map all of X to the identity in GL_d. This mapping is a homomorphism as

$$X(g)X(h) = \mathbf{1}_d\mathbf{1}_d = X(gh)$$

for all $g, h \in G$. We often use 1_G or 1 to represent the trivial representation of G. Moreover, here we wrote $[1]_d$ to denote that this is the identity matrix for a matrix in GL_d.

Example 5.6. An important 1 degree representation that is commonly seen for S_n is the sign representation, $X : G \to GL_1$. That is, given $\pi \in S_n$, then

$$X(\pi) = (\operatorname{sgn}(\pi))$$

In fact, the reader should have already seen a matrix representation for the symmetric group while studying permutations.

Example 5.7. Consider S_3, then we have the matrix representation $X : S_3 \to GL_3$ given by

$$X(e) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad X((12)) = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$X((13)) = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad X((23)) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

$$X((123)) = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \quad X((132)) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

Notice that the result of multiplying the permutations yields the same result as multiplying their corresponding matrices. In fact, it is always easy to construct a matrix representation for S_n. If $i \mapsto j$ under a permutation, simply go to column i and row j and place a 1 there. Do this for the entire permutation $\pi \in S_n$ and the remaining entries are 0. Moreover, notice each row and column has exactly one 1 in them. Finally, the det of the corresponding permutation matrix is either 1 or -1 depending on if the permutation is even or odd, respectively.

Example 5.8. The defining representation of S_n is also an important representation. If $\pi \in S_n$ then $X(\pi) = (x_{i,j})_{n\times n}$, where

$$x_{i,j} = \begin{cases} 1, & \text{if } \pi(j) = i \\ 0, & \text{otherwise} \end{cases}$$

This is called the defining representation and it has degree n.

Example 5.9. We are going to find all 1-dimensional representations X of the cyclic group of order n, C_n. That is, we are going to describe them all. Suppose g is the generator of C_n,

$$C_n = \{g, g^2, g^3, \cdots, g^n = e\}$$

Now suppose that $X(g) = (c)$, $c \in \mathbb{C}$, then the matrix for every element of C_n is determined by g since $X(g^k) = (c^k)$, then

$$(c^n) = X(g^n) = X(e) = (1)$$

Therefore, c is a root of unity. But then there are exactly n representations with degree 1. For example, take $n = 4$ and $C_4 = \{e, g, g^2, g^3\}$. The fourth roots of unity are $1, i, -1, -i$. Then we have the table

	e	g	g^2	g^3
$X^{(1)}$	1	1	1	1
$X^{(2)}$	1	i	-1	$-i$
$X^{(3)}$	1	-1	1	-1
$X^{(4)}$	1	$-i$	-1	i

Notice the trivial representation appearing in the first row of the table. We can create other representations of a cyclic group. For example, we can take the example

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

But notice this is the same as a combination of $X^{(1)}$ and $X^{(2)}$. So the above representation is reducible with irreducible components $X^{(1)}$ and $X^{(2)}$. In fact, every representation of C_n can be built using n representations of degree 1 as building blocks.

5.3. **G-Modules and Group Algebras.** Now that we have represented a group as a matrix, we can formalize the language we need to think of representations in terms of linear transformations. This is called a G-module. Because matrix multiplication corresponds to linear transformations this is fairly simple. Suppose that V is a complex vector space, we will always assume that V has finite dimension. $GL(V)$ is the set of all invertible linear transformations of V to itself, the general linear group of V. Note that if $\dim V = d$ then GL_d and $GL(V)$ are isomorphic groups.

Definition 5.4. (G-Module)
Let V be a vector space and G be a group. Then V is a G-module if there is a group homomorphism

$$\rho : G \to GL(V)$$

Equivalently, V is a G-module if there is a multiplication, $G\mathbf{v}$, of elements of V by elements of G such that

1. $g\mathbf{v} \in V$
2. $g(c\mathbf{v} + d\mathbf{w}) = c(g\mathbf{v}) + d(g\mathbf{w})$
3. $(gh)\mathbf{v} = g(h\mathbf{v})$
4. $e\mathbf{v} = \mathbf{v}$

for all $g, h \in G$, $\mathbf{v}, \mathbf{w} \in V$, and scalars $c, d \in \mathbb{C}$. Often G-module is shortened to just module when the group used is clear. Moreover, often one says that the space V carries a representation of G.

The equivalence of the definitions above is easy to see and the interested reader can check that they are. Moreover, one can go back and forth between the definitions. If one is given a representation X with degree d, let V be the vector space $\mathbb{C}^d$ of all possible columns of length d. Then using

$$g\mathbf{v} \stackrel{\text{def}}{=} X(g)\mathbf{v}$$

Then choose a basis β for V, where V is a G-module. Then $X(g)$ is a matrix (linear transformation) for $g \in G$ in terms of the basis β. We can multiply $\mathbf{v} \in V$ by $g \in G$ and the multiplication on the right is matrix multiplication. Moreover, we have emphasized the importance of group actions many times already and it doesn't lose any importance here. Suppose that G is a group and S is a nonempty finite set such that G acts on S. Then we can take S and turn it into a G-module. Let $S = \{s_1, s_2, \cdots, s_n\}$ and let $\mathbb{C}\mathbf{S} = \mathbb{C}\{\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_n\}$ denote the vector space generated by S over $\mathbb{C}$, i.e. it consists of all linear combinations of $\{s_1, s_2, \cdots, s_n\}$ with coefficients in $\mathbb{C}$. Addition and scalar multiplication in $\mathbb{C}\mathbf{S}$ are defined as they are in normal vector operations. Now we extend the action of G on S to an action on $\mathbb{C}\mathbf{S}$ by linearity

$$g(c_1\mathbf{s}_1 + c_2\mathbf{s}_2 + \cdots + c_n\mathbf{s}_n) = c_1(g\mathbf{s}_1) + c_2(g\mathbf{s}_2) + \cdots + c_n(g\mathbf{s}_n)$$

where $c_i \in \mathbb{C}$. Then $\mathbb{C}\mathbf{S}$ is a G-module with dimension $|S| = n$. This associated module,$\mathbb{C}\mathbf{S}$, is known as the permutation representation. The elements of $\mathbf{S}$ form a basis for $\mathbb{C}\mathbf{S}$ called the standard basis.

Definition 5.5. (Permutation Representation)
(Sagan) If a group G acts on a set S, then the associated module described above, $\mathbb{C}\mathbf{S}$, is called the permutation representation associated with S. Also, the elements of S form a basis for $\mathbb{C}\mathbf{S}$ called the standard basis.

Example 5.10. We can create an S_n-module from the set $S = \{1, 2, \cdots, n\}$ as follows,

$$\mathbb{C}\mathbf{S} = \{c_1\mathbf{1} + c_2\mathbf{2} + \cdots + c_n\mathbf{n} \,|\, c_i \in \mathbb{C} \text{ for all } i\}$$

with action

$$\pi(c_1\mathbf{1} + c_2\mathbf{2} + \cdots + c_n\mathbf{n}) = c_1\pi(\mathbf{1}) + c_2\pi(\mathbf{2}) + \cdots + c_n\pi(\mathbf{n})$$

for all $\pi \in S_n$. Of course, we can create these using permutation matrices as described in Example 5.7.

Example 5.11. Here is one of the most important representations for a group G, the (left) regular representation - the right is constructed analogously. Let G be a group, then G acts on itself through left multiplication, using the usual product in the group. Almost all of the representation properties follow from the group axioms of G. So if G is finite, i.e. $G = \{g_1, g_2, \cdots, g_n\}$ then the G-module

$$\mathbb{C}[\mathbf{G}] = \{c_1\mathbf{g}_1 + c_2\mathbf{g}_2 + \cdots + c_n\mathbf{g}_n \,|\, c_i \in \mathbb{C} \text{ for all } i\}$$

is called the group algebra of G. The square brackets indicate that this is an algebra and not simply a vector space. If $g_ig_j = g_k \in G$ then $\mathbf{g}_i\mathbf{g}_j = \mathbf{g}_k \in \mathbb{C}[\mathbf{G}]$. So the action of G on the group algebra is given by

$$g(c_1\mathbf{g}_1 + c_2\mathbf{g}_2 + \cdots + c_n\mathbf{g}_n) = c_1(\mathbf{g}\mathbf{g}_1) + c_2(\mathbf{g}\mathbf{g}_2) + \cdots + c_n(\mathbf{g}\mathbf{g}_n)$$

for all $g \in G$. Similarly, we can construct a coset representation analogously as well.

Definition 5.6. (Group Algebra) Let G be a finite group. Then the group algebra of G is

$$\mathbb{C}[\mathbf{G}] = \{c_1\mathbf{g}_1 + c_2\mathbf{g}_2 + \cdots + c_n\mathbf{g}_n \,|\, c_i \in \mathbb{C} \text{ for all } i\}$$

Example 5.12. We can easily find the regular representation for the cyclic group C_3. Let g be a generator of C_3, then

$$\mathbb{C}[\mathbf{C}_3] = \{c_1\mathbf{e} + c_2\mathbf{g} + c_3\mathbf{g}^2 \,|\, c_i \in \mathbb{C} \text{ for all } i\}$$

Then the matrices are easily found in the standard basis. For example,

$$g\mathbf{e} = \mathbf{g} \qquad g\mathbf{g} = \mathbf{g}^2 \qquad g\mathbf{g}^2 = \mathbf{e}$$

and then

$$X(g) = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

Remark 5.1. If G acts on a vector space V, then so does $\mathbb{C}[\mathbf{G}]$.

Notice that the left regular representation for a finite group G embeds G into the symmetric group on a set of $|G|$ elements. However, notice that we have shown this before in the Introduction and the reader know this as Cayley's Theorem.

5.4. **Reducibility of Representations.** When working with any complex mathematical object, the most fruitful method of attack is to break the object up into manageable pieces. Once these pieces are understood, we can take them and build larger representations. The reader should be familiar with this from the study of normal groups and the work we did earlier with quotient rings previously. With representations, this concept is reducibility.

Definition 5.7. (Submodule) Let V be a G-module. A submodule of V is a subspace W that is closed under the action of G, i.e.

$$\mathbf{w} \in W \to g\mathbf{w} \in W \text{ for all } g \in G$$

Equivalently, W is a subset of V that is a G-module under the same action. We say that W is a G-invariant subspace. If W is submodule of V, then we write $W \leq V$.

Example 5.13. Every G-module comes equipped with two submodules, $W = \{\mathbf{0}\}$, called the trivial submodule, and $W = V$.

Example 5.14. Take $G = S_n$ for $n \geq 2$ and $W = \mathbb{C}\{\mathbf{1} + \mathbf{2} + \cdots + \mathbf{n}\}$. Define W as

$$W = \mathbb{C}\{\mathbf{1} + \mathbf{2} + \cdots + \mathbf{n}\} = \{c(\mathbf{1} + \mathbf{2} + \cdots + \mathbf{n}) \,|\, c \in \mathbb{C}\}$$

Then W is a 1-dimensional submodule spanned by $\mathbf{1} + \mathbf{2} + \cdots + \mathbf{n}$.

We can now define what exactly it means to be able to break a representation into smaller pieces.

Definition 5.8. (Reducibility) A nonzero G-module V is reducible if it contains a non-trivial submodule W. Otherwise, V is said to be irreducible. Equivalently, V is reducible if it has a basis B in which every $g \in G$ is assigned a block matrix of the form

$$C = \left(\begin{array}{c|c} \text{A(g)} & \text{B(g)} \\ \hline 0 & \text{C(g)} \end{array}\right)$$

where the $A(g)$ are square matrices, all of the same size, and 0 is a nonempty matrix of zeros.

Remark 5.2. The equivalence of the two definitions for reducibility is not immediately clear. To see it, suppose that V has dimension d and has a non-zero proper submodule W with dimension f. Then let

$$\mathfrak{B} = \{\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_f, \mathbf{v}_{f+1}, \mathbf{v}_{f+2}, \cdots, \mathbf{v}_d\}$$

with the first f $\mathbf{w}$ vectors are a basis for W. Since W is a submodule of V, $g\mathbf{w}_i \in W$ for all i, $1 \leq i \leq f$. Then the last $d - f$ coordinates of $g\mathbf{w}_i$ will all be zero. This is why there is a 0 in the bottom left corner of the block matrix. Now note that we also know that $A(g)$ for all $g \in G$ are matrices of the restriction of G to W and hence must all be square and of the same size.

Now if $X(g)$ has the form given by the definition with every $A(g)$ being $f \times f$ in size. Then let $V = \mathbb{C}^d$ and consider

$$W = \mathbb{C}\{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_f\}$$

where $\mathbf{e}_i$ is the column vector with a 1 in the ith row and zeros everywhere else, i.e. the standard basis for $\mathbb{C}^d$. The zero placement guarantees that $X(g)\mathbf{e}_i \in W$ for $1 \leq i \leq f$ and all $g \in G$. Then W is a G-module. It must be nontrivial because the matrix of zeros is nonempty.

It is not easy to determine when a representation is reducible. Without doubt, the 1-dimensional representation is irreducible. However, given a representation of degree d, is it reducible? One could find all subspaces of a vector space V and check to see if they are also submodules would be computationally inefficient. In fact, with the tools we have now, we cannot answer this question. But later we will develop the tools needed to check the reducibility of a module. For now, we will extend the concept of reducibility.

Definition 5.9. (Complements) Let V be a vector space with subspaces U and W. Then V is the internal direct sum of U and W, $V = U \oplus V$, if every $\mathbf{v} \in V$ can be written uniquely as a sum

$$\mathbf{v} = \mathbf{u} + \mathbf{w} \quad \mathbf{u} \in U, \mathbf{w} \in W$$

If both are G-modules, say that U and W are complements of each other. However, if X is a matrix, then X is the direct sum of matrices A and B, written $X = A \oplus B$, if X has the block diagonal form

$$X = \left(\begin{array}{c|c} \mathrm{A} & 0 \\ \hline 0 & \mathrm{B} \end{array}\right)$$

Remark 5.3. Again, we shall show the equivalency of the definitions. Suppose V is a G-module with $V = U \oplus W$, where $U, W \leq V$. So given a basis $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_f\}$ and $\{\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_{d-f}\}$ are a basis for U and W, respectively, then we can construct a basis for V

$$\mathfrak{B} = \{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_f, \mathbf{w}_{f+1}, \mathbf{w}_2, \cdots, \mathbf{w}_d\}$$

Since U, W are submodules, then

$$g\mathbf{u}_i \in U \text{ and } g\mathbf{w}_j \in W$$

for all $g \in G$, $\mathbf{u}_i \in U$, and $\mathbf{w}_j \in W$. Then any matrix of $g \in G$ in the basis $\mathfrak{B}$ is

$$X = \left(\begin{array}{c|c} \mathrm{A} & 0 \\ \hline 0 & \mathrm{B} \end{array}\right)$$

where $A(g)$ and $B(g)$ are matrices of the action of G restricted to U and W, respectively.

However, we are left with the problem of finding these irreducible parts. This problem is not as simple as finding how to break apart the vector space. For example, consider the defining representation of S_3. It is clear that

$$V = \mathbb{C}\{\mathbf{1}, \mathbf{2}, \mathbf{3}\} = \mathbb{C}\{\mathbf{1} + \mathbf{2} + \mathbf{3}\} \oplus \mathbb{C}\{\mathbf{2}, \mathbf{3}\}$$

as vector spaces. But $\mathbb{C}\{\mathbf{1}+\mathbf{2}+\mathbf{3}\}$ is an S_3 submodule while $\mathbb{C}\{\mathbf{2}, \mathbf{3}\}$ is not an S_3 submodule. as $(12)\mathbf{2} = \mathbf{1} \notin \mathbb{C}\{\mathbf{2}, \mathbf{3}\}$. So in order to completely reduce S_3, we need to find a submodule U such that

$$\mathbb{C}\{\mathbf{1}, \mathbf{2}, \mathbf{3}\} = \mathbb{C}\{\mathbf{1} + \mathbf{2} + \mathbf{3}\} \oplus U$$

It is clear that U will be the complement of $\mathbb{C}\{\mathbf{1}+\mathbf{2}+\mathbf{3}\}$ in $\mathbb{C}\{\mathbf{1},\mathbf{2},\mathbf{3}\}$. To find such a complement, we will need the concept of the inner product.

Definition 5.10. (Inner Product) Suppose that $\mathbf{v}=(v_1,v_2,\cdots,v_n)$ and $\mathbf{w}=(w_1,w_2,\cdots,w_n)$. Then their inner product, $\langle\mathbf{v},\mathbf{w}\rangle$, is given by

$$\langle\mathbf{v},\mathbf{w}\rangle=\sum_{i=1}^{n} v_i\overline{w_i}$$

where $\bar{\cdot}$ is complex conjugation. Equivalently, define

$$\langle\mathbf{v},\mathbf{w}\rangle=\delta_{i,j}$$

where $\delta_{i,j}$ is the Kronecker delta and extend the linearity in the first variable and conjugate linearity in the second.

Both these definitions satisfy the axioms for an inner product, as the reader can easily check. Moreover, these inner products are invariant under the action of G, that is

$$\langle g\mathbf{v},g\mathbf{w}\rangle=\langle\mathbf{v},\mathbf{w}\rangle$$

for all $g\in G$ and $\mathbf{v},\mathbf{w}\in V$. To show that the inner product is invariant on V, one need only check these for the basis elements. But now given a submodule, we can find its orthogonal complement.

Definition 5.11. (Orthogonal Complement) Given a subspace W of a vector space V, the orthogonal complement of W, $W^{\perp}$ is

$$W^{\perp}=\{\mathbf{v}\in V\,|\,\langle\mathbf{v},\mathbf{w}\rangle=0 \text{ for all } \mathbf{w}\in W\}$$

Notice then that $V=W\oplus W^{\perp}$. This is not to say if we take a subspace that we are then always able to reduce the space to an inner product. If the orthogonal complement of W is $\{\mathbf{0}\}$, then this is trivial and we haven't reduced the space. This is certainly the case when the space is already reduced. Moreover, the orthogonal complement itself may be reducible! But this process allows us to slowly reduce the space - when possible. However, there are circumstances where we can extend this further, when the inner product is G-invariant and $W\leq V$.

Theorem 5.3. *Let V be a G-module, W a submodule, and $\langle\cdot,\cdot\rangle$ an inner product that is invariant under the action of G, then $W^{\perp}$ is also a G-submodule.*

Proof: Our goal is to show that $g\mathbf{u}\in W^{\perp}$ holds for all $g\in G$ and $\mathbf{u}\in W^{\perp}$. So suppose that $\mathbf{w}\in W$, then

$$\begin{aligned}\langle g\mathbf{u},\mathbf{w}\rangle=&\langle(g^{-1}g)\mathbf{u},g^{-1}\mathbf{w}\rangle\\=&\langle\mathbf{u},g^{-1}\mathbf{w}\rangle\\=&0\end{aligned}$$

But then $W^{\perp}$ is closed under the action of G. □

Example 5.15. Take the example we discussed earlier of $\mathbb{C}\{\mathbf{1}+\mathbf{2}+\mathbf{3}\}$ in $\mathbb{C}\{\mathbf{1},\mathbf{2},\mathbf{3}\}$. We find the orthogonal complement.

$$\begin{aligned}\mathbb{C}\{\mathbf{1}+\mathbf{2}+\mathbf{3}\}^{\perp} =& \{\mathbf{v}=a\mathbf{1}+b\mathbf{2}+c\mathbf{3} \,|\, \langle \mathbf{v}, \mathbf{1}+\mathbf{2}+\mathbf{3}\rangle = 0\} \\ =& \{\mathbf{v}=a\mathbf{1}+b\mathbf{2}+c\mathbf{3} \,|\, a+b+c=0\}\end{aligned}$$

Choose for $\mathbb{C}\{\mathbf{1}+\mathbf{2}+\mathbf{3}\}$ the obvious basis $\{\mathbf{1}+\mathbf{2}+\mathbf{3}\}$ and for $\mathbb{C}\{\mathbf{1}+\mathbf{2}+\mathbf{3}\}^{\perp}$ the basis $\{\mathbf{2}-\mathbf{1}, \mathbf{3}-\mathbf{1}\}$.

Then simple calculation yields the matrices

$$X(e)=\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix} \qquad X((12))=\begin{pmatrix}1&0&0\\0&-1&-1\\0&0&1\end{pmatrix}$$
$$X((13))=\begin{pmatrix}1&0&0\\0&1&0\\0&-1&-1\end{pmatrix} \qquad X((23))=\begin{pmatrix}1&0&0\\0&0&1\\0&1&0\end{pmatrix}$$
$$X((123))=\begin{pmatrix}1&0&1\\0&-1&-1\\0&1&0\end{pmatrix} \qquad X((132))=\begin{pmatrix}1&0&0\\0&0&1\\0&-1&-1\end{pmatrix}$$

Notice then that all the matrices we have calculated are direct sums of matrices of the form

$$X(g)=\left(\begin{array}{c|cc} A(g) & 0 & 0 \\ \hline 0 & & \\ & & \\ 0 & & B(g) \end{array}\right)$$

One immediately notices that $A(g)$ must be irreducible as it has degree 1. So we have reduced the defining representation of S_3 into irreducible parts.

First, notice that we don't reduce S_3 but rather we reduce a particular representation of S_3, in our example the defining representation. Second, notice nothing in this method is case specific to S_n. So can this method be applied to any group? The answer is yes, so long as the group is finite. This is the idea of Maschke's Theorem.

Theorem 5.4. *(Maschke's Theorem) Let G be a finite group and let V be a nonzero G-module, then*

$$V = W^{(1)} \oplus W^{(2)} \oplus \cdots \oplus W^{(k)}$$

where each $W^{(i)}$ is an irreducible G-submodule of V. This can also be equivalently stated as every representation of a finite group having positive dimension is completely reducible.

Proof: We prove this by induction on $\dim V$. If $\dim V = 1$, it is immediate that V is irreducible and we are done (as $k=1$ and then $W^{(1)}=V$). Now assume this is true for all V' with $1 < \dim V' < \dim V' + 1 = \dim V$. Now if V is irreducible, we have again that $k=1$ and $W^{(1)}=V$. Suppose then V is reducible, then there is a nontrivial G-submodule, say W. It remains to show that there is a submodule complement for W.

Let $d = \dim V$ and $\mathfrak{B} = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_d\}$ be a basis for V. Then the inner product

$$\langle \mathbf{v}_i, \mathbf{v}_j \rangle = \delta_{i,j}$$

on all the elements of $\mathfrak{B}$. This inner product may not be G-invariant. If it is not, we will "fix" it so that it is. Define a new inner product, if necessary, $\langle \cdot, \cdot \rangle$, defined by

$$\langle \mathbf{v}, \mathbf{w} \rangle' = \sum_{g \in G} \langle g\mathbf{v}, g\mathbf{w} \rangle$$

for all $\mathbf{v}, \mathbf{w} \in V$. It is simple to check that $\langle \cdot, \cdot \rangle'$ satisfies the axioms of an inner product (it essentially inherits these properties from $\langle \cdot, \cdot \rangle$). Now we show that $\langle \cdot, \cdot \rangle'$ is G-invariant, which means we need to show that

$$\langle g\mathbf{v}, g\mathbf{w} \rangle' = \langle \mathbf{v}, \mathbf{w} \rangle'$$

for all $g, h \in G$ and $\mathbf{v}, \mathbf{w} \in V$. However,

$$\begin{aligned} \langle g\mathbf{v}, g\mathbf{w} \rangle' &= \sum_{g \in G} \langle gh\mathbf{v}, gh\mathbf{w} \\ &= \sum_{f \in G} \langle \mathbf{fv}, \mathbf{fw} \rangle' \rangle' \\ &= \langle \mathbf{v}, \mathbf{w} \rangle' \end{aligned}$$

where $\mathbf{f} = \mathbf{gh}$. This shows that $\langle \cdot, \cdot \rangle'$ is G-invariant. Now let

$$W^{\perp} = \{ \mathbf{v} \in V \mid \langle \mathbf{v}, \mathbf{w} \rangle' = 0 \}$$

Then we have $W^{\perp}$ is a G-submodule of V with

$$V = W \oplus W^{\perp}$$

Then simply applying induction to W and $W^{\perp}$, writing each as a direct sum of irreducibles, and putting the decompositions together, we have V in the form of the theorem. □

Corollary 5.1. *(Masche's Diagonal) Let G be a finite group and let X be a matrix representation of G of dimension $d > 0$. Then there is a fixed matrix T such that every matrix $X(g)$, $g \in G$, has the form*

$$TX(g)T^{-1} = \begin{pmatrix} X^{(1)}(g) & 0 & \cdots & 0 \\ 0 & X^{(2)}(g) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X^{(k)}(g) \end{pmatrix}$$

where each $X^{(i)}$ is an irreducible matrix representation of G.

Proof: Let $V = \mathbb{C}^d$ with action

$$g\mathbf{v} = X(g)\mathbf{v}$$

for all $g \in G$ and $\mathbf{v} \in V$. Then using Maschke's Theorem,

$$V = W^{(1)} \oplus W^{(2)} \oplus \cdots \oplus W^{(k)}$$

with each $W^{(i)}$ being irreducible with dimension d_i. Take some basis $\mathfrak{B}$ for V such that the first d_1 vectors are a basis for $W^{(1)}$, the next d_2 are a basis for $W^{(2)}$, and so forth.

Then our matrix T transforms the standard basis for $\mathbb{C}^d$ into $\mathfrak{B}$ will work, as conjugating T expresses each $X(g)$ in the new basis $\mathfrak{B}$. $\square$

Definition 5.12. (Complete Reducibility) A representation is completely reducible if it can be written as a direct sum of irreducibles.

Of course then we are able to state Mascke's Theorem as, "every representation of a finite group having positive dimension is completely reducible."

5.5. G**-Homomorphisms.** A key tool in Mathematics in studying objects is studying functions which preserve their structures. For G-modules, the corresponding function is called a G-homomorphism.

Definition 5.13. (G-Homomorphism) Let V and W be G-modules. Then a G-homomorphism is a linear transformation $\theta : V \to W$ such that

$$\theta(g\mathbf{v}) = g\theta(\mathbf{v})$$

for all $g \in G$ and $\mathbf{v} \in V$. We say that θ preserves or respects the action of G.

Example 5.16. Suppose that $G = S_n$ and let $V = \mathbb{C}\{\mathbf{v}\}$ with the trivial action of S_n. Finally, let $W = \mathbb{C}\{\mathbf{1}, \mathbf{2}, \cdots, \mathbf{n}\}$ with the defining action of S_n. Now define a transformation $\theta : V \to W$ by

$$\theta(c\mathbf{v}) = c(\mathbf{1} + \mathbf{2} + \cdots + \mathbf{n})$$

for all $c \in \mathbb{C}$. The last two equalities follow because π is a permutation. To check that θ is a G-homomorphism, it suffices to check the action of G is preserved on a basis of V. For all $\pi \in S_n$,

$$\theta(\pi\mathbf{v}) = \theta(\mathbf{v}) = \sum_{\mathbf{i}=1}^{n} \mathbf{i} = \pi \sum_{\mathbf{i}=1}^{n} \mathbf{i} = \pi\theta(\mathbf{v})$$

Example 5.17. Similar to the previous example, let G be an arbitrary group acting on $V = \mathbb{C}\{\mathbf{v}\}$ and let $W = \mathbb{C}[\mathbf{G}]$ be the group algebra. We can create a G-homomorphism $\theta : V \to W$ given by extending

$$\theta(\mathbf{v}) = \sum_{g \in G} \mathbf{g}$$

linearly.

Definition 5.14. (G-Equivalency) Let V and W be G-modules. A G-isomorphism is a G-homomorphism $\theta : V \to W$ that is bijective. Then we say that V and W are G-isomorphic, or G-equivalent. This is written $V \cong W$. If V and W are not G-equivalent, we say that V and W are G-inequivalent.

Of course with a homomorphism, we immediately have two subspaces that are of particular interest: the kernel and image under a homomorphism θ. However, when θ is a G-homomorphism, the kernel and image have the standard properties:

Theorem 5.5. *Let $\theta : V \to W$ be a G-homomorphusm. Then*

1. *ker θ is a G-submodule of V.*
2. *im θ is a G-submodule of W.*

Both of these properties are easily shown; in fact, they should be expected to be so! But we are now able to characterize G-homomorphisms for irreducible modules.

Theorem 5.6. *(Schur's Lemma) Let V and W be two irreducible G-modules. If $\theta : V \to W$ is a G-homomorphism, then one of the following*

1. *θ is a G-isomorphism.*
2. *θ is the zero map.*

Proof: Since V is irreducible and $\ker\theta$ is a submodule, it must be the case that $\ker\theta = \{\mathbf{0}\}$ or $\ker\theta = V$. The irreducibility of W equally implies that $\text{im}\,\theta = \{\mathbf{0}\}$ or $\text{im}\,\theta = W$. If $\ker\theta = V$ or $\text{im}\,\theta = \{\mathbf{0}\}$ then θ must be the zero map. But if $\ker\theta = \{\mathbf{0}\}$ or $\text{im}\,\theta = W$, θ must be an isomorphism. □

Corollary 5.2. *(Schur's Matrix Lemma) Let X and Y be two irreducible matrix representations of G. If T is any matrix such that $TX(g) = Y(g)T$ for all $g \in G$, then one of the following*

1. *T is invertible.*
2. *T is the zero matrix.*

and where the range module is not irreducible,

Corollary 5.3. *Let V and W be two G-modules with V being irreducible. Then dim $Hom_G(V, W) = 0$ if and only if W contains no submodule isomorphic to V.*

From Schur's lemma, we know that for G-homomorphisms between irreducible G-modules must either be trivial or a G-isomorphism. Why? If φ is a G-homomorphism between two irreducible G-modules which are moreover matrix representations, then if φ is an isomorphism, the mapping must be a linear mapping which is invertible, or if φ is the zero map then the linear matrix representing the transformation must be the zero matrix. Also, Corollary 5.3 follows from Corollary 5.2 and Theorem 5.2. These corollaries give us more than we would expect! If the field is $\mathbb{C}$ and T is a matrix such that

$$TX(g) = X(g)T$$

for all $g \in G$, clearly

$$(T - cI_n)X = X(T - cI_n)$$

where I_n is the appropriately sized identity matrix and $c \in \mathbb{C}$ is a scalar. But because $\mathbb{C}$ is algebraically closed, take c to be an eigenvalue of T. But then $T - cI_n$ satisfies the corollary with $X = Y$. But $T - cI_n$ is not invertible by the choice of c. Hence, it must be the case that $T - cI_n = 0_n$, where 0_n is the appropriately sized zero matrix. But then we have $T = cI_n$.

Corollary 5.4. *Let X be an irreducible matrix representation of G over the complex numbers. Then the only matrices T that commute with $X(g)$ for all $g \in G$ are those of the form $T = cI_n$, scalar multiplies of the identity.*

5.6. **Special Algebras.** In the previous section, we placed a special emphasis on those matrices with commute with the representation. These matrices correspond in some way to the set of G-homomorphisms from a G-module to itself. Here we shall extend these ideas. These concepts will play a critical role in constructing the irreducible representations for the symmetric group later.

Definition 5.15. (Commutant Algebra) Given a matrix representation $X : G \to GL_d$, the corresponding commutant algebra is

$$\text{Com } X = \{T \in \text{Mat}_d \,|\, TX(g) = X(g)T \text{ for all } g \in G\}$$

where Mat_d is the set of $d \times d$ matrices with entries in $\mathbb{C}$.

Definition 5.16. (Endomorphism Algebra) Given a G-module V, the corresponding endomorphism algebra is

$$\text{End } V = \{\theta : V \to V \,|\, \theta \text{ is a } G\text{-homomorphism}\}$$

Example 5.18. Suppose we have a matrix representation X, where

$$X = \begin{pmatrix} X^{(1)} & 0 \\ 0 & X^{(2)} \end{pmatrix} = X^{(1)} \oplus X^{(2)}$$

where $X^{(1)}$ and $X^{(2)}$ are inequivalent and irreducible of degree d_1, d_2, respectively. Then one can check that the commutant algebra is

$$\text{Com } X = \{c_1 I_{d_1} \oplus c_2 I_{d_2} \,|\, c_1, c_2 \in \mathbb{C}\}$$

where $d_1 = \deg X^{(1)}$ and $d_2 = \deg X^{(2)}$.

In fact, in general, if $X = \oplus_{i=1}^{k} X^{(i)}$, where the $X^{(i)}$ are pairwise inequivalent irreducibles, then it is the case that

$$\text{Com } X = \{\oplus_{i=1}^{k} c_i I_{d_i} \,|\, c_i \in \mathbb{C}\}$$

where again $d_i = \deg X^{(i)}$. This follows from the corollaries to Schur's lemma. Of course, the degree of X is $\sum_{i=1}^{k} d_i$ and the dimension of Com X is k. To help simplify the notation, we write

$$mX \stackrel{\text{def}}{=} X \oplus X \oplus \cdots \oplus X$$

where m is a nonnegative integer and the direct sum on the right is taken m times. The integer m is called the multiplicity of X. But of course then that gives us the general case that if

$$X = m_1 X^{(1)} \oplus m_2 X^{(2)} \oplus \cdots \oplus m_k X^{(k)}$$

again where the $X^{(i)}$ are pairwise inequivalent irreducible with deg $X^{(i)} = d_i$, then the degree of X is given by

$$\deg X = \sum_{i=1}^{k} \deg(m_i X^{(i)})$$

We shall not say much about the general properties of the commutant and endomorphism algebras as they will only play a structural role in what will come. However, we shall summarize their general properties.

Theorem 5.7. *Let X be a matrix representation of G such that*

$$X = m_1X^{(1)} \oplus m_2X^{(2)} \oplus \cdots \oplus m_kX^{(k)}$$

where the $X^{(i)}$ are pairwise inequivalent and irreducible with deg $X^{(i)} = d_i$. Then we have

1. *deg $X = \sum_{i=1}^{k} deg(m_iX^{(i)})$*
2. *Com $X = \{\oplus_{i=1}^{k}(M_{m_i} \otimes I_{d_i}) \mid M_m \in Mat_{m_i}$ for all $i\}$*
3. *$dim(Com\ X) = m_1^2 + m_2^2 + \cdots + m_k^2$*
4. *$Z_{Com\ X} = \{\oplus_{i=1}^{k} c_i I_{m_i d_i} \mid c_i \in \mathbb{C}$ for all $i\}$*
5. *dim $Z_{Com\ X} = k$*

where $Z_{\text{Com } X}$ is the center of Com X. An observant reader should have noticed the careful assumption that X is already decomposed into a list of irreducibles of some multiplicity. What, if anything, would change if some of the $X^{(i)}$ representations were reducible? Using Maschke's Theorem (specifically the matrix version), if we had any such module X with one of its parts reducible, then it has to be equivalent to some reduced representation that is of the form from the theorem statement. Moreover, saying that if X is a matrix representation not broken up into irreducibles and $X = MYM^{-1}$ for some fixed matrix M, then the map

$$T \mapsto MTM^{-1}$$

is an algebra isomorphism from Com Y to Com X. But if the commutant algebras are isomorphic, then their centers too must be isomorphic. Then the above theorem holds with the realization that the equalities are not true equalities but rather are isomorphisms. As one would expect, there is a module version as well.

Theorem 5.8. *Let V be a G-module, such that*

$$V \cong m_1V^{(1)} \oplus m_2V^{(2)} \oplus \cdots \oplus m_kV^{(k)}$$

where the $V^{(i)}$ are pairwise inequivalent irreducibles and dim $V^{(i)} = d_i$. Then we have

1. *deg $V = \sum_{i=1}^{k} deg(m_iV^{(i)})$*
2. *End $V \cong \oplus_{i=1}^{k} Mat_{m_i}$*
3. *$dim(End\ V) = m_1^2 + m_2^2 + \cdots + m_k^2$*
4. *$Z_{End\ V}$ is isomorphic to the algebra of diagonal matrices of degree k*
5. *dim $Z_{End\ V} = k$*

Of course then we have

Proposition 5.1. *Let V and W be G-modules with V irreducible. Then dim $Hom(V, W)$ is the multiplicity of V in W.*

5.7. **Group Characters.** Our goal has been to learn about a group through representations. To understand a representation, it is easiest to break the representation into smaller, more manageable parts, i.e. the irreducible components of a representation. To understand these irreducible parts is to understand the entire representation. As it turns out, much of the information about a representation is encoded in the trace of the corresponding matrices.

Definition 5.17. (Group Characters) Let $X(g)$, $g \in G$, be a matrix representation. Then the character of X is

$$\chi(g) = \text{tr } X(g)$$

where tr denotes the trace of a matrix. Stated differently, χ is the map

$$G \xrightarrow{\text{tr } X} \mathbb{C}$$

If V is a G-module, then its character is the character of a matrix representation X corresponding to V.

Of course, since there are many representations for a G-module, one first needs to check that the module character is well-defined. However, this is a simple matter. If X and Y both correspond to a vector space V, then $Y = TXT^{-1}$ for some fixed T. But then for all $g \in G$,

$$\text{tr } Y(g) = \text{tr } TX(g)T^{-1} = \text{tr } X(g)$$

since the trace of a matrix is invariant under conjugation. But then X and Y have the same character and the definition is well-defined. The language we have created for modules will remain the same for characters. For example, if X has character χ, then χ is said to be irreducible whenever X is irreducible.

Example 5.19. Let G be an arbitrary group and X a representation of degree 1. Then the character $\chi(g)$ is the sole entry of $X(g)$ for each $g \in G$. These characters are called linear characters.

Example 5.20. Consider the defining representation of S_n with its character χ. If $n = 3$, then the character values can be computed by taking the traces of the matrices from Example 5.7.

$$\begin{array}{lll} \chi(\,e\,) = 3 & \chi(\,(12)\,) = 1 & \chi(\,(13)\,) = 1 \\ \chi(\,(23)\,) = 1 & \chi(\,(123)\,) = 0 & \chi(\,(132)\,) = 0 \end{array}$$

From the previous example, it is clear that for a permutation $\pi \in S_n$, that $\chi(\pi)$ is the number of 1s along the diagonal of $X(\pi)$. But this is precisely the number of fixed points of π.

Proposition 5.2. *(Group Characters) Let X be a matrix representation of a group G of degree d with character χ*

1. $\chi(e) = d$
2. *If K is a conjugacy class of G, then*

$$g, h \in K \rightarrow \chi(g) = \chi(h)$$

3. *If Y is a representation of G with character ψ, then*

$$X \cong Y \rightarrow \chi(g) = \psi(g)$$

for all $g \in G$

Proof:

1. Since $X(e) = I_d$, we have
$$\chi(e) = \text{tr } I_d = d$$
2. By hypothesis, $g = khk^{-1}$. Therefore,
$$\chi(g) = \text{tr } X(g) = \text{tr } X(k)X(h)X(k)^{-1} = \text{tr } X(h) = \chi(h)$$
3. This follows from the fact that the group character is well-defined. □

In fact, in the preceding proposition, the converse of number three is also true.

Definition 5.18. (Class Function) A class function on a group G is a mapping $f : G \to \mathbb{C}$ such that $f(g) = f(h)$, whenever g and h are in the same conjugacy class. The set of all class functions on G is denoted by $R(G)$.

In fact, since sums and scalar multiples of class functions are again class functions, $R(G)$ is a vector space over $\mathbb{C}$. Moreover, $R(G)$ has a natural basis consisting of all functions with value 1 on a given conjugacy class and 0 elsewhere. Then

$$\dim R(G) = \text{ the number of conjugacy classes of } G$$

But then if K is a conjugacy class and χ is a character then since χ is a class function, we can define χ_k to be the value of the given character on the given class:

$$\chi_k = \chi(g)$$

for any $g \in K$.

Definition 5.19. (Character Table) Let G be a group. The character table of G is an array with rows indexed by the inequivalent irreducible characters of G and columns indexed by the conjugacy classes. The table entry in row χ and column K is χ_k:

	$\cdots$	K	$\cdots$
$\vdots$		$\vdots$	
χ	$\cdots$	χ_k	
$\vdots$			

By convention, the first row corresponds to the trivial character and the first column corresponds to the class of the identity, $K = \{e\}$.

As it turns out, we can always write a character table as the character table is always finite. But why? There may be an infinite number of irreducible characters of G! As it turns out, the number of inequivalent irreducible representations of G is equal to the number of conjugacy classes, so the character table is always finite. Furthermore, the character table is always square.

5.8. **Young Subgroups and Tableaux.** Here we will finally construct all irreducible representations of the symmetric group. The number of such representations must be the number of conjugacy classes, which for S_n is the number of partitions of n. We will need to find how to associate an irreducible submodule with each partition λ. Then we produce the right number of representations by inducing the trivial representation on each S_λ up to S_n. Then if M^λ is a module for $1\uparrow_{S_\lambda}^{S_n}$, we cannot expect that these modules will be irreducible. But we find an ordering $\lambda^{(1)}, \lambda^{(2)}, \cdots$ of all partitions of n with a nice property. The first module $M^{\lambda^{(1)}}$ will be irreducible, call it $S^{\lambda^{(1)}}$. The next one, $M^{\lambda^{(2)}}$ will contain only copies of $S^{\lambda^{(1)}}$ plus a single copy of a new irreducible $S^{\lambda^{(2)}}$. In general, $M^{\lambda^{(l)}}$ will decompose into $S^{\lambda^{(k)}}$ for $k < i$ and a unique new irreducible $S^{\lambda^{(i)}}$ called the ith Specht module. Then the matrix giving the the multiplicities for expressing $M^{\lambda^{(i)}}$ as a direct sum of the $S^{\lambda^{(i)}}$ will be lower triangular with ones down the diagonal. It is then easy to compute the irreducible characters of S_n. The first thing we will need to do is build the modules, M^λ. But first it is necessary to introduce the necessary language.

Definition 5.20. (Parition) If $\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_l)$ is a partition of n, we write $\lambda \vdash n$. Moreover, the notation $|\lambda| = \sum_i \lambda_i$ so that $|\lambda| = n$.

Definition 5.21. (Ferrers Diagram) Suppose that $\lambda = (\lambda_1, \cdots, \lambda_l) \vdash n$. The Ferrers diagram, or shape, of λ is an array of n dots having l left-justified rows with row i containing λ_i dots for $1 \leq i \leq l$.

Example 5.21. Take the partition $\lambda = (5, 5, 4, 2, 1)$. Then Ferrers diagrams for λ are

. . . . .
. . . . .
. . . .
. .
.

or of course

–	–	–	–	–
–	–	–	–	–
–	–	χ	–	
–	–			
–				

where the χ is in the $(3, 3)$ position.

However, this is the "English" notation. Some will represent their Ferrers diagram inverted vertically to what we have here. Now we can associate with λ a subgroup of S_n.

Definition 5.22. (Young Subgroup) Let $\lambda = (\lambda_1, \cdots, \lambda_l) \vdash n$. The corresponding Young subgroup of S_n is

$$S_\lambda = S_{\{1,2,\cdots,\lambda_l\}} \times S_{\{\lambda_1+1,\lambda_1+2,\cdots,\lambda_1+\lambda_2\}} \times \cdots \times S_{\{n-\lambda_l+1,n-\lambda_l+2,\cdots,n\}}$$

Example 5.22. Consider the partition $\lambda = (3,3,2,1)$ in S_9. Then

$$\begin{aligned} S_{(3,3,2,1)} =& S_{\{1,2,3\}} \times S_{\{4,5,6\}} \times S_{\{7,8\}} \times S_{\{9\}} \\ \cong& S_3 \times S_3 \times S_2 \times S_1 \end{aligned}$$

Definition 5.23. (Young Tableaux) Suppose that $\lambda \vdash n$. A Young tableau of shape λ is an array t obtained by replacing the dots of the Ferrers diagram of λ with the numbers $1, 2, \cdots, n$, bijectively. Let $t_{i,j}$ stand for the entry of t in position (i,j). A Young tableaux of shape λ is also called a λ-tableau and denoted by t^λ. We will also write sh $t = \lambda$. There must be $n!$ shapes.

Moreover, what we will ultimately be interested in is equivalence classes of tableaux.

Definition 5.24. (λ-tabloid) Two λ-tableaux t_1 and t_2 are said to be row equivalent, $t_1 \sim t_2$, if the corresponding rows of the two tableaux contain the same elements. A tabloid of shape λ, or λ-tabloid, is then

$$\{t\} = \{t_1 \,|\, t_1 \sim t\}$$

where sh $t = \lambda$.

Of course we already know that if $\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_l) \vdash n$, then the number of tableaux in any given equivalence class is $\lambda_1!\lambda_2! \cdots \lambda_l! \stackrel{\text{def}}{=} \lambda!$ making the total number of λ-tabloids $\frac{n!}{\lambda!}$. But how exactly do permutations act on a tableaux? If πS_n acts on a tableau $t = (t_{i,j})$ of shape $\lambda \vdash n$ as follows:

$$\pi t = (\ \pi(t_{i,j})\ )$$

That is, π acts on the elements of the rows and columns of a λ-tabloid. This induces an action on tabloids by defining

$$\pi\{t\} = \{\pi t\}$$

Although it is not immediately evident, this definition is well defined. With this definition of a tabloid action, we give rise to an S_n-module.

Definition 5.25. (Permutation Module) Suppose that $\lambda \vdash n$. Let

$$M^\lambda = \mathbb{C}\{\ \{t_1\}, \{t_2\}, \cdots, \{t_k\}\ \}$$

where $\{t_1\}, \cdots, \{t_k\}$ is a complete list of λ-tabloids. Then M^λ is called the permutation module corresponding to λ.

Example 5.23. If $\lambda = (n)$, then

$$M^{(n)} = \mathbb{C}\{\overline{\mathbf{1\ 2\ \cdots\ n}}\}$$

with the trivial action, where $\overline{\mathbf{1\ 2\ \cdots\ n}}$ is a tabloid.

Example 5.24. Consider $\lambda = (1^n)$. Each equivalence class $\{t\}$ consists of a single tableau and this tableau can be then identified with a permutation in one-line notation. Because the action of S_n is preserved,

$$M^{(1^n)} \cong \mathbb{C}S_n$$

and the regular representation presents itself.

Example 5.25. If $\lambda = (n-1\ \ 1)$ then each λ-tabloid is uniquely determined by the element in its second row, which is a number from 1 to n.

$$M^{(n-1\ \ 1)} \cong \mathbb{C}\{\mathbf{1}, \mathbf{2}, \cdots, \mathbf{n}\}$$

which is the defining representation.

Moreover, M^λ enjoy the general properties of modules.

Definition 5.26. Any G-module M is cyclic if there is a $\mathbf{v} \in M$ such that

$$M = \mathbb{C}G\mathbf{v}$$

where $G\mathbf{v} = \{g\mathbf{v} \mid g \in G\}$. We say that M is generated by $\mathbf{v}$.

Proposition 5.3. *If $\lambda \vdash n$ then M^λ is cyclic and is generated by any given λ-tabloid. In addition,* $\dim M^\lambda = \frac{n!}{\lambda!}$*, the number of λ-tabloids.*

Theorem 5.9. *Consider $\lambda \vdash n$ with Young subgroup S_λ and tabloid $\{t^\lambda\}$. Then $V^\lambda = \mathbb{C}S_n\mathbf{S}_\lambda$ and $M^\lambda = \mathbb{C}S_n\{t^\lambda\}$ are isomorphic as S_n modules.*

Proof: Let $\pi_1, \pi_2, \cdots, \pi_k$ be transversal for S_λ. Define a map

$$\theta : V^\lambda \to M^\lambda$$

by $\theta(\pi_i\mathbf{S}_\lambda) = \{\pi_i\mathbf{t}^\lambda\}$ for $i = 1, 2, \cdots, k$ and linear extension. It is then a simple matter of calculation to show that θ is the desired S_n-isomorphism of modules. □

5.9. **Lexicographic Ordering.** Now we need to find an ordering of partitions of λ such that the M^λ have the nice property we discussed at the start of our discussion of Young Tableaux. We will consider two possible important orderings on partitions n, one of which will be a partial order.

Definition 5.27. (Partial Ordering) If S is a set, then a partial order on S is a relation $\leq$ such that

1. $a \leq a$
2. $a \leq b$ and $b \leq a$ implies that $a = b$
3. $a \leq b$ and $b \leq c$ implies that $a \leq c$

for all $a, b, c \in S$. We write $(A, \leq)$ and say that A is a partial ordered set, or poset. If we have either $a \leq b$ or $b \leq a$, then $\leq$ is a total order and $(A, \leq)$ is a totally ordered set.

Pairs of elements, say $a, b \in A$, such that neither $a \leq b$ and $b \leq a$ hold are called incomparable. The set $\{0, 1, 2, \cdots, n\}$ with the normal ordering is a totally ordered set called the n-chain and is denoted by C_n. The particular partial order which will be of great interest to us will be the following:

Definition 5.28. (Domination) Suppose that $\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_l)$ and $\mu = (\mu_1, \mu_2, \cdots, \mu_m)$ are partitions of n. Then λ dominates μ, written $\lambda \trianglerighteq \mu$ if

$$\sum_{k=1}^{i} \lambda_i \geq \sum_{k=1}^{i} \mu_i$$

for all $i \geq 1$. If $i > l, m$, then we take λ_i to be zero.

In a visual sense, λ is great than μ in the domination order if the Ferrers diagram of λ is short and fat while the one for μ is long and skinny. This concept is encoded in the Hasse diagram for partitions.

Definition 5.29. (Hasse Diagram) If $(A, \leq)$ is a poset and $b, c \in A$, then we say that b is covered by c (or c covers b), written $b \prec c$ (or $c \succ b$), if $b < c$ and there is no $d \in A$ with $b < d < c$. The Hasse diagram of A consists of vertices representing the elements of A with an arrow from vertex b up to vertex c if b is covered by c.

Hasse diagrams greatly resemble a lattice of subgroups or lattice of subfields.

Lemma 5.2. *(Dominance Lemma for Partitions) Let t^λ and s^μ be tableaux of shape λ and μ, respectively. If for each index i the elements of row i of s^μ are all in different columns in t^λ, then $\lambda \trianglerighteq \mu$.*

Proof: By hypothesis, we can sort the entries in each column of t^λ so that the elements of rows 1, 2, $\cdots$, i of s^μ all occur in the first i rows of t^λ. Thus,

$$\begin{aligned}\lambda_1 + \lambda_2 + \cdots + \lambda_i &= \text{the number of elements in the first } i \text{ rows of} t^\lambda \\ &\geq \text{number of elements of } s^\mu \text{ in the first } i \text{ rows of } t^\lambda \\ &= \mu_1 + \mu_2 + \cdots + \mu_i\end{aligned}$$

□

Definition 5.30. (Lexicographic Order) Let $\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_l)$ and $\mu = (\mu_1, \mu_2, \cdots, \mu_m)$ be partitions of n. Then $\lambda < \mu$ in the lexicographic order if for some index i,

$$\lambda_j = \mu_j \text{ for } j < i \text{ and } \lambda_i < \mu_i$$

This is a total ordering on partitions. The lexicographic ordering is then a refinement of the dominance order in the following sense:

Proposition 5.4. *If $\lambda, \mu \vdash n$ with $\lambda \trianglerighteq \mu$, then $\lambda \geq \mu$.*

Proof: If $\lambda \neq \mu$, then find the first index i where they differ. Then $\sum_{j=1}^{i-1} \lambda_j = \sum_{j=1}^{i-1} \mu_j$ and $\sum_{j=1}^{i} \lambda_j > \sum_{j=1}^{i} \mu_j$, since $\lambda \trianglerighteq \mu$, $\lambda_i > \mu_i$. □

5.10. **Specht Modules.** We can now construct all irreducible modules of S_n. These are the Specht modules, S^λ.

Definition 5.31. Suppose that the tableau t has rows $R_1, R_2, \cdots, R_l$ and columns $C_1, C_2, \cdots, C_k$, then

$$R_t = S_{R_1} \times S_{R_2} \times \cdots \times S_{R_l}$$

and

$$C_t = S_{C_1} \times S_{C_2} \times \cdots \times S_{C_k}$$

are the row-stabilizer and column-stabilizer of t, respectively.

Example 5.26. Suppose we have

$$t = \begin{array}{ccc} 4 & 1 & 2 \\ 3 & 5 & \end{array}$$

then we have

$$R_t = S_{\{1,2,4\}} \times S_{\{3,5\}}$$

and

$$C_t = S_{\{3,4\}} \times S_{\{1,5\}} \times S_{\{2\}}$$

Note in the previous example, the equivalence classes on the tabloids can be expressed as $\{t\} = R_t t$, this is true in general. Moreover, these groups are associated with certain elements of $\mathbb{C}[S_n]$. In general, given a subset $H \subseteq S_n$, we can form a group algebra sum

$$H^+ = \sum_{\pi \in H} \pi$$

and

$$H^- = \sum_{\pi \in H} \operatorname{sgn}(\pi)\pi$$

The group algebra is acting on the permutation modules M^λ. Now if t has columns $C_1, C_2, \cdots, C_k$, the the κ_t factors as

$$\kappa_t = \kappa_{C_1}\kappa_{C_2}\cdots\kappa_{C_k}$$

Now we will be able to pass from t to an element of the module M^λ using the following definition.

Definition 5.32. (Polytabloid) If t is a tableau, then the associated polytabloid is

$$\mathbf{e}_t = \kappa_t\{t\}$$

Example 5.27. Take $\lambda = (3,2)$. We can then compute $\kappa_t = (e-(34))(e-(15))$. Therefore,

$$\mathbf{e}_t = \begin{array}{ccc} \mathbf{4} & \mathbf{1} & \mathbf{2} \\ \mathbf{3} & \mathbf{5} & \end{array} - \begin{array}{ccc} \mathbf{3} & \mathbf{1} & \mathbf{2} \\ \mathbf{4} & \mathbf{5} & \end{array} - \begin{array}{ccc} \mathbf{4} & \mathbf{5} & \mathbf{2} \\ \mathbf{3} & \mathbf{1} & \end{array} + \begin{array}{ccc} \mathbf{3} & \mathbf{5} & \mathbf{2} \\ \mathbf{4} & \mathbf{1} & \end{array}$$

We now described what happens when we pass from t to πt.

Lemma 5.3. *Let t be a tableau and π be a permutation. Then*
1. $R_{\pi t} = \pi R_t \pi^{-1}$

2. $C_{\pi t} = \pi C_t \pi^{-1}$
3. $\kappa_{\pi t} = \pi \kappa_t \pi^{-1}$
4. $\mathbf{e}_{\pi t} = \pi \mathbf{e}_t$

We can now fully define the Specht modules

Definition 5.33. (Specht Module) For any partition λ, the corresponding Specht Module, S^λ, is the submodule of M^λ spanned by the polytabloids $\mathbf{e}_t$, where t is the shape of λ.

But then using the previous lemma, we immediately arrive at

Theorem 5.10. *The S^λ are cyclic modules generated by any given polytabloid.*

Example 5.28. Suppose that we have $\lambda = (n)$. Then $\mathbf{e}_{1\ 2\ \dots\ n} = \mathbf{1\ 2\ \cdots\ n}$ must be the only polytabloid. Then $S^{(n)}$ carries the trivial representation. Moreover, S_n acts trivially.

Example 5.29. Let $\lambda = (1_n)$ and fix

$$t = \begin{array}{c} 1 \\ 2 \\ \vdots \\ \mathrm{n} \end{array}$$

Then we have,

$$\kappa_t = \sum_{\sigma \in S_n} (\mathrm{sgn}(\sigma))\sigma$$

and $\mathbf{e}_t$ must be the signed sum of all possible $n!$ permutations regarded as tabloids. Now given any permutation $\pi \in S_n$, we have

$$\mathbf{e}_{\pi \mathbf{t}} = \pi \mathbf{e}_{\mathbf{t}} = \sum_{\sigma \in S_n} (\mathrm{sgn}(\sigma))\pi\sigma\{\mathbf{t}\}$$

but upon replacing $\pi\sigma$ by τ, we have

$$\mathbf{e}_{\pi \mathbf{t}} = \sum_{\tau \in S_n} (\mathrm{sgn}(\pi^{-1}\tau))\tau\{\mathbf{t}\} = (\mathrm{sgn}(\pi^{-1}) \sum_{\tau \in S_n} (\mathrm{sgn}(\tau)\tau\{\mathbf{t}\} = (\mathrm{sgn}(\pi))\mathbf{e}_t$$

Then every polytabloid is a scalar multiple of $\mathbf{e}_t$, where t is given as before. Therefore,

$$S^{(1^n)} = \mathbb{C}\{\mathbf{e}_t\}$$

with the action $\pi \mathbf{e}_t = (\mathrm{sgn}(\pi))\mathbf{e}_t$. THis is the sign representation.

5.11. **The Submodule Theorem.** We will now show that the S^λ constitute a full set of irreducible S_n-modules. Moreover, with a simple change of substitution of a bilinear form for the inner product, these results here will be true for an field. First, recall that H^- is given by

$$H^- = \sum_{\pi \in H} \mathrm{sgn}(\pi)\pi$$

for any subset $H \subseteq S_n$. So if $H = \{\pi\}$, then we write π^- for H^-. Moreover, recall the unique inner product on M^λ for which

$$\langle \{\mathbf{t}\}, \{\mathbf{s}\} \rangle = \delta_{\{t\},\{s\}}$$

Lemma 5.4. *Let $H \leq S_n$ be a subgroup. Then*

(1) If $\pi \in H$, then

$$\pi H^- = H^- \pi = sgn(\pi) H^-$$

(2) For any $\mathbf{u}, \mathbf{v} \in M^\lambda$

$$\langle H^- \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, H^- \mathbf{v} \rangle$$

(3) If the transposition $(bc) \in H$, then we can factor

$$H^- = k(e - (bc))$$

where $k \in \mathbb{C}[S_n]$.

(4) If t is a tableau with b, c in the same row of t and $(bc) \in H$, then

$$H^- \{\mathbf{t}\} = \mathbf{0}$$

Proof:

1. This follows similarly to the "proof" in Example 5.29.
2. Since our form is S_n-invariant,

$$\langle H^- \mathbf{u}, \mathbf{v} \rangle = \sum_{\pi \in H} \langle \mathrm{sgn}(\pi) \pi \mathbf{u}, \mathbf{v} \rangle = \sum_{\pi \in H} \langle \mathbf{u}, \mathrm{sgn}(\pi) \pi^{-1} \mathbf{v} \rangle$$

   Then replacing π by π^{-1} and given the fact this does not change the sign, the last sum is equivalent to $\langle \mathbf{u}, H^- \mathbf{v} \rangle$.
3. Consider the subgroup $K = \{e, (bc)\}$ of H. Then we can find a transversal and write $H = \uplus_i k_i K$. But then $H^- = (\sum_i k_i^-)(e - (bc))$.
4. By hypothesis, $(bc)\{\mathbf{t}\} = \{\mathbf{t}\}$. Then

$$H^- \{\mathbf{t}\} = k(e - (bc))\{\mathbf{t}\} = k(\{\mathbf{t}\} - \{\mathbf{t}\}) = \mathbf{0}$$

□

Corollary 5.5. *Let $t = t^\lambda$ be a λ-tableau and $s = s^\mu$ be a μ-tableau, where $, \lambda, \mu \vdash n$. If $\kappa_t \{\mathbf{s}\} \neq \mathbf{0}$, then $\lambda \trianglerighteq$. Moreover, if $\lambda = \mu$, then $\kappa_t \{\mathbf{s}\} = \pm \mathbf{e}_t$.*

Proof: Suppose that b and c are two elements in the same row of s^μ. Then they cannot be in the same column of t^λ because then $\kappa_t = k(e - (bc))$ and $\kappa_t \{\mathbf{s}\} = \mathbf{0}$ by the previous lemma. Then by the dominance lemma, we have $\lambda \trianglerighteq \mu$. Now if $\lambda = \mu$, then it must be the case that $\{s\} = \pi\{t\}$ for some $\pi \in C_t$, by the same argument in the proof of the Dominance Lemma. Then using the previous part we have

$$\kappa_t \{\mathbf{s}\} = \kappa_t \pi \{\mathbf{t}\} = \mathrm{sgn}(\pi) \kappa_t \{\mathbf{t}\} = \pm \mathbf{e}_t$$

□

Corollary 5.6. *If $\mathbf{u} \in M^\mu$ and sh $t = \mu$, then $\kappa_t \mathbf{u}$ is a multiple of $\mathbf{e}_t$.*

Proof: We are able to write $\mathbf{u} = \sum_i c_i\{s_i\}$, where s_i are μ-tableaux. By the previous corollary, $\kappa_t \mathbf{u} = \sum_i \pm c_i \mathbf{e}_i$. □

Theorem 5.11. *(Submodule Theorem) Let U be a submodule of M^μ, then*

$$U \supseteq S^\mu \text{ or } U \subseteq S^{\mu\perp}$$

In particular, when the field is $\mathbb{C}$, the S^μ are irreducible.

Proof: Consider $\mathbf{u} \in U$ and a μ-tableau t. Using the previous corollary, we have that $\kappa_t \mathbf{u} = f\mathbf{e}_t$ for some $f \in k$, where k is a field. There are two possible cases, depending on the multiplies.

First, assume that there exists a $\mathbf{u}$ and a t with $f \neq 0$. Since $\mathbf{u}$ is in the submodule U, we have $f\mathbf{e}_t = \kappa_t \mathbf{u} \in U$. Therefore, $\mathbf{e}_t \in U$ as f is nonzero and $S^\mu \subseteq U$ because S^μ is cyclic.

Now suppose that $\kappa_t \mathbf{u} = \mathbf{0}$ and consider any $\mathbf{u} \in U$ and an arbitrary μ-tableau t, applying the second part of the sign lemma,

$$\langle \mathbf{u}, \mathbf{e}_t \rangle = \langle \mathbf{u}, \kappa_t\{\mathbf{t}\} \rangle = \langle \kappa_t \mathbf{u}, \{\mathbf{t}\} \rangle = \langle \mathbf{0}, \{\mathbf{t}\} \rangle = \mathbf{0}$$

Notice up till now, our field has been arbitrary. However, to get the final result, we need our field to have stronger properties than an arbitrary field will allow, i.e. begin algebraically closed or having special inner product properties.

Proposition 5.5. *Suppose the field of scalars is $\mathbb{C}$ and $\theta \in Hom(S^\lambda, M^\mu)$ is nonzero. Then $\lambda \trianglerighteq \mu$ and if $\lambda = \mu$, then θ is multiplication by a scalar.*

Proof: Since $\theta \neq 0$, there is some basis vector $\mathbf{e}_t$ such that $\theta(\mathbf{e}_t) \neq \mathbf{0}$. But because $\langle \cdot, \cdot \rangle$ is a inner product with complex scalars, $M^\mu = S^\lambda \oplus S^{\lambda\perp}$. Therefore, we can extend θ to an element of $\mathrm{Hom}(M^\lambda, M^\mu)$ by setting $\theta(S^{\lambda\perp} = \mathbf{0}$. Then

$$\mathbf{0} \neq \theta(\mathbf{e}_t) = \theta(\kappa_t\{\mathbf{t}\}) = \kappa_t \theta(\{\mathbf{t}\}) = \kappa_t \sum_i c_i\{\mathbf{s}_i\}$$

where the s_i are μ-tableaux. Then by Corollary 5.5, we have $\lambda \trianglerighteq \mu$.

Now in the case where $\lambda = \mu$, Corollary 5.6 implies that $\theta(\mathbf{e}_t) = c\mathbf{e}_t$ for some constant $c \in \mathbb{C}$. But then for any permutation π,

$$\theta(\mathbf{e}_{\pi t}) = \theta(\pi\, \mathbf{e}_t) = \pi\theta(\mathbf{e}_t) = \pi(c\, \mathbf{e}_t) = c\, \mathbf{e}_{\pi t}$$

But then θ is a multiplication by some scalar $c \in \mathbb{C}$. □

Recall that our goal here was to verify the claim that the Specht modules, S^λ, form a full set of irreducible S_n-modules. Here, we can finally show this. Moreover, notice how simple the proof is now!

Theorem 5.12. *(Specht Modules $\sim$ Set of Irreducibles)*
The Specht Modules, S^λ, for $\lambda \vdash n$, form a complete list of irreducible S_n-modules over the complex field.

Proof: The S^λ are irreducible by the Submodule Theorem and the fact that $S^\lambda \cap S^{\lambda\perp} = \mathbf{0}$ for $\mathbb{C}$. So thus far we have shown that we have the correct number of modules. However, we need to show that these modules are pairwise equivalent. But if $S^\lambda \cong S^\mu$, then there must be a nonzero homomorphism $\theta \in \mathrm{Hom}(S^\lambda, M^\mu)$, since $S^\mu \subseteq M^\mu$. Therefore, by Proposition 5.5, $\lambda \trianglerighteq \mu$. But similarly we have $\mu \trianglerighteq \lambda$. Hence, $\lambda = \mu$. □

Notice that our proof relies heavily on the fact that the field is $\mathbb{C}$. For an arbitrary field of characteristic p, the Specht modules need not be irreducible. However, the Submodule Theorem gives us the quotient $S^\lambda/(S^\lambda \cap S^{\lambda\perp})$ is irreducible. One can work with these instead for S^λ when looking at p-modular representations of S_n. In any case, we have shown that S^λ form the full set of irreducible S_n-modules. However, we have not demonstrated the decomposition of the module.

Corollary 5.7. *The permutation module decomposes as*

$$M^\mu = \bigoplus_{\lambda \trianglerighteq \mu} m_{\lambda_\mu} S^\lambda$$

with diagonal multiplicity $m_{\mu\mu} = 1$.

Proof: This follows from Proposition 5.5. Suppose that S^λ appears in M^μ with a nonzero coefficient, then $\lambda \trianglerighteq \mu$. Now if $\lambda = \mu$, we have that

$$m_{\mu\mu} = \dim \mathrm{Hom}(S^\mu, M^\mu) = 1$$

□

5.12. **Standard Tableaux.** Given an arbitrary set of polytabloids which generate S^λ, the set will most likely not be independent. It would be more convenient to have some subset which can form a basis. This would be especially useful for computing the matrices of the representation or for computing the characters of the representation. This is done by using standard tableaux.

Definition 5.34. (Standard Tableaux) A tableau t is standard if the rows and columns of t are increasing sequences. The corresponding tabloid and polytabloids are also said to be standard.

Though we will not prove it, as it turns out, we have already seen a basis forming set for S^λ.

Theorem 5.13. *(S^λ Basis) The set*

$$\{\mathbf{e}_t \,|\, t \text{ is a standard } \lambda - \text{tableau}\}$$

is a basis for S^λ.

To show this, one needs to create a way of comparing tabloids. In fact, the comparison is exactly what we did before with dominating partitions. In this manner it is simple to show that the $\{\mathbf{e}_t\}$ set is an independent set. However, to show that it is a basis, it

would remain to show that all the standard polytabloids of shape λ span S^λ. To do this, take an arbitrary tableau t. We need to show that $\mathbf{e}_t$ is a linear combination of standard polytabloids. Assume that the columns of t are increasing because if not then there must be a $\sigma \in C_t$ such that $s = \sigma t$ has increasing columns. Then from part 4 of Lemma 5.3 and from part 1 of the Sign Lemma, it is the case that

$$\mathbf{e}_s = \sigma \mathbf{e}_t = \operatorname{sgn}(\sigma)\mathbf{e}_t$$

This says that $\mathbf{e}_t$ is some linear combination of polytabloids whenever $\mathbf{e}_s$ is. Now if there is a permutation, say π, such that in each tableau πt, a pair of adjacent, out-of-order elements in a row has been eliminated and the element of the group algebra, $g = e + \sum_\pi \operatorname{sgn}(\pi)\pi$, must satisfy $g\mathbf{e}_t = \mathbf{0}$. But then $\mathbf{e}_t$ must have the form

$$\mathbf{e}_t = -\sum_\pi \mathbf{e}_{\pi t}$$

The process of eliminating the pair of adjacent out-of-order elements in a row is called row descent. Now we have expressed $\mathbf{e}_t$ in terms of a polytabloid which is somehow closer to being standard. After some finite number of iterations of this process, we can find appropriate group elements g such that we can express $\mathbf{e}_t$ as a some linear combination of polytabloids. These special group elements are called the Garnir elements.

Definition 5.35. (Garnir Elements) Let A and B be two disjoint sets of positive integers and choose a permutation π such that

$$S_{A\cup B} = \biguplus_\pi \pi(S_A \times S_B)$$

Then a corresponding Garnir element is

$$g_{A,B} = \sum_\pi \operatorname{sgn}(\pi)\pi$$

As one would expect, the Garnir element is dependent on A and B. However, the Garnir element also depend on the transversal as well.

Definition 5.36. Let t be a tableau and let A and B be subsets of the jth and $(j+1)st$ columns of t, respectively. Then the Garnir element associated with t, A, B is

$$g_{A,B} = \sum_\pi \operatorname{sgn}(\pi)\pi$$

where the π have been chosen so that the elements of $A \cup B$ are increasing down the columns of πt.

While the Garnir elements allow us to introduce row descents, it adds the problem that while inducing a row descent in one location, it can also result in descents elsewhere as well. The question becomes to these extra descents help or hinder our progress to "standardness"? In order to answer that question, one applies a partial ordering this time to a column equivalence class, called the column tabloid, as one did before with rows. This

finally allows one to put everything together and prove the major result of this section - that the set

$$\{\mathbf{e}_t \,|\, t \text{ is a standard } \lambda - \text{tableau}\}$$

is a basis for S^λ.

5.13. **Branching Rules.** The results we have seen should be shocking. We have taken a group and mapped it into some linear space. That is, we have created a representation for a group. This allows us to study some of the properties of the group more easily than having to study the structure of its binary operation or group table. However, if one recalls Cayley's Theorem, every groups binary operation creates a permutation of the group and hence every possible group is a subgroup of some permutation group. So if we can study groups using their representations, in a sense, we need only understand the representations of the symmetric group to understand the representations of all groups.

This should come as no surprise as we have already said the same statement in our discussion on Cayley's Theorem. Surprisingly, to study these representations, it suffices to consider ways to arrange numbers into rows and columns corresponding to a particular representation! These diagrams, the Young Tableau, are easy to graphically represent, simple to understand, and easy to manipulate. The matrices for the module S^λ in the standard basis are known as Young's natural representation. Furthermore, we need not even compute all these matrices! As we have mentioned before, S_n has several generating subsets, namely the $(k\ \ k+1)$ transpositions. To generate the whole representation, the only computation needed is to find the matrices representing these transpositions. As a short summary , we have

Theorem 5.14. *Let f^λ be the number of standard λ-tableaux. For any partition, λ,*

1. $\{\mathbf{e}_t \,|\, t$ *is a standard* λ*-tableau*$\}$ *is a basis for* S^λ.
2. $\dim S^\lambda = f^\lambda$
3. $\sum_{\lambda \vdash n} (f^\lambda)^2 = n!$

Of course the last part of the theorem follows from the fact that for any group G,

$$|G| = \sum_V (\dim V)^2$$

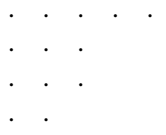
where the sum is to be taken over all irreducible G-modules.

The other useful aspect of these Young diagrams is that they easily allow us to see what happens when we restrict or induce an irreducible representation S^λ to S_{n-1} or S_{n+1}. In fact, through the construction of the Young diagrams, we have already provided a path to easily demonstrate the induction/restriction. These inductions or restrictions will correspond to adding or removing nodes (or squares) to the rows or columns of the Ferrers diagram.

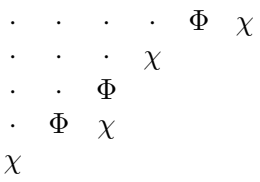
Definition 5.37. (Inner/Outer Corner) If λ is a diagram, then the inner corner of λ is a node $(i,j) \in \lambda$ whose removal leaves the Ferrers diagram of a partition. An partition obtained by such a removal is denoted λ^-. An outer corner of λ is a node $(i,j) \notin \lambda$ whose

addition produces a Ferrers diagram of a partition. Any partition obtained by such an addition is denoted λ^+.

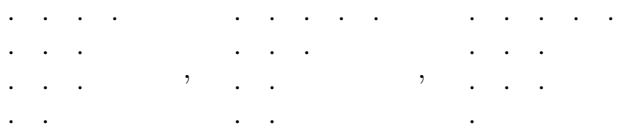
Example 5.30. Take the partition $\lambda = (5, 3, 3, 2)$. Then the Ferrers diagram of λ is



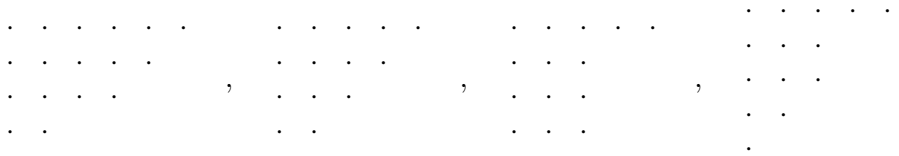
Now if we denote the inner corners with Φ's and the outer corners with χ's, we have



which means that the possible λ^- are



while the possible λ^+ are



Given these arrays it should come as no surprise that

$$S^{(5,3,3,2)}\downarrow_{S_{12}} \cong S^{(4,3,3,2)} \oplus S^{(5,3,2,2)} \oplus S^{(5,3,3,1)}$$

while

$$S^{(5,3,3,2)}\uparrow^{S_{14}} \cong S^{(6,3,3,2)} \oplus S^{(5,4,3,2)} \oplus S^{(5,3,3,3)} \oplus S^{(5,3,3,2,1)}$$

In fact, we can extend this to more than this example. Since every standard tableau of some shape $\lambda \vdash n$ must consist of n in some inner corner with a standard tabluea of shape $\lambda^- \vdash n - 1$, we have

$$f^\lambda = \sum_{\lambda^-} f^{\lambda^-}$$

Which is essential to establish the general rule, known as the branching rule.

Theorem 5.15. *(Branching Rule) If $\lambda \vdash n$, then*

$$1.\quad S^\lambda \downarrow_{S_{n-1}} \cong \bigoplus_{\lambda^-} S^{\lambda^-}$$

$$2.\quad S^\lambda \uparrow^{S_{n+1}} \cong \bigoplus_{\lambda^+} S^{\lambda^+}$$

This should be regarded as an incredible result! One could then quickly teach even a middle schooler to manipulate Young diagrams to tell one what the inductions/restrictions of the S^λ modules are! We have taken very deep representation problems and turned them into problems of drawing pictures.

Finally, we shall consider again Corollary 5.7,

$$M^\mu = \bigoplus_{\lambda \trianglerighteq \mu} m_{\lambda_\mu} S^\lambda$$

The only thing we have not shown is how to calculate the multiplicities of the S^λ. We do this again by considering another type of tableau "shape".

Definition 5.38. (Generalized Young Tableau) A generalized Young tableau of shape λ is an array T obtained by replacing the nodes of λ with positive integers with repetitions allowed. The type or content of T is the composition $\mu = (\mu_1, \mu_2, \cdots, \mu_m)$, where μ_i equals the number of i's in T. Let

$$T_{\lambda\mu} = \{T \,|\, T \text{ has shape } \lambda \text{ and content } \mu\}$$

Example 5.31. The array

$$T = \begin{array}{cccc} 3 & 3 & 4 & 1 \\ 5 & 1 & 4 & \\ 1 & & & \end{array}$$

has shape $(4, 3, 1)$ and content $(3, 0, 2, 2, 1)$.

Though we will not show it, for any given partition λ, the modules M^μ and $\mathbb{C}[T_{\lambda\mu}]$ are isomorphic. But the multiplicity of S^μ in M^μ is $\dim \mathrm{Hom}(S^\lambda, M^\mu)$. So we can contract homomorphisms from M^λ to M^μ in terms of generalized tableaux and then restrict this to S^λ.

Definition 5.39. For each $T \in T_{\lambda\mu}$, the homomorphism corresponding to T is the map $\theta_T \in \mathrm{Hom}(M^\lambda, M^\mu)$ given by

$$\{\mathbf{t}\} \mapsto \sum_{S \in \{T\}} \mathbf{S}$$

and extension by using the cyclicity of M^λ. In fact, θ is also a homomorphism into $\mathbb{C}[T_{\lambda\mu}]$.

Then we consider yet another type of tableau.

Definition 5.40. (Semistandard Tableau) A generalized tableau is semistandard if its rows weakly increase and its columns strictly increase. We let $T^0_{\lambda\mu}$ denote the set of semistandard λ-tableaux of type μ.

Example 5.32. The tableau

$$T = \begin{array}{cccc} 1 & 1 & 1 & 3 \\ 1 & 3 & 3 & \\ 2 & 4 & & \end{array}$$

is semistandard.

These turn out to form a basis for $\mathrm{Hom}(S^\lambda, M^\mu)$.

Theorem 5.16. *The set*

$$\{\overline{\theta}_T \mid T \in T^0_{\lambda\mu}\}$$

is a basis for Hom(S^λ, M^μ).

Using fairly trivial combinatorics, one can count the $T^0_{\lambda\mu}$. This number is called the Kostka number.

Definition 5.41. (Kostka Numbers) The Kostka numbers are

$$K_{\lambda\mu} = |T^0_{\lambda\mu}|$$

Then we finally get a better formula for Corollary 5.7,

Theorem 5.17. *The multiplicity of S^λ in M^μ is equal to the number of semistandard tableaux of shape λ and content μ,*

$$M^\mu \cong \bigoplus_\lambda K_{\lambda\mu} S^\lambda$$

6. Lie Representations

Introduction. After looking at general definitions and properties of representations, it is natural to turn our attention to a specific type of representation, namely Lie representations. Lie Algebras are intimately connected with symmetries of special geometric objects, namely Lie groups and ultimately many types of differentiable manifolds. Moreover, many important objects in theoretical Physics, especially in relativity, quantum mechanics, and string theory are Lie groups. Each Lie group has an associated Lie algebra that is used to understand the group. Here we focus on representing this important algebra. Eventually, we will show that we can represent these algebras using Weyl groups and Dynkin diagrams that are clearly visually connected with symmetries of geometric figures. First, we will discuss Lie groups to set the stage and because many of the concepts of Lie groups are simplier to understand and we see the same concepts for Lie groups later in their algebras and representations.

6.1. **Lie Groups.** We need to define what we mean when we say Lie group or Lie algebra. However, we remind the reader of the definition of a manifold. A topological space X for which every $x \in X$ has a neighborhood that is locally homeomorphic to a Euclidean space E^n for some $n \in \mathbb{Z}$ is called locally Euclidean. This simply means that when one looks "closely enough" at the space, it looks like a typical Euclidean space. Then a (topological) manifold is simply a topological space which is a locally Euclidean. Since the topological space X is locally Euclidean, it comes with a built in metric and hence is a metric space. Therefore, X is also Hausdorff. A smooth manifold is a manifold which it is possible to do Calculus on. Specifically, a smooth manifold is a second countable Hausdorff space that is locally homeomorphic to a linear space. Equivalently, a smooth manifold is a manifold where the maps (often called charts) between X and E^n are smooth. Sometimes differentiable manifold is used in place of smooth manifold when no confusion exists that the smoothness is required, though we will make the distinction here between a smooth and differentiable manifolds.

Definition 6.1. (Lie Group)
A Lie Group G is a group which is also a n-dimensional smooth manifold, i.e. the operations $G \times G \to G : (g, h) \mapsto gh$ and $G \to G : g \mapsto g^{-1}$ for all $g, h \in G$ are smooth. (For complex manifolds we require the operations be analytic.)

Most often when we say Lie group, we will mean a matrix Lie group, that is a set of $n \times n$ matrices that is closed under products, closed under inverses, and closed under nonsingular limits (meaning that if L is the set of matrices and $A_i \in L$, implies that given $A_1, A_2, \cdots$ and $A = \lim_{i\to\infty} A_i$ has an inverse then $A \in L$). But why the limit requirement for the matrix set? In fact, this is what guarantees smoothness for the group. Although, we will not go into the specifics of what exactly we meet by the action being smooth beyond construction of tangent spaces. The following examples are more than just that, they are some of the classic Lie groups are of great importance and one sees them time and time again when studying Lie Theory.

Example 6.1. The general linear group, denoted, GL_n, is the group of invertible $n \times n$ matrices under the operation of ordinary matrix multiplication. When the field is not clear, it often written $GL(n, \mathbb{C})$, that is if the field is $\mathbb{C}$. Here, we will assume the entries are complex. So in the case where $n = 2$

$$GL_2 = GL(2, \mathbb{C}) = \{M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mid ad - bc \neq 0\}$$

Example 6.2. The special linear group, denoted SL_n or $SL(n, \mathbb{C})$, is a subgroup of $GL(n, \mathbb{C})$ of all $n \times n$ matrices with determinant 1. In the case where $n = 2$, we have

$$SL_2 = SL(2, \mathbb{C}) = \{M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mid ad - bc = 1\}$$

Example 6.3. The orthogonal group, denoted $O(n, F)$, is the set of $n \times n$ orthogonal matrices with entries in a field, in our case $\mathbb{C}$. This is a subgroup of the general linear group over the same field. This is often denoted

$$O(n, \mathbb{C}) = \{M \in GL(n, F) \mid M^T M = M M^T = \mathbf{I}_n\}$$

where M^T is the transpose of T and $\mathbf{I}_n$ is the $n \times n$ identity matrix. Furthermore, the orthogonal group is not path-connected as one cannot connect matrices with positive determinant to those with negative determinant without going through determinant zero, which is not in $O(n)$.

Example 6.4. The special orthogonal group, denoted $SO(n, F)$, is the subgroup of $O(n, \mathbb{C})$ of matrices that have determinant 1. Moreover, $SO(3)$ is the group that represents all possible rotations in 3-dimensional space. Furthermore, for all n, $SO(n)$ is path-connected, which we will not show.

Example 6.5. The unitary group, denoted $U(n)$, is the set of all $n \times n$ unitary matrices under the group operation of matrix multiplication. That is,

$$U(n) = \{U \in M_n \mid U^* U = U U^* = \mathbf{I}_n\}$$

where M_n is the set of all $n \times n$ matrices and U^* is the conjugate transpose of U. Of course then the orthogonal group is the real analogue of the unitary group. The case of $U(1)$ is the the circle group, or $\{e^{i\theta}\}$ for all possible θ. Of course, the unitary group is a subgroup of $GL(n, \mathbb{C})$.

Example 6.6. The special unitary group, denoted $SU(n)$, is the subgroup of $U(n)$ consisting of all unitary matrices with determinant 1. The special unitary groups are commonly found in mathematical physics as $SU(2)$ is found in explaining the electroweak force interaction and $SU(3)$ in the explanation of quantum chromodynamics.

Example 6.7. The symplectic group, denoted $Sp(n)$, is the analogue of the orthogonal group for $\mathbb{H}^n$, the quaternions, and consists of all matrices of $GL(n, \mathbb{H})$ which preserve the hermitian form on $\mathbb{H}^n$:

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

That is the matrices have the condition $M\overline{M}^T = 1_n$, where $\overline{M}^T$ is the conjugate transpose of M and $\mathbf{1}_n$ is the $n \times n$ identity matrix. In the case where $n = 1$, we have

$$Sp(1) = \{\begin{pmatrix} a+id & -b-ic \\ b-ic & a-id \end{pmatrix} \mid a^2+b^2+c^2+d^2 = 1\} = SU(2)$$

So $Sp(1)$ consists of all unit quaternions. Moreover, $Sp(n)$ is path-connected for all n.

6.2. **Maximal Tori and Centers.** As ideals and normal groups help us understand the ring/group structure, we need some object in the Lie group that will help us to understand the structure of the group. It turns out the two most useful subgroups of a Lie group for this purpose are the center and maximal tori of a Lie group.

Definition 6.2. (Maximal Tori) The largest subgroup of a group G that is isomorphic to

$$\mathbb{T}^k = \mathbb{S}^1 \times \mathbb{S}^1 \times \cdots \times \mathbb{S}^1$$

contained in G, where the product on the right is taken k times.

Notice that the maximal torus is an abelian group. One would then expect it to be connected in some way to the center of the group. Indeed, the maximal torus can help us find the center of a group. However, the center of a group need not contain its maximal torus nor does the maximal torus need contain the center, see Example 6.9

Example 6.8. In the group $SO(2) = \mathbb{S}^1$, the maximal torus is $\mathbb{S}^1 = \mathbb{T}^1$. Therefore, $SO(2)$ is its own maximal torus.

Example 6.9. Here we find the maximal torus of $SO(3)$. Consider an element of the group $SO(3)$ as a rotation in $\mathbb{R}^3$ with $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ being the standard basis vectors. Now notice that

$$R'_\theta = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

is an obvious $\mathbb{T}^1 = \mathbb{S}^1$ in $SO(3)$. If $\mathbb{T}$ is a torus in G that contains $\mathbb{T}^1$ then any $A \in \mathbb{T}$ commutes with all $R'_\theta \in \mathbb{T}^1$, as any torus is abelian. It is easy enough to show that if

$$AR'_\theta = R'_\theta A \text{ for all } R'_\theta \in \mathbb{T}^1$$

then $A \in \mathbb{T}^1$ and so $\mathbb{T} = \mathbb{T}^1$ and $\mathbb{T}^1$ is then maximal. This can be done by showing that

$$A(\mathbf{e}_1), A(\mathbf{e}_e) \in (\mathbf{e}_1, \mathbf{e}_2)\text{-plane}$$

(this is because R'_θ is a rotation leaving the $\mathbf{e}_3$ axis fixed). However, notice that an element $A \in Z(SO(3))$ that commutes with all the elements of $SO(3)$. Using the same argument above, A must fix the basis vectors $\mathbf{e}_1$ and $\mathbf{e}_2$. Then A must be the identity vector $\mathbf{1}$. So $Z(SO(3)) = \{\mathbf{1}\}$ and is contained in but clearly distinct from the maximal torus.

Proposition 6.1. *(Maximal Tori) The maximal tori in $SO(2m)$, $SO(2m+1)$, $U(n)$, $SU(n)$, $Sp(n)$ are as follows:*

1. *In $SO(2m)$, the maximal torus $\mathbb{T}^m$ consist of all matrices $R_{\theta_1,\theta_2,\cdots,\theta_m}$, where*

$$R_{\theta_1,\theta_2,\cdots,\theta_m} = \begin{pmatrix} \cos\theta_1 & -\sin\theta_1 & & & & & & \\ \sin\theta_1 & \cos\theta_1 & & & & & & \\ & & \cos\theta_2 & -\sin\theta_2 & & & & \\ & & \sin\theta_2 & \cos\theta_2 & & & & \\ & & & & \ddots & & & \\ & & & & & \cos\theta_m & -\sin\theta_m \\ & & & & & \sin\theta_m & \cos\theta_m \end{pmatrix}$$

   where all blank entries are 0.
2. *In $SO(2m+1)$, $\mathbb{T}^m$ consists of all matrices $R_{\theta_1,\theta_2,\cdots,\theta_m}$, where*

$$\begin{pmatrix} \cos\theta_1 & -\sin\theta_1 & & & & & & \\ \sin\theta_1 & \cos\theta_1 & & & & & & \\ & & \cos\theta_2 & -\sin\theta_2 & & & & \\ & & \sin\theta_2 & \cos\theta_2 & & & & \\ & & & & \ddots & & & \\ & & & & & \cos\theta_m & -\sin\theta_m & \\ & & & & & \sin\theta_m & \cos\theta_m & \\ & & & & & & & 1 \end{pmatrix}$$

3. *In $U(n)$, $\mathbb{T}^n$ consists of all matrices of the form $Z_{\theta_1,\theta_2,\cdots,\theta_n}$.*
4. *In $SU(n)$, $\mathbb{T}^{n-1}$ consists of all matrices of the form $Z_{\theta_1,\theta_2,\cdots,\theta_n}$, where $\theta_1+\theta_2+\cdots+\theta_n = 0$.*
5. *In $Sp(n)$, $\mathbb{T}^n$ consists of all matrices of the form $Q_{\theta_1,\theta_2,\cdots,\theta_n}$.*

Proposition 6.2. *(Lie Centers) The centers of the classical Lie groups are as follows:*

1. $Z(SO(2m)) = \{\pm\mathbf{1}\}$
2. $Z(SO(2m+1)) = \{\mathbf{1}\}$
3. $Z(U(n)) = \{\omega\mathbf{1} \,|\, |\omega| = 1\}$
4. $Z(SU(n)) = \{\omega\mathbf{1} \,|\, \omega^n = 1\}$
5. $Z(Sp(n)) = \{\pm\mathbf{1}\}$

Often the center is important because we wish to decompose the Lie group and this is of course most often done using a normal subgroup. Notice that the center is a normal subgroup. So a Lie group cannot be simple unless the center $Z(G)$ is trivial, which of course disqualifies all of the classical Lie groups except $SO(2m+1)$. To decide whether $SO(2m+1)$ has any other normal subgroups, one uses the concept of discreteness.

Definition 6.3. (Discrete) A subgroup H of a matrix Lie group G is called discrete if there is a positive lower bound to the distance between any two members of H, the distance between matrices (a_{ij}) and (b_{ij}) being

$$\sqrt{\sum_{i,j} |a_{ij} - b_{ij}|^2}$$

Notice by the definition of the matrix distance, any finite subgroup of a Lie group G is discrete. Moreover, it turns out for certain types of Lie groups, it suffices to consider only the center for most discrete normal subgroups.

Theorem 6.1. *If G is a path-connected matrix Lie group with a discrete normal subgroup H, then H is contained in the center of $Z(G)$ of G.*

Proof: Since H is normal, $BAB^{-1} \in H$ for each $A \in H$ and $B \in G$. Thus, $B \mapsto BAB^{-1}$ defines a continuous map from G into discrete set H. Since G is path connected and a continuous map sends paths to paths, the image of the map must be a single point of H. The point is necessarily A because $\mathbf{1} \mapsto \mathbf{1}A\mathbf{1}^{-1} = A$. So each $A \in H$ has the property that $BA = AB$ for all $B \in G$. That is, $A \in Z(G)$. □

6.3. **Exponential Map.** Consider the Lie group $SO(2) = \mathbb{S}^1$. We can view this Lie group as the image of the line $\mathbb{R}i = \{i\theta \,|\, \theta \in \mathbb{R}\}$ under the exponential function as

$$e^{i\theta} = \cos\theta + i\,\sin\theta$$

is a line in some sense. Moreover, this line is tangent to the circle at the identity. In fact, every Lie group has some linear space with equal dimension as the tangent space at the identity. But do we care? Certainly, it is interesting that the tangent space has the same dimension and touches the Lie group at its identity. But does this do anything for us? Indeed, the tangent space can be very informative about the properties of the Lie group. The exponential function is the gateway from Lie groups to its tangent space. Furthermore, the exponential function can be generalized to square matrices and maps the tangent space of any matrix Lie group G into G. Many of the times, this mapping is onto G and the structure of G is similar to the structure of the tangent space. This tangent space is called the Lie algebra of G. The conjugation operation on G, which tells us how "far" from commutativity the operation of G is, corresponds to the Lie bracket in the tangent space.

First, we consider the famous observation of Euler

$$e^{i\theta} = \cos\theta + i\sin\theta$$

This relationship is easily derived from comparison of the Taylor series for e^x and those for $\cos x$ and $\sin x$. The exponential function maps the imaginary axis, $\mathbb{R}i$ onto the circle, $\mathbb{S}^1$, of points $\cos\theta + i\sin\theta$ in the plane of complex numbers. Then the operations of addition and negation on $\mathbb{R}i$ correspond to multiplication and inverses on $\mathbb{S}^1$:

$$e^{i\theta_1}e^{i\theta_2} = e^{i(\theta_1+\theta_2)} \quad \& \quad \left(e^{i\theta}\right)^{-1} = e^{-i\theta}$$

Furthermore, as we have previously stated, the line of points $i\theta$ mapped onto $\mathbb{S}^1$ by the exponential function can be viewed as the tangent to $\mathbb{S}^1$ at the identity, though the points on the tangent line look like $1 + i\theta$ (the constant 1 can be easily ignored). The power of the complex numbers to map lines to curves and its ability to preserve the lengths of small arcs proves essential in our study of Lie groups and algebras.

Now recall that operation in the tangent space if the Lie bracket. This originated from the Norwegian mathematician Sophus Lie, credited with many of the original great

discoveries in Lie Theory from his work on continuous groups. As we have seen before with Galois Theory, it is often had to discern the properties of a group operation, especially for larger or infinite groups where the Cayley table is difficult/impossible to write down. The Lie algebra is the tangent space at the identity of the Lie group and reflects many of the properties of the group. Being a linear space, it is often easier to work with than the Lie group itself and can yield many of its properties. This is exactly what we did in Galois Theory by taking group questions into field questions. This was precisely Lie's idea - to take group elements "infinitesimally close to the identity" (that is close to the tangent space) to approximate the behavior of all group elements. The commutative part of the group operation is reflected by the "commutative" part of the exponential function:

$$e^X e^Y = e^{X+Y}$$

The non-commutative properties of the group is reflected by the Lie bracket of the group. Before calculating a few examples, we go through some of the properties of the exponential function extended to matrices.

Definition 6.4. (Matrix Absolute Value) The absolute value of a matrix $A = (a_{ij})$ is defined by

$$|A| = \sqrt{\sum_{i,j} |a_{ij}|^2}$$

Equivalently, for a real $n \times n$ matrix A, the absolute value is the distance from the origin O in $\mathbb{R}^{n^2}$ to the point

$$(a_{1,1}, a_{1,2}, \cdots, a_{1,n}, a_{2,1}, a_{2,2}, \cdots, a_{2,n}, \cdots, a_{n,1}, \cdots, a_{n,n})$$

For entries in the complex plane, we can view them as being in $\mathbb{R}^2$ and define the distance similarly.

From this definition of absolute value, we able to define the distance between matrices. If A and B are matrices, then the distance between A and B is $|A - B|$. We then get a useful inequality on the distances.

Lemma 6.1. *(Submultiplicative Property) For any two real $n \times n$ matrices A and B,*

$$|AB| \leq |A|\ |B|$$

Proof: If $A = (a_{ij})$ and $B = (b_{ij})$, we define $AB = \left(ab_{ib}\right)$ and have

$$\begin{aligned}
|(ab)_{ij}| =& |a_{i,1}b_{1,j} + a_{i,2}b_{2,j} + \cdots + a_{i,m}b_{n,j}| \\
\leq& |a_{i,1}b_{1,j}| + |a_{i,2}b_{2,j}| + \cdots + |a_{i,n}b_{n,j}| \\
=& |a_{i,1}|\,|b_{i,j}| + |a_{i,2}|\,|b_{2,j}| + \cdots + |a_{i,n}|\,|b_{n,j}| \\
\leq& \sqrt{|a_{i,1}|^2 + \cdots + |a_{i,n}|^2}\sqrt{|b_{1,j}|^2 + \cdots + |b_{n,j}|^2}
\end{aligned}$$

Then summing over both sides,

$$\begin{aligned}|AB|^2 =& \sum_{i,j} |(ab)_{i,j}|^2 \\ \leq& \left(|a_{i,1}|^2 + \cdots + |a_{i,n}|^2\right)\left(|b_{1,j}|^2 + \cdots + |b_{n,j}|^2\right) \\ =& \sum_i \left(|a_{i,1}|^2 + \cdots + |a_{i,n}|^2\right) \sum_j \left(|b_{i,j}|^2 + \cdots + |b_{n,j}|^2\right) \\ =& |A|^2|B|^2\end{aligned}$$

□

Corollary 6.1. *If A is a matrix, then $|A^m| \leq |A|^m$.*

Corollary 6.2. *If A and B are matrices, then $|A + B| \leq |A| + |B|$.*

Of course, the definition of the exponential function for matrices does very little if it does not converge.

Theorem 6.2. *(Exponential Convergence) If A is any $n \times n$ matrix, then*

$$\mathbf{1} + \frac{A}{1!} + \frac{A^2}{2!} + \cdots +$$

where $\mathbf{1}$ is the $n \times n$ identity matrix, is convergent in $\mathbb{R}^{n^2}$.

Proof: We only need to show that the series is absolutely convergent, which means we need to show the convergence of

$$|\mathbf{1}| + \frac{|A|}{1!} + \frac{|A^2|}{2!} + \cdots$$

This is a series of positive real numbers, each terms less than the next with the exception of the first (by the submultiplicative property) to the corresponding terms of

$$1 + \frac{|A|}{1!} + \frac{|A^2|}{2!} + \cdots$$

But we know this sequence is convergent. □

But then once we know the matrix exponential is convergent, we can properly define it.

Definition 6.5. (Matrix Exponential) The exponential of any $n \times n$ matrix A is given by the series

$$e^A = \mathbf{1} + \frac{A}{1!} + \frac{A^2}{2!} + \frac{A^3}{3!} + \cdots$$

Remark 6.1. While the matrix exponential has many of the same properties as its real and complex counterparts, it does not share all their same properties. Namely, the matrix exponential does not share the addition formula because matrices to not necessarily commute. So if A and B are matrices,

$$e^{A+B} \neq e^A e^B$$

However, this holds in special cases.

Theorem 6.3. *If A and B are matrices and $AB = BA$, then*

$$e^{A+B} = e^A e^B$$

Proof: We want

$$\mathbf{1}+\frac{A+B}{1!}+\frac{(A+B)^2}{2!}+\cdots+\frac{(A+B)^n}{n!}+\cdots = \left(\mathbf{1}+\frac{A}{1!}+\cdots+\frac{A^n}{n!}+\cdots\right)\left(\mathbf{1}+\frac{B}{1!}+\cdots+\frac{B^n}{n!}+\cdots\right)$$

We can do this by expanding both sides and showing the coefficients of $A^l B^m$ is the same for both sides of the equation. However, if $AB = BA$ then the calculation is exactly that for $e^{x+y} = e^x e^y$ for real or complex numbers. Since it is correct for those, it is correct commuting matrices. □

Theorem 6.4. *If A is a matrix with complex entries, then $det(e^A) = e^{trace(A)}$.*

Proof: For any complex matrix A, there is an invertible complex matrix B and upper triangular matrix T such that $A = BTB^{-1}$. One can easily show using induction that

$$A^n = (BTB^{-1})^n = BT^nB^{-1}$$

Then

$$e^A = \sum_{m\geq 0}\frac{A^m}{m!} = B\left(\sum_{m\geq 0}\frac{T^m}{m!}\right)B^{-1} = Be^TB^{-1}$$

Then we only need to show that $\text{set}(e^T) = e^{\text{trace}(T)}$ for T upper triangular as

$$\det(e^A) = \det(Be^TB^{-1}) = \det(e^T) = e^{\text{trace}(T)} = e^{\text{trace}(BTB^{-1})} = e^{\text{trace}(A)}$$

But we also know that $\text{trace}(AB) = \text{trace}(BA)$. But then $\text{trace}(ABA^{-1}) = \text{trace}(B)$. Now suppose that

$$T = \begin{pmatrix} t_{1,1} & \sim & \sim & \cdots & \sim \\ 0 & t_{2,2} & \sim & \cdots & \sim \\ 0 & 0 & t_{2,2} & \cdots & \sim \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & t_{n,n} \end{pmatrix}$$

where the $\sim$ entries are arbitrary. We then know 3 things about T:

1. T^2 is upper triangular with the ith diagonal entry equivalent to $t_{i,i}^2$.
2. T^m is upper triangular with the ith diagonal entry equivalent to $t_{i,i}^m$.
3. e^T is upper triangular with the ith diagonal entry equal to $e_{i,i}^t$.

But then

$$\det(e^T) = e^{t_{1,1}}e^{t_{2,2}}\cdots e^{t_{n,n}} = e^{t_{1,1}+t_{2,2}+\cdots+t_{n,n}} = e^{\text{trace}(T)}$$

□

Example 6.10. Recall that every number $a + bi = z \in \mathbb{C}$ can be represented as a 2×2 matrix

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

Here we check the matrix exponential agrees with the standard complex exponential. Since $e^{i\theta} = \cos\theta + i\sin\theta$, we have

$$e^z = e^{x+iy} = e^x\left(\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n)!}y^{2n} + i\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n+1)!}y^{2n+1}\right)$$

So then suppose we have e^{iZ} with Z being the matrix form of the complex number $z = x + iy$,

$$\begin{pmatrix} x & -y \\ y & x \end{pmatrix}$$

Notice that we can express Z as the sum of matrices X and Y

$$X = \begin{pmatrix} x & 0 \\ 0 & x \end{pmatrix} \quad , \quad Y = \begin{pmatrix} 0 & -y \\ y & 0 \end{pmatrix}$$

Notice that $XY = YX$, so we can use the addition formula. Then it will suffice to calculate e^X and e^Y separately.

$$e^X = e^{\begin{pmatrix} x & 0 \\ 0 & x \end{pmatrix}} = \mathbf{1} + \frac{X}{1!} + \frac{X}{2!} + \cdots + \frac{X^n}{n!} + \cdots$$

Now notice that

$$X^n = \begin{pmatrix} x^n & 0 \\ 0 & x^n \end{pmatrix}$$

But then

$$e^X = \sum_{i=0}^{\infty}\frac{X^i}{i!} = \begin{pmatrix} \sum_{i=0}^{\infty}\frac{x^i}{i!} & 0 \\ 0 & \sum_{i=0}^{\infty}\frac{x^i}{i!} \end{pmatrix} = \begin{pmatrix} e^x & 0 \\ 0 & e^x \end{pmatrix}$$

And now we calculate e^Y, observe that

$$Y^n = \begin{cases} \begin{pmatrix} y^n & 0 \\ 0 & y^n \end{pmatrix} & \text{if } n \equiv 0 \mod 4 \\ \begin{pmatrix} 0 & -y^n \\ y^n & 0 \end{pmatrix} & \text{if } n \equiv 1 \mod 4 \\ \begin{pmatrix} -y^n & 0 \\ 0 & -y^n \end{pmatrix} & \text{if } n \equiv 2 \mod 4 \\ \begin{pmatrix} 0 & y^n \\ -y^n & 0 \end{pmatrix} & \text{if } n \equiv 3 \mod 4 \end{cases}$$

Notice then we can clump even and odd powers together. Then using the fact that

$$yi = y\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -y \\ y & 0 \end{pmatrix}$$

We can write

$$\begin{aligned} e^Y =& \mathbf{1} + \frac{Y}{1!} + \frac{Y^2}{2!} + \cdots + \frac{Y^n}{n!} + \cdots \\ =& \left(\mathbf{1} + \frac{Y^2}{2!} + \cdots + \frac{Y^{2n}}{(2n)!} + \cdots\right) + \left(\frac{Y}{1!} + \frac{Y^3}{3!} + \cdots + \frac{Y^{2n+1}}{(2n+1)!} + \cdots\right) \\ =& \left(\mathbf{1} + \frac{Y^2}{2!} + \cdots + \frac{Y^{2n}}{(2n)!} + \cdots\right) + \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\left(\frac{y}{1!} + \frac{y^3}{3!} + \cdots + \frac{y^{2n+1}}{(2n+1)!} + \cdots\right) \\ =& \cos y + \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \sin y \\ =& \cos y + i \sin y \end{aligned}$$

putting everything together, we have

$$e^Z = e^{X+iY} = e^X e^{iY} = e^x\big(\cos y + i \sin y\big)$$

Example 6.11. Here we give a list of the tangent spaces for the classic matrix groups.

Lie Group	Tangent Space
$SO(n)$	All $n \times n$ vectors such that $X + X^T = \mathbf{0}$.
$O(n)$	All $n \times n$ vectors such that $X + X^T = \mathbf{0}$.
$U(n)$	All $n \times n$ complex matrices with $X + \overline{X}^T = \mathbf{0}$.
$SU(n)$	All $n \times n$ complex matrices with $X + \overline{X}^T = \mathbf{0}$ and trace$(X) = 0$.
$Sp(n)$	All $n \times n$ quaternion matrices with $X + \overline{X}^T = \mathbf{0}$.

We have stated usefulness of the tangent space is that you need only look at the tangent space around the identity in the group. So we will need to define what we mean by that.

Definition 6.6. (Matrix Path) In a space S of matrices, a path is a continuous function $t \mapsto A(t) \in S$, where $t \in [a, b] \subset \mathbb{R}$ and the entires $a_{i,j}(t)$ of $A(t)$ are continuous functions of t. The path is called smooth if the functions $a_{i,j}(t)$ are smooth. We can of course define the derivative in the usual manner:

$$\lim_{\Delta t \to 0} \frac{A(t + \Delta t) - A(t)}{\Delta t}$$

Notice that by this definition, $A'(t)$ is then the matrix with $a'_{i,j}(t)$ as its entries. Then the tangent vectors at $\mathbf{1}$ are the matrices X with the form $X = A'(0)$. Using a bit of Physics language, we can think of the $A'(t)$ then as "velocity vectors" that move through a matrix space and go through the identity matrix $\mathbf{1}$ smoothly, which corresponds to the group idemtity, $e \in G$.

Definition 6.7. (Identity Tangent Space) If G is any matrix group, we define its tangent space at the identity, $T_{\mathbf{1}}(G)$ to be the set of matrix of the form $X = A'(0)$, where $A(t)$ is a smooth path in G with $A(0) = \mathbf{1}$.

Now we show that the tangent space is a linear space.

Theorem 6.5. *$T_{\mathbf{1}}(G)$ is a vector space over $\mathbb{R}$. That is for all $X, Y \in T_{\mathbf{1}}(G)$, we have $X + Y \in T_{\mathbf{1}}(G)$ and $rX \in T_{\mathbf{1}}(G)$ for $r \in \mathbb{R}$.*

Proof: Suppose $A(t), B(t) \in G$ are smooth paths with $\mathbf{1} = A'(0) = X$ and $\mathbf{1} = B'(0) = Y$. Then $X, Y \in T_{\mathbf{1}}(G)$. Then $C(t) = A(t)B(t)$ is a smooth path in G with $C(0) = \mathbf{1}$. But then $C'(0) \in T_{\mathbf{1}}(G)$. Finally, we compute $C'(0)$,

$$C'(0) = \frac{d}{dt}\Big(A(t)B(t)\Big) \mid_{t=0} = \Big(A'(t)B(t) + A(t)B'(t)\Big) \mid_{t=0} = A'(0)B(0) + A(0)B'(0) = X + Y$$

But then $X, Y \in T_{\mathbf{1}}(G)$ implies that $X + Y \in T_{\mathbf{1}}(G)$. Now consider the smooth path $D(t) = A(rt)$. We then have $D(0) = A(0) = \mathbf{1}$, then it's the case $D'(0) \in T_{\mathbf{1}}(G)$ and $D'(0) = rA'(0) = rX$. But then if $X \in T_{\mathbf{1}}(G)$ then $rX \in T_{\mathbf{1}}(G)$. □

The vector sum in the tangent space in some way corresponds to the product operation in G. However, this correspondence cannot be faithful as the vector sum is commutative and the product in G is generally not. But we can improve the faithfulness of the correspondence by focusing on smooth paths through the identity. That is the idea of the Lie bracket.

Theorem 6.6. *(Lie Bracket) $T_{\mathbf{1}}(G)$ is closed under the Lie bracket, that is if $X, Y \in T_{\mathbf{1}}(G)$, then $[X, Y] \in T_{\mathbf{1}}(G)$, where $[X, Y] = XY - YX$.*

Proof: Suppose that $A(0) = B(0) = \mathbf{1}$ with $A'(0) = X, B'(0) = Y$, meaning that $X, Y \in T_{\mathbf{1}}(G)$. Now consider the path

$$C_s(t) = A(s)B(t)A(s)^{-1} \text{ for some fixed } s$$

Then $C_s(t)$ is a smooth path and $C_s(0) = \mathbf{1}$, so $C_s'(0) \in T_{\mathbf{1}}(G)$. Moreover,

$$C_s'(0) = A(s)B'(0)A(s)^{-1} = A(s)YA(s)^{-1}$$

is a smooth function of s as $A(s)$ is. Then $A(s)YA(s)^{-1} \in T_{\mathbf{1}}(G)$. Now let $D(s) = A(s)YA(s)^{-1}$, then differentiating respect to s and using the fact that $A(0) = \mathbf{1}$

$$\begin{aligned} D'(0) =& A'(0)YA(0)^{-1} + A(0)Y(-A'(0)) \\ =& XY - YX = [X, Y] \end{aligned}$$

as $A'(0) = X$ and $A(0) = \mathbf{1}$. Then $X, Y \in T_{\mathbf{1}}(G)$ implies $[X, Y] \in T_{\mathbf{1}}(G)$. □

So the tangent space of a Lie matrix group G is a vector space that is closed under $[\cdot, \cdot]$ and is called the Lie algebra of G.

Definition 6.8. (Matrix Lie Algebra) A matrix Lie algebra is a vector space of matrices that is closed under the Lie bracket $[X, Y] = XY - Y$.

This is then later generalized to a bracket of arbitrary action so long as $[\cdot, \cdot]$ is bilinear, anti commutative, and satisfies a special identity (called the Jacobi identity). As a final note, though the Lie algebra is useful in generalizing the properties of a Lie group, they do not complete the entire story nor do they associate with a unique Lie group. For example,

both $O(n)$ and $SO(n)$ have the same tangent space at $\mathbf{1}$. Nevertheless, Lie algebras provide valuable information about the structure about the "curved" objects Lie groups by looking instead at a linear space.

6.4. **Logarithmic Map.** Though a Lie algebra may help inform us about a Lie group, it will be of little use if we cannot transfer this information back into the world of Lie groups. To do this we need to make use of an inverse function. Lie the classical exponential function, the inverse of the matrix exponential is the matrix logarithm. Similar to the matrix exponential, we will define the matrix logarithm through a power series. However, the matrix logarithm has more convergence issues than its exponential counterpart. In fact, the matrix logarithm converges only in a neighborhood of the identity. In this section, we will show that there is a homeomorphism between a neighborhood of $\mathbb{1}$ in a matrix Lie group G and a neighborhood of $\mathbf{0}$ in its Lie algebra $\mathfrak{g} = T_{\mathbf{1}}(G)$. But then the matrix logarithm must produce tangents to Lie groups. First, we need to define the matrix logarithm.

Definition 6.9. (Matrix Logarithm) If A is a square matrix, then the logarithm of A is

$$\log(\mathbf{1} + A) = A - \frac{A^2}{2} + \frac{A^3}{3!} - \frac{A^4}{4!} + \cdots \frac{(-1)^n A^n}{n!} + \cdots$$

The matrix logarithm is absolutely convergent, as one can easily confirm by comparing it to the geometric series for $|A| < 1$. Then the $\log(\mathbf{1} + A)$ is a well-defined continuous function in this neighborhood of $\mathbf{1}$. Now we need to show that the matrix logarithm is really the inverse of the matrix exponential.

Theorem 6.7. *For any matrix e^X with distance at most 1 of the identity,*

$$\log(e^X) = X$$

Proof: Since $e^X = \mathbf{1} + \frac{X}{1!} + \frac{X^2}{2!} + \cdots$ and $|e^X - \mathbf{1}| < 1$ and we can write

$$\begin{aligned}\log(e^X) =& \log\left(\mathbf{1} + \left(\frac{X}{1!} + \frac{X^2}{2!} + \cdots\right)\right.\\ =& \left(\frac{X}{1!} + \frac{X^2}{2!} + \cdots\right) - \frac{1}{2}\left(\frac{X}{1!} + \frac{X^2}{2!} + \cdots\right)^2 + \frac{1}{3}\left(\frac{X}{1!} + \frac{X^2}{2!} + \cdots\right)^3 - \cdots\end{aligned}$$

Now because the series is absolutely convergent, we can rearrange the terms

$$\log(e^X) = X + \left(\frac{1}{2!} - \frac{1}{2}\right)X^2 + \left(\frac{1}{3!} - \frac{1}{2} + \frac{1}{3}\right)X^3 - \cdots$$

But these are the same terms for the expansion of the real valued logarithm when $|e^x - 1| < 1$, hence their sum must too be zero as $\log e^x = x$ in that case. But then we have $\log e^X = X$ as expected. □

Theorem 6.8. *If A, B are matrices such that $AB = BA$ and $\log A, \log B$, and $\log(AB)$ are defined then*

$$\log(AB) = \log(A) + \log(B)$$

Proof: Suppose that $\log(A) = X$ and $\log(B) = Y$, then

$$e^X = A \quad , \quad e^Y = B$$

Now because $XY = YX$,

$$X = \log\big(\mathbf{1} + (A - \mathbf{1})\big) = (A - \mathbf{1}) - \frac{(A-1)^2}{2!} + \frac{(A-\mathbf{1})^3}{3!} - \cdots$$

$$Y = \log\big(\mathbf{1} + (B - \mathbf{1})\big) = (B - \mathbf{1}) - \frac{(B-1)^2}{2!} + \frac{(B-\mathbf{1})^3}{3!} - \cdots$$

the series commute because A and B do. It then follows that

$$AB = e^X e^Y = e^{X+Y}$$

Then taking the log of both sides yields

$$\log(AB) = X + Y = \log(A) + \log(B)$$

□

Though we certainly hope it is the case that the logarithmic function maps G into $T_{\mathbf{1}}(G)$, it is not immediately obvious that any points, except $\mathbf{1}$, are mapped into $T_{\mathbf{1}}(G)$. But we finally have enough language to rigorously define a matrix group and prove that $T_{\mathbf{1}}(G)$ is mapped into G.

Definition 6.10. (Matrix Lie Grouo) A matrix Lie group is a group of matrices that is closed under nonsingular limits. That is, if $A_1, A_2, \cdots, A_n, \cdots$ is a convergent sequence of matrices in G with limit A and $\text{set}(A) \neq 0$, then $A \in G$.

Theorem 6.9. *If $A'(0)$ is the tangent vector at $\mathbf{1}$ to a matrix Lie group, then $e^{A'(0)} \in G$. That is, the exponential function maps the tangent space $T_{\mathbf{1}}(G)$ into G.*

Proof: Suppose that $A(t)$ is a smooth path in G such that $A(0) = \mathbf{1}$ and that $A'(0)$ is the corresponding tangent vector at $\mathbf{1}$. Then using the derivative definition,

$$A'(0) = \lim_{\Delta t \to 0} \frac{A(\Delta t) - \mathbf{1}}{\Delta t} = \lim_{n \to \infty} \frac{A(\frac{1}{n}) - \mathbf{1}}{\frac{1}{n}}$$

where n is a natural number greater than some n_0. Then we can compare the formula with the definition of $\log A(\frac{1}{n})$,

$$\begin{aligned} n \log A(\frac{1}{n}) =& \frac{\log A(\frac{1}{n})}{\frac{1}{n}} \\ =& \frac{A(\frac{1}{n}) - \mathbf{1}}{\frac{1}{n}} - \frac{A(\frac{1}{n}) - \mathbf{1}}{\frac{1}{n}} \left(\frac{A(\frac{1}{n}) - \mathbf{1}}{2} - \frac{(A(\frac{1}{n}) - \mathbf{1})^2}{3} + \cdots \right) \end{aligned}$$

Then taking n_0 large enough that $|A(\frac{1}{n}) - \mathbf{1}| < \epsilon < \frac{1}{2}$, the series in the parenthesis is less than $\epsilon + \epsilon^2 + \epsilon^3 + \cdots < 2\epsilon$. Which means the sum tends to $\mathbf{0}$ as $n \to \infty$. Then then follows the right side has the limit

$$A'(0) - A'(0)[\mathbf{0}] = A'(0)$$

as $n \to \infty$. Then the left side has the same limit, so

$$A'(0) = \lim_{n\to\infty} n \log A(\frac{1}{n})$$

Now exponentiating each side yields

$$\begin{aligned} e^{A'(0)} =& e^{\lim_{n\to\infty} n \log A(\frac{1}{n})} \\ =& \lim_{n\to\infty} e^{n \log A(\frac{1}{n})} \\ =& \lim_{n\to\infty} \left(e^{\log A(\frac{1}{n})} \right)^n \\ =& \lim_{n\to\infty} A(\frac{1}{n})^n \end{aligned}$$

But because $A(\frac{1}{n}) \in G$ by assumption, so $A(\frac{1}{n})^n \in G$ because G is closed under products. Then we have a convergent sequence of elements of G and the limit $e^{A'(0)}$ must be nonsingular because it has inverse $e^{-A'(0)}$. Then $e^{A'(0)} \in G$, by the closure of G under nonsingular limits. But then the exponential function maps $T_{\mathbf{1}}(G) = \mathfrak{g}$ into G. □

To show the opposite direction, that log maps G into $T_{\mathbf{1}}(G)$, requires much more language and careful construction, for which we will need limits.

Definition 6.11. (Sequential Tangent) X is a sequential tangent to G at $\mathbf{1}$ if there is a sequence $\langle A_m \rangle$ of members of G and a sequence $\langle \alpha_m \rangle$ of real numbers such that $A_m \to \mathbf{1}$ and $\frac{A_m - \mathbf{1}}{\alpha_m} \to X$ as $m \to \infty$.

So if $A(t)$ is a smooth path in G with $A(0) = \mathbf{1}$, then the sequence of points $A_m = A(\frac{1}{m})$ tends to $\mathbf{1}$ and

$$A'(0) = \lim_{m\to\infty} \frac{A_m - \mathbf{1}}{\frac{1}{m}}$$

Then any ordinary tangent vector $A'(0)$ is a sequential tangent vector. In fact, sometimes it is easier to consider sequential tangents rather than considering smooth paths. In fact, all sequential tangents are smooth.

Theorem 6.10. *Suppose that $\langle A_m \rangle$ is a sequence in a matrix Lie group G such that $A_m \to \mathbf{1}$ as $m \to \infty$ and that $\langle \alpha_m \rangle$ is a sequence of real numbers such that*

$$\frac{A_m - \mathbf{1}}{\alpha_m} \to X$$

as $m \to \infty$. Then $e^{tX} \in G$ for all real t. Therefore, X is the tangent at $\mathbf{1}$ to the smooth path e^{tX}.

Proof: Let $X = \lim_{m\to\infty} \frac{A_m - \mathbf{1}}{\alpha_m}$. We prove that $e^X \in G$. Given that $\frac{A_m - \mathbf{1}}{\alpha_m} \to X$ as $m \to \infty$, it follows that $\alpha_m \to 0$ as $A_m \to \mathbf{1}$. Hence, $\frac{1}{\alpha_m} \to \infty$. Now set

$$a_m = \text{ closest integer to } \frac{1}{\alpha_m}$$

Moreover, we have $a_m(A_m - \mathbf{1}) \to X$ as $m \to \infty$. Since a_m is an integer

$$\begin{aligned}\log(A_m^{a_m} =& a_m \log(A_m) \\ =& a_m(A_m - \mathbf{1}) - a_m(A_m - \mathbf{1})\left(\frac{A_m - \mathbf{1}}{2} - \frac{(A_m - \mathbf{1})^2}{3} + \cdots\right)\end{aligned}$$

since $A_m \to \mathbf{1}$, the series in the parenthesis tends to zero. Now since $\lim_{m\to\infty} a_m(A_m - \mathbf{1}) = X$, we have

$$X = \lim_{m\to\infty} \log(A_m^{a_m})$$

Then by the inverse property of the logarithm and the continuity of the exponent,

$$e^X = \lim_{m\to\infty} A_m^{a_m}$$

Since a_m is an integer, $A_m^{a_m} \in G$ by the closure of G under products. By the closure of G under nonsingular limits,

$$e^X = \lim_{m\to\infty} A_m^{a_m} \in G$$

To prove that $e^{tX} \in G$ for any real t, one replaces $\frac{1}{\alpha_m}$ above by $\frac{t}{\alpha_m}$. Now if

$$b_m = \text{ nearest integer to } \frac{t}{\alpha_m}$$

Similarly, $b_m(A_m - \mathbf{1}) \to tX$ as $m \to \infty$. Then consider the series of

$$\log(A_m^{b_m}) = b_m \log(A_m)$$

and then we find that

$$e^{tX} = \lim_{m\to\infty} A_m^{b_m} \in G$$

by the closure of G under nonsingular limits. □

This is critical in our proof that the log maps a neighborhood of $\mathbf{1}$ in G onto a neighborhood of $\mathbf{0}$ in $T_{\mathbf{1}}(G)$. Moreover, this general idea was a central idea of von Neumann when he showed that matrix Lie groups are smooth manifolds. We see this in the above proof in the passing from $\langle A_m \rangle$ to the curve e^{tX}.

Theorem 6.11. *For any matrix Lie group G, there is a neighborhood $N_\delta(\mathbf{1})$ mapped into $T_{\mathbf{1}}(G)$ by* log.

Proof: Suppose to the contrary that no $N_\delta(\mathbf{1})$ is mapped into $T_{\mathbf{1}}(G)$ by log. Then we are able to find $A_1, A_2, \cdots \in G$ with $A_m \to \mathbf{1}$ as $m \to \infty$ and with each $\log A_m \notin T_{\mathbf{1}}(G)$. Now G is contained in some $M_n(\mathbb{C})$, so each $\log A_m$ is in $M_n(\mathbb{C})$

$$\log A_m = X_m + Y_m$$

where X_m is the component of $\log A_m$ in $T_{\mathbf{1}}(G)$ and $Y_m \neq \mathbf{0}$ is the component in $T_{\mathbf{1}}(G)^\perp$, the orthogonal component of $T_{\mathbf{1}}(G)$ in $M_n(\mathbb{C})$. Moreover, $X_m, Y_m \to \mathbf{0}$ as $m \to \infty$ because $A_m \to \mathbf{1}$ and the log is continuous.

Now consider the matrices $\frac{Y_m}{|Y_m|} \in T_{\mathbf{1}}(G)^\perp$. All these matrices have absolute value 1 and hence lie on a sphere $\mathfrak{L}$ of radius 1 centered at $\mathbf{0}$ in $M_n(\mathbb{C})$. By the boundedness of $\mathfrak{L}$, the

sequence $\langle \frac{Y_m}{|Y_m|} \rangle$ has a convergent subsequence and the limit Y of this subsequence is also a vector in $T_{\mathbf{1}}(G)^{\perp}$ with length 1. Furthermore, $Y \notin T_{\mathbf{1}}(G)$. Taking a subsequence of limit Y in place of our original sequence,

$$\lim_{m\to\infty} \frac{Y_m}{|Y_m|} = Y$$

Finally, we need to consider the terms of the sequence

$$T_m = e^{-X_m} A_m$$

Now each $T_m \in G$ because $-X_m \in T_{\mathbf{1}}(G)$ so $e^{-X_m} \in G$ by the exponentiation of tangent vectors and the fact that $A_m \in G$.

On the other hand, $A_m = e^{X_m+Y_m}$ by the inverse property of log

$$\begin{aligned} T_m =& e^{-X_m} e^{X_m+Y_m} \\ =& \left(\mathbf{1} - X_m + \frac{X_m^2}{2!} + \cdots\right)\left(\mathbf{1} + X_m + Y_m + \frac{(X_m+Y_m)^2}{2!} + \cdots\right) \\ =& \mathbf{1} + Y_m + \text{ higher-order terms} \end{aligned}$$

These higher-order terms include X_m^2 and other powers of X_m. These cannot be guaranteed to be smaller than Y_m. However, these powers of X_m are those in

$$\mathbf{1} = e^{-X_m} e^{X_m}$$

so they sum to 0. Therefore,

$$\lim_{m\to\infty} \frac{T_m - \mathbf{1}}{|Y_m|} = \lim_{m\to\infty} \frac{Y_m}{|Y_m|} = Y$$

Then since each $T_m \in G$, it follows that the sequential tangent

$$\lim m \to \infty \frac{T_m - \mathbf{1}}{|Y_m|} = Y$$

which is in $T_{\mathbf{1}}(G)$ by the smoothness of sequential tangents. But $Y \notin T_{\mathbf{1}}(G)$. □

Corollary 6.3. *The log function gives a homeomorphism between $N_\delta(\mathbf{1})$ in G and $\log N_\delta(\mathbf{1}) \in T_{\mathbf{1}}(G) = \mathfrak{g}$.*

Proof: The continuity of the log and its inverse gives the necessary bijection and all the necessary continuity. □

We end with a few remarks about Lie groups and their algebras before focusing on the algebra side of Lie group questions. Schreier's Theorem tells us that any discrete normal subgroup of a path-connected group lies in its center. The question remains if there are any nondiscrete normal subgroups. We use normal subgroups to study the group structure by breaking it apart. However, if there are only discrete normal subgroups, we are still left with an infinite number of elements and hence a possible infinite structure left to break

down. So nondiscrete normal subgroups could be more useful but the hard part will be to find them. In fact, Lie algebras will prove very useful in find such subgroups.

Theorem 6.12. *If G is a path-connected matrix Lie group with discrete center and a nondiscrete normal subgroup H, then $\mathfrak{g} = T_{\mathbf{1}}(H) \neq \{\mathbf{0}\}$.*

Corollary 6.4. *If H is a nontrivial normal subgroup of G under the above conditions, then $T_{\mathbf{1}}(H)$ is a nontrivial ideal of $T_{\mathbf{1}}(G)$.*

So examining the tangent space gives us a way of "seeing" normal subgroups. That is, we can hunt down the normal subgroups of G by hunting for ideals of $\mathfrak{g}$. This is one of the many insights that Lie algebras offers to Lie group questions. Lastly, we will talk about a deep result in Lie Theory: Campbell-Baker-Hausdorff. We know that any elements of G have the form e^X. Then the product of any two elements in G have the same form, that is e^Z for some Z. But we can say more about the properties of Z. This is Campbell-Baker-Hausdorff Theorem. If

$$e^X e^Y = e^Z$$

then we can show that Z has the form

$$Z = X + Y + \frac{1}{2}[X, Y] + \text{ higher-order terms}$$

The genius of Campbell-Baker-Hausdorff is showing that all these higher-order terms are composed of nested Lie brackets. Notice for a commutative Lie group, since the bracket is always 0, we have the normal addition formula. However, the general case needs Campell-Baker-Hausdorff. This proof took a decade to complete and over the last century many more proofs have been given using various approaches, most highly technical and long. However, after reading through several proofs, the author agrees with Stillwell that the proof given b Eichler in 1968 is the most brilliant for its unique approach and brevity. For a complete walkthrough through the proof, see Stillwell's *Naive Lie Theory*, as we will not prove it here.

6.5. **Lie Algebras.**

Definition 6.12. (Lie Algebra)
A Lie algebra is a vector space L over a field k with a binary operation $L \times L \to L$ given by $(x, y) \mapsto [x, y]$, called the (Lie) bracket or commutator of x and y, that has the following properties:

1. Bilinearity: That is, $[ax + by, z] = a[x, z] + b[y, z]$ for all $a, b \in k$ and $x, y, z \in L$.
2. Anticommutativity: $[x, x] = 0$ for all $x \in L$.
3. Jacobian Identity: $[[x, y], z] + [[z, x], y] + [[y, x], z] = 0$ for all $x, y, z \in L$.

However, notice that property one and two of a Lie algebra gives

$$\begin{aligned}
[x+y, x+y] &= 0 \\
[x, x+y] + [y, x+y] &= 0 \\
[x, x] + [x, y] + [y, x] + [y, y] &= 0 \\
[x, y] + [y, x] &= 0 \\
[x, y] &= [-y, x]
\end{aligned}$$

So if the char $k \neq 2$, one could replace property 2 with $[x, y] = -[y, x]$. Moreover, notice that that the Lie bracket is not in general associative, that is $[x, [y, z]] \neq [[x, y], z]$ for some $x, y, z \in L$. Therefore, a Lie algebra is typically a non-associative algebra. We call a Lie algebra L commutative $[x, y] = 0$ (that is x and y commute) for all $x, y \in L$. However, in general Lie algebras are not commutative, nor are they associative. The definition of a subalgebra is trite:

Definition 6.13. (Subalgebra)
A vector subspace H of a Lie algebra L is called a subalgebra of L if $[x, y] \in H$ for all $x, y \in H$. Moreover, H is a lie algebra with respect to the operation of L restricted to H.

Example 6.12. Any vector space L with bracket defined by $[u, v] = 0$ for all $u, v \in L$ is easily checked to be a trivial (Abelian) Lie algebra.

Example 6.13. The simplest nontrivial example of a Lie algebra is well known to physicists as the (vector) cross product $[u, v] = u \times v$ for $u, v \in \mathbb{R}^3$. It is clear that $[\cdot, \cdot]$ is bilinear and anti-commutative. We need check the Jacobi identity

$$\begin{aligned}
[[x, y], z] &= (x \times y) \times z \\
&= (x \cdot z)y - (y \cdot z)x \\
&= (z \cdot x)y - (z \cdot y)x \\
&= [[z, y], x] + [[x, z], y]
\end{aligned}$$

where $\cdot$ is the usual dot product in $\mathbb{R}^3$. If we look at the standard basis e_1, e_2, e_3 for $\mathbb{R}^3$,

$$\begin{aligned}
[e_1, e_2] &= e_3 \\
[e_2, e_3] &= e_1 \\
[e_3, e_1] &= e_2
\end{aligned}$$

Furthermore, the bracket has the following properties:

1. If $w = [u, v]$, then $|w| = |u|\,|v| \sin\theta$, where $|\cdot|$ is the vector norm and θ is the angle between u and v.
2. w is orthogonal to both u and v. That is $w \times u = w \times v = \mathbf{0}$.
3. u, v, w form a standard triple of basis vectors that can be rotated to the standard basis vectors for $\mathbb{R}^3$.

4. If $u = (u_1, u_2, u_3)$ and $v = (v_1, v_2, v_3)$ for the standard orthogonal basis $i, j, k \in \mathbb{R}^3$, then one can equally compute the bracket $w = [u, v]$ as a determinant.

$$w = [u, v] = u \times v = \det \begin{vmatrix} i & j & k \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix}$$

Example 6.14. Another typical way ones sees a Lie bracket defined is in Linear Algebra, often called the commutator, $[a, b] = ab - ba$. It is easy to see when the algebra is commutative, the commutator is trivial. The bracket can equally be defined when $a, b \in M_n(k)$, where $n \in \mathbb{Z}_+$ and k is a field. This Lie algebra, denoted $\mathfrak{gl}(n, k)$

Example 6.15. The center of a Lie algebra L, defined by $Z(L) = \{z \in L \,|\, [x, y] = 0 \text{ for all } x \in L\}$, is an ideal of L. Moreover, L is abelian if and only if $[L, L] = 0$ as one can easily check.

Example 6.16. Perhaps the most important introductory Lie algebra is the endomorphism algebra. Take V as a finite dimensional vector space over a field k. Denote as usual End V the set of all possible linear transformations from $V \to V$. When viewed as a vector space over k, End V has dimension n^2. Now End V is a ring relative to the usual product operation in End V. Now define a new operation in End V by $[x, y] = xy - yx$ (the commutator of Example 6.14). Under this new operation, End V is a Lie algebra over k. This Lie algebra is denoted $\mathfrak{gl}(k)$ for the general linear algebra because of its close relation to the general linear group $GL(V)$ (all invertible endomorphisms of V). Any subalgebra of $\mathfrak{gl}(k)$ is called a linear Lie algebra.[1]

The Lie algebra of Example 6.16 is of great importance as a result of the following theorem

Theorem 6.13. *(Ado's Theorem)*
Any finite dimensional Lie algebra L over a field k of characteristic 0 is isomorphic to a subalgebra of some Lie algebra $\mathfrak{gl}(n, k)$. (That is, it is a linear Lie algebra.)

Ideals of Lie Algebras. We have seem the usefulness of ideals when we studied polynomials previously. Ideals give us a useful way of breaking an object apart into simpler parts that are easier to study but still give us useful information about the structure of the original object. We again will want to pay particularly close attention to ideals in our study of Lie algebras.

Definition 6.14. (Lie algebra Ideal)
A subspace of I of a Lie algebra L is called an ideal if $[x, y] \in I$ for all $x \in L$ and $y \in I$. Moreover, a Lie ideal is a subalgebra of the Lie group L.

Lemma 6.2. *(Ideals are Subalgebras)*
If I is an ideal of a Lie algebra L, then I is a subalgebra of L.

[1] See Appendix XXXX for more important examples of Lie algebras.

Proof: Since I is an ideal of L, $[x, y] \in I$ for all $x \in L$ and $y \in I$. If we take $x \in I$, then clearly $x \in L$. But then for all $x, y \in I$, $[x, y] \in I$. Hence, I is a subalgebra of L. □

Furthermore, it follows the Jacobian identity property of Lie algebras that all ideals are necessarily two-sided; i.e., there is no difference between right and left ideals for Lie algebras.

Example 6.17. The two easiest ideals to identify are the trivial vector subspaces **0** and the whole Lie algebra L. These are called the trivial ideals.

Example 6.18. The center of a Lie algebra is an ideal. That is, if L is a Lie algebra, then $Z(L) = \{z \in L \mid [x, z] = 0 \text{ for all } x \in L\}$. Notice if the Lie algebra is commutative then the ideal composed by the center is a trivial ideal.

As with ring ideals, the product of two Lie ideals is again a Lie ideal. However, it is not as easy to see as in the ring case.

Theorem 6.14. *(Products of Lie ideals are Ideals)*
Let L be a Lie algebra and i, J two ideals of L. Then the Lie product

$$[I, J] = Span\{[x, y] \mid x \in I, y \in J\}$$

is an ideal of L. Moreover, $[L, L] \subset L$ is an ideal of L.

Proof: Suppose L is a Lie algebra with two ideal I, J. First, observe that $[I, J]$ is a subspace of L. Moreover, since $[x, y] = -[y, x]$, we have $[I, J] = [J, I]$. Now suppose that $x \in I$ and $y \in J$ with $a \in L$. Then applying the Jacobi identity,

$$[[x, y], a] + [[y, a], x] + [[a, x], y] = 0$$

But since $[[y, a], x] \in [J, I] = [I, J]$ and $[[a, x], y] \in [I, J]$, it must be the case that $[[x, y], a] \in [I, J]$. Now any element $u \in [I, J]$ has the form $u = \sum c_{ij}[x_i, y_i]$, where $x_i \in I$ and $y_j \in J$ and the c_{ij} are scalars. Then for any $v \in L$, we have

$$[u, v] = [\sum c_{ij}[x_i, y_j], v] = \sum c_{ij}[[x_i, y_j], v]$$

where $[[x_i, y_j], v] \in [I, J]$. Hence, we have $[u, v] \in [I, J]$ and therefore $[I, J]$ is an ideal in L. □

We construct the definition of a ring being simple similarly, (note the nonabelian condition is to avoid importance being given to the one dimensional algebra since if L is simple then $Z(L) = 0$ and $L = [L, L]$):

Definition 6.15. (Simple Lie Algebra)
If L is a Lie algebra with no nontrivial ideals and $[L, L] \neq 0$, then L is simple.

As with rings and groups previously, we use ideals to break apart Lie algebras into manageable pieces which can be studied to understand the structure of the entire algebra. As before, we do this by creating equivalence classes through a quotient.

Definition 6.16. (Quotient Lie Algebra)
Let L be a Lie algebra and I an ideal of L. Take a coset space $L/I = \{x + I \,|\, x \in L\}$ with multiplication defined by $[x + I, y + I] = [x, y] + I$. The coset is called the quotient Lie algebra of L by I.

6.6. **Lie Homomorphisms.** Naturally, to observe the properties of the entire Lie algebra, we need a way of relating the structure of its ideals or other well understood Lie algebras to the entire Lie algebra in question. To do this, we examine the properties of the homomorphism from the quotient Lie algebra or an alternate algebra and the Lie algebra in question.

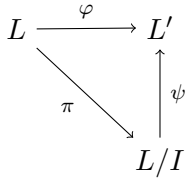
Definition 6.17. (Lie Homomorphism)
If L and H are Lie algebras, then a linear map $\varphi : L \to H$ is called a Lie homomorphism if

$$\varphi([x, y]) = [\varphi(x), \varphi(y)] \text{ for all } x, y \in L$$

We again have the normal properties of morphisms; that is, a given morphism $\varphi : L \to H$ is called a monomorphism if $\ker(\varphi) = 0$ and an epimorphism if $\text{im}(\varphi) = H$. A morphism which is both a monomorphism and an epimorphism is called an isomorphism. An isomorphism from a Lie algebra L to itself is called an automorphism. Moreover, observe that $\ker(\varphi)$ is an ideal of L and $\text{im}(\varphi)$ is a subalgebra of H. We remind the reader that there is a natural injection between the homomorphisms and ideals of a Lie algebra. The correspondence constructed by associating φ with $\ker \varphi$. The associated ideal I is called the canonical map $\varphi : L \to L/I$, defined by $x \mapsto x + I$. Now we observe important properties of Lie homomorphisms that immediately follow:

Theorem 6.15. *(Lie Mappings)*

1. *If L and H and are Lie algebras and $\varphi : L \to H$ is a homomorphism of Lie algebras, then $L/\ker \varphi \cong im\ \varphi$. If I is an ideal of L in $\ker \varphi$, then there is a unique homomorphism $\psi : L/I \to H$ that makes the following diagram commute, where π is the canonical map:*



2. *If I and H are ideals of a Lie algebra L such that $I \subset J$ then J/I is an ideal of L/I and $(L/I)(J/I) \cong L/J$.*
3. *If I and J are ideals of a Lie algebra L, there is a natural isomorphism between $(I+J)/J$ and $I/(I \cap J)$.*

Proof:

1. Since the ker is an ideal, we can create the quotient ring $L/\ker \varphi$. Let $K = \ker \varphi$. Now we create a mapping $\psi : L/K \to \text{im}\ \varphi$ given by $x + K \mapsto \varphi(x)$. First, we check that φ is

well defined. Suppose that $x + K = x' + K$, then $(x - x') + K = K$ so $x - x' \in K$. But then $\psi(x + K) = \psi(x')$. Moreover, this shows that ψ is injective. Surjectivity should be clear from the definition of ψ. Now

$$\psi(x + K)\psi(y + K) = \varphi(x)\varphi(y) = \varphi(xy) = \psi(xy + K)$$

and ψ is a homomorphism. Furthermore, ψ is also linear. But then ψ is an isomorphism of Lie algebras.

2.

Example 6.19. If L is an orthogonal Lie algebra such that $sx = -x^t s$ for all $x, s \in L$ and g is an orthogonal matrix in that g is invertible and $g^t s g = s$. We show that the mapping $x \mapsto gxg^{-1}$ is an automorphism of L.

$$\begin{aligned} sgxg^{-1} =& (g^{-1})^t s x g^{-1} \\ =& -(g^{-1})^t x^t s g^{-1} \\ =& -(g^{-1})^t x^t g^t s \\ =& -(gxg^{-1})s \end{aligned}$$

so the map $x \mapsto gxg^{-1}$ is a linear automorphism of L. Now we need only show that it preserves the operation.

$$[gxg^{-1}, gyg^{-1}] = gxyg^{-1} - gyxg^{-1} = g[x, y]g^{-1}$$

6.7. **Solvability and Nilpotent Lie Algebras.** In order to study Lie algebras, we break it into its smallest parts via quotients of ideals (as previously mentioned). But when can we break apart Lie algebras in such a way? As we did with rings, we use the concept of simplicity.

Definition 6.18. (Solvable)
If L is a Lie algebra, define a sequence of ideals of L by $L^{(0)} = L$, $L^{(1)} = [L, L]$, $L^{(2)} = [L^{(1)}, L^{(1)}]$, $\cdots$, $L^{(i)} = [L^{(i-1)}, L^{(i-1)}]$. Then L is a solvable Lie algebra if $L^{(n)} = 0$ for some nonnegative integer n.

This definition of course immediately implies that all simple Lie algebras are nonsolvable. Moreover, an Abelian Lie algebra is immediately solvable as $L^{(i)} = [L^{(i-1)}, L^{(i-1)}] = 0$ for any nonnegative integer i. The structure of a solvable Lie algebra is like that of a solvable ring or field:

Theorem 6.16. *(Solvable Lie Algebras)*
Let L be a Lie algebra, then

1. *If L is solvable then so are all subalgebras of L.*
2. *If L is solvable then all homomorphic images of L are solvable.*
3. *If I is a solvable ideal of L with L/I solvable, then L must be a solvable Lie algebra.*
4. *If I, J are solvable ideals of L then so is $I + J$.*

Proof: Suppose that L is a Lie algebra:

1. Suppose that K is a subalgebra of L, then it must be that $K^{(i)} \subset L^{(i)}$. Since there is an i_0 such that $L^{(i_0)} = 0$ for all elements in L and all elements of K are in L, $K^{(i_0)} = 0$. (However, this does not mean that a smaller i such that this holds does not exist for K).
2. Suppose that $\varphi : L \to M$ is a epimorphism from L to a Lie algebra M. We will apply induction. First, consider the case where $L^{(0)} = 0$. This implies L is Abelian and since φ is a homomorphism, M must be Abelian as well; hence, $M^{(0)} = 0$. Moreover, suppose that $L^{(1)} = 0$. Then we have
$$0 = \varphi(L^{(1)}) = \varphi([L, L]) = [\varphi(L), \varphi(L)] = [M, M] = M^{(1)}$$
Now assume the statement is true for all i up to some $j \in \mathbb{Z}_+$. Now observe that if $L^{(j+1)} = 0$, then
$$0 = \varphi(L^{(j+1)}) = \varphi([L^{(j)}, L^{(j)}]) = [\varphi(L^{(j)}), \varphi(L^{(j)})] = [M^{(j)}, M^{(j)}] = M^{(j+1)}$$
by induction, the homomorphic image of a solvable Lie algebra is solvable.
3. Suppose that I is a solvable ideal of L such that L/I is solvable. Then $(L/I)^{(n)} = 0$ for some n. Consider the typical canonical homomorphism $\pi : L \to L/I$, then $\pi(L^{(n)}) = 0$ by the previous part. Hence, $L^{(n)} \subset \ker \pi = I$. Since I is solvable, there is an m such that $I^{(m)} = 0$. Applying the previous part again and using the obvious fact that $(L^{(i)})^{(j)} = L^{(i+j)}$, we obtain that $L^{(n+m)} = 0$.
4. Assume that I, J are solvable ideals of L. By the Second Isomorphism Theorem, there is an isomorphism between $(I + J)/J$ and $I/(I \cap J)$. However, notice that $I/(I \cap J)$ is a homomorphic image of I. By one of the previous parts, $I/(I \cap J)$ must then be solvable since I is solvable. However, then $(I + J)/J$ must be solvable since it is the holomorphic image of $I/(I \cap J)$. Since J is solvable and $(I + J)/J$ is solvable, by the previous part, $I + J$ must be solvable. □

Notice the previous theorem seems to build on itself into a proof that if I and J are solvable ideals of a Lie algebra then $I + J$ is a solvable ideal. This becomes useful as it allows us to prove the existence of a maximal solvable ideal.

Theorem 6.17. *(Maximal Solvable Ideal)*
If L is a Lie algebra, there exists a unique solvable ideal of L.

1. *Existence: Every Lie algebra has a solvable (trivial) ideal. That is, $\{0\} \subset L$ is both an ideal and solvable. If there is*
2. *Uniqueness: Suppose that S, S' are distinct maximal solvable ideals of a LIe algebra L.*

Proof: L.T.R. □

Example 6.20. We show that L is solvable if and only if there exists a chain of subalgebras $L = L_0 \supset L_1 \supset L_2 \supset \cdots \supset L_k = 0$ such that L_{i+1} is an ideal of L_i and each quotient L_i/L_{i+1} is abelian. First, assume that L is solvable. Now the derived series
$$L = L^{(0)} \supset L^{(1)} \supset \cdots \supset L^{(k)} = 0$$

with $L^{(m+1)} = [L^{(m)}, L^{(m)}]$ is an ideal of $L^{(m)}$ and $L^{(m)}/L^{(m+1)}$ is abelian. Now assume that there is a chain of subalgebras

$$L = L_0 \supset L_1 \supset L_2 \supset \cdots \supset L_k = 0$$

such that L_{i+1} is an ideal of L and that each quotient L_i/L_{i+1} is abelian. Now if I is an ideal of L and L/I is abelian, then $I \subseteq [L, L]$ because L/I is abelian then

$$[x, y] \in I$$

for all $x, y \in L$. That is, that $[L, L] \subset I$. Then we have by induction that $L^{(m)} \subset L_m$. Finally, we then have

$$L^{(m+1)} = [L^{(m)}, L^{(m)}] \subseteq L_{m+1}$$

By the induction hypothesis, $L^{(k)} = 0$ and L is solvable.

This maximal solvable ideal is called the radical of L. This is also how we define the notion of semisimplicity of a Lie algebra,

Definition 6.19. (Radical/Semisimple)
If L is a Lie algebra, then there is a maximal solvable ideal K, called the radical of L and is denoted Rad L. If K is the trivial ideal, i.e. Rad $L = 0$, then L is called semisimple.

Of course, any simple Lie algebra L automatically fits the requirements for being semisimple as L has no ideal expect for $\mathbf{0}$ and L. Moreover, using the idea of Theorem 6.17, $L/\text{Rad } L$ is semisimple.

While solvable Lie algebras are useful to study the properties of there exist more recent useful types of Lie algebras which merit special attention. Moreover, this property is closely related to the solvability of a Lie algebra. In fact, it is a stronger property.

Definition 6.20. (Nilpotent)
Let L be a Lie algebra and define a sequence given by $L^0 = L$,$L^1 = [L, L]$, $L^2 = [L, L^1]$, $\cdots$, $l^i = [L, L^{i-1}]$. If there is an i such that $L^i = 0$ then L is called nilpotent.

Similar to solvable Lie algebras, all abelian Lie algebras are necessarily nilpotent. One should notice that $L^0 = L = L^{(0)}$ and $L^1 = [L, L] = L^{(1)}$. Furthermore, we have $L^{(i)} \subset L^i$. This then implies that if a Lie algebra L is nilpotent it is necessarily solvable. However, the converse does not hold. Since the definition of nilpotency is so close to that of solvability, one expects them to have similar properties. In fact, they do.

Theorem 6.18. *(Nilpotency) Let L be a Lie algebra, then*

1. *If L is nilpotent then so are all subalgebras of L.*
2. *If L is nilpotent then so are all homomorphic images of L.*
3. *If $L/Z(L)$ is nilpotent and not zero then L is nilpotent.*
4. *If L is nilpotent and not zero then $Z(L)$ is nontrivial.*

Proof: This proof is similar to that of Theorem 6.16 and follows from the ring homomorphism theorems. □

Of course we can rephrase the condition for L to be nilpotent as ad x_1 ad $x_2 \cdots$ ad $x_n(y) = 0$ for some n for all $x_i, y \in L$. Then in particular, it's the case that $(\text{ad } x)^n = 0$ for all $x \in L$.

Definition 6.21. (Ad-Nilpotent) If L is a Lie algebra and $x \in L$, we call x ad-nilpotent if ad x is a nilpotent endomorphism.

Therefore, we can rephrase the above result as: "if L is nilpotent then all the elements of L are ad-nilpotent." However, one wonders if the converse holds. Indeed, luckily the converse does hold. However, first we need a lemma.

Lemma 6.3. *Let $x \in \mathfrak{gl}(V)$ be a nilpotent endomorphism, then ad x is also nilpotent.*

Proof: Let λ_x and ρ_x be endomorphisms given by $\lambda_x(y) = xy$ and $\rho_x(y) = yx$, i.e. left and right translation, respectively. Both λ_x and ρ_x are nilpotent because x is nilpotent. Now λ_x and ρ_x commute as

$$(\lambda_x\rho_x)(y) = \lambda_x(\rho_x(y)) = \lambda_x(yx) = xyx$$
$$(\rho_x\lambda_x)(y) = \rho_x(\lambda_x(y)) = \rho_x(xy) = xyx$$

Now in End (End (V)), indeed any ring, the sum and difference of two commuting nilpotents is nilpotent. But then ad $x = \lambda_x - \rho_x$ is clearly nilpotent. □

Example 6.21. Here we show that L is solvable/nilpotent if and only if ad L is solvable/nilpotent. First, notice that ad : $L \to$ ad L is a homomorphism, then we know that ad $L \cong L/Z(L)$ because ker ad $= Z(L)$. Then $[Z(L), Z(L)] = 0$, so $Z(L)$ is a solvable ideal. But we know that L is solvable if and only if ad L is solvable.

Example 6.22. We will show that the sum of nilpotent ideals of a Lie algebra L is again a nilpotent ideal and that L possesses a unique maximal nilpotent ideal. Suppose that I, J are nilpotent ideals of $L, I^m = 0$, and $J^n = 0$.

$$[I + J, I + J] \subset [I, I] + [J, J] + [I, J] \subset [I, I] + [J, J] + I \cap J$$

Then one can show by induction that

$$(I + J)^k \subset I^k + J^k + I \cap J$$

Now if $x > \max(m, n)$, then $I^x = 0, J^x = 0$ and $(I + J)^x \subset I \cap J$

$$(I + J)^{x+y} = [I + J, \cdots, [I + J, I \cap J], \cdots] \subset I^y \cap J + I \cap J^y$$

but then $(I + J)^{x+y} = 0$ and $I + J$ is a nilpotent ideal of L. □

Example 6.23. If L is a nilpotent and K is a proper subalgebra of L, then $N_L(K)$ properly includes K. Why? Assume this isn't so. Let $L^0 = L, L^1 = [L, L], \cdots, L^n = 0$ be a descending central series of L. It is clear that K is proper subalgebra of L. But then there must be a n such that $L^{n+1} \subset K$. However, this contradicts the fact that K is a proper subalgebra.

Example 6.24. If L is nilpotent then L must has some ideal of codimension 1. To see why this is so. Let L be a nilpotent Lie algebra. Then $[L, L] \not\geq L$ and we have a natural homomorphism $\pi : L \to L/[L, L]$. Clearly, $L/[L, L]$ is a nonzero abelian algebra. It then must have a subspace I of codimension 1. Call this subspace I'. Because $L/[L, L]$ is abelian, I' is an ideal of $L/[L, L]$. Then $\pi^{-1}(I')$ is an ideal of L with codimension 1.

Theorem 6.19. *(Lie Eigenvectors) Let L be a subalgebra of $\mathfrak{gl}(V)$ and V be of finite dimension. If L consists of nilpotent endomorphisms and $V \neq 0$, then there exists nonzero $v \in V$ for which $L\mathbf{v} = \mathbf{0}$.*

Proof: We prove this by induction. The base cases where dim $L = 0$ and dim $L = 1$ are trivial. Assume the statement of the theorem is true for all Lie algebras of dimension less than L. If K is any proper subalgebra of L, then according to Lemma 6.3, K acts as a Lie algebra of nilpotent linear transformations on the vector space L, through the adjoint. Hence, it also acts similarly on the vector space L/K. By the induction hypothesis, dim $K <$ dim L, there is a vector $x + K \neq K \in L/K$ that is killed by the image of K in $\mathfrak{gl}(L/K)$. Therefore, $[y, x] \in K$ for all $y \in K$ but $x \notin K$. This is equivalent to $K \subset N_L(K)$, the normalizer of K in L.

Since V is finite dimensional, there must be a maximal proper subalgebra of L, even if is the trivial subalgebra. If we let K be the maximal proper subalgebra of L, since $K \subset N_L(K)$, it must be the case that $N_L(K) = L$. Then K is an ideal of L. We have two cases, dim $L/K = 1$ or dim $L/K > 1$. If dim $L/K > 1$, then the inverse image of any 1-dimensional subalgebra of L/K, which obviously must exist, would be a proper subalgebra which contains K, which would contradict the fact that K is maximal. Therefore, dim $L/K = 1$ and we can write $L = K + Fz$ for any $z \in L - K$.

Finally, $W = \{v \in V \mid K\mathbf{v} = 0\}$ contains more than the zero vector. Because K is an ideal, if $x \in L, y \in K$, and $w \in W$ then $yx\mathbf{w} = xy\mathbf{w} - [x, y]\mathbf{w} = \mathbf{0}$. So simply choose $z \in L - K$ then the nilpotent endomorphism z acting on W has an eigenvector, i.e. $z\mathbf{v} = 0$. But since $z \in L \subset \mathfrak{gl}(V)$, there is a vector v in L such that $L\mathbf{v} = \mathbf{0}$. □

Theorem 6.20. *(Engel's Theorem) If all elements of L are ad-nilpotent, then L is nilpotent.*

Proof: Suppose that L is a Lie algebra such that x is ad-nilpotent for all $x \in L$. Therefore, ad $L \subset \mathfrak{gl}(L)$. If $L = 0$, the result is trivial. Assume $L \neq 0$. Then by Theorem 6.19, there is a nonzero $x \in L$ such that $[L, x] = 0$. This means that x commutes with all of L and then $Z(L) \neq 0$. Since the center is a normal subgroup, $L/Z(L)$ exists and must consist of all ad-nilpotent elements. Moreover, $L/Z(L)$ must have dimension less than L. It follows easily from induction on dim L that $L/Z(L)$ must be nilpotent. But if $L/Z(L)$ is nilpotent then L is nilpotent. □

From Engel's Theorem, we can make a powerful statement about an important concept in Lie algebras, flags.

Definition 6.22. (Flag) If V is a finite dimensional vector space then a flag in V is a chain of subspaces $0 = V_0 \subset V_1 \subset V_2 \subset \cdots \subset V_n = V$, with $\dim V_i \subset V_{i-1}$ for all i. This is equivalent to saying that there is a basis of V relative to which al the matrices of L are in $\mathfrak{n}(n, F)$.

Corollary 6.5. *($\mathfrak{n}(n, F)$ Basis) If L is a Lie algebra with all its elements ad-nilpotent, then there exists a flag (V_i) in V stable under L with $xV_i \subset V_{i-1}$ for all i. (There exists a basis of V relative to which the matrices of L are all in $\mathfrak{n}(n, F)$.)*

Proof: Suppose $v \in V$ is nonzero and is killed by L. We know such a $v \in V$ exists by Engel's Theorem. Set $V_1 = Fv$ and $W = V/V_1$. Notice that the induced action of L on W is by nilpotent endomorphisms. Then by induction on $\dim V$, W must have a flag stabilized by L, the inverse image in V will work. □

Engel's Theorem holds many applications but an important one which we shall see is the following.

Lemma 6.4. *Let L be nilpotent and K an ideal of L. If $K \neq 0$ and $K \cap Z(L) \neq 0$. Moreover, $Z(L) \neq 0$.*

Proof: L acts on K through the adjoint representation. Therefore, by Engel's Theorem yields a $x \in K$ that is nonzero and is killed by L. Therefore, $x \in K \cap Z(L)$. □

Example 6.25. Suppose that L is a Lie algebra and K is an ideal of L with L/K being nilpotent with $\text{ad } x \mid_K$ for all $x \in L$. Because L/K is nilpotent for all $x \in L$, $\text{ad } x$ is a nilpotent endomorphism in End L/K. That is to say, there must be an $n \in \mathbb{Z}_+$ such that $(\text{ad } x)^n(y) \in K$ for all $y \in L$. However, $\text{ad } x \mid_K$ is nilpotent. This implies that there is an $m \in \mathbb{Z}_+$ such that $(\text{ad x})^m(\text{ad } x)^n(y) = (\text{ad } x)^{m+n}(y) = 0$. This shows that $\text{ad } x$ is a nilpotent endomorphism in $\mathfrak{gl}(L)$. Applying Engel's Theorem yields that L is nilpotent.

6.8. **Semisimple Lie Algebras.** Though we have obtained many results thus far, we have not made much probes. Much of this results from the fact that we are allowing F to be an arbitrary field. From now on, we shall focus on fields F with characteristic 0. Moreover, we will also make the assumption that F is algebraically closed (usually to guarantee the existence of the eigenvalue we seek). We will focus on Lie algebras that are composed of simple Lie algebras. So if we can take this larger Lie algebra and break it up into its simple parts, we can understand the larger object through these components. The power of Engel's Theorem is it allows us to find a common eigenvector for a Lie algebra that consists of nilpotent endomorphisms. We now prove a similar theorem.

Theorem 6.21. *Let L be a solvable subalgebra of $\mathfrak{gl}(V)$ and V be a finite dimensional. If V is nonzero then V contains a common eigenvector for all endomorphisms in L.*

Proof: We shall prove this using induction on $\dim L$. If $\dim L = 0$, then the result follows trivially. Now we follow closely the ideas of Engel's Theorem. We follow the following plan:

1. Find an ideal of codimension 1.
2. Show through induction that there is a common eigenvalue for K.
3. Verify that L stabilizes a space consisting of such eigenvalues.
4. Find in that space an eigenvalue for a single $z \in L$ satisfying $L = K + Fz$.

Now we proceed along our path.

1. Since L is solvable and of positive dimension, L properly contains $[L, L]$. Therefore, as $L/[L, L]$ is abelian, any subspace is automatically an ideal. Now take a subspace of codimension 1. Its inverse image K is an ideal of codimension 1 in L.
2. Now we use induction to find a common eigenvector $v \in V$ for K. If $K = 0$ then L is abelian of dimension 1 and an eigenvalue for a basis vector of L completes the proof. So for $x \in K$, $x\mathbf{v} = \lambda(x)\mathbf{v}$, where $\lambda : K \to F$ is some linear function. Fix this function λ and let W be the subspace
$$W = \{w \in V \mid x\mathbf{w} = \lambda(x)\mathbf{w} \text{ for all } x \in K\}$$
Note that W is nonzero.
3. We need to show that L leaves W invariant. First, we need to test if $x\mathbf{w}$ is in W. Take an arbitrary $y \in K$. Examine
$$yx\mathbf{w} = xy\mathbf{w} - [x, y]\mathbf{w} = \lambda(y)x\mathbf{w} - \lambda([x, y])\mathbf{w}$$
If this is to be in W, we need to show that $\lambda([x, y]) = 0$. Fix $w \in W$ and $x \in L$. Let $n > 0$ be the smallest integer for which $\mathbf{w}, x\mathbf{w}, \cdots, x^n\mathbf{w}$ are linearly independent. Now let W_i be the subspace of V spanned by $\mathbf{w}, x\mathbf{w}, \cdots, x^{i-1}\mathbf{w}$ (or the set $W_0 = 0$). Then $\dim W_n = n$ and $W_n = W_{n+1}$ for $i \geq 0$ and x maps W_n into W_n. Each $y \in K$ leaves each W_i invariant. Now relative to the basis $\mathbf{w}, x\mathbf{w}, \cdots, x^{n-1}\mathbf{w}$ of W_n, $y \in K$ can be represented by a upper triangular matrix whose diagonal entries are all equal to $\lambda(y)$. This follows from
$$yx^i\mathbf{w} \equiv \lambda(y)x^i\mathbf{w} \mod W_i$$
We prove this statement by induction on i. If $i = 0$, the statement is obvious. Now we write
$$yx^i\mathbf{w} = yxx^{i-1}\mathbf{w} = xyx^{i-1}\mathbf{w} - [x, y]x^{i-1}\mathbf{w}$$
By the induction hypothesis, $yx^{i-1}\mathbf{w} = \lambda(y)x^{i-1}\mathbf{w} + w'$, where $w' \in W_{i-1}$. Now by construction, x maps W_{i-1} into W_i, therefore it must be the case that
$$yx^i\mathbf{w} \equiv \lambda(y)x^i\mathbf{w} \mod W_i$$
holds for all i. Finally, because of the way $y \in K$ acts on W_n, $\text{trace}_{W_n}(y) = n\lambda(y)$. But this is true for elements of K of the form $[x, y]$ with $y \in K$ and x of the form $x\mathbf{v} = \lambda(x)\mathbf{v}$. However, both x, y stabilize W_n. Therefore, $[x, y]$ acts on W_n as the commutator of two endomorphisms of W_n, which implies its trace is 0. Then $n\lambda([x, y]) = 0$. Because char $F = 0$, this forces $\lambda([x, y]) = 0$.
4. Write $L = K + Fz$. Since F is algebraically closed, we can find an eigenvector $v_0 \in W$ of z, for some eigenvalue z. But then v_0 is obviously a common eigenvalue for L. Moreover, λ can be extended to a linear function on L such that $x\mathbf{v}_0 = \lambda(x)\mathbf{v}_0$ for all $x \in L$. $\square$

We then get the following powerful corollaries.

Corollary 6.6. *(Lie's Theorem) Let L be a solvable subalgebra of $\mathfrak{gl}(V)$ with dim $V = n < \infty$. Then L stabilizes some flag in V. Equivalently, the matrices of L relative to a suitable basis of V are upper triangular.*

Proof: This follows immediately from 6.21 with induction applied to dim V. □

Corollary 6.7. *Let L be a solvable Lie algebra. Then there exists a chain of ideals of L,*

$$0 = L_0 \subset L_1 \subset L_2 \subset \cdots \subset L_n = L$$

such that dim $L_i = i$.

Proof: Let L be any solvable Lie algebra and $\phi : L \to \mathfrak{gl}(V)$ a finite dimensional representation of L. Then $\phi(L)$ must be solvable. Then it must stabilize a flag from Lie's Theorem. If ϕ is the adjoint representation, a flag of subspaces is just a chain of ideals of L, each of codimension 1 in the next. □

Corollary 6.8. *Let L be a solvable Lie algebra. Then $x \in [L, L]$ implies that $ad_L\ x$ is nilpotent. In particular, $[L, L]$ is nilpotent.*

Proof: Find a flag of ideals as described in Corollary 6.7. Relative to some basis $\{x_1, x_2, \cdots, x_n\}$ of L, for which $\{x_1, x_2, \cdots, x_n\}$ spans L_i, the matrices of ad L must be in $\mathfrak{t}(n, F)$. Therefore, the matrices of $[\text{ad } L, \text{ad } L] = \text{ad}_L[L, L]$ lie in $\mathfrak{n}(n, F)$, the derived algebra of $\mathfrak{t}(n, F)$. Therefore, $\text{ad}_L\ x$ is nilpotent for $x \in [L, L]$ and $\text{ad}_{[L,L]}\ x$ is nilpotent. Therefore, by Engel's Theorem, $[L, L]$ is nilpotent. □

6.9. **Jordan-Chevalley Decomposition.** The reader should be familiar with the Jordan canonical form. For a single endomorphism, x over an algebraically closed field, we can express x as matrix as a sum of blocks

$$\begin{pmatrix} a & 1 & & & \\ & a & 1 & & \\ & & \ddots & \ddots & \\ & & & & 1 \\ & & & & a \end{pmatrix}$$

We can use this to create a very useful decomposition for the endomorphism matrices. Why? The a-diagonal commutes with a nilpotent matrix with 1's above the diagonal and zero everywhere else. But given the Jordan canonical form, x must be the sum of a diagonal and a nilpotent matrix which commute! But let's be more precise.

Definition 6.23. (Semisimple) Let $x \in \text{End } V$, where V is a finite dimensional vector space. Then x is semisimple if the roots of its minimal polynomial over a field F, with arbitrary characteristic, are all distinct.

Of course, we could have also just then said that x is semisimple if and only if x is diagonalizable. But if two semisimple endomorphisms commute, then we can diagonalize them simultaneously. This means that their sum/difference must be semisimple.

Proposition 6.3. *Let V be a finite dimensional vector space over F with $x \in End\, V$. Then we have*

1. *There exist unique $x_s, x_n \in End\, V$ satisfying the condition: $x = x_s + x_n$, where x_s is semisimple, x_n is nilpotent, and x_s, x_n commute.*
2. *There exist polynomials $p(T), q(T)$ in one indeterminate, without a constant term, such that $x_s = p(x), x_n = q(x)$. In particular, x_s and x_n commute with any endomorphism commuting with x.*
3. *If $A \subset B \subset V$ are subspaces and x maps b into A, then $x + s$ and x_n also map B into A.*

Proof: Let $a_1, \cdots, a_k$, with multiplicities $m_1, \cdots, m_k$, be the distinct eigenvalues of x, so the characteristic polynomial is

$$\Pi\ (T - a_i)^{m_i}$$

If $V_i = \ker\ (x - a_i \cdot 1)^{m_i}$, then V is a direct sum of the subspaces $V_1, \cdots, V_k$, each stable under x. On V_i, x clearly has characteristic polynomial $(T - a_i)^{m_i}$. Now applying the Chinese Remainder Theorem for the ring $F[T]$ to locate a polynomial $p(T)$ satisfying the congruences, with mapirwise relatively prime moduli:

$$p(T) \equiv a_i \mod (T - a_i)^{m_i} \quad , \quad p(T) \equiv 0 \mod T.$$

Now set $q(T) = T - p(T)$. Each of $p(T), q(T)$ has a zero constant term, since $p(T) \equiv 0 \mod T$. Now set $x_s = p(x)$ and $x_n = q(x)$. Since they are polynomials in x, x_s, x_n and commute with each other, as well as with all endomorphisms which commute with x. They also must stabilize all subspaces of V stabilized by x, in particular the V_i. Hence, we have shown (2). The congruence $p(T) \equiv a_i \mod (T - a_i)^{m_i}$ shows that the restriction of $x_s - a_i \cdot 1$ to V_i is zero for all i. Hence, x_s acts diagonally on V_i with single eigenvalue a_i. By definition, $x_n = x - x_s.$, so x_n is nilpotent. But because $p(T)$ and $q(T)$ have no constant term, we have shown (3). We only need to show (1) at this point. Let $x = s + n$ be another such decomposition. Then we have $x_s - s = n - x_n$. But because of (2), all the endomorphisms in sight commute. Sums of commuting semisimple/nilpotent endomorphisms are again semisimple/nilpotent, whereas only 0 can be both semisimple and nilpotent. Therefore, $s = x_s$ and $n = x_n$. □

The decomposition $x = x_s + x_n$ is called the (additive) Jordan-Chevalley decomposition of x. This is also sometimes called the Jordan decomposition. The x_s is called the semisimple part of x and x_n is called the nilpotent part of x.

Corollary 6.9. *Let $x \in End\, V$, with V being finite dimensional and $x = x_s + x_n$ being its Jordan decomposition. Then ad x = ad x_s + ad x_n is the Jordan decomposition of ad x in End(End V).*

Proof: We know that ad x_s and ad x_n are semisimple and nilpotent. Moreover, they must commute (as $[\text{ad } x, \text{ad } x_n] = \text{ad } [x_s, x_n] = 0$. Then the result follows from (2) of Proposition 6.3. □

Corollary 6.10. *Let $\mathfrak{U}$ be a finite dimensional F-algebra. Then der $\mathfrak{U}$ contains the semisimple and nilpotent parts off all its elements in End $\mathfrak{U}$.*

Proof: If $\delta \in \text{der } \mathfrak{U}$, suppose that $\sigma, \upsilon \in \text{End } \mathfrak{U}$ be its semisimple and nilpotent parts, respectively. It is sufficient to show that $\sigma \in \text{der } \mathfrak{U}$. If $a \in F$, set

$$\mathfrak{U}_a = \{x \in \mathfrak{U} \,|\, (\delta - a \cdot 1)^k x = 0\}$$

for some k that depends on x. Then $\mathfrak{U}$ is the direct sum of those $\mathfrak{U}_a$ for which a is an eigenvalue of δ or σ. Also, σ acts on $\mathfrak{U}_a$ as a scalar multiplication by a. It is easily checked by induction that for $a, b \in F$ that $\mathfrak{U}_a, \mathfrak{U}_b \subset \mathfrak{U}_{a+b}$ using the formula

$$\sum_{i=0}^{n} \binom{n}{i} \big((\delta - a \cdot 1)^{n-i}\big) \cdot \big((\delta - b \cdot 1)^i y\big)$$

for $x, y \in \mathfrak{U}$ (which can be shown by induction on n). Now if $x \in \mathfrak{U}_a$ and $y \in \mathfrak{U}_b$, then $\sigma(xy) = (a+b)xy$ as $xy \in \mathfrak{U}_{a+b}$, which could be 0. Moreover,

$$(\sigma x)y + x(\sigma y) = (a+b)xy$$

By the directness of the sum

$$\mathfrak{U} = \coprod_a \mathfrak{U}_a$$

if follows that σ is a derivation. □

6.10. **Cartan's Criterion.** Ultimately, our goal is to represent a Lie algebra. Though we will eventually focus having to only look at simple diagrams which classify the Lie algebras, as we did in the Representation Theory section, we want to do be thinking of the simplest parts of a Lie algebra - that is, the semisimple parts. Ultimately, there is a deep connection between solvability and simplicity. Here, we develop a criterion for solvability of a Lie algebra based on the trace of special endomorphisms of the Lie algebra. This is called the Cartan criterion.

Lemma 6.5. *Let $A \subset B$ be two subspaces of $\mathfrak{gl}(V)$ with V being a finite dimensional vector space. Let*

$$M = \{x \in \mathfrak{gl}(V) \,|\, [x, B] \subset A\}$$

Suppose that $x \in M$ satisfied trace $(xy) = 0$ for all $y \in M$, then x is nilpotent.

Proof: Let $x = s + n$, where $s = x_s$ and $n = x_n$, be the Jordan decomposition of x. Now fix a basis $v_1, \cdots, v_m$ of V relative to which s has diagonal matrix with $a_1, \cdots, a_m$ along the diagonal. Let E be the vector subspace of F over the prime field $\mathbb{Q}$ spanned by the eigenvalues $a_1, \cdots, a_m$. We need to show that $s = 0$. Equivalently, we can show that

$E = 0$. Because E has finite dimension over $\mathbb{Q}$, it suffices to that the dual space E^* is 0. That is, any linear function $f : E \to \mathbb{Q}$ is 0.

Now given any f, let y be the element of $\mathfrak{gl}(V)$ whose matrix relative to our given basis is a diagonal matrix with $f(a_1), \cdots, f(a_m)$ along the diagonal. Now if $\{e_{ij}\}$ is the corresponding basis of $\mathfrak{gl}(V)$,

$$\text{ad } s(e_{ij}) = (a_i - a_j)e_{ij} \quad , \quad \text{ad } y(eij) = \big(f(a_i) - f(a_j)\big)e_{ij}$$

Now let $r(T) \in F[T]$ be a polynomial with a constant term satisfying $r(a_i - a_j) = f(a_i) - f(a_j)$ for all pairs i, j. Such a $r(T)$ exists, as we can show from Lagrange interpolation. Moreover, there is no ambiguity in the assigned values because if $a_i - a_j = a_k - a_l$ then by the linearity of f, we have $f(a_i) - f(a_j) = f(a_k) - f(a_l)$. Then ad $y = r$ ad s.

By Lemma 6.9, we know that ad s is the semisimple part of ad x. We can then write it as a polynomial in ad x without constant term. Therefore, ad y is a polynomial in ad x without a constant term. By assumption, ad x send B into A. Then we also have ad $y(B) \subset A$. Using the assumption that $\text{trace}(xy) = 0$, we obtain

$$\sum a_i f(a_i) = 0$$

The left side is a $\mathbb{Q}$-linear combination of elements of E. Now applying f, we obtain

$$\sum f(a_i)^2 = 0$$

But the numbers $f(a_i)$ are rational. Therefore, all the $f(a_i)$ must be 0. Then f must identically be 0 as a_i spans E. □

Theorem 6.22. *(Cartan's Criterion) Let L be a subalgebra of $\mathfrak{gl}(V)$ with V begin a finite dimensional vector space. Suppose then that the trace$(xy) = 0$ for all $x \in [L, L]$ and $y \in L$. Then L is solvable.*

Proof: It suffices to prove that $[L, L]$ is nilpotent or that all $x \in [L, L]$ are nilpotent endomorphisms by Engel's Theorem and Lemma 6.3. Now applying Lemma 6.3 with V given, $A = [L, L]$ and $B = L$, we have

$$M = \{x \in \mathfrak{gl}(V) \,|\, [x, L] \subset [L, L]\}$$

We also have $l \subset M$. Now if $[x, y]$ is a generator of $[L, L]$ and if $z \in M$, then using the identity

$$\text{trace}([x, y]z) = \text{trace}(x[y, z])$$

(this identity only holds for x, y, z which are endomorphisms in a finite dimensional vector space) we know that $\text{trace}([x, y]z) = \text{trace}(x[y, z]) = \text{trace}([y, z]x)$. But by the definition of M, $[y, z] \in [L, L]$, so the right side is 0 by hypothesis. □

Corollary 6.11. *Let L be a Lie algebra such that trace(ad x, ad y) $= 0$ for all $x \in [L, L]$ and $y \in L$, then L is solvable.*

Proof: We can apply Cartan's Criterion to the adjoint representation of L, which tells us that ad L is solvable. But then because ker ad $= Z(L)$ is solvable, L is solvable. □

6.11. **Killing Form.** Similarly to the previous section, we can develop a criterion, this time not for solvability, but rather semisimplicity.

Definition 6.24. (Killing Form) Let L be a Lie algebra. Let $x, y \in L$. Now define

$$\kappa(x, y) = \text{trace}(\text{ad } x, \text{ad } y)$$

Then κ is a symmetric bilinear form on L, called the Killing form. Furthermore, κ is associative in the sense that

$$\kappa([x, y], z) = \kappa(x, [y, z])$$

Lemma 6.6. *Let I be an ideal of L. If κ is the Killing form of L and κ_I the killing form of I view as a Lie algebra, then $\kappa_I = \kappa \mid_{I \times I}$.*

Proof: First, if W is a subspace of a finite dimensional vector space V and φ is an endomorphism of V mapping V into W, then trace(ϕ) = trace$(\phi \mid_W)$. Now if $x, y \in I$, then ad $x$ad y is an endomorphism of L mapping L into I, so its trace $\kappa(x, y)$ coincides with the trace of $\kappa_I(x, y)$ of ad $x$ad y $\mid_I = \text{ad}_I\ x\text{ad}_I\ y$. □

Example 6.26. If L is nilpotent then the Killing form of L is zero. Because L is nilpotent, there must be an $n \in \mathbb{Z}_+$ such that $L^{2n+1} = 0$. Then we have

$$(\text{ad } x \text{ ad } y)^n \in L^{2n+1} = 0$$

for all $x, y \in L$. But ad x ad y is a nilpotent endomorphism of L. Then it is the case that

$$\kappa(x, y) = \text{trace}(\text{ad } x \text{ ad } y) = 0$$

Definition 6.25. (Nondegenerate) A symmetric bilinear form $\beta(x, y)$ is called a nondegenerate if its radical S is 0, where

$$S = \{x \in L \mid \beta(x, y) = 0 \text{ for all } y \in L\}$$

Theorem 6.23. *Let L be a Lie algebra. Then L is semisimple if and only if its Killing form is nondegenerate.*

Proof: Suppose that rad $L = 0$. Let S be the radical of κ. By definition, trace(ad x, ad y) = 0 for all $x \in S$ and $y \in L$. According to Cartan's Criterion, $\text{ad}_L S$ is solvable. Therefore, S is solvable. But when S is an ideal of L, $S \subset$ rad $L = 0$ and κ is nondegenerate.

Now suppose that $S = 0$. To prove that L is semisimple, it suffices to prove that every abelian ideal I of L is included in S. Suppose that $x \in I$ and $y \in L$. Then ad x ad y maps $L \to L \to I$ and (ad x ad $y)^2$ maps L into $[I, I] = 0$. This means that ad x ad y is nilpotent. Therefore, trace(ad y ad $y) = \kappa(x, y)$. Then $I \subset S = 0$. □

A Lie algebra is said to be the direct sum of of ideal $I_1, \cdots, I_t$ provided that $L = I_1 + \cdots + I_t$. This forces $[I_i, I_j] \subset I_i \cap I_j = 0$ if $i \neq j$. We write

$$L = I_1 \oplus \cdots \oplus I_t$$

Theorem 6.24. *Let L be semisimple. Then there exist ideal $L_!, \cdots, L_t$ of L which are simple as Lie algebras, such that*

$$L = L_1 \oplus \cdots \oplus L_t$$

Every simple ideal of L coincides with one of the L_i. Moreover, the Killing form of L_i is the restriction of κ to $L_i \times L_i$.

Proof: Let I be an arbitrary ideal of L. Then $I^\perp = \{x \in L \,|\, \kappa(x, y) = 0 \text{ for all } y \in I\}$ is also an ideal because of the associativity of κ. Applying Cartan's Criterion to the Lie algebra I yields that the ideal $I \cap I^\perp$ of L is solvable (therefore, it is 0). But since $\dim I + \dim I^\perp = \dim L$, we must have

$$L = I \oplus I^\perp$$

Now we apply induction to the dim L to obtain a decomposition into a direct sum of simple ideals. If L has no nonzero proper ideal, then L is simple already. Otherwise, let L_1 be a minimal nonzero ideal. Then we have $L = L_1 \oplus L_1^\perp$. Moreover, any ideal of L_1 is also an ideal of L, so L_1 is semisimple. By induction, it splits into a direct sum of simple ideals, which are also ideals of L. Then the corresponding decomposition of L follows.

Now we prove uniqueness. If I is any simple ideal of L, then $[I, L]$ is also an ideal of I. This ideal is also nonzero as $Z(L) = 0$. This forces $[I, L] = I$. However, $[I, L] = [I, L_1] \oplus \cdots \oplus [I, L_t]$. So all but one of the summands must be 0. Suppose that $[I, L_i] = I$. Then $I \subset L_i$ and $I = L_i$ because of Lemma 6.6. □

Corollary 6.12. *If L is semisimple, then $L = [L, L]$ and all ideals and homomorphic images of L are semisimple. Moreover, each ideal of L is a sum of certain simple ideals of L.*

The nondegeneracy of the Killing form has many implications. The following gives an important structural aspect of semisimple Lie algebras.

Theorem 6.25. *If L is semisimple then ad L = der L. (Every derivation of L is an inner derivation.)*

Proof: If L is semisimple, then $Z(L) = 0$. Therefore, $L \to \text{ad } L$ is an isomorphism of Lie algebras. In particular, $M = \text{ad } L$ itself has a nondegenerate Killing form, following from Theorem 6.23. If $D = \text{der } L$, $[D, M] \subset M$. Then by Lemma 6.6, κ_M is the restriction to $M \times M$ of the Killing form κ_D of D. In particular, if $I = M^\perp$ is the subspace of D orthogonal to M under κ_D, then the non degeneracy of κ_M forces $I \cap M = 0$. Both I and M are ideals of D, so we have $[I, M] = 0$. If $\delta \in I$, this forces ad $\delta x = 0$ for al l$x \in L$, given the fact that

$$[\delta, \text{ad } x] = \text{ad } \delta x \text{ for all } x \in L, \delta \in \text{der } L$$

Therefore, $\delta x = 0$ for all $x \in L$ because $\delta = 0$ and ad x is injective. But then $I = 0$ and der $L = M =$ ad L. $\square$

Example 6.27. We will show that L is solvable if and only if $[L, L]$ is in the radical of the Killing form. First, suppose that $[L, L]$ is in the radical of the Killing form. Then for $x \in [L, L]$ and $y \in L$, we have

$$\kappa(x, y) = \text{trace}(\text{ad } x \text{ ad } y) = 0$$

But then L is solvable. Now on the other hand, suppose that L is solvable. By Lie's Theorem, L must have some basis $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$ such that for all $x \in L$, ad x is an upper triangular matrix relative to the chosen basis. Therefore,

$$\text{ad}[x, y] = \text{ad } x \text{ ad } y - \text{ad } y \text{ ad } x$$

is a upper triangular matrix. Furthermore, we also know that ad$[x, y]$ ad z is an upper triangular matrix for all $x, y, z \in L$. That finally yields that trace(ad$[x, y]$ ad $z) = 0$ then $[L, L] \subset \text{rad}(L)$.

6.12. **Reducibility of Representations and the Casimir Element.** We will not bother to restate all the definitions for G-modules (though here they will be L-modules), G-homomorphisms, or reducibility that we defined in the sections 5.3 and 5.4. If one lacks these definitions or has forgotten them, go to these sections now.

Moreover, the theorems of those sections, especially Schur's Lemma, still hold. So given a representation $\varphi : L \to \mathfrak{gl}(V)$ be irreducible, then the only endomorphisms of V commuting with all $\varphi(x)$ for all $x \in L$ are scalar.

Definition 6.26. (Dual) Let V be a finite dimensional L-module. Then the dual vector space V^* becomes an L-module, called the dual or contragredient, if we define for all $f \in V^*$, $v \in V$, and $x \in L$

$$(x \cdot f)(v = -f(x \cdot v)$$

One can check that the dual is an L-module. Before we used Cartan's trace criterion to determine conditions for solvability to prove that every semisimple Lie algebra L has a nondegenerate Killing form. As before, a semisimple representation $\varphi : L \to \mathfrak{gl}(V)$ is said to be faithful, or injective, representation of L. We can define a symmetric bilinear form

$$\beta(x, y) = \text{trace}(\varphi(x)\varphi(y))$$

on L. The form β is associative, as the reader can check. Moreover, the form β is also nondegenerate. We know that $\varphi(S) \cong S$ is solvable and then $S = 0$. In fact, the form β is the Killing form in the special case $\varphi =$ ad.

Now if we suppose that L is semisimple and β is any nondegenerate symmetric associative bilinear form on L and $x_1, x_2, \cdots, x_n$ is a basis of L, then there is a uniquely determined dual basis $y_1, y_2, \cdots, y_n$ relative to β satisfying

$$\beta(x_i, y_i) = \delta_{ij}$$

If $x \in L$, we can write $[x, x_i] = \sum_j a_{ij}x_j$ and $[x, y_i] = \sum_j b_{ij}y_j$. Using the associativity of β, we compute

$$a_{ik} = \sum_j a_{ij}\beta(x_j, y_k) = \beta([x, x_i], y_k) = \beta(-[x_i, x], y_k) = \beta(x_i, -[x, y_k]) = -\sum_j b_{kj}\beta(x_i, y_j) = -b_{ki}$$

Now if $\varphi : L \to \mathfrak{gl}(V)$ is any representation of L, we write

$$c_\varphi(\beta) = \sum_i \varphi(x_i)\varphi(y_i) \in \text{End } V$$

where the x_i, y_i run over dual bases relative to β. Using the identity in End V

$$[x, yz] = [x, y]z + y[x, z]$$

and the fact that $a_{ik} = -_{ki}$ for $x \in L$. Then we obtain

$$[\varphi(x), c_\varphi(\beta)] = \sum_i [\varphi(x), \varphi(x_i)]\varphi(y_i) + \sum_i \varphi(x_i)[\varphi(x), \varphi(y_i)] = \sum_{i,j} a_{ij}\varphi(x_j)\varphi(y_i) + \sum_{ij} b_{ij}\varphi(x_i)\varphi(y_j) = 0$$

In other words, $c_\varphi(\beta)$ is an endomorphism of V commuting with $\varphi(L)$.

Definition 6.27. (Casimir Element) Let $\varphi : L \to \mathfrak{gl}(V)$ be a faithful representation with nondegenerate trace form $\beta(x, y) = \text{trace}(\varphi(x)\varphi(y))$. Having a fixed basis $x_1, x_2, \cdots, x_n$ of L, we write c_φ for $c_\varphi(\beta)$ and call this the Casimir element of φ. Its trace is

$$\sum_i \text{trace}(\varphi(x_i)\varphi(y_i)) = \sum_i \beta(x_i, y_i) = \dim L$$

In the case that φ is also irreducible, Schur's Lemma implies that c_φ is a scalar. Then c_φ is independent of the basis of L chosen. However, in the case where φ is no longer faithful, we need to make several small changes. Because the ker φ is an ideal of L, it is the sum of certain simple ideals. Now let L' denote the sum of the remaining simple ideals. Then the restriction of φ to L' is a faithful representation of L'. Then the resulting element of End V is again called the Casimir element of φ and again denoted c_φ. Furthermore, it commutes wotj $\varphi(L) = \varphi(L')$. Often, it is convenient to assume that we have a faithful representation of L, which is equivalent to studying the representations of certain semisimple ideals of L. If L is simple, the only one dimensional module on which L acts trivially or the module 0 will fail to be faithful.

6.13. **Weyl's Theorem and Jordan Decomposition Preservation.**

Lemma 6.7. *Let $\varphi : L \to \mathfrak{gl}(V)$ be a representation of a semisimple Lie algebra of L. Then $\varphi(L) \subset \mathfrak{sl}(V)$. In particular, L acts trivially on any one dimensional L-module.*

Proof: This is easily shown using $L = [L, L]$ together with the fact that $\mathfrak{sl}(V)$ is the derived algebra of $\mathfrak{gl}(V)$. □

Theorem 6.26. *(Weyl's Theorem) Let $\varphi : L \to \mathfrak{gl}(V)$ be a finite dimensional representation of a semisimple Lie algebra. Then φ is completely reducible.*

Proof: First, we prove the special case where V has an L-submodule W with codimension 1. Because L acts trivially on V/W, by the preceding lemma, we can denote this module F with the short exact sequence

$$0 \to W \to V \to F \to 0$$

Then using induction on dim W, we reduce this to the case where W is an irreducible L-module, as follows. Suppose that W' is a proper nonzero submodule of W. That then yields the short exact sequence

$$0 \to W/W' \to V/W' \to F \to 0$$

By induction, there must exist a 1-dimensional L-submodule of V/W', say $\tilde{W}/W'$, complementary to W/W', i.e. the sequence splits. But then we get yet another short exact sequence

$$0 \to W' \to \tilde{W} \to F \to 0$$

But this is the same as the original short exact sequence with the exception that dim $W' <$ dim W, then induction yields a 1-dimensional submodule X complementary to W' in $\tilde{W} : \tilde{W} = W' \oplus X$. But then $V/W' = W/W' \oplus \tilde{W}/W'$. It follows that $V = W \oplus X$, since the dimensions add up to dim V and since $W \cap X = 0$.

Now assume that W is irreducible and without loss of generality that L acts faithful on V. Then let $c = c_\varphi$ be the Casimir element of φ. Since c commutes with $\varphi(L)$, c is an L-module endomorphism of V. In particular, $c(W) \subset W$ and ker c is an L-submodule of V. Because L acts trivially on V/W, that is $\varphi(L)$ send V to W, c must likewise act as a linear combination of products on the elements of $\varphi(x)$. Then c has trace 0 on V/W. However, c also acts as a scalar on the irreducible L-submodule W by Schur's Lemma. Since this scalar cannot be 0 as this would force $\text{trace}_v(c) = 0$, it follows that ker c is a 1-dimensional L-submodule of V which intersects W trivially This is the complement of W.

Now we consider the general case. Let W be a nonzero submodule of V:

$$0 \to W \to V \to V/W \to 0$$

Now let $\text{Hom}(V, W)$ be the space of linear maps $V \to W$ viewed as an L-module. Let $\mathfrak{V}$ be the subspace of $\text{Hom}(V, W)$ consisting of those maps whose restriction to W is a scalar multiplication Now say that $f \mid_W = a \cdot 1_W$ then for $x \in L$, $w \in W$

$$(x \cdot f)(w) = x \cdot f(w) - f(x \cdot w) = a(x \cdot w) - a(x \cdot w) = 0$$

Therefore, $x \cdot f \mid_W = 0$. Let $\mathfrak{W}$ be the subspace of $\mathfrak{V}$ consisting of those f whose restrictions to W is zero. The preceding calculation shows that $\mathfrak{W}$ is also an L-submodule and that L maps $\mathfrak{V}$ into $\mathfrak{W}$. Moreover, $\mathfrak{V}/\mathfrak{W}$ has dimensions one since each $f \in \mathfrak{V}$ is determined module $\mathfrak{W}$ by the scalar $f \mid_W$. But then this brings us to the case where

$$0 \to \mathfrak{W} \to \mathfrak{V} \to F \to 0$$

From the first part of the proof, we know that $\mathfrak{V}$ has a 1-dimensional submodule complementary to $\mathfrak{W}$. Let $f : V \to W$ span t. Then after multiplication by a nonzero scalar,

we can assume that $f\mid_W = 1_W$. But then saying that L kills f is just to say that

$$0 = (x \cdot f)(v) = x \cdot f(v) - f(x \cdot v)$$

or that f is an L-homomorphism. Therefore, $\ker f$ is an L-submodule of V. But since f maps V into W and acts as 1_W on W, we conclude that

$$V = W \oplus \ker f$$

□

Weyl's theorem can be considered the fundamental theorem for representations of semisimple Lie algebras L. This theorem has many broad reaching applications and implications. One such is showing that the abstract Jordan decomposition in Proposition 6.3 is compatible with various linear representations of L.

Theorem 6.27. *Let $L \subset \mathfrak{gl}(V)$ be a semisimple linear Lie algebra with V being finite dimensional. Then L contains the semisimple and nilpotent parts in $\mathfrak{gl}(V)$ of all its elements. In particular, the abstract and usual Jordan decompositions in L coincide.*

Proof: First, observe that the second implication of the theorem follows from the first as the Jordan decomposition is unique. To it will suffice to show the first part. Let $x \in L$ be arbitrary with Jordan decomposition $x = x_s + x_n$ in $\mathfrak{gl}(V)$. We need to show that x_s, x_n are in L. Now since $\operatorname{ad} x(L) \subset L$, from P(3) of Proposition 6.3, we know that $\operatorname{ad} x_s(L) \subset L$ and that $\operatorname{ad} x_n(L) \subset L$, where $\operatorname{ad} = \operatorname{ad}_{\mathfrak{gl}(V)}$. That is to say, $x_s, x_n \in N_{\mathfrak{gl}(V)}(L) = N$, which is a Lie subalgebra of $\mathfrak{gl}(V)$ with L as an ideal. It cannot be the case that $N = L$ as $L \subset \mathfrak{sl}(V)$ but the scalars lie in N but not in L. Therefore, we need to show that x_s, x_n are in smaller subalgebras than N, which are equal to L. Assume that W is any L-submodule of V and define

$$L_W = \{y \in \mathfrak{gl}(V) \mid y(W) \subset W \bigwedge \operatorname{trace}(y\mid_W) = 0\}$$

But since $L = [L, L]$, we know that L lies in all such L_W. Let L' be the intersection of N with all such L_W. Clearly, L' is a subalgebra of N that includes L as an ideal. Moreover, if $x \in L$ then x_s, x_n also lie in L_W and therefore in L'.

We only need to show now that $L = L'$. But since L' is a finite dimensional L-module, Weyl's Theorem says we can write $L' = L \oplus M$ for some L-submodule M. But $[L, L'] \subset L$ because $L' \subset N$. Then the action of L on M is trivial. Let W be any irreducible L-submodule of V. If $y \in M$ then $[L, y] = 0$. Then by Schur's Lemma, this implies that y acts on W as a scalar. On the other hand, $\operatorname{trace}(y\mid_W) = 0$ as $y \in L_W$. Therefore, y acts on W a zero. Then V can be written as a direct sum of irreducible L-submodules by Weyl's Theorem. So $y = 0$ meaning that $M = 0$ and $L = L'$. □

Corollary 6.13. *Let L be a semisimple Lie algebra with $\varphi : L \to \mathfrak{gl}(V)$ a finite dimensional representation of L. If $x = s + n$ is the abstract Jordan decomposition of $x \in L$, then $\varphi(x) = \varphi(s) + \varphi(n)$ is the usual Jordan decomposition of $\varphi(x)$.*

Proof: The algebra $\varphi(L)$ is spanned b the eigenvectors of $\mathrm{ad}_{\varphi(L)}\varphi(s)$, since L has this property relative to ad s. Therefore, $\mathrm{ad}_{\varphi(L)}\varphi(s)$ is semsimple. Similarly, $\mathrm{ad}_{\varphi(L)}\varphi(n)$ is nilpotent and commutes with $\mathrm{ad}_{\varphi(L)}\varphi(s)$. Thus, $\varphi(x) = \varphi(s) + \varphi(n)$ is the abstract Jordan decomposition of $\varphi(x)$ in the semisimple Lie algebra of $\varphi(L)$. Then using the above theorem, we obtain the desired result. □

6.14. **Weights and Maximal Vectors.** Here we will present several important concepts in the representations of Lie algebras via creating the representations for $\mathfrak{sl}(2, F)$, where F is an algebraically closed field. But why $\mathfrak{sl}(2, F)$? In fact, $\mathfrak{sl}(2, F)$ appears in many other Lie algebras in one form or another. One can often take representations of other Lie algebras and "rephrase" them in terms of representations of $\mathfrak{sl}$. So throughout this section, when we refer to a Lie algebra L, we mean $\mathfrak{sl}(2, F)$. First, recall that $\mathfrak{sl}(2, F)$ has a basis

$$x = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad , \quad y = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \quad , \quad h = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

and notice $[h, x] = 2x$, $[y, h] = 2y$, and $[x, y] = h$. Now let V be an arbitrary L-module. Because h must be semisimple, h must act diagonally on V. Because F is algebraically closed, all possible eigenvalues lie in F. Then we can decompose V as a direct sum of eigenspaces

$$V_\lambda = \{v \in V \,|\, h \cdot v = \lambda v\}$$

with $\lambda \in F$. When λ is not an eigenvalue for the endomorphism of V representing h, $V_\lambda = \{\mathbf{0}\}$. But when V_λ is nonzero, we call λ a weight of $h \in V$ and call V_λ a weight space.

Definition 6.28. (Weight Space) Let V be a representation of a Lie algebra L over a field F. A weight on L over F is a linear map, $\lambda : L \to F$ such that $\lambda([x, y]) = 0$ for all $x, y \in L$. Then the weight space of the representation V with weight λ is the subspace

$$V_\lambda = \{v \in V \,|\, r \cdot v = \lambda_r v \text{ for all } r \in L\}$$

If V happens to be the direct sum of its weight spaces, i.e. if

$$V = \bigoplus_{\lambda \in L^*} V_\lambda$$

then V is called a weight module.

Lemma 6.8. *If $v \in V_\lambda$, then $x \cdot v \in V_{\lambda+2}$ and $y \cdot v \in V_{\lambda-2}$.*

Proof: Notice that

$$h \cdot (x \cdot v) = [h, x] \cdot v + x \cdot h \cdot v = 2 \cdot x \cdot v + \lambda x \cdot v = (\lambda + 2)x \cdot v$$

This holds similarly for y. □

Lemma 6.8 tells us that x, y are represented by nilpotent endomorphisms of V. Now because V is finite dimensional and the sum is direct,

$$V = \bigsqcup_{\lambda \in F} V_\lambda$$

there must be a $V_\lambda \neq 0$ with $V_{\lambda+2} = 0$. For any λ, we call any such nonzero vector $v \in V_\lambda$ a maximal vector of weight λ. We will now classify all the irreducible modules for $\mathfrak{sl}(2, F)$. Accordingly, assume that V is an irreducible L-module. Then choose a maximal vector, say $v_0 \in V_\lambda$ and set $v_{-1} = 0$ and $v_i = \frac{1}{i!} y^i \cdot v_0$ for $i \geq 0$.

Lemma 6.9. *For all $i \geq 0$,*

1. $h \cdot v_i = (\lambda - 2i)v_i$
2. $y \cdot v_i = (i+1)v_{i+1}$
3. $x \cdot v_i = (\lambda - i + 1)v_{i-1}$

Proof:

1. This follows from repeated use of the previous lemma.
2. This follows from the definition.
3. We use induction on i. If $i = 0$, this is evident since $v_{-1} = 0$. Then

$$\begin{aligned} ix \cdot v_i =& x \cdot y \cdot v_{i-1} \\ =& [x, y] \cdot v_{i-1} + y \cdot x \cdot v_{i-1} \\ =& h \cdot v_{i-1} + y \cdot x \cdot v_{i-1} \\ =& (\lambda - 2(i-1))v_{i-1} + (\lambda - i + 2)y \cdot v_{i-2} \\ =& (\lambda - 2i + 2)v_{i-1} + (i-1)(\lambda - i + 2)v_{i-1} \\ =& i(\lambda - i + 1)v_{i-1} \end{aligned}$$

But division by i yields the necessary result. □

From the lemma, we know that the nonzero v_i are linearly independent. Now because V is finite dimensional, if we let m be the smallest integer for which $v_m \neq 0$ and $v_{m+1} = 0$, $v_{m+i} = 0$ for all $i > 0$. Then combining all the parts from our lemma, the subspace of V with basis $(v_0, v_1, \cdots, v_m)$ is an L-submodule that is not the zero subspace. But because V is an irreducible L-module by assumption, the subspace must be all of V. Furthermore, relative to the ordered basis $(v_0, v_1, \cdots, v_m)$, the matrices of the endomorphisms representing x, y, h can be written down explicitly with h yielding a diagonal matrix while x, y yield an upper and lower triangular nilpotent matrix, respectively.

Furthermore, (3) implies that for $i = m+1$, $ix \cdot v_i = 0$ while the right side is $(\lambda - m)v_m$. Since $v_m \neq 0$, we can then conclude that $\lambda = m$. This says that the weight of a maximal vector is nonnegative integers that is less than the dim V. Such a λ is called the highest weight of V. If each weight μ occurs with multiplicity 1, that is dim $V_\mu = 1$ if $V_\mu \neq 0$, since V determines λ uniquely ($\lambda =$ dim $V - 1$), the maximal vector v_0 is the only possible one in V with the exception of nonzero scalar multiples.

Theorem 6.28. *Let V be an irreducible module for $L = \mathfrak{sl}(2, F)$, then*

1. *Relative to h, V is the direct sum of weight spaces V_μ, $\mu = m, m-2, \cdots, -(m-2), -m$, where $m + 1 =$ dim V and dim $V_\mu = 1$ for each μ.*
2. *V has, up to nonzero scalar multiples, a unique maximal vector whose weight, called the heightest weight of V, is m.*

3. The action of L on V is given explicitly by the formulas from Lemma 6.9 - if the basis is chosen in the prescribed fashion. In particular, there exists at most one irreducible L-module, up to isomorphism, of each possible dimension $m+1$, with $m \geq 0$. □

Corollary 6.14. *Let V be any finite dimensional L-module with $L = \mathfrak{sl}(2, F)$. Then the eigenvalues of h on V are all integers and each occurs along with its negative with equal multiplicity. Moreover, in any decomposition of V into direct sums of irreducible submodules, the number of summands is precisely dim V_0 + dim V_1.*

Proof: If $V = 0$, then there remains nothing more to show. So suppose that V is nonzero and finite dimensional. Using Weyl's Theorem, we write V as a direct sum of irreducible submodules. Then the first assertion is clear. The second follows from the observation that each of the irreducible L-modules has a unique occurrence of either weight 0 or weight 1, but not both. □

We have shown the properties of the irreducible submodules of $\mathfrak{sl}(2, F)$. However, we have not answered the question if there are modules with highest weight $m = 0, 1, 2, \cdots$. We can construct some of these easily. For example, the trivial module with dimension 1, the natural representation with dimension 2, the adjoint representation with dimension 3. But Lemma **??** actually allows us to go further. We can use the lemma to define an irreducible representation of L on an $m+1$ dimensional vector space over F with basis $(v_0, v_1, \cdots, v_m)$, called $V(m)$. Moreover, there is deep symmetry in this $V(m)$. Let $\varphi : L \to \mathfrak{gl}(V(m))$ be the irreducible representation of highest weight m. We define an automorphism of $V(m)$ by

$$\tau = e^{\varphi(x)} e^{\varphi(-y)} e^{\varphi(x)}$$

Assuming that $m > 0$ makes the representation faithful as L is simple. Then conjugating $\varphi(h)$ by τ produces the same effect as applying

$$e^{\text{ad } \varphi(x)} e^{\text{ad } \varphi(-y)} e^{\text{ad } \varphi(x)}$$

to $\varphi(h)$. But because $\varphi(L) \cong L$, we can calculate this action explicitly. We then conclude that

$$\tau\varphi(h)\tau^{-1} = -\varphi(h) \quad \text{or} \quad \tau\varphi(h) = -\varphi(h)\tau$$

So τ sends the basis vector v_i with weight $m - 2i$ to the basis vector v_{m-i} with weight $-(m - 2i)$. Furthermore, if V is any finite dimensional L-module, then τ interchanges positive and negative weight spaces.

6.15. **Root Space Decompositions.** If L is a nonzero semisimple Lie algebra, it will be our goal to further explore the structure that L must have. We will approach this via its adjoint representation. The most powerful tool in this approach will be the Killing form. So first, suppose that L consists entirely of nilpotent, that is ad-nilpotent, elements. Then by Engel's Theorem, L too must be nilpotent. If L is not entirely nilpotent, then we can find some $x \in L$ with semisimple part x_s in the abstract Jordan decomposition being

nonzero. Then L has some nonzero subalgebra with the span of x_s consisting of semisimple elements. Such a subalgebra is called toral.

Definition 6.29. (Toral) Let L be a finite dimensional semisimple Lie algebra. Then the toral subalgebra T is the subalgebra consisting of all semisimple elements of L.

The toral algebra comes from the corresponding concept of Lie groups is the maximal torus. In fact, the toral is the corresponding Lie algebra of the maximal torus. Just as with a maximal torus, we use the toral to understand the properties of the larger structure. We start with the following, not surprising, lemma.

Lemma 6.10. *A toral subalgebra of L is abelian.*

Proof: Let T be toral. We need show that $\mathrm{ad}_T\ x = 0$ for all $x \in T$. But because ad x is diagonalizable, ad x being semisimple and F being algebraically closed. This amounts to showing that $\mathrm{ad}_T\ x$ has no nonzero eigenvalues. Suppose to the contrary that $[x, y] = ay$ with $a \neq 0$ for some nonzero $y \in T$. Then $\mathrm{ad}_T\ y(x) = -ay$ is itself an eigenvector of $\mathrm{ad}_T\ y$ with eigenvalue 0. ON the other hand, we can write x as some linear combination of eigenvectors of $\mathrm{ad}_T\ y$, with y being semisimple. Then after applying $\mathrm{ad}_t\ y$ to x, all that remains is a combination of eigenvectors which belong to nonzero eigenvalues, if any. However, this is a contradiction. □

Example 6.28. The maximal toral subalgebra, T, of $\mathfrak{sl}(n, F)$ is the set of all diagonal matrices with trace 0.

Now fix some maximal toral subalgebra T of L, that is a toral subalgebra that is not properly included in any other. Because T is abelian, $\mathrm{ad}_L\ T$ is a commuting family of semisimple endomorphisms of L. Now that means that $\mathrm{ad}_L\ T$ is also diagonalizable. That is, L is the direct sum of the subspaces

$$L_\alpha = \{x \in L \,|\, [t, x] = \alpha(h)x \text{ for all } t \in T\}$$

where α ranges over T^*. Notice that L_0 is simply $C_L(T)$, the centralizer of H. Because of the previous lemma, it must include T. Then the set of all nonzero $\alpha \in T^*$ for which $L_\alpha \neq 0$, is denoted by Φ. The elements of Φ are called the roots of L relative to T. Moreover, the number of such roots is finite. In this notation, we have a root space decomposition, also called the Cartan decomposition:

$$L = C_L(T) \bigoplus \bigsqcup_{\alpha \in \Phi} L_\alpha$$

This space and the reason for its name (including why the roots are finite in number) will appear later. It will be our goal to prove that $T = C_L(T)$ and then give some properties of the roots, or at least describe them.

Proposition 6.4. *For all $\alpha, \beta \in H^*$, $[L_\alpha, L_\beta] \subset L_{\alpha+\beta}$. If $x \in L_\alpha$, then ad x is nilpotent. If $\alpha, \beta \in H^*$ and $\alpha + \beta \neq 0$, then L_α is orthogonal to L_β, relative to the Killing form κ of L.*

Proof: The first part of the proposition follows quite trivially from the Jacobi Identity. Moreover, the second statement of the proposition follows from the first. We then only need show the final statement of the proposition. First, find $t \in T$ for which $(\alpha+\beta)(t) \neq 0$. Then if $x \in L_\alpha$ and $y \in L_\beta$, associativity of the form allows us to write $\kappa([t,x],y) = -\kappa([x,t],y) = -\kappa(x,[t,y])$ or $\alpha(t)\kappa(x,y) = -\beta(t)\kappa(x,y)$ or $(\alpha+\beta)(t)\kappa(x,y) = 0$. This then forces $\kappa(x,y) = 0$. □

Corollary 6.15. *The restriction of the Killing form to $L_0 = C_L(T)$ is nondegenerate.*

Proof: We know from Theorem 6.24 that κ is nondegenerate. On the other hand, L_0 is orthogonal to all L_α for all $\alpha \in \Phi$ from the previous proposition. If $z \in L_0$ is orthogonal to L_0 as well, then $\kappa(z, L) = 0$, forcing $z = 0$. □

To prove that the toral subalgebra is it its own centralizer (when it is maximal), we need the following basic lemma.

Lemma 6.11. *If x, y are commuting endomorphisms of a finite dimensional vector space with y nilpotent, then xy is nilpotent. In particular, trace$(xy) = 0$.*

Proposition 6.5. *Let T be a maximal toral subalgebra of L. Then $T = C_L(T)$.*

Proof:

1. C contains the semisimple and nilpotent parts of its elements: To say that $x \in C_L(T)$ is equivalent to ad x maps the subspace T of L into the subspace 0. But then by the Jordan-Chevalley decomposition, $(\text{ad x})_s = \text{ad } x_s$ and $(\text{ad } x)_n = \text{ad } x_n$.
2. All semisimple elements of C lie in T: If x is semisimple and centralizes T, then $T + Fx$, which must be a abelian subalgebra of L, is toral. The sum of commuting semisimple elements is again semisimple. But by the maximality of T, $T + Fx = T$, so $x \in T$.
3. The restriction of κ to T is nondegenerate: Let $\kappa(t, T) = 0$ for some $t \in T$. We need to show that $t = 0$. If $x \in C$ is nilpotent, then the fact that $[x, T] = 0$ and the fact that ad x is nilpotent together imply with the above lemma that trace(ad x, ad y) $= 0$ for all $y \in T$ or $\kappa(x, T) = 0$. But then (1) and (2) imply that $\kappa(t, C) = 0$. Then $t = 0$, the restriction of κ to C being nondegenerate by 6.15.
4. C is nilpotent: If $x \in C$ is semisimple, then by $x \in T$ by (2) and $\text{ad}_C\, x$ is nilpotent. On the other hand, if $x \in C$ is nilpotent, then $\text{ad}_C\, x$ is of course nilpotent. Let $x \in C$ be an arbitrary element. Then $x = x_s + x_n$. Since both $x_s, x_n \in C$, $\text{ad}_C\, x$ is the sum of commuting nilpotents and is therefore itself nilpotent. But then by Engel's Theorem, C is nilpotent.
5. $T \cap [C, C] = 0$: Since κ is associative and $[T, C] = 0$, $\kappa(T, [C, C]) = 0$, we can then apply (3) to complete the fact that $T \cap [C, C] = 0$.
6. C is abelian: If not, then $[C.C] \neq 0$. Since C is nilpotent, by (4), we have $Z(C) \cap [C, C] \neq 0$. Then $z \neq 0$ lie in the intersection. Then by (2) and (5), z cannot be semisimple. Its nilpotent part n is therefore nonzero and lies in C by (1). But then it also lies in $Z(C)$ by the Jordan-Chevalley decomposition. Then the preceding lemma implies that $\kappa(n, C) = 0$, contradicting Corollary 6.15.

7. $C = H$: If this is not the case, then C contains a nonzero nilpotent element, x, by (1) and (2). Then by the preceding lemma and (6), $\kappa(x, y) = \text{trace}(\text{ad } x, \text{ad } y) = 0$ for all $y \in C$, contradicting Corollary 6.15. □

Corollary 6.16. *The restriction of κ to H is nondegenerate.*

But this corollary lets us identify T with T^*. But $\varphi \in T^*$ corresponds to the unique element $t_\varphi \in H$ that satisfy $\varphi(t) - \kappa(t_\varphi, t)$ for all $t \in T$. Particularly, Φ corresponds to the subset $\{t_\alpha \,|\, \alpha \in \Phi\}$ of T.

6.16. **Orthogonality/Integrality Properties.** Here our goal will be to obtain more precise information about the root space decomposition by examining the Kiling form. We already know that $\kappa(L_\alpha, L_\beta) = 0$ if $\alpha, \beta \in H^*$ then $\alpha + \beta \neq 0$. Moreover, $\kappa(H, L_\alpha) = 0$ for all $\alpha \in \Phi$, then by Proposition kasdfkjasbjdglakjlsdg, the restriction of κ to H is nondegenerate.

Proposition 6.6. *1. Φ spans H^*.*
2. Let $\alpha \in \Phi$, $x \in L_\alpha$, $y \in L_{-\alpha}$, then $[x, y] = \kappa(x, y)t_\alpha$.
3. If $\alpha \in \Phi$, then $[L_\alpha, L_{-\alpha}]$ is 1-dimensional with basis t_α.
4. $\alpha(t_\alpha) - \kappa(t_\alpha, t_\alpha) \neq 0$ for $\alpha \in \Phi$.
5. If $\alpha \in \Phi$ and x_α is any nonzero element of L_α, then there exists $y_\alpha \in L_{-\alpha}$ such that $x_\alpha, y_\alpha, h_\alpha = [x_\alpha, y_\alpha]$ span a 3-dimensional simple subalgebra of L isomorphic to $\mathfrak{sl}(2, F)$ via

$$x_\alpha \mapsto \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad , \quad y_\alpha \mapsto \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \quad , \quad h_\alpha \mapsto \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

6. $h_\alpha = \frac{2t_\alpha}{\kappa(t_\alpha, t_\alpha)}$; $h_\alpha = -h_\alpha$.

Proof:

1. If Φ fails to span H^*, then by duality there exists nonzero $h \in H$ such that $\alpha(h) = 0$ for all $\alpha \in \Phi$. But this implies that $[h, L_\alpha] = 0$ for all $\alpha \in \Phi$. Since $[h, H] = 0$, this forces $[h, L] = 0$ or $h \in Z(L) = 0$. But this is impossible.
2. Let $\alpha \in \Phi$. If $-\alpha \notin \Phi$, that is $L_{-\alpha} = 0$, then $\kappa(L_\alpha, L_\beta) = 0$ for all $\beta \in H^*$. Therefore, $\kappa(L_\alpha, L) = 0$, contradicting the non degeneracy which is absurd.
3. Let $\alpha \in \Phi$, $x \in L_\alpha$, and $y \in L_{-\alpha}$. Let $h \in H$ be arbitrary. The associativity of κ implies:

$$\kappa(h, [x, y]) = \kappa([h, x], y) = \alpha(h)\kappa(x, y) = \kappa(t_\alpha, h)\kappa(x, y) = \kappa(\kappa(x, y)t_\alpha, h) = \kappa(h, \kappa(x, y)t_\alpha)$$

   This implies that H is orthogonal to $[x, y] - \kappa(x, y)t_\alpha$, forcing $[x, y] = \kappa(x, y)t_\alpha$.
4. From (3), we have that t_α spans $[L_\alpha, L{-}\alpha]$, provided that $[L_\alpha, L_{-\alpha}] \neq 0$. Let $0 \neq x \in L_\alpha$. If $\kappa(x, L_{-\alpha}) = 0$, then $\kappa(x, L) = 0$. But this is impossible as κ is nondegenerate. Therefore, we can find $0 \neq y \in L_{-\alpha}$ for which $\kappa(x, y) \neq 0$. But then by (3), we have $[x, y] \neq 0$.
5. Suppose that $\alpha(t_\alpha) = 0$, so that $[t_\alpha, x] = 0 = [t_\alpha, y]$ for all $x \in L_\alpha$, $y \in L_{-\alpha}$. Then as in (4), we can then find a x, y satisfying $\kappa(x, y) \neq 0$. Modifying one or the other by a scalar,

we can assume that $\kappa(x,y)=1$. Then $[x,y]=t_\alpha$ by (3), it follows that the subspace S of L spanned by x,y,t_α is a 3-dimensional solvable algebra, $S\cong \mathrm{ad}_L\, S\subset \mathfrak{gl}(L)$. In particular, $\mathrm{ad}_L\, s$ is nilpotent for all $s\in[S,S]$, so $\mathrm{ad}_L\, t_\alpha$ is both semisimple and nilpotent, that is to say $\mathrm{ad}_L\, t_\alpha=0$. This then implies that $t_\alpha\in Z(L)=0$, contrary to the assumption on t_α.

6. Given $0\neq x_\alpha\in L_\alpha$, find $y_\alpha\in L_{-\alpha}$ such that
$$\kappa(x_\alpha,y_\alpha)=\frac{2}{\kappa(t_\alpha,t_\alpha)}$$
This is possible given (5) and the fact that $\kappa(x_\alpha,L_{-\alpha})\neq 0$. Now set $h_\alpha=\frac{2t_\alpha}{\kappa(t_\alpha,t_\alpha)}$. Then we have $[x_\alpha,y_\alpha]=h_\alpha$ by (3). Moreover, we have
$$[h_\alpha,x_\alpha]=\frac{2}{\alpha(t_\alpha)}[t_\alpha,x_\alpha]=\frac{2\alpha(t_\alpha)}{\alpha(t_\alpha)}x_\alpha=2x_\alpha$$
Similarly, $[h_\alpha,y_\alpha]=-2y_\alpha$. So $x_\alpha,y_\alpha,h_\alpha$ spans a 3-dimensional subalgebra of L with the same multiplication table as $\mathfrak{sl}(2,F)$.
7. Recall that t_α is defined by $\kappa(t_\alpha,h)=\alpha(h)$ for all $h\in H$. This shows that $t_\alpha=-t_{-\alpha}$ and because of how h_α is defined, the assertion follows. □

Let α be a root. in light of Proposition 6.6 (2), $-\alpha$ is also a root. Now let $S_\alpha\cong\mathfrak{sl}(2,F)$ be the subalgebra of L constructed in Proposition 6.6 (5). Using Weyl's Theorem and the orthogonality properties, we actually have a full description of of all finite dimensional S_α-modules. So now fix a $\alpha\in\Phi$. Then consider the subspace M of L spanned by H along with all root spaces of the form $L_{c\alpha}$ for all $c\in F^*$. This is an S_α-submodule of L. The weights of h_α on M are integers 0 and $2c=c\alpha(h_\alpha)$ for nonzero c with $L_{c\alpha}\neq 0$. Then all c here must be integral multiplies of $\frac{1}{2}$. Then S_α acts trivially on $\ker\alpha$ which is a subspace of codimension 1 in T complementary to Fh_α. Furthermore, S_α is an irreducible S_α-module of M. Then considered together, $\ker\alpha$ and S_α must exhaust the occurrences of the weight 0 for h_α. Then the only even weights occurring in M are 0 and ±2. Then it must be the case that twice a root is not a root, i.e. if α is a root then 2α is not a root. This also implies that $\frac{1}{2}\alpha$ is also not a root. Therefore, 1 cannot occur as a weight of h_α in M. It must then be the case that
$$M=H\oplus S_\alpha$$
Because $\dim L_\alpha=1$, S_α is uniquely determined as the subalgebra of L generated by L_α and $L_{-\alpha}$ and the only multiplies of a root α which are also themselves roots must be $\pm\alpha$.

Now let $K=\sum L_{\beta+i\alpha}$. Each root space is 1-dimensional and non of the $\beta+i\alpha$ can be 0. Then K is an S_α-submodule of L with a 1-dimensional weight space for the distinct integral weights $\beta(h_\alpha)+2i$ with $i\in\mathbb{Z}$ and $\beta+i\alpha\in\Phi$. Now both 1 and 0 can occur as weights in this form. This implies then that K is irreducible. The the highest weight is $\beta(h_\alpha)+2q$ and the lowest weight is $\beta(h_\alpha)-2r$ if q,r are the largest integer for which $\beta+q\alpha,\beta-r\alpha$ is a root, respectively. The weights on K then form an arithmetic progression with difference 2. Then the roots $\beta+i\alpha$ form a string, the α string through β. Finally, notice that if

$\alpha, \beta, \alpha+\beta \in \Phi$, then ad L_α maps L_β onto $L_{\alpha+\beta}$. That is, $[L_\alpha, L_\beta] = L_{\alpha+\beta}$. We summarize our results here.

Proposition 6.7. *Let L be a nonzero semisimple Lie algebra, then*

1. *If $\alpha \in \Phi$ then dim $L_\alpha = 1$. In particular, $S_\alpha = L_\alpha + L_{-\alpha} + T_\alpha$, where $T_\alpha = [L_\alpha, L_{-\alpha}]$ and for given nonzero $x_\alpha \in L_\alpha$, there exists a unique $y_\alpha \in L_{-\alpha}$ satisfying $[x_\alpha, y_\alpha] = h_\alpha$.*
2. *If $\alpha \in \Phi$, then the only scalar multiples of α which are roots are α and $-\alpha$.*
3. *If $\alpha, \beta \in \Phi$ then $\beta(h_\alpha) \in \mathbb{Z}$ and $\beta - \beta(h_\alpha)\alpha \in \Phi$. The numbers $\beta(h_\alpha)$ are called the Cartan integers.*
4. *If $\alpha, \beta, \alpha+\beta \in \Phi$, then $[L_\alpha, L_\beta] = L_{\alpha+\beta}$.*
5. *Let $\alpha, \beta \in \Phi$ and $\beta \neq \pm\alpha$. Let r, q be the largest integers for which $\beta - r\alpha$ and $\beta + q\alpha$ are roots, respectively. Then all $\beta + i\alpha \in \Phi$ for $-r \leq i \leq q$ and $\beta(h_\alpha) = r - q$.*
6. *L is generated as a Lie algebra by the root spaces L_α.* □

6.17. **Root Systems.** Similarly to Young Tableaux, we can create diagrams to help us identify and classify most Lie algebras. To do this we look at vector spaces over $\mathbb{R}$. Fix a Euclidean space $\mathbb{E}$. Geometrically, a reflection in $\mathbb{E}$ is an invertible linear transformation leaving pointwise fixed a hyperplane and sending any orthogonal vector to the plane to its negative - that is, some subspace of $\mathbb{E}$ of codimension 1 and any normal defining it. So these reflections are orthogonal in the sense that they preserve the inner product on $\mathbb{E}$. We can take any nonzero vector α and determine along with a reflecting hyperplane

$$P_\alpha = \{\beta \in \mathbb{E} \,|\, (\beta, \alpha) = 0\}$$

a reflection σ_α . We can explicitly write a formula representing the reflection

$$\sigma_\alpha(\beta) = \beta - \alpha \frac{2(\beta, \alpha)}{(\alpha, \alpha)}$$

Notice that $-\alpha$ determines the same reflection. Moreover, the reflection fixes all the points of P_α.

Lemma 6.12. *Let Φ be a finite set which spans $\mathbb{E}$. Suppose all reflections $\sigma_\alpha (\alpha \in \Phi)$ leave Φ invariant. If $\sigma \in GL(\mathbb{E})$ leaves Φ invariant, fixes pointwise a hyperplane P of $\mathbb{E}$, and sends some nonzero $\alpha \in \Phi$ to its negative, then $\sigma = \sigma_\alpha$ and $P = P_\alpha$.*

Proof: Let $\tau = \sigma\sigma_\alpha = \sigma\sigma_\alpha^{-1}$. Then $\tau(\Phi) = \Phi$ and $\tau(\alpha) = \alpha$ and τ acts as the identity on the subspace of $\mathbb{R}_\alpha$ as well as on the quotient $\mathbb{E}/\mathbb{R}_\alpha$. So all the eigenvalues of τ are 1. Therefore, the minimal polynomial of τ divides $(T-1)^l$ where $l = \dim \mathbb{E}$. But since Φ is finite, not all the vectors $\beta, \tau(\beta), \tau^2(\beta), \cdots, \tau^k(\beta)$ for $\beta \in \Phi$ and $k \geq \text{card } \Phi$ can be distinct. So Φ spans $\mathbb{E}$ and forces $\tau^k = 1$. Then the minimal polynomial of τ divides $T^k - 1$. Then this shows that τ has a minimal polynomial

$$T - 1 = \gcd(T^k - 1, (T-1)^l$$

Definition 6.30. (Root System) A subset Φ of the Euclidean space $\mathbb{E}$ is called a root system in $\mathbb{E}$ if Φ has the following:

1. Φ is finite, spans $\mathbb{E}$, and does not contain 0.
2. If $\alpha \in \Phi$ then the only multiples of $\alpha \in \Phi$ are $\pm\alpha$.
3. If $\alpha \in \Phi$ then the only reflection σ_α leaves Φ invariant.
4. If $\alpha, \beta \in \Phi$ then $\frac{2(\beta,\alpha)}{(\alpha,\alpha)} \in \mathbb{Z}$.

Of course not all these axioms for a root system are needed. Both condition (2) and condition (3) imply that $\Phi = -\Phi$. We call $l = \dim \mathbb{E}$ from above the rank of the root system Φ.

Remark 6.2. In the definition of root system, sometimes condition (2) is omitted in the definition of a root system. When this is the case, what we have defined as a root system is known as a reduced system, which makes sense because given a reflection α we only get $\pm\alpha$.

Suppose that Φ is a root system in $\mathbb{E}$ and let $\mathfrak{W}$ be the subgroup of $\mathrm{GL}(\mathbb{E})$ generated by the reflections $\sigma_\alpha(\alpha \in \Phi)$. But then by axiom (3), $\mathfrak{W}$ is a permutation on the set Φ. Then we can identify $\mathfrak{W}$ with a subgroup of the symmetric group on Φ. But since Φ is finite, $\mathfrak{W}$ is also finite. This group $\mathfrak{W}$ has a special name: the Weyl group, which we shall return to later.

Definition 6.31. (Weyl Group) Suppose that Φ is a root system in $\mathbb{E}$ and let $\mathfrak{W}$ be the subgroup of $\mathrm{GL}(\mathbb{E})$ generated by the reflections $\sigma_\alpha(\alpha \in \Phi)$. Then $\mathfrak{W}$ is called

Lemma 6.13. *Let Φ be a root system in $\mathbb{E}$ with Weyl group $\mathfrak{W}$. If $\sigma \in GL(\mathbb{E})$ leaves Φ invariant, then $\sigma\sigma_\alpha\sigma^{-1} = \sigma_{\sigma(\alpha)}$ for all $\alpha \in \Phi$ and $\frac{2(\beta,\alpha)}{(\alpha,\alpha)} = \frac{2(\sigma(\beta),\sigma(\alpha)}{(\sigma(\alpha),\sigma(\alpha))}$ for all $\alpha, \beta \in \Phi$.*

Proof: First, observe that

$$\sigma\sigma_\alpha\sigma^{-1}(\sigma(\beta)) = \sigma\sigma_\alpha(\beta) \in \Phi$$

since $\sigma_\alpha(\beta) \in \Phi$. But this is equivalent to

$$\sigma(\beta - \frac{2(\beta,\alpha)}{(\alpha,\alpha)}\alpha) = \sigma(\beta) - \frac{2(\beta,\alpha)}{(\alpha,\alpha)}\sigma(\alpha)$$

But since $\sigma(\beta)$ runs over Φ as β runs over Φ, we conclude that $\sigma\sigma_\alpha\sigma^{-1}$ leaves Φ invariant while leaving pointwise the hyperplane $\sigma(P_\alpha)$ and sending $\sigma(\alpha)$ to $-\sigma(\alpha)$. Then by Lemma 6.12, $\sigma\sigma_\alpha\sigma^{-1} = \sigma_{\sigma(\alpha)}$. Then comparing the equation above with the equation $\sigma_{\sigma(\alpha)}(\sigma(\beta)) = \sigma(\beta) - \frac{2(\sigma(\beta),\sigma(\alpha))}{(\sigma(\alpha),\sigma(\alpha))}\sigma(\alpha)$, we get the second assertion. □

Since the Weyl group and root systems are what we will use to classify the representations of Lie algebras in a way, we need to know when such representations are unique. Here, there is a natural notion of isomorphisms between Φ and Φ' with respective Euclidean spaces $\mathbb{E}$ and $\mathbb{E}'$. We say that $(\Phi, \mathbb{E})$ are $(\Phi', \mathbb{E}')$ called isomorphic if there is a vector space isomorphism (though not necessarily an isometry) $\phi : \mathbb{E} \to \mathbb{E}'$ sending Φ to Φ' such that

$$\frac{2(\phi(\beta),\phi(\alpha))}{(\phi(\alpha),\phi(\alpha))} = \frac{2(\beta,\alpha)}{(\alpha,\alpha)}$$

for each pair of roots $\alpha, \beta \in \Phi$. It then follows that $\sigma_{\phi(\alpha)}(\phi(\beta)) = \phi(\sigma_\alpha(\beta))$. Therefore, an isomorphism of root systems induces a natural isomorphism $\sigma \mapsto \phi \circ \sigma \circ \phi^{-1}$ of Weyl groups. So in view of the previous lemma, any automorphism of Φ is the same thing as an automorphism of $\mathbb{E}$ leaving Φ invariant. We then regard $\mathfrak{W}$ as a subgroup of aut Φ.

Definition 6.32. (Dual) Let Φ be a root system for a Euclidean space $\mathbb{E}$. Then

$$\Phi^v = \{\alpha^v \,|\, \alpha \in \Phi\}$$

is the dual or inverse of Φ. Moreover, Φ^v is a root system in $\mathbb{E}$ whose Weyl group is canonically isomorphic to $\mathfrak{W}$.

Example 6.29. When $l \leq 2$, we can represent Φ pictorially. When $l = 1$, there is only one possible Φ show below.

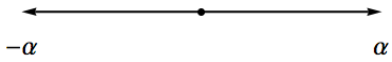
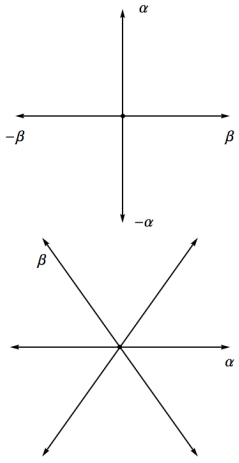
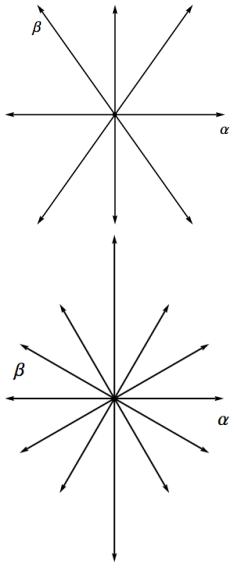


FIGURE 4. A representation of the root system Φ, where $l = 1$.

This root system is called A_1. This root system has a Weyl group of order 2. As a Lie algebra, it belongs to $\mathfrak{sl}(2, F)$.

Example 6.30. If we consider a root system of rank 2, we have 4 possible root systems.





In the definition of a root system, axiom (4) limits the possible angles between pairs of roots. We can use the formula $\|\alpha\|\|\beta\|\cos\theta = (\alpha,\beta)$. Then we have

$$\frac{2(\beta,\alpha)}{(\alpha,\alpha)} = 2\frac{\|\beta\|}{\|\alpha\|}$$

and

$$\frac{2(\beta,\alpha)}{(\alpha,\alpha)}\frac{2(\alpha,\beta)}{(\beta,\beta)} = 4\cos^2\theta$$

the last number is a nonnegative integers. Moreover, we know that $0 \leq \cos^2\theta \leq 1$. Since $\frac{2(\beta,\alpha)}{(\alpha,\alpha)}$ and $\frac{2(\alpha,\beta)}{(\beta,\beta)}$ have the same sign, the only possibilities are when $\alpha \neq \pm\beta$ and $\|\beta\| \geq \|\alpha\|$.

$\frac{2(\beta,\alpha)}{(\alpha,\alpha)}$	$\frac{2(\alpha,\beta)}{(\beta,\beta)}$	θ	$\frac{\|\beta\|^2}{\|\alpha\|^2}$
0	0	$\frac{\pi}{2}$	undetermined
1	1	$\frac{\pi}{3}$	1
-1	-1	$\frac{2\pi}{3}$	1
1	2	$\frac{\pi}{4}$	2
-1	-2	$\frac{3\pi}{4}$	2
1	3	$\frac{\pi}{6}$	3
-1	-3	$\frac{5\pi}{6}$	3

Lemma 6.14. *Let α, β be non proportional roots. If $(\alpha,\beta) > 0$ then $\alpha - \beta$ is a root. If $(\alpha,\beta) < 0$ then $\alpha + \beta$ is a root.*

Proof: The second part follows from the first when one applies to $-\beta$ in place of β. So we only need to show the first. Since (α,β) is positive if and only if $\frac{2(\alpha,\beta)}{(\beta,\beta)}$. Now either $\frac{2(\alpha,\beta)}{(\beta,\beta)}$ or $\frac{2(\beta,\alpha)}{(\alpha,\alpha)}$ is 1. If $\frac{2(\alpha,\beta)}{(\beta,\beta)} = 1$ then $\sigma_\beta(\alpha) = \alpha - \beta \in \Phi$. Similarly, if $\frac{2(\beta,\alpha)}{(\alpha,\alpha)} = 1$ then

$\beta - \alpha \in \Phi$. Therefore, $\sigma_{\beta-\alpha}(\beta - \alpha) = \alpha - \beta \in \Phi$. □

In addition, take a pair of non proportional roots α, β. Consider all roots of the form $\beta + z\alpha$ with $z \in \mathbb{Z}$, called the α-string through β. Let $r, q \in \mathbb{Z}_+$ be the largest integers for which $\beta - r\alpha \in \Phi$ and $\beta + q\alpha \in \Phi$. If some $\beta + z\alpha \notin \Phi$ for $-r \leq z \leq q$, we can find $p < s$ in this interval such that $\beta + p\alpha \in \Phi$

$$\beta + (p+1)\alpha \notin \Phi \quad , \quad \beta + (s-1)\alpha \in \Phi \quad , \quad \beta + s\alpha \in \Phi$$

Then the previous lemma implies that both $(\alpha, \beta + p\alpha) \geq 0$ and $(\alpha, \beta + s\alpha) \leq 0$. But this isn't possible since $p < s$ and $(\alpha, \alpha) > 0$. Then the α-string through β is unbroken from $\beta - r\alpha$ to $\beta + q\alpha$. So σ_α just adds or subtracts a multiple of α to any root, so the string is invariant under σ_α. Geometrically, σ_α reverses the string. Now since $\sigma_\alpha(\beta + q\alpha) = \beta - r\alpha$ and the left side is $\beta - \frac{2(\beta,\alpha)}{(\alpha,\alpha)}\alpha - q\alpha$, we know that $r - q = \frac{2(\beta,\alpha)}{(\alpha,\alpha)}$ and so every root strings have length at most 4.

6.18. **Simple Roots and the Weyl Group.**

Definition 6.33. (Base) A subset Δ of Φ is called a base if:

1. Δ is a basis of $\mathbb{E}$.
2. Each root β can be written as $\beta = \sum k_\alpha \alpha$ with $\alpha \in \Delta$ with integral coefficients k_α all nonnegative or all nonpositive.

The roots in Δ are called simple.

Given axiom (1) for a base, card $\Delta = l$ and the expression for β must be unique.

Definition 6.34. (Height) Given a root β in a base Δ, we define the height, relative to Δ, by

$$\text{ht } \beta = \sum_{\alpha \in \Delta} k_\alpha$$

If all the k_α are nonnegative, then we call β positive and write $\beta \succ 0$. If all the k_α are nonpositive, we call k_α negative and write $\beta \prec 0$.

The collection of positive roots, relative to Δ is denoted Φ^+ and the collection of negative roots, relative to Δ is denoted Φ^-. Clearly, it is the case that

$$\Phi^- = -\Phi^+$$

Now if α, β are positive roots and $\alpha + \beta$ is a root, then $\alpha + \beta$ must also be positive. Moreover, Δ defines a partial order on $\mathbb{E}$ that is compatible with the notation $\alpha \prec 0$. So now define $\beta \succ \alpha$ if and only if $\alpha - \beta$ is a sum of positive (equivalently simple roots) or if $\beta = \alpha$. It is important to note that this definition in no way implies existence.

Lemma 6.15. *If Δ is a base of Φ, then $(\alpha, \beta) \leq 0$ for all $\alpha \neq \beta$ in Δ and $\alpha - \beta$ is not a root.*

Proof: Suppose that $(\alpha, \beta) > 0$. Because $\alpha \neq \beta$ and $\alpha \neq -\beta$, then Lemma 6.14 gives that $\alpha - \beta$ is a root. But then this contradicts axiom (2) of a base. □

Definition 6.35. (Regular/Singular) Given a Euclidean space $\mathbb{E}$ and a vector $\gamma \in \mathbb{E}$, we call γ regular if $\gamma \in \mathbb{E} - \bigcup_{\alpha \in \Phi} P_\alpha$. Otherwise, γ is called singular.

Definition 6.36. (Decomposable/Indecomposable) We call $\alpha \in \Phi^+(\gamma)$ decomposable if $\alpha = \beta_1 + \beta_2$ for some $\beta_i \in \Phi^+(\gamma)$. Otherwise, α is called indecomposable.

Theorem 6.29. *Let $\gamma \in \mathbb{E}$ be regular. Then the set $\Delta(\gamma)$ of all indecomposable roots in $\Phi^+(\gamma)$ is a base of Φ and every base is obtainable in this manner.*

Proof: We prove this in steps:

1. Each root in $\Phi^+(\gamma)$ is a nonnegative $\mathbb{Z}$-linear combination of $\Delta(\gamma)$: Otherwise there is some $\alpha \in \Phi^+(\gamma)$ cannot be written as such. Choose α so that (γ, α) is as small as possible. Obviously, α itself cannot be in $\Delta(\gamma)$, so $\alpha = \beta_1 + \beta_2$ with $\beta_i \in \Phi^+(\gamma)$. Then $(\gamma, \alpha) = (\gamma, \beta_1) + (\gamma, \beta_2)$. But each of the (γ, β_i) is positive, so β_1 and β_2 must each be a nonnegative $\mathbb{Z}$-linear combination of $\Delta(\gamma)$, whence α is also. This contradicts the assumption that α is minimal.
2. If $\alpha, \beta \in \Delta(\gamma)$, then $(\alpha, \beta) \leq 0$ unless $\alpha = \beta$: Otherwise, $\alpha - \beta$ is a root, since β cannot be $-\alpha$, either $\alpha - \beta$ or $\beta - \alpha$ is in $\Phi^+(\gamma)$. If the first, $\alpha = \beta + (\alpha - \beta)$, meaning α is decomposable. If the second, $\beta = \alpha + (\beta - \alpha)$ is decomposable. This contradicts the assumption.
3. $\Delta(\gamma)$ is a linearly independent set: Suppose
$$\sum_\alpha r_\alpha \alpha = 0$$
for all $\alpha \in \Delta(\gamma)$ and $r_\alpha \in \mathbb{R}$. Separating the indices α for which $r_\alpha > 0$ from those for which $r_\alpha < 0$, we are then able to rewrite this as
$$\sum s_\alpha \alpha = \sum t_\beta \beta$$
for all $s_\alpha, t_\beta > 0$, the sets of α's and β's being disjoint. Now call $\epsilon = \sum s_\alpha \alpha$. Then
$$(\epsilon, \epsilon) = \sum_{\alpha,\beta} s_\alpha t_\beta (\alpha, \beta) \leq 0$$
This then forces ϵ to be 0. Then $0 = (\gamma, \epsilon) = \sum s_\alpha (\gamma, \alpha)$, forcing all the s_α to be 0. Similarly, all the t_β must also be 0. Moreover, this shows that any set of vectors lying strictly on one side of a hyperplane in $\mathbb{E}$ and forming pairwise obtuse angles must be linearly independent.
4. $\Delta(\gamma)$ is a base of Φ: Since $\Phi = \Phi^+(\gamma) \cup -\Phi^+(\gamma)$, we have axiom (2) satisfied from the first step. It then follows that $\Delta(\gamma)$ spans $\mathbb{E}$. Then by step 3, we have axiom (1).
5. Each base Δ has the form $\Delta(\gamma)$ for some regular $\gamma \in \mathbb{E}$: Given Δ, we choose a $\gamma \in \mathbb{E}$ such that $(\gamma, \alpha) > 0$ for all $\alpha \in \Delta$. Then from axiom (2), γ is regular and $\Phi^+ \subset \Phi^+(\gamma)$ and $\Phi^- \subset -\Phi^+(\gamma)$. But since $\Phi^+ = \Phi^+(\gamma)$, Δ consists of indecomposable elements, i.e. $\Delta \subset \Delta(\gamma)$. But we have card δ = card $\Delta(\gamma)$. Therefore, $\Delta = \Delta(\gamma)$. □

Definition 6.37. (Weyl Chambers) The hyperplanes P_α for all $\alpha \in \Phi$ partition $\mathbb{E}$ into finitely many regions, the connected components of $\mathbb{E} - \bigcup_\alpha P_\alpha$ and are called the (open) Weyl chambers of $\mathbb{E}$.

Each of the regular $\gamma \in \mathbb{E}$ belong then to one Weyl chamber, denoted $\mathfrak{C}(\gamma)$. Then saying that $\mathfrak{C}(\gamma) = \mathfrak{C}(\gamma')$ is to say that γ, γ' lie on the same side of each hyperbplane P_α for all $\alpha \in \Phi$, i.e. that $\Phi^+(\gamma) = \Phi^+(\gamma')$ or $\Delta(\gamma) = \Delta(\gamma')$. But then the Weyl chambers are in a natural one-to-one correspondence with bases. Writing $\mathfrak{C}(\Delta) = \mathfrak{C}(\gamma)$ if $\Delta = \Delta(\gamma)$, this is called the fundamental Weyl chamber relative to Δ. Moreover, $\mathfrak{C}(\Delta)$ is an open convex set of all $\gamma \in \mathbb{E}$ which satisfy the inequality $(\gamma, \alpha) > 0$ for all $\alpha \in \Delta$.

The Weyl group always sends one Weyl chamber into another. Said differently, $\sigma(\mathfrak{C}(\gamma)) = \mathfrak{C}(\sigma\gamma)$, if $\sigma \in \mathfrak{W}$ and γ is regular. However, $\mathfrak{W}$ permutes bases: σ sends Δ to $\sigma(\Delta)$, again a base. These two actions of $\mathfrak{W}$ are in fact compatible with the above correspondence between Weyl chambers and bases. It is the case that

$$\sigma(\Delta(\gamma)) = \Delta(\sigma\gamma)$$

Now if we let Δ be a fixed base of Φ. We then obtain several useful lemmas of the behavior of simple roots.

Lemma 6.16. *If α is a positive but not simple root, then $\alpha - \beta$ is a root (not necessarily positive) for some $\beta \in \Delta$.*

Proof: If $(\alpha, \beta) \leq 0$ for all $\beta \in \Delta$, it is the case that $\Delta \cup \{\alpha\}$ is a linearly independent set (see step (3) in our previous proof). But this is impossible as Δ already has a basis of $\mathbb{E}$. Then $(\alpha, \beta) > 0$ for some $\beta \in \Delta$ and then $\alpha - \beta \in \Phi$. Then we write

$$\alpha = \sum_{\gamma \in \Delta} k_\gamma \gamma$$

for all $k_\gamma \geq 0$ and some $k_\gamma > 0$ for all $\gamma \neq \beta$. Subtracting β from α yields a $\mathbb{Z}$-linear combination of simple roots with at least one positive coefficient. This forces all the coefficients to be positive, due to the uniqueness of the expression. □

Corollary 6.17. *Each $\beta \in \Phi^+$ can be written in the form $\alpha_1 + \alpha_2 + \cdots + \alpha_k$, where $\alpha_i \in \Delta$, though not necessarily distinct, in such a way that each partial sum $\alpha_1 + \cdots + \alpha_i$ is a root.*

Proof: This follows trivially from induction on ht β and the previous lemma. □

Lemma 6.17. *Let α be simple. Then σ_α permutes the positive roots other than α.*

Proof:

6.19. **Irreducible Root Systems.**

Definition 6.38. (Irreducible) Φ is called irreducible if it cannot be partitioned into the union of two proper subset such that each root in one set is orthogonal to each root in the other.

Example 6.31. Notice that A_1, A_2, B_2, G_2 are irreducible but $A_1 \times A_1$ is not.

Suppose that Δ is a base of Φ. Then Φ is irreducible if and only if Δ cannot be partitioned in the way just stated. Why? In one direction, let $\Phi = \Phi_1 \cup \Phi_2$ with $(\Phi_1, \Phi_2) = 0$. Unless Δ is wholly contained in Φ_1 and Φ_2, this induces a similar partition of Δ. However, $\Delta \subset \Phi_1$ implies that $(\Delta, \Phi_2) = 0$ or $(\mathbb{E}, \Phi_2) = 0$, because Δ spans $\mathbb{E}$. Conversely, suppose that let Φ be irreducible but $\Delta = \Delta_1 + \Delta_2$ with $(\Delta_1, \Delta_2) = 0$. Each root is conjugate to a simple root. Therefore, $\Phi = \Phi_1 \cup \Phi_2$, Φ_i is the set of roots having a conjugate in Δ_i. Recalling that $(\alpha, \beta) = 0$ implies that $\sigma_\alpha \sigma_\beta = \sigma_\beta \sigma_\alpha$. Because $\mathfrak{W}$ is generated by the σ_α for all $\alpha \in \Delta$. Using the formula for the reflection, we know that Φ_i lies in the subspace $\mathbb{E}_i$ of $\mathbb{E}$ spanned by Δ_i and we see that $(\Phi_1, \Phi_2) = 0$. This forces $\Phi_1 = \emptyset$ or $\Phi_2 = \emptyset$, then $\Delta_1 = \emptyset$ or $\Delta_2 = \emptyset$.

Lemma 6.18. *Let Φ be irreducible. Relative to the partial ordering $\prec$, there is a unique maximal root β. If $\beta = \sum k_\alpha \alpha$ for all $\alpha \in \Delta$ then all $k_\alpha > 0$.*

Proof: Let $\beta = \sum k_\alpha \alpha$ for all $\alpha \in \Delta$ be maximal in the ordering, $\beta \succ 0$. If

$$\Delta_1 = \{\alpha \in \Delta \,|\, k_\alpha > 0\}$$

and

$$\Delta_2 = \{\alpha \in \Delta \,|\, k_\alpha = 0\}$$

then $\Delta = \Delta_1 \cup \Delta_2$ is partition of Δ. No suppose that Δ_2 is nonempty. Then $(\alpha, \beta) \leq 0$ for all $\alpha \in \Delta_2$. Since Φ is irreducible, at least one of the $\alpha' \in \Delta_1$, then $(\alpha, \beta) < 0$. This implies from Lemma 6.13 that $\beta + \alpha$ is a root, contradicting the maximality of β. Therefore, Δ_2 is empty and all $k_\alpha > 0$. This also shows that $(\alpha, \beta) \geq 0$ for all $\alpha \in \Delta$ with $(\alpha, \beta) > 0$ for at least one $\alpha \in \Delta$ for which $(\alpha, \beta) > 0$. It then follows that $(\beta', \beta) > 0$ and $\beta - \beta'$ is a root, unless $\beta = \beta'$. But if $\beta - \beta'$ is a root then either $\beta \prec \beta'$ or $\beta' \prec \beta$, which is impossible. Therefore, β is unique. □

Lemma 6.19. *Let Φ is irreducible. Then $\mathfrak{W}$ acts irreducibly on $\mathbb{E}$. In particular, the $\mathfrak{W}$-orbit of a root α spans $\mathbb{E}$.*

Proof: The span of $\mathfrak{W}$-orbit of a nonzero root is $\mathfrak{W}$-invariant subspace of $\mathbb{E}$. Then the second statement of the theorem follows from the first. It suffices to prove the first. Let $\mathbb{E}'$ be a nonzero subspace of $\mathbb{E}$ invariant under $\mathfrak{W}$. The orthogonal complement $\mathbb{E}''$ of $\mathbb{E}'$ is also $\mathfrak{W}$-invariant, and

$$\mathbb{E} = \mathbb{E}' \oplus \mathbb{E}''$$

It is trivial that for all $\alpha \in \Phi$, either $\alpha \in \mathbb{E}'$ or $\mathbb{E} \subset P_\alpha$ since $\sigma_\alpha(\mathbb{E}') = \mathbb{E}'$. Therefore, $\alpha \notin \mathbb{E}'$ implies that $\alpha \in \mathbb{E}''$. So each the roots lie in one subspace or the other. This then partitions Φ into orthogonal subsets, forcing one or the other to be empty. Since Φ spans $\mathbb{E}$, we can conclude that $\mathbb{E}' = \mathbb{E}$. □

Lemma 6.20. *Let Φ be irreducible. Then at most two root lengths occur in Φ and all roots of a given length are conjugate under $\mathbb{W}$.*

Proof: Suppose that α, β are roots. Then not all $\sigma(\alpha)$ can be orthogonal to β for all $\sigma \in \mathfrak{W}$ because $\sigma(\alpha)$ span $\mathbb{E}$, from Lemma 6.19. Now if $(\alpha, \beta) \neq 0$, we know that the possible ratios of squared root lengths of α, β are 1, 2, 3, $\frac{1}{2}$, and $\frac{1}{3}$. This implies the first assertion as the presence of three root lengths would yield a ratio of $\frac{3}{2}$. Now suppose that α, β have equal length. Upon replacing one of these by $\mathfrak{W}$-conjugate, we can assume them to be nonorthogonal and distinct. Then we have

$$\frac{2(\alpha, \beta)}{(\beta, \beta)} = \frac{2(\beta, \alpha)}{(\alpha, \alpha)}$$

Therefore, we have

$$(\sigma_\alpha \sigma_\beta \sigma_\alpha)(\beta) = \sigma_\alpha \sigma_\beta(\beta - \alpha) = \sigma(-\beta - \alpha + \beta) = \alpha$$

□

If Φ is irreducible with two distinct root lengths, one often refers to long and short roots. If all the roots are of equal length, they are conventionally called long.

Lemma 6.21. *Let Φ be irreducible with two distinct root lengths. Then the maximal root β of Lemma 6.18 is long.*

Proof: Let $\alpha \in \Phi$. It suffices to show that $(\beta, \beta) \geq (\alpha, \alpha)$. We may replace α by a $\mathfrak{W}$-conjugate lying in the closure of the fundamental Weyl chamber, relative to Δ. Since $\beta - \alpha \succ 0$ by Lemma 6.18, we have $(\gamma, \beta - \alpha) \geq 0$ for any $\gamma \overline{\mathfrak{C}(\Delta)}$. Applying the to the cases $\gamma = \beta$ and $\gamma = \alpha$, yields $(\beta, \beta) \geq (\beta, \alpha) \geq (\alpha, \alpha)$. □

6.20. **Cartan Matrix, Coxeter Graphs, and Dynkin Diagrams.**

Definition 6.39. (Cartan Matrix) Fix an ordering $(\alpha_1, \alpha_2, \cdots, \alpha_l)$ of the simple roots. Then the matrix $\left(\frac{2(\alpha_i, \alpha_j)}{(\alpha_j, \alpha_j)} \right)$ is called the Cartan matrix of Φ. The entries of the matrix are called the Cartan integers.

Example 6.32. For systems of rank 2, the matrices are

$$A_1 \times A_1 = \begin{matrix} 2 & 0 \\ 0 & 2 \end{matrix} \quad , \quad A_2 = \begin{matrix} 2 & -1 \\ -1 & 2 \end{matrix} \quad , \quad B_2 = \begin{matrix} 2 & -2 \\ -1 & 2 \end{matrix} \quad , \quad G_2 = \begin{matrix} 2 & -1 \\ -3 & 2 \end{matrix}$$

The matrix depends on the chosen ordering of the simple roots. However, this selection does not have serious implications. The most important feature the Cartan matrix is that it is independent of the choice of Δ, due to the fact that $\mathfrak{W}$ acts transitively on the collection of bases. Moreover, the Cartan matrix is nonsingular since Δ is a basis of $\mathbb{E}$. Indeed, Δ characterizes the Cartan matrix entirely.

Proposition 6.8. *Let $\Phi' \subset \mathbb{E}'$ be another root system with base $\Delta' = \{\alpha_1, \cdots, \alpha_l\}$. If*

$$\frac{2(\alpha_i, \alpha_j)}{(\alpha_j, \alpha_j)} = \frac{2(\alpha_i', \alpha_j')}{(\alpha_j', \alpha_j')}$$

for $1 \leq i$ and $j \leq l$, then the bijection $\alpha_i \mapsto \alpha_i'$ extends uniquely to an isomorphism $\phi : \mathbb{E} \to \mathbb{E}'$ mapping Φ onto Φ' and satisfying

$$\frac{2(\phi(\alpha), \phi(\beta)}{(\phi(\beta), \phi(\beta))} = \frac{2(\alpha, \beta)}{(\beta, \beta)}$$

for all $\alpha, \beta \in \Phi$. Therefore, the Cartan matrix of Φ determines Φ up to isomorphism.

Proof: Since Δ, respectively Δ', is a basis of $\mathbb{E}$, and again respectively $\mathbb{E}'$, there is a unique vector space isomorphism $\phi : \mathbb{E} \to \mathbb{E}'$ sending α_i to α_i' for all $1 \leq i \leq l$. If $\alpha, \beta \in \Delta$, the hypothesis insures that

$$\sigma_{\phi(\alpha)}(\phi(\beta)) = \sigma_{\alpha'}(\beta') = \beta' - \frac{2(\beta', \alpha')}{(\alpha', \alpha')}\alpha' = \phi(\beta) - \frac{2(\beta, \alpha)}{(\alpha, \alpha)}\phi(\alpha) = \phi\left(\beta - \frac{2(\beta, \alpha)}{(\alpha, \alpha)}\alpha\right) = \phi(\sigma_\alpha(\beta))$$

That is, the following diagram commutes for all $\alpha \in \Delta$:

$$\begin{array}{ccc} \mathbb{E} & \xrightarrow{\phi} & \mathbb{E}' \\ {\scriptstyle \sigma_\alpha}\downarrow & & \downarrow{\scriptstyle \sigma_{\phi(\alpha)}} \\ \mathbb{E} & \xrightarrow[\phi]{} & \mathbb{E}' \end{array}$$

The respective Weyl groups $\mathfrak{W}, \mathfrak{W}'$ are generated by simple reflections. So it follows that the map $\sigma \mapsto \phi \circ \sigma \circ \phi^{-1}$ is an isomorphism of $\mathfrak{W}$ onto $\mathfrak{W}'$, sending σ_α to $\sigma_{\phi(\alpha)}$ for all $\alpha \in \Delta$. But for each $\beta \in \Phi$ is conjugate under $\mathfrak{W}$ to a simple root. This forces

$$\phi(\beta) = (\phi \circ \sigma \circ \phi^{-1})(\phi(\alpha)) \in \Phi'$$

It then follows that ϕ maps Φ onto Φ'. Then the formula for a reflection shows that ϕ preserves all Cartan integers. □

Therefore, it is theoretically possible to recover Φ from a knowledge of the Cartan integers. In fact, we can do this algorithmically. We can do this by considering the root strings. Begin with the roots of height 1 - that is, the simple roots. For any pair $\alpha_i \neq \alpha_j$, the integer r for the α_j-string through α_i is 0. Then the integer q equals

$$-\frac{2(\alpha_i, \alpha_j)}{(\alpha_j, \alpha_j)}$$

This allows us to write a list of all roots α of height 2, i.e all the integers $\frac{2(\alpha,\alpha_j)}{(\alpha_j,\alpha_j)}$. Then for each root α of height 2, the integer r for the α_j-string through α can be determined easily because the α_j can be subtracted at most once and then q is found because we have $r - q = \frac{2(\alpha,\alpha_j)}{(\alpha_j,\alpha_j)}$. Once this process is repeated a sufficient number of times, all positive roots are obtained. This is guaranteed by Lemma 6.16.

If α, β are distinct positive roots, we have

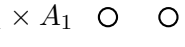
$$\frac{2(\alpha, \beta)}{(\beta, \beta)} \; \frac{2(\beta, \alpha)}{(\alpha, \alpha)} = 0, 1, 2, \text{ or } 3$$

Definition 6.40. (Coxeter Graph) The Coxeter graph of Φ be a graph having l vertices, the ith joined to the jth for $i \neq j$ by

$$\frac{2(\alpha_i, \alpha_j)}{(\alpha_j, \alpha_j)} \ \frac{2(\alpha_j, \alpha_i)}{(\alpha_i, \alpha_i)}$$

edges.

Example 6.33. We have the following examples of a few Coxeter graphs:

$A_1 \times A_1$ 

A_2 

B_2 

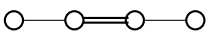
G_2 

The Coxeter graph determines the numbers $\frac{2(\alpha_i,\alpha_j)}{(\alpha_j,\alpha_j)}$ in the case that all roots have equal length because

$$\frac{2(\alpha_i, \alpha_j)}{(\alpha_j, \alpha_i)} = \frac{2(\alpha_j, \alpha_i)}{(\alpha_i, \alpha_i)}$$

If more than root length occurs, the graph fails to tell us which of a pair of vertices should correspond to a short simple root and which to a long simple root. However, one can show that the Coxeter graph completely determines the Weyl group completely as it determines the orders of products of generators of $\mathfrak{W}$. If double/triple edges occur in the Coxeter graph of Φ, we add an arrow pointing to the short of the two roots. This allows us to recover the Cartan integers. The resulting figure is the Dynkin diagram of Φ, which will depend on the numbering of the simple roots.

Example 6.34. Given the diagram

F_4 

we are able to recover the Cartan matrix

$$\begin{pmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -2 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{pmatrix}$$

Recall that Φ is irreducible if and only if Φ (or Δ) cannot be partitioned into two proper orthogonal subsets. It is then apparent that Φ is irreducible if and only if its Coxeter graph is connected. In general, there will be a number of connected components of the Coxeter graph. Let

$$\Delta = \Delta_1 \cup \cdots \cup \Delta_t$$

be the corresponding partition of Δ into mutually orthogonal subsets. If $\mathbb{E}_i$ is the span of Δ_i, it is clear that $\mathbb{E} = \mathbb{E}_1 \oplus \cdots \oplus \mathbb{E}_t$. The $\mathbb{Z}$-linear combination of Δ_i, which are the roots of a set, say Φ_i. This forms a root system in $\mathbb{E}_i$, whose Weyl group is the restriction to $\mathbb{E}_i$

of the subgroup of $\mathfrak{W}$ generated by all σ_α for all $\alpha \in \Delta_i$. Finally, each $\mathbb{E}_i$ is $\mathfrak{W}$-invarient since $\alpha \notin \Delta_i$ implies that σ_α acts trivially on $\mathbb{E}_i$. It then follows that each root lies in one of the $\mathbb{E}_i$, that is

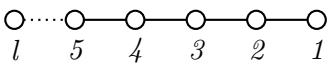
$$\Phi = \Phi_1 \cup \cdots \cup \Phi_t$$

Proposition 6.9. *Φ decomposes uniquely as the union of irreducible root systems Φ_i in subspaces $\mathbb{E}_i$ of $\mathbb{E}$ such that $\mathbb{E} = \mathbb{E}_1 \oplus \cdots \mathbb{E}_t$, i.e. the orthogonal direct sum.*

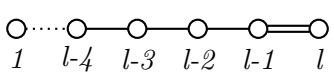
But this shows that it is sufficient to classify the irreducible root systems. This is equivalent to classifying the connected Dynkin diagrams.

Theorem 6.30. *If Φ is an irreducible root system of rank l, its Dynkin diagram is one of the following (l vertices in each case):*

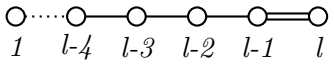
A_l *for* $l \geq 1$:



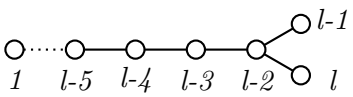
B_l *for* $l \geq 2$:



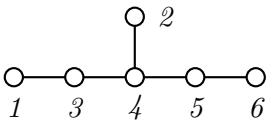
C_l *for* $l \geq 3$:



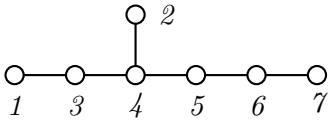
D_l *for* $l \geq 4$:



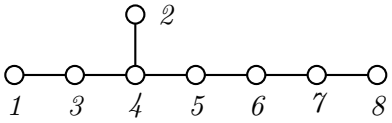
E_6 :



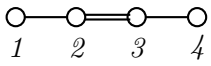
E_7 :



E_8 :



F_4 :



G_2 :

Where the double loop in B_l is directed to the right while the C_l is directed to the left. The corresponding Cartan matrices are...

$$A_l : \begin{pmatrix} 2 & -1 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ -1 & 2 & -1 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & -1 & 2 & -1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \cdot & \cdot & \cdot & \cdot & -1 & 2 \end{pmatrix}$$

$$B_l : \begin{pmatrix} 2 & -1 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ -1 & 2 & -1 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & \cdot & -1 & 2 & -2 \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & \cdot & 0 & -1 & 2 \end{pmatrix}$$

$$C_l : \begin{pmatrix} 2 & -1 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ -1 & 2 & -1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & -1 & 2 & -1 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & \cdot & -1 & 2 & -1 \\ 0 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & -2 & 2 \end{pmatrix}$$

$$D_l : \begin{pmatrix} 2 & -1 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ -1 & 2 & -1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & \cdot & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & \cdot & \cdot & \cdot & \cdot & -1 & 2 & -1 & -1 \\ 0 & 0 & \cdot & \cdot & \cdot & \cdot & 0 & -1 & 2 & 0 \\ 0 & 0 & \cdot & \cdot & \cdot & \cdot & 0 & -1 & 0 & 2 \end{pmatrix}$$

$$E_6 : \begin{pmatrix} 2 & 0 & -1 & 0 & 0 & 0 \\ 0 & 2 & 0 & -1 & 0 & 0 \\ -1 & 0 & 2 & -1 & 0 & 0 \\ 0 & -1 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & -1 & 2 \end{pmatrix}$$

$$E_7 : \begin{pmatrix} 2 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & -1 & 0 & 0 & 0 \\ -1 & 0 & 2 & -1 & 0 & 0 & 0 \\ 0 & -1 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & 0 & -1 & 2 \end{pmatrix}$$

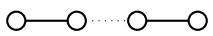
$$E_8 : \begin{pmatrix} 2 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & -1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 2 & -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 \end{pmatrix}$$

$$F_4 : \begin{pmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -2 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{pmatrix}$$

$$G_2 : \begin{pmatrix} 2 & -1 \\ -3 & 2 \end{pmatrix}$$

Though we will not go through the lengthy proof of this, it can be found in the Humphrey reference. The steps of the proof are as follows: Let $\mathbb{E}$ be a Euclidean space of arbitrary dimension and $\mathfrak{U} = \{\epsilon_1, \cdots, \epsilon_n\}$ a set of n linearly dependent unit vectors which satisfy $(\epsilon_i, \epsilon_j) \leq 0$ for $i \neq j$ and $4(\epsilon_i, \epsilon_j)^2 = 0, 1, 2,$ or 3 for $i \neq j$. Such a set is called admissible. We then show...

1. If some of the ϵ_i are discarded, the remaining ones still form an admissible set, whose graph is obtained from Γ by omitting the corresponding vertices and all incident edges.
2. The number of pairs of vertices in Γ connected by at least one edge is strictly less than n.
3. Γ contains no cycles.
4. No more than three edges can originate at a given vertex of Γ.
5. The only connected graph Γ of an admissible set $\mathfrak{U}$ which can contain a triple edge is the Coxeter graph of G_2.
6. Let $\{\epsilon_1, \cdots, \epsilon_k\} \subset \mathfrak{U}$ have subgraph

   which is a simple chain in Γ. If $\mathfrak{U}' = (\mathfrak{U} - \{\epsilon_1, \cdots, \epsilon_k\}) \cup \{\epsilon\}$, where $\epsilon = \sum_{i=1}^{k} \epsilon_i$, then $\mathfrak{U}'$ is admissible.
7. Γ contains no subgraph of 3 possible special forms.
8. Any connected graph Γ of an admissible set has one of 4 possible forms.
9. The only connected Γ of a certain type from the previous step is the Coxeter graph D_n or the Coxeter graph E_n for $n = 6, 7,$ or 8.

6.21. **Root System Automorphisms.** Our goal here is to give a complete description of aut Φ for each root system Φ. We know that $\mathfrak{W}$ is a normal subgroup of autΦ. Now let

$$\Gamma = \{\sigma \in \text{aut } \Phi \,|\, \sigma(\Delta) = \Delta\}$$

with Δ a fixed base of Φ. Evidently, Γ is a subgroup of aut Φ. If $\tau \in \Gamma \cap \mathfrak{W}$, then $\tau = 1$ from the simple transitivity of $\mathfrak{W}$. If $\tau \in$ aut Φ, then $\tau(\Delta)$ is another base of Δ, so there exists $\sigma \in \mathfrak{W}$ such that $\sigma\tau(\Delta) = \Delta$. Therefore, $\tau \in \Gamma\mathfrak{W}$. It then follows that aut Φ is the semi direct product of Γ and $\mathfrak{W}$.

Now for each $\tau \in$ aut Φ and $\alpha, \beta \in \Phi$, we have

$$\frac{2(\alpha, \beta)}{(\beta, \beta)} = \frac{2(\tau(\alpha), \tau(\beta))}{(\tau(\beta), \tau(\beta)}$$

Therefore, each $\tau \in \Gamma$ determines an automorphism of the Dynkin diagram of Φ. If τ acts trivially on the diagram, then $\tau = 1$ because Δ spans $\mathbb{E}$. However, each automorphism of the Dynkin diagram obviously determines an automorphism of Φ. So Γ may be identified with the group of diagram automorphisms.

Type	Number of Positive Roots	Order of $\mathfrak{W}$	Structure of $\mathfrak{W}$	Γ
A_l	$\binom{l+1}{2}$	$(l+1)!$	$\mathfrak{L}_{l+1}$	$\mathbb{Z}/2\mathbb{Z}, l \geq 2$
B_l, C_l	l^2	$2^l l!$	$(\mathbb{Z}/2\mathbb{Z})^l \rtimes \mathfrak{L}_l$	1
D_l	$l^2 - l$	$2^{l-1} l!$	$(\mathbb{Z}/2\mathbb{Z})^{l-1} \rtimes \mathfrak{L}_l$	$\begin{cases} \mathfrak{L}_3, & \text{if } l = 4 \\ \mathbb{Z}/2\mathbb{Z}, & \text{if } l > 4 \end{cases}$
E_6	36	$2^7 \cdot 3^4 \cdot 5$		$\mathbb{Z}/2\mathbb{Z}$
E_7	63	$2^{10} \cdot 3^4 \cdot 5 \cdot 7$		1
E_8	120	$2^{14} \cdot 3^5 \cdot 5^2 \cdot 7$		1
F_4	24	$2^7 \cdot 3^2$		1
G_2	6	$2^2 \cdot 3$	$\mathfrak{D}_6$	1

When the Dynkin diagram and Coxeter graph coincide, the term graph automorphism may also be used.

6.22. **Weights.** Here, our final goal will be to describe the representations of semisimple Lie algebras that depend on only their root system. This is the simplest case and was our purpose in developing the theory of root systems.

Definition 6.41. (Weights) Let Λ be the set of all $\lambda \in \mathbb{E}$ for which $\frac{2(\lambda,\alpha)}{(\alpha,\alpha)} \in \mathbb{Z}$ for all $\alpha \in \Phi$. The elements of Λ are called weights. But $\frac{2(\lambda,\alpha)}{(\alpha,\alpha)}$ depends only on the linearity of λ, Λ is a subspace of $\mathbb{E}$ including Φ.

Definition 6.42. (Root Lattice) Let Λ_r be the subgroup of Λ generated by Φ. Λ_r is a lattice in $\mathbb{E}$: it is the $\mathbb{Z}$-span of an $\mathbb{R}$-base of $\mathbb{E}$ (any set of simple roots). Fixing a base $\Delta \subset \Phi$, define $\lambda \in \Lambda$ to be dominant if all the integers $\frac{2(\lambda,\alpha)}{(\alpha,\alpha)}$ for $\alpha \in \Delta$ are nonnegative. We say it is strongly dominant if these integers are positive.

If we let Δ^+ be the set of all dominant weights. Then Δ^+ is the set of all weights lying in the closure of the fundamental Weyl chamber $\mathfrak{C}(\Delta)$, while $\Lambda \cap \mathfrak{C}(\Delta)$ is the set of all strongly dominant weights.

Suppose that $\Delta = \{\alpha_1, \alpha_2, \cdots, \alpha_l\}$, then the vectors $\frac{2\alpha_i}{(\alpha_i,\alpha_i)}$ form a basis of $\mathbb{E}$. Now let $\lambda_1, \lambda_2, \cdots, \lambda_l$ be the dual basis relative to the inner product on $\mathbb{E}$:

$$\frac{2(\lambda_i, \alpha_j)}{(\alpha_j, \alpha_j)} = \delta_{ij}$$

Now since all $\frac{2(\lambda_i,\alpha)}{(\alpha,\alpha)}$ are nonnegative integers for all $\alpha \in \Delta$, the λ_i are dominant weights.

Definition 6.43. (Fundamental Dominant Weights) The λ_i from above are called the fundamental dominant weights relative to Δ.

Note that $\sigma_i\lambda_j = \lambda_j - \delta_{ij}\alpha_i$. Now if $\lambda \in \mathbb{E}$ then

$$\frac{2(\lambda - \sum m_i\lambda_i, \alpha)}{(\alpha,\alpha)}$$

for each simple root α. This implies that

$$(\lambda - \sum m_i\lambda_i, \alpha) = 0$$

or simply that $\lambda = \sum m_i\lambda_i$. Therefore, Λ is a lattice with basis λ_i for $1 \le i \le l$ and $\lambda \in \Delta^+$ if and only if all $m_i \ge 0$.

The lattice Λ/Λ_r is actually the fundamental group of Φ (see Definition 7.4). From well known properties of such groups, we know that Λ/Λ_r must be finite.

Because the Weyl group $\mathfrak{W}$ preserves the inner product on $\mathbb{E}$, it must leave Λ invariant.

Lemma 6.22. *Each weight is conjugate under $\mathfrak{W}$ to one and only one dominant weight. if λ is dominant, then $\sigma\lambda \prec \lambda$ for all $\sigma \in \mathfrak{W}$ and if λ is strongly dominant, then $\sigma\lambda = \lambda$ only when $\sigma = 1$.*

Lemma 6.23. *Let $\lambda \in \Lambda^+$. Then the number of dominant weights $\mu \prec \lambda$ is finite.*

Proof: Since $\lambda + \mu \in \Lambda^+$ and $\lambda - \mu$ is a sum of positive roots

$$0 \le (\lambda + \mu, \lambda - \mu) = (\lambda,\lambda) - (\mu,\mu)$$

Thus, μ lies in the compact set

$$\{x \in \mathbb{E} \,|\, (x,x) \le (\lambda,\lambda)\}$$

whose intersection with the discrete set Λ^+ is finite. □

Lemma 6.24. *$\delta = \sum_{j=1}^{l} \lambda_j$, so δ is a (strongly) dominant weight.*

Proof: Since $\sigma_i\delta = \delta - \alpha_i$, $(\delta - \alpha_i) = (\sigma_i^2\delta, \sigma_i\alpha_i) = (\delta, -\alpha_i)$, or $2(\delta,\alpha_i) = (\alpha_i,\alpha_i)$, or $\frac{2(\delta,\alpha_i)}{(\alpha,\alpha)} = 1$ for $1 \le i \le l$. But then

$$\delta = \sum_i \frac{2(\delta,\alpha_i)}{(\alpha_i,\alpha_i)}\lambda_i$$

and the lemma follows. □

Lemma 6.25. *Let $\mu \in \Lambda^+$, $\upsilon = \sigma^{-1}\mu$ for all $\sigma \in \mathfrak{W}$. Then $(\upsilon + \delta, \upsilon + \delta) \le (\mu + \delta, \mu + \delta)$, with equality only if $\upsilon = \mu$.*

Proof:

$$(\upsilon + \delta, \upsilon + \delta) = \big(\sigma(\upsilon + \delta), \sigma(\upsilon + \delta)\big) = (\mu + \sigma\delta, \mu + \sigma\delta) = (\mu + \delta, \mu + \delta) - 2(\mu, \delta - \sigma\delta)$$

Since $\mu \in \Lambda^+$ and $\delta - \sigma\delta$ is a sum of positive roots, the right side is

$$\le (\mu + \delta, \mu + \delta)$$

with equality if and only if $(\mu, \delta - \sigma\delta) = 0$. But if $\mu - \upsilon$ is a sum of positive roots and δ is strongly dominant, so $\mu = \upsilon$. □

Definition 6.44. (Saturated) A subset Π of Λ is called saturated if for all $\lambda \in \Pi$, $\alpha \in \Pi$, and i between 0 and $\frac{2(\lambda,\alpha)}{(\alpha,\alpha)}$, the weight $\lambda - i\alpha$ also lies in Π. We say that a saturated set Π has highest weight $\lambda \in \Lambda^+$ if $\lambda \in \Pi$ and $\mu \prec \lambda$ for all $\mu \in \Pi$.

Notice that any saturated set is automatically stable under $\mathfrak{W}$, since $\sigma_\alpha\lambda = \lambda - \frac{2(\lambda,\alpha)}{(\alpha,\alpha)}\alpha$ and $\mathfrak{W}$ is generated by reflections.

Example 6.35. The set consisting of 0 alone is saturated with highest weight 0.

Example 6.36. The set of all Φ of all roots of a semisimple Lie algebra, including 0, is saturated. If Φ is irreducible, there is a unique highest root relative to the fixed base Δ of Φ, so Π has this root as its highest weight.

Lemma 6.26. *A saturated set of weights having highest weight λ must be finite.*

Proof: This follows from Lemma 6.23. □

Lemma 6.27. *Let Π be saturated with heights weight λ. If $\mu \in \Lambda^+$ and $\mu \prec \lambda$ then $\mu \in \Pi$.*

Proof: Suppose that

$$\mu' = \mu + \sum_{\alpha\in\Delta} k_\alpha\alpha \in \Pi$$

for $k_\alpha \in \mathbb{Z}_+$. We need to show how to reduce any of the k_α by one left over in Π. Then we hope to arrive at the conclusion that $\mu \in \Pi$. We start at the fact that λ is such a μ'. Now assuming that $\mu' \neq \mu$, some k_α is positive. Starting with

$$(\sum_\alpha k_\alpha\alpha, \sum_\alpha k_\alpha\alpha) > 0$$

we know that $(\sum_\alpha k_\alpha\alpha, \beta) > 0$ for some $\beta \in \Delta$ with $k_\beta > 0$. In particular, $\frac{2(\sum_\alpha k_\alpha\alpha,\beta)}{(\beta,\beta)}$ is positive. Now since μ is dominant, $\frac{2(\mu,\beta)}{(\beta,\beta)}$. Therefore, $\frac{2(\mu',\beta)}{(\beta,\beta)}$. From the definition of a saturated set, it is possible to subtract β once from μ' without leaving Π. then this reduces k_β by 1. □

Lemma 6.28. *Let Π be saturated with highest weight λ. If $\mu \in \Pi$, then $(\mu + \delta, \mu + \delta) \leq (\lambda + \delta, \lambda + \delta)$, with equality only if $\mu = \lambda$.*

Proof: From Lemma 6.25, it is sufficient to show that the case when μ is dominant. Write $\mu = \lambda - \pi$, where π is a sum of positive roots. Then

$$(\lambda+\delta, \lambda+\delta)-(\mu+\delta, \mu+\delta) = (\lambda+\delta, \lambda+\delta)-(\lambda+\delta-\pi, \lambda+\delta-\pi) = (\lambda+\delta, \pi)+(\pi, \mu+\delta) \geq (\lambda+\delta, \pi) \geq 0$$

follows because $\mu + \delta$ and $\lambda + \delta$ are dominant. Of course equality holds only if $\pi = 0$ because $\lambda + \delta$ is strongly dominant. □

7. Topology

Introduction. Some of most important questions in Mathematics are those which are not easily defined: what numbers do this, what are the properties of things that look like this, why do these do this? For example, example the symmetries of shapes that look like squares. Group theory turned out to be very efficient at answering many of these questions, i.e. the dihedral group. However, there are many other ways to answer these types of questions. Topology originally grew out of attempts to generalize the problems of Analysis. However, Algebraic Topology seeks to find properties of topological spaces which classify them up to homeomorphism by looking at the deeper fundamental properties of a space: connectedness, compactness, Hausdorffness, et cetera. These investigations often involve careful examinations of the symmetries of the space. Of course, these spaces need not be abstract mathematical creations. To understand crystals, one needs only to understand the structural symmetries within the crystal to understand some of its more fundamental properties, similar to understanding aspects like compactness of a topological spaces help to classify it. But why is this important?

In the late 20^{th} century, physicists noticed that when certain chemical compounds where at extremely low temperatures, their electrical resistance dropped to zero - not near zero, exactly zero. These compounds are called superconductors. In fact, many important power delivery systems and internet cables around the world are delivered through such medium. Understanding these materials would help to find a room temperature superconductor. To examine these questions with Topology, we study the physical structure of the space. The reader is of course familiar that there are three common phases of matter: solids, liquids, and gases. Each of these phases of matter have various levels of symmetry. Though high school textbooks suggest much more perfect structures than are actually typically found in nature. Based on the amount of kinetic energy in the system, the particles exhibit more or less symmetry that we identify with the object being solid, liquid, or gas.

However, the story is not that simple. These three phases of matter are not exhaustive. For example, glass is not a solid but rather a supercooled liquid - it flows slowly over time. Take silly putty which acts as a solid over short periods but a liquid over long periods of time or over saturated liquids, which at the slightest disturbance cause solids to drop from the liquid. Furthermore, water mixed with corn starch acts like a solid when forces are applied over a short period of time but as a liquid when applied over a long period of time. These examples alone shows that there are some substances whose structure is time dependent. In fact, there are many substances that don't fit the standard solid, liquid, gas model: liquid crystals, supercooled liquids, spin glasses, gels, foams, superconductors, polymers, colloids, etc. Our goal here is to examine these local structure and properties of some of these substances to explain their global behavior. First, we will need to introduce ways to look at mappings between spaces and paths in those spaces. Then we use this to establish a fundamental (with some pun intended) property of the space which we will use to classify these substances later. Finally, before talking about the properties of matter we are interested in, we will talk about manifolds because much of the physical structure of these objects form a manifold.

7.1. **Homotopy.** To see if two spaces are the same, we can see if we can continuously deform one space into another. Imagine taking a ball with a hole in it and sticking ones finger into the hole and stretching it out a bit. Then one has a larger circle which one can then stretch out in all directions to infinity, flattening it out as one goes. The result is the plane, which should be of no surprise given stereographic projection; given the missing point to be the north pole of the unit sphere, a point on the sphere (x, y, z) and a point on the plane (X, Y), then their relationship is given by

$$(X, Y) = \left(\frac{x}{1-z}, \frac{y}{1-z} \right)$$
$$(x, y, z) = \left(\frac{2X}{1+(X^2+Y^2)}, \frac{2Y}{1+(X^2+Y^2)}, \frac{(X^2+Y^2)-1}{1+(X^2+Y^2)} \right)$$

The stereographic projection is a classic example of a homeomorphism, here between the plane and the punctured sphere. However, do the properties of a space restrict the types of maps which are homeomorphisms between two topological spaces? Of course, the answer is yes. In general, it is very difficult to resolve the question whether two general topological spaces are homeomorphic. Necessary conditions come easily and cheaply while sufficient conditions are sparse and difficult to find. However, there are three major fields of study in topology which create sufficient conditions for topological spaces to be homeomorphic. Each topological space comes a variety of groups associated with it: homology groups, cohomology groups, homotopy groups, and cohomotopy groups. Should two spaces be homeomorphic their corresponding associated groups are also homeomorphic. This provides important necessary conditions which also reveal many of the properties of the space. Homotopy is the study of the types of continuous maps from one topological space to another. Of course, one would be interested in this because the properties of a given topological spaces restricts the types of mappings that might send it into another topological space.

Definition 7.1. (Homotopy)
If f and f' are continuous maps of the topological space X into the topological space Y, we say that f is homotopic to f' if there is a continuous map $F : X \times I \to Y$ such that

$$F(x, 0) = f(x) \qquad \text{and} \qquad F(x, 1) = f'(x)$$

for each $x \in X$. The map F is called a homotopy between f and f'. If f is homotopic to f', we write $f \simeq f'$. If f' is a constant map, we say that f is nulhomotopic.

Example 7.1. Take the homotopy in Figure 5. The top red line is the result of applying a composition of homomorphisms such that $[0, 1] \mapsto [\frac{7\pi}{10}, \pi]$ and then applying the continuous map $f(x) = 1 - x^2 + x^3$. Similarly, we send $[0, 1] \mapsto [\frac{7\pi}{10}, \pi]$ for the bottom line but then applying the mapping $g(x) = (1 - x^2 + x^3)^2 = f(x)^2$. So notice then that both the top and bottom line are homeomorphic with $[0, 1]$. Simple define a map $h(x) = \frac{3\pi}{10}x + \frac{7\pi}{10}$, which is obviously a continuous map from $[0, 1] \to [\frac{7\pi}{10}, \pi]$. Then the top line is homeomorphic to $[0, 1]$ under the mapping $(f \circ h)(x)$ and the bottom line is homeomorphic to $[0, 1]$ under

the mapping $(g \circ h)(x)$. Moreover, these spaces are homotopic since the continuous map $F(x,t) = f(x)^{\frac{3}{2}t+1/2}$ for all x in our domain and $t \in [0,1]$ serves as a homotopy between the spaces.

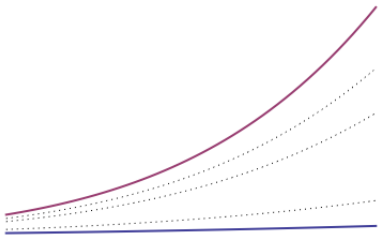


FIGURE 5. An example of a homotopy between two topological spaces. Both lines are equivalent to $[0,1]$ as topological spaces and the dotted lines are intermediate steps of the homotopy.

Example 7.2. Any map $f : \mathbb{S}^1 \to X$, where X is a topological space, is null-homotopic when f is extendable to the unit disk, D^2, that agrees with f on the boundary of D^2.

Example 7.3. If $f, g : X \to \mathbb{R}^2$ are maps, then a homotopy between them is given by

$$F(x,t) = (1-t)\, f(x) + t\, g(x)$$

called the straight-line homotopy.

However, if f is a path that is in the topological space X, we get an even stronger relation - path homotopy.

Definition 7.2. (Path Homotopic)
Two paths f and f', mapping the interval $I = [0,1]$ into the topological space X are said to be path homotopic if they have the same initial point x_0 and the same final point x_1 and if there is a continuous map $F : I \times I \to X$ such that

$$\begin{aligned} F(s,0) &= f(s) \quad \text{and} \quad F(s,1) = f'(s) \\ F(0,t) &= x_0 \quad \text{and} \quad F(1,t) = x_1 \end{aligned}$$

for each $s \in I$ and each $t \in I$. We call F a path homotopy between f and f'. We write $f \simeq_p f'$.

This definition says two things:

1. F is a continuous function that deforms f to f'. Moreover, F is a homotopy between f and f'.
2. The path f_t given by $f_t(s) = F(s,t)$ is a path from x_0 to x_1. Moreover, the initial and final points of the paths f, f' remain fixed throughout the deformation.

Now while these definitions and examples seem interesting enough, they have yet to prove themselves more useful than an ordinary homeomorphism. But indeed, they are surprisingly useful. We can turn these homotopies into a groupoid then they will begin to be of some use. First, we show that they form an equivalence class and them create an operation on the classes to form the groupoid.

Lemma 7.1. *The relations $\simeq$ and $\simeq_p$ form equivalence classes.*

Proof: We need only show this for $\simeq$ as $\simeq_p$ is just a special case of $\simeq$. Let f, f' be paths and $[f], [f']$ denote their equivalence class.

1. Reflexivity: It is trivial that $f \simeq f$ as $F(x,t) = f(x)$ is the required homotopy.
2. Symmetry: Let F be the homotopy given by the fact that $f \simeq f'$. Then the required homotopy necessary for $f' \simeq f$ is $G(x,t) = F(x, 1-t)$. It is also clear if F is a path homotopy then so is G.
3. Transitivity: Suppose that $f \simeq f'$ and $f' \simeq f''$. Let F be the homotopy between f and f' and F' be the homotopy between f' and f''. Now the required homotopy between f' and f'' is
$$G(x,t) = \begin{cases} F(x,2t) & \text{for } t \in [0, \frac{1}{2}] \\ F'(x, 2t-1) & \text{for } t \in [\frac{1}{2}, 1] \end{cases}$$
G is well defined as $G(x, \frac{1}{2}) = F(x,1) = F'(x,1) = f'(x)$. Moreover, since it is continuous at $t = \frac{1}{2}$ and given that F and F' are continuous, G is continuous. Then G is truly the required homotopy. It is again simple to check that if F and F' are path homotopies than so is G. □

Now we can induce an operation on path-homotopy classes that will allow us to form the groupoid.

Definition 7.3. (Path Products)
If f is a path in X from x_0 to x_1 and if g is a path in X from x_1 to x_2, we define the product $f * g$ to be the path h given by the equations
$$h(s) = \begin{cases} f(2s), & \text{for } s \in [0, \frac{1}{2}] \\ g(2s-1), & \text{if } s \in [\frac{1}{2}, 1] \end{cases}$$
The function h is well-defined and continuous and is a path in X from x_0 to x_2, whose first half path is f and whose second half path is g. The product operations on a well-defined operation in path homotopy classes, defined by the equation
$$[f] * [g] = [f * g]$$
The operation $*$ satisfies the property of a groupoid, which we will now show.

Theorem 7.1. *The operation $*$ satisfies all the properties of a groupoid: associativity, right and left identities, and the existence of inverses.*

Proof:

1. Associativity: We describe $f * g$ is a bit of a different way. Suppose $[a, b], [c, d] \in \mathbb{R}$ are intervals. Then there is a map $\rho : [a, b] \to [c, d]$ with form $\rho(x) = mx + k$ that carries a to c and b to d. This map ρ is called the positive linear map of $[a, b]$ to $[c, d]$. Then we can describe $f * g$ as follows: If $t \in [0, \frac{1}{2}]$ then it is the positive linear map of $[0, \frac{1}{2}]$ to $[0, 1]$, then followed by f and if $t \in [\frac{1}{2}, 1]$ it is the positive linear map of $[\frac{1}{2}, 1]$ to $[0, 1]$ followed by g.

   Now suppose that $f, g, h \in X$ and the products $f * (g * h)$ and $(f * g) * h$ are defined when $f(1) = g(0)$ and $g(1) = h(0)$. If we assume this, then we can define the triple product above as follows: let $a, b \in I$ so that $0 < a < b < 1$. Now define a path $k_{a,b} \in X$ so that on $[0, a]$ it is the positive linear map of $[0, a]$ to I followed by f and on $[a, b]$ it is the positive linear map $[a, b]$ to I followed by g and finally on $[b, 1]$ it is the positive linear map $[b, 1]$ to I followed by h. Of course, the path $k_{a,b}$ depends on the choice of a, b. However, notice its path homotopy does not depends on the choice of a or b. It only remains to show that if c, d are any pair of points in I with $0 < c < d < 1$, then $k_{c,d}$ must be path homotopic to $k_{a,b}$.

   Suppose that $p : I \to I$ is a map with $p(a) = c$, $p(b) = d$, and $p(1) = 1$. Then when restricted to $[0, a]$, $[a, b]$ and $[b, 1]$, p is the positive linear map of these intervals onto $[0, c]$, $[c, d]$, and $[d, 1]$, respectively. But then it follows that $k_{c,d} \circ p$ must be $k_{a,b}$. But then p is a path in I from 0 to 1 and so is the identity map $i : I \to I$. Therefore, there must be a path homotopy P in I between p and i. Finally, $k_{c,d} \circ P$ is a path homotopy in X between $k_{c,d}$ and $k_{a,b}$.
2. Right/Left Identity: Let e_0 denote the constant path in I at 0 and let $i : I \to I$ denote the identity map. Then $e_0 * i$ is a path in I from 0 to 1. However, because I is convex, there is a path homotopy G in I between i and $e_0 * i$. Then $f \circ G$ is a path homotopy in X between the paths $f \circ i = f$ and

$$f \circ (e_0 * i) = (f \circ e_0) * (f \circ i) = e_{x_0} * f$$

   Similarly, an argument can be demonstrated to show that if e_1 is the constant path at 1, then $i * e_1$ is path homotopic in I to the path i, showing that $[f] * [e_{x_1}] = [f]$.
3. Inverses: First, note the reverse of the path i is $\overline{i}(s) = 1 - s$. Then $i * \overline{i}$ is a path in I that is a path that both begins and ends at 0, and so it the constant path e_0. However, because I is convex, there is a path homotopy H in I between e_0 and $i * \overline{i}$. Then $f \circ H$ is a path homotopy between $f \circ e_0 = e_{x_0}$ and

$$(f \circ i) * (f \circ \overline{i}) = f * \overline{f}$$

   A similar argument shows that $\overline{i} * i$ is path homotopic in I to e_1, showing that $[\overline{f}] * [f] = [e_{x_1}]$. □

It would seem that $*$ is a group from the above. Why does it only form a groupoid? This is because $[f] * [g]$ is not defined for every pair of classes. For example, what if $f(1) \neq g(0)$, the case where the start of one path is not the end of the other.

7.2. **Fundamental Group.** Though certainly not developed homotopy very deeply, it clearly presents itself as a powerful tool for studying topological spaces. However, we can still do better. Recall that path-homotopy classes result in only a groupoid, we can improve this to form a group. If X is a topological space, simply choose a point $x_0 \in X$ as a base point and only observe paths whose initial and final points are x_0, i.e. all loops at x_0. Then our original groupoid satisfies all the properties of a group and is known as the fundamental group.

Definition 7.4. (Fundamental Group)
Let X be a topological space; let x_0 be a point of X. A path in X that begins and ends at x_0 is called a loop based at x_0. The set of path homotopy classes of loops based at x_0 with operation $*$ is called the fundamental group of X relative to the base point x_0. It is denoted $\pi_1(X, x_0)$. This is also known as the first homotopy group of X and is denoted $\pi_1(X, x_0)$.

There are groups $\pi_n(X, x_0)$ for all $n \in \mathbb{Z}$ but discussing these homotopy groups are well beyond the scope of our discussion but the interested reader can look into homotopy theory for such discussions. For example, one could read Cornell professor Allen Hatcher's wonderful free book Algebraic Topology.

Remark 7.1. As with splitting fields, one cannot simply say, "the fundamental group of X", but rather one says, "the fundamental group of X relative to the base point x_0". A base point must be specified except in special circumstances.

Example 7.4. If f is any loop in $\mathbb{R}^n$ at some point x_0, then the straight-line homotopy is a path homotopy between f and some constant path at x_0. Therefore, $\pi_1(\mathbb{R}^n, x_0)$ is trivial (consisting of the identity alone). Indeed, this is also true for any convex region in $\mathbb{R}^n$ as one can see with a bit of thought.

Example 7.5. The fundamental group of the the circle is $\mathbb{Z}$. Though we won't prove this until later, we can show this using some thought. Any loop at $x_0 \in \mathbb{S}^1$ that does not loop around the whole circle can be deformed back to the point x_0 and hence are trivial. Consider a loop which goes around the circle n times counter-clockwise. This loop can only be deformed back to the $n-1$st loop without breaking the loop or stretched to the $n+1$st loop without changing the loop. For example, a loop which goes around the circle can only be squished back to the 1st loop or stretched to a loop that goes around the circle twice without changing the loop. Therefore, any loop that wraps around the unit circle i times is equivalent to to a loop which goes around $\lfloor i \rfloor$ times. But then there is a distinct loop for each $n \in \mathbb{Z}$ and nonnegative. Similarly, there are as many loops that go clockwise, i.e. the negative direction. Hence, the fundamental group of the circle is $\mathbb{Z}$.

Example 7.6. The fundamental group of the sphere is trivial since every loop on the surface of the sphere can be deformed back to the base point x_0.

The first thing one observes about the fundamental group is that it depends on the choice of base point $x_0 \in X$. How "fundamental" can this group be if it depends on the

choice of base point? How dependent is the group on the choice of base point? Here we will clarify these questions more.

Definition 7.5. (Fundamental Group Mappings)
Let α be a path in X from x_0 to x_1. Define a map

$$\hat{\alpha} : \pi_1(X, x_0) \to \pi_1(X, x_1)$$

by the equation

$$\hat{\alpha}([f]) = [\overline{\alpha}] * [f] * [\alpha]$$

It is clear that $\hat{\alpha}$ is well defined as $*$ is. Then if f is any loop at x_0 then $\overline{\alpha} * (f * \alpha)$ is a loop based at x_1. Therefore, $\hat{\alpha}$ maps $\pi_1(X, x_0)$ to $\pi_1(X, x_1)$ as desired and only depends on the path-homotopy class of α. Moreover, $\hat{\alpha}$ is an isomorphism between $\pi_1(X, x_0)$ and $\pi_1(X, x_1)$.

Theorem 7.2. *The map $\hat{\alpha}$ is a group isomorphism.*

Proof: This can be done through simple computation.

$$\begin{aligned}\hat{\alpha}([f]) * \hat{\alpha}([g]) =& ([\overline{\alpha}] * [f] * [\alpha]) * ([\overline{\alpha} * [g] * [\alpha]) \\ =& [\overline{\alpha}] * [f] * [g] * [\alpha] \\ =& \hat{\alpha}([f] * [g])\end{aligned}$$

which shows that $\hat{\alpha}$ is a homomorphism. Now let β denote the path $\overline{\alpha}$, the reverse of α. Then $\hat{\beta}$ is the inverse of $\hat{\alpha}$. Given $[h] \in \pi_1(X, x_1)$ then

$$\begin{aligned}\hat{\beta}([h]) =& [\overline{\beta}] * [h] * [\beta] = [\alpha] * [h] * [\overline{\alpha}] \\ \hat{\alpha}(\hat{\beta}([h])) =& [\overline{\alpha}] * ([\alpha] * [h] * [\overline{\alpha}]) * [\overline{\alpha}] = [h]\end{aligned}$$

It is trivial to show also that $\hat{\beta}(\hat{\alpha}([f])) = [f]$ for all $[f] \in \pi_1(X, x_0)$. □

It then immediately follows that in any path connected topological space that the choice of base point x_0 is irrelevant in some sense.

Corollary 7.1. *If X is a path connected topological space and x_0 and x_1 are two points of X then $\pi_1(X, x_0)$ is isomorphic to $\pi_1(X, x_1)$.*

One can study some properties of topological spaces using the fundamental group through its path components that contain x_0. But usually one restricts the majority of the fundamental groups usefulness to those spaces that are path-connected spaces. However, despite Corollary 7.1, one cannot leave out the choice of base point x_0. Indeed, there is no natural way of identifying $\pi_1(X, x_0)$ with $\pi_1(X, x_1)$ in general since different paths α and β from x_0 to x_1 could result to different isomorphisms between the groups. As it turns out, the isomorphism between $\pi_1(X, x_0)$ and $\pi_1(X, x_1)$ is independent of the path if and only if the fundamental group is abelian, which is really a deep requirement on the topological space X itself.

Theorem 7.3. *Let x_0 and x_1 be points of the path-connected space X. Then $\pi_1(X,x_0)$ is abelian if and only if for every pair α and β of paths from x_0 and x_1, we have $\hat{\alpha} = \hat{\beta}$.*

Proof: First, suppose that $\pi_1(X, x_0)$ is abelian, α and β are paths from x_0 to x_1, and $[f] \in \pi_1(X, x_0)$. It is obvious that $\alpha * \overline{\beta}$ is a loop at x_0. Now

$$[f] * [\alpha] * [\overline{\beta}] = [f] * [\alpha * \overline{\beta}] = [\alpha * \overline{\beta}] * [f] = [\alpha] * [\overline{\beta}] * [f]$$

Then,

$$\begin{aligned}\hat{\alpha}([f]) =& [\overline{\alpha}] * [f] * [\alpha] \\ =& [\overline{\alpha}] * [f] * [\alpha] * [\overline{\beta}] * [\beta] \\ =& [\overline{\alpha}] * [\alpha] * [\overline{\beta}] * [f] * [\beta] \\ =& [\overline{\beta}] * [f] * [\beta] \\ =& \hat{\beta}([f])\end{aligned}$$

□

Furthermore, X being path connected induces restrictions on the homomorphisms of the fundamental group relative to base point x_0 as well.

Theorem 7.4. *If X is path connected, the homomorphism induced by a continuous map is independent of bases point, up to isomorphism of the groups involved. More precisely, let $h : X \to Y$ be continuous, with $h(x_0) = y_0$ and $h(x_1) = y_1$. Let α be a path in X from x_0 to x_1 and let $\beta = h \circ \alpha$, then*

$$\hat{\beta} \circ (h_{x_0})_* = (h_{x_1})_* \circ \hat{\alpha}$$

Proof: We can show this through simple computation. Suppose that $f \in \pi_1(X, x_0)$. Then we have

$$\begin{aligned}\hat{\beta}((h_{x_0}) * ([f])) =& \hat{\beta}([h \circ f]) \\ =& [\overline{h \circ \alpha}] * [h \circ f] * [h \circ \alpha] \\ =& [(h \circ \overline{\alpha}) * ((h \circ f) * (h \circ \alpha))] \\ =& [h \circ (\overline{\alpha} * (f * \alpha))] \\ =& (h_{x_1}) * ([\overline{\alpha} * [f] * [\alpha]) \\ =& (h_{x_1}) * (\overline{\alpha}([f]))\end{aligned}$$

□

In fact, this stringent requirement on the space (that the fundamental group be abelian) forces the topological space X to be abelian.

Definition 7.6. (Simply Connected)
A topological space X is said to be simply connected if it is a path-connected space and if

$\pi_1(X, x_0)$ is the trivial (one-element) group for some $x_0 \in X$ and hence for every $x_0 \in X$. We often express the fact that $\pi_1(X, x_0)$ is the trivial group by writing $\pi_1(X, x_0) = 0$.

Lemma 7.2. *If X is a simply connected topological space, then any two paths with the same initial and final points are path homotopic.*

Proof: Suppose α and β are two paths from x_0 to x_1 then $\alpha * \overline{\beta}$ is a loop based at x_0. Since X is simply connected, this loop must be path homotopic to a constant loop at x_0. But then

$$[\alpha] = [\alpha * \overline{\beta}] * [\beta] = [e_{x_0}] * [\beta] = [\beta]$$

□

Example 7.7. The spheres $\mathbb{S}^n$ are all simply connected for $n \geq 2$.

Example 7.8. The plane $\mathbb{R}^2$ is simply connected. Furthermore, every convex subset of $\mathbb{R}^2$ is simply connected. However, the punctured plane $\mathbb{R}^2 - \{0\}$ is not simply connected.

Example 7.9. The special orthogonal group $SO(n, \mathbb{R})$ is simply connected for $n \geq 2$. The special unitary group $SU(n)$ is always simply connected.

Example 7.10. All topological vector spaces are simply connected. Importantly, this implies then that the Banach and Hilbert spaces are simply connected.

It should be clearer now that forming a group out of loop equivalence classes is a powerful tool for classifying topological spaces to some extent. Indeed, the fundamental group is clearly a topological invariant of a space. We show this using the notion of homomorphisms induced by a continuous map.

Definition 7.7. (h-Homomorphisms)
Let $h : (X, x_0) \to (Y, y_0)$ be a continuous map. Define

$$h_* : \pi_1(X, x_0) \to \pi_1(Y, y_0)$$

by the equation

$$h_*([f]) = [h \circ f]$$

The map h_* is called the homomorphism induced by h, relative to the base point x_0.

Notice that h_* is well defined because if F is a path homotopy between two paths f and f', then $h \circ F$ must be a path homotopy between the paths $h \circ f$ and $h \circ f'$. Clearly, h_* is a homomorphism from the fact

$$(h \circ f) * (h \circ g) = h \circ (f * g)$$

But what is h_* doing? Let f be some loop in X based at the point x_0 and $y_0 \in Y$. Suppose that $h : X \to Y$ is a continuous map that carries x_0 to y_0, that is $h : (X, x_0) \to (Y, y_0)$. Then the composition $h \circ f : I \to Y$ is a loop in Y based at the point y_0. But then the correspondence $f \to h \circ f$ gives a map carrying $\pi_1(X, x_0)$ to $\pi_1(Y, y_0)$ and is exactly the map h_*. Furthermore, notice that the homomorphism h_* depends not only on the topological spaces X, Y but also the choice of base point x_0. This also makes it clear that

if $x_0, x_1 \in X$ are distinct, they may not yield the same homomorphism h_*, even when X is path connected, though then the groups are certainly isomorphic. Because of this, the notation

$$(h_{x_0})_* : \pi_1(X, x_0) \to \pi_1(Y, y_0)$$

is often used in context unless the base point under consideration remains fixed. We now show that the fundamental group is an invariant property of a topological space.

Theorem 7.5. *If $h : (X, x_0) \to (Y, y_0)$ and $k : (Y, y_0) \to (Z, z_0)$ are continuous, then $(k \circ h)_* = k_* \circ h_*$. If $i : (X, x_0) \to (X, x_0)$ is the identity map, then i_* is the identity homomorphism.*

Proof: The proof follows trivial from simple computation.

$$\begin{aligned}(k \circ h)_*([f]) =& [(k \circ h) \circ f] \\ (k_* \circ h_*)([f]) =& k_*(h_*([f])) = k_*([h \circ f]) = [k \circ (h \circ f)]\end{aligned}$$

We similarly find $i_*([f]) = [i \circ f] = [f]$. □

Corollary 7.2. *If $h : (X, x_0) \to (Y, y_0)$ is a homeomorphism of X with Y, then h_* is an isomorphism of $\pi_1(X, x_0)$ with $\pi_1(Y, y_0)$.*

Proof: Let $k : (Y, y_0) \to (X, x_0)$ be the inverse of h. Then $k_* \circ h_* = (k \circ h)_* = i_*$, where i is the identity map of (X, x_0) and $h_* \circ k_* = (h \circ k)_* = j_*$, where j is the identity map of (Y, y_0). Since i_* and j_* are the identity homomorphisms of the groups $\pi_1(X, x_0)$ and $\pi_1(Y, y_0)$, respectively, k_* is the inverse of h_*. □

Hence, a necessary condition for two topological spaces to be homeomorphic is that their fundamental groups need to be isomorphic. Sadly however, this is not a sufficient condition.

7.3. **Covering Spaces.** Though the fundamental group is quite useful, it is far less so if one cannot calculate the fundamental group. Though this is often easy for fundamental groups which are trivial, those which are not are much more difficult in general. This is where covering spaces come in handy. Indeed, they are essential in the study of Riemannian surfaces and manifolds.

Definition 7.8. (Evenly Covered)
Let $p : E \to B$ be a continuous surjective map. The open set U of B is said to be evenly covered by p if the inverse image $p^{-1}(U)$ can be written as the union of disjoint open sets V_α in E such that for each α, the restriction of p to V_α is a homeomorphism of V_α onto U. The collection $\{V_\alpha\}$ will be called a partition of $p^{-1}(U)$ into slices.

The image often associated with being evenly covered is the "pancake stack". Suppose U is an open set that is evenly covered by p, then $p^{-1}(U)$ can be visualized as a "stack of pancakes" above U. The map p then smashes all these pancakes down onto U, see Figure 6.

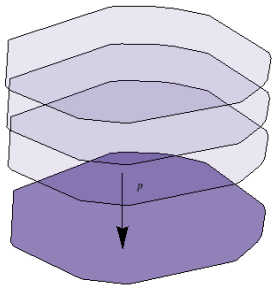


FIGURE 6. The standard representation of being evenly covered by a pancake stack. The lower darker stack is of course U and the upper stacks are $p^{-1}(U)$.

Definition 7.9. (Covering Space)
Let $p : E \to B$ be continuous and surjective. If every point $b \in B$ has a neighborhood U that is evenly covered by p, then p is called a covering map and E is said to be the covering space of B.

It is a useful observation to make that if $p : E \to B$ is a covering map then p is also necessarily an open map.

Example 7.11. The map $p : \mathbb{R} \to \mathbb{S}^1$ given by

$$p(x) = (\cos 2\pi x, \sin 2\pi x)$$

is a covering map. This map wraps all of $\mathbb{R}$ continuously around the circle and maps each interval $[n, n+1]$ onto $\mathbb{S}^1$.

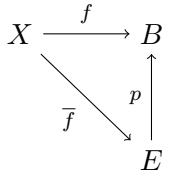
Moreover, one should notice that if we place restrictions on the mapping $p : E \to B$, then we get a local homeomorphism of E with B. That is to say, each point $e \in E$ has a neighborhood that is mapped homeomorphically by p onto an open subset of B. Why? Because this open set wraps part of the positive real line entirely around the circle. But then any open set which contains it must cover the circle slightly more than one, meaning we have lost injectivity and therefore cannot be a homeomorphism. However, it is important to know that this is not sufficient for a mapping to be a covering map. Take the mapping $p : \mathbb{R}_+ \to \mathbb{S}^1$ given by

$$p(x) = (\cos 2\pi x, \sin 2\pi x)$$

Notice that this is surjective and a local homeomorphism. However, it is not a covering map as the point $(1, 0)$ has no neighborhood U that is evenly covered by p.

We have yet to explain exactly how covering maps relate to the fundamental group. Here, we rectify this missing connection.

Definition 7.10. (Lift)
Let $p : E \to B$ be a map. If f is a continuous mapping of some topological space X into B, a lifting of f is a map $\overline{f} : X \to E$ such that $f = p \circ \overline{f}$.



Notice that a lift exists when p is a covering map, which we shall show. Moreover, path homotopies can also be lifted.

Lemma 7.3. *Let $p : E \to B$ be a covering map, let $p(e_0) = b_0$. Any path $f : [0, 1] \to B$ beginning at b_0 has a unique path lifting to a path $\overline{f}$ in E beginning at e_0.*

Proof:

1. Existence: First, cover B by open sets U, each of which is evenly covered by p. Then using the Lebesgue Number Lemma, find a subdivision, $s_0, s_1, \cdots, s_n$, of $[0, 1]$ such that for each i $f([s_i, s_{i+1}])$ lies in an open set U. Now we create the lift $\overline{f}$ as follows:

   First, let $\overline{f}(0) = e_0$. Supposing that $\overline{f}(s)$ is defined for all $0 \leq s \leq s_i$ and define $\overline{f}$ on $[s_i, s_{i+1}]$ as follows: The set $f([s_i, s_{i+1}])$ lies in some open set U that is evenly covered by p. Let $\{V_\alpha\}$ be a partition of $p^{-1}(U)$ into slices, each set V_α is mapped homeomorphically onto U by p. Then $\overline{f}(s_i)$ lies in one of these sets. Suppose it lies in V_0. Now define $\overline{f}(s)$ for $s \in [s_i, s_{i+1}]$ by

   $$\overline{f}(s) = (p|V_0)^{-1}(f(s))$$

   Since $p|V_0 : V_0 \to U$ is a homeomorphism, $\overline{f}$ is continuous on $[s_i, s_{i+1}]$. Doing this for each of the s_i, we can define $\overline{f}$ on all of $[0, 1]$. The continuity of $\overline{f}$ follows from the pasting lemma and the fact that $p \circ \overline{f} = f$ follows from the definition of $\overline{f}$.
2. Uniqueness: Suppose that $\tilde{\overline{f}}$ is another lifting of f starting at e_0. Then $\tilde{\overline{f}}(0) = e_0 = \overline{f}(0)$. Suppose that $\tilde{\overline{f}}(s) = \overline{f}(s)$ for all s such that $0 \leq s \leq s_i$. Let V_0 as in the preceding step. Then for all $s \in [s_i, s_{i+1}]$, $\overline{f}(s)$ is defined as $(p|V_0)^{-1}(f(s))$. Since $\tilde{\overline{f}}$ is a lifting of f, it must carry the interval $[s_i, s_{i+1}]$ into the set $p^{-1}(U) = \bigcup V_\alpha$. The slices V_α are then open and disjoint. Because the set $\tilde{\overline{f}}([s_i, s_{i+1}])$ is connected, it must lie entirely in one of the sets V_α. Moreover, because $\tilde{\overline{f}}(s_i) = \overline{f}(s_i)$, which of course is in V_0, $\tilde{\overline{f}}(s)$ must equal some point of $y \in V_0$ lying in $p^{-1}(f(s))$. But there is only one such point y, $(p|V_0)^{-1}(f(s))$. Therefore, $\tilde{\overline{f}}(s) = \overline{f}(s)$ for $s \in [s_i, s_{i+1}]$. □

Lemma 7.4. *Let $p : E \to B$ be a covering map; let $p(e_0) = b_0$. Let the map $F : I \times I \to B$ be continuous, with $F(0,0) = b_0$. There is a unique lifting of F to a continuous map*

$$\overline{F} : I \times I \to E$$

such that $\overline{F}(0,0) = e_0$. If F is a path homotopy, then $\overline{F}$ is a path homotopy.

Proof:

1. Existence: Given a path homotopy F, define $\overline{F}(0,0) = e_0$. Now use Lemma 7.3 to extend $\overline{F}$ to the left-hand edge $0 \times I$ and the bottom edge $I \times 0$ of $I \times I$. Then we can extend $\overline{F}$ to all of $I \times I$ as we will describe.

   Using the Lebesgue Number Lemma, choose subdivisions $s_0 < s_1 < \cdots < s_m$ and $t_0 < t_1 < \cdots < t_n$ of I fine enough so that each rectangle

   $$I_i \times J_j = [s_{i-1}, s_i] \times [t_{j-1}, t_j]$$

   is mapped by F into an open set of B that is evenly covered by p. Now define the lifting $\overline{F}$ slowly, beginning with the rectangle $I_1 \times J_1$, continuing with all other rectangles, $I_i \times J_1$, then the rectangles $I_i \times J_2$, et cetera. In general, if given i_0 and j_0, assume that $\overline{F}$ is defined on the set A which is the union of $0 \times I$ and $I \times 0$ and all the rectangles before $I_{i_0} \times J_{j_0}$. Assume also that $\overline{F}$ is continuous liftings of $F|A$. Define $\overline{F}$ on $I_{i_0} \times J_{j_0}$. Then choose an open set U of B that is evenly covered by p and contains the set $F(I_{i_0} \times J_{j_0})$. Let $\{V_\alpha\}$ be a partition of $p^{-1}(U)$ into slices. Each set V_α is mapped homeomorphically onto U by p.

   Now $\overline{F}$ is defined on the set $C = A \cap (I_{i_0} \times J_{j_0})$. This set is the union of the left and bottom edges of the rectangle $I_{i_0} \times J_{j_0}$ and so is connected. Therefore, $\overline{F}(C)$ is connected and must lie entirely within on of the sets V_α. Suppose it lies in V_0. Let $p_0 : V_0 \to U$ denote the restriction of p to V_0. Since $\overline{F}$ is a lifting of $F|A$, we know that for any $x \in C$,

   $$p_0(\overline{F}(x)) = p(\overline{F}(x)) = F(x)$$

   so that $\overline{F}(x) = p_0^{-1}(F(x))$. Therefore, we may extend $\overline{F}$ be defining

   $$\overline{F}(x) = p_0^{-1}(F(x))$$

   for $x \in I_{i_0} \times J_{j_0}$. The extended map will be continuous by the pasting lemma. Simply continue this process to define $\overline{F}$ on all of I^2.
2. Uniqueness: At each step of the construction of $\overline{F}$, $\overline{F}$ can only be extended continuously in one way. So once $\overline{F}(0,0)$ is defined, $\overline{F}$ is completely determined. Suppose then that F is a path homotopy. We need to show that $\overline{F}$ is a path homotopy. The map F carries the entire left edge $0 \times I$ of I^2 into a single point, $b_0 \in B$. Because $\overline{F}$ is a lifting of F, it must carry this edge into the set $p^{-1}(b_0)$. But this set has the discrete topology as a subspace of E. Since $0 \times I$ is connected and $\overline{F}$ is continuous, $\overline{F}(0 \times I)$ is connected and thus must equal a one-point set. Similarly, $\overline{F}(1 \times I)$ must be a one-point set. Therefore, $\overline{F}$ is a path homotopy. □

We are now fully equipped to prove the statement earlier: that when p is a covering map then there exists a lift.

Theorem 7.6. *(Covering Maps Induce Lifts) Let $p : E \to B$ be a covering map; let $p(e_0) = b_0$. Let f and g be two paths in B from b_0 to b_1 and let $\overline{f}$ and $\overline{g}$ be their respective liftings to paths in E beginning at e_0. If f and g are path homotopic, then $\overline{f}$ and $\overline{g}$ end at the same point of E and are path homotopic.*

Proof: Let $F : I \times I \to B$ be the path homotopy between f and g. Then $F(0,0) = b_0$. Let $\overline{F} : I \times I \to E$ be the lifting of F to E such that $\overline{F}(0,0) = e_0$. By Lemma 7.4, $\overline{F}$ is a path homotopy. Then $\overline{F}(0 \times I) = \{e_0\}$ and $\overline{F}(1 \times I)$ is a one-point set, namely $\{e_1\}$.

Then the restriction $\overline{F}|I \times 0$ of $\overline{F}$ to the bottom edge of $I \times I$ is a path in E, beginning at e_0, that is a lifting of $F|I \times 0$. By the uniqueness of path liftings, we must have $\overline{F}(s,0) = \overline{f}(s)$. Similarly, $\overline{F}|I \times I$ is a path on E that lifts $F|I \times 1$ and begins at e_0 because $\overline{F}(0 \times I) = \{e_0\}$. By the uniqueness of path liftings, $\overline{F}(s,1) = \overline{g}(s)$. Therefore, both $\overline{f}$ and $\overline{g}$ end at e_1 and $\overline{F}$ is a path homotopy between them. □

Definition 7.11. (Lifting Correspondence)
Let $p : E \to B$ be a covering map; let $b_0 \in B$. Choose e_0 so that $p(e_0) = b_0$. Given an element $[f]$ of $\pi_1(B, b_0)$, let $\overline{f}$ be the lifting of f to a path in E that begins at e_0. Let $\phi([f])$ denote the end point $\overline{f}(1)$ of $\overline{f}$. Then ϕ is a well-defined set map

$$\phi : \pi_1(B, b_0) \to p^{-1}(b_0)$$

We call ϕ the lifting correspondence derived from the covering map p. Notice that ϕ still depends on the choice of the point e_0.

Theorem 7.7. *Let $p : E \to B$ be a covering map; let $p(e_0) = b_0$. If E is path connected then the lifting correspondence*

$$\phi : \pi_1(B, b_0) \to p^{-1}(b_0)$$

is surjective. If E is simply connected, it is bijective.

Proof: If E is path connected, then given $e_1 \in p^{-1}(b_0)$, there is a path $\overline{f}$ in E from e_0 to e_1. But then $f = p \circ \overline{f}$ is a loop in B at b_0 and $\phi([f]) = e_1$ by definition.

Now suppose that E is simply connected. Let $[f]$ and $[g]$ be two elements of $\pi_1(B, b_0)$ such that $\phi([f]) = \phi([g])$ and let $\overline{f}, \overline{g}$ be the liftings of f and g, respectively, to paths in E that begin at e_0. Then $\overline{f}(1) = \overline{g}(1)$. Since E is simply connected there is a path homotopy $\overline{F}$ in E between $\overline{f}$ and $\overline{g}$. Then $p \circ \overline{F}$ is a path homotopy in B between f and g. □

Example 7.12. We can provide not only an interesting example of a lifting correspondence but prove that the fundamental group of $\mathbb{S}^1$ is indeed the additive group $\mathbb{Z}$, as previously suggested. Let $p : \mathbb{R} \to \mathbb{S}^1$ be the covering map given by

$$p(x) = (\cos 2\pi x, \sin 2\pi x)$$

Let $e_0 = 0$ and $b_0 = p(e_0)$. Then $p^{-1}(b_0)$ is the set $\mathbb{Z}$ of integers. Because $\mathbb{R}$ is simply connected, the lifting correspondence

$$\phi : \pi_1(\mathbb{S}^1, b_0) \to \mathbb{Z}$$

is bijective. To show that ϕ is an isomorphism, it remains to show that ϕ is a homomorphism. Suppose that $[f], [g] \in \pi_1(B, b_0)$ and let $\overline{f}, \overline{g}$ be their respective liftings to paths on $\mathbb{R}$ at 0. Notice that $\phi([f]) = \overline{f}(1)$ and $\phi([g]) = \overline{g}(1)$ by definition. Now let $\tilde{\overline{g}}$ be the path

$$\tilde{\overline{g}}(s) = \overline{f}(1) + \overline{g}(s)$$

on $\mathbb{R}$. Because $p(x) = p(\overline{f}(1) + x)$ for all $x \in \mathbb{R}$, the path $\tilde{\overline{g}}$ is a lifting of g as it begins at $\overline{f}(1)$. Then the product $\overline{f} * \tilde{\overline{g}}$ is defined and must be the lifting of $f * g$ that begins at 0. The end point of the path is $\tilde{\overline{g}}(1) = \overline{f}(1) + \overline{g}(1)$. But then it follows that

$$\phi([f] * [g]) = \overline{f}(1) + \overline{g}(1) = \phi([f]) + \phi([g])$$

7.4. **Manifolds.** Since we are going to be applying the symmetric groups role in topology to very real objects, it only remains to explain what these objects are. We follow Basener's "Topology and Its Applications" [?] for its wonderful description of the uses of Topology, except where indicated.

Definition 7.12. (Manifold) An n-dimensional manifold is a topological space M such that for any $x \in M$, there exists a neighborhood $U \subseteq M$ of x and an open set $V \subseteq \mathbb{R}^n$ such that U and V are homeomorphic. We also assume that a manifold is Hausdorff and second countable (it has a countably dense subset). We usually denote the homeomorphism by $\phi : U \to V$. A compact manifold is called a closed manifold.

In essence, manifolds are objects which when one looks "closely" enough at one, they appear to be a Euclidean space. For example, the sphere is a manifold as are the ripples in a sheet as flaps in the wind. Moreover, when one stands on the sphere, or for example the Earth, one would think that they are on a flat 2-dimensional plane when in reality one is on the surface of the sphere. Think about standing up with this paper and walking across the room. Though one seems to be walking in a straight line, since one is on the surface of the Earth, one is really walking on an arc. Despite the fact it seems like a straight line and my environment flat, it is actually curved. Manifolds are a higher dimensional analogue of what we image curves and surfaces to be. Moreover, they are of the greatest important in Applied Mathematics and the sciences as most of the objects we observe in nature are manifolds. However, many common objects are not manifolds. For example, the figure eight cannot be a manifold as any neighborhood of the self intersection of the curve is not homeomorphic with $\mathbb{R}$.

Definition 7.13. (Chart/Atlas) If M is a topological manifold, then the subsets $U \subseteq M$ are called the charts and the maps ϕ are called the chart maps. The collection of the charts that cover M with their associated chart maps is called an atlas for M and can be written $\{U_\alpha, \phi_\alpha\}$.

Notice if the manifold M is compact, then only finitely many charts are required. Thinking of the example where M is the sphere, we can take this definition literally. If you know roughly where you are on the sphere, i.e. the world, you know your U. Then you know what ϕ or chart map, what in the real world would be your literal flat map you could buy at the store, to tell you where you are locally on the sphere. But one must be careful not to picture "nice" spaces when thinking of manifolds. The reader may have been imagining "nice" spaces thus far, such as the sphere. But a manifold need not even have a boundary. A manifold with a boundary needs more specification.

Definition 7.14. (Manifold with Boundary) A manifold with boundary is a second countable Hausdorff topological space M such that for any $x \in M$, there is a neighborhood $U \subseteq M$ of x and an open set $V \subseteq \mathbb{R}^{n-1} \times \mathbb{R}_+$ such that U and V are homeomorphic. We denote the homomorphism $\phi : U \to V$. The set of points $x \in M$ with $\phi(x) \in \mathbb{R}^{n-1} \times \{0\}$ is called the boundary of M and the complement of the boundary of M in M is called the interior of M.

Moreover, all of the manifolds we have thus far mentioned are already subsets of a Euclidean space $\mathbb{R}^k$. But notice the definition only requires that the topological space be locally homeomorphic to a Euclidean space. Nowhere do we require that the space originate in Euclidean space. For example, consider the manifold as defined by

$$C = I/(0 \sim 1)$$

Here we have created the circle using an identification space and is not in $\mathbb{R}^2$. Though a technicality, this shows that the topological space doesn't need to be in $\mathbb{R}^k$ to begin with. But this technicality holds great importance. Notice that the circle divides the plane into two sets while the sphere does not divide $\mathbb{R}^3$ into two pieces. To talk about these differences, we need more definitions.

Definition 7.15. (Embedding/Immersion) A map $i : X \to Y$ is called an embedding if i is a homeomorphism between X and $i(X)$, where the topology on $i(X)$ is the subspace topology inherited from Y. To indicate that a map is an embedding, we write $i : X \hookrightarrow Y$. A map $f : X \to Y$ is called an immersion if given any point $x \in X$, there is a neighborhood U of x such that $f : U \to Y$ is a homeomorphism between U and $f(U)$. One thinks of an immersion as being locally an embedding.

First, note that every embedding is an immersion. However, the converse fails. For an immersion to be an embedding it needs to be injective and the inverse needs to be continuous.

Definition 7.16. (Topological Properties) A property of a topological space X is said to be intrinsic property if it depends only on the topological space X. A property of a topological space X embedded in Y is called extrinsic if it depends on the embedding.

These topological properties are the difference between the circle and sphere in how they break up the space they are embedded in. Moreover, we have one more important property of a surface.

Definition 7.17. (Orientable) A surface S is said to be orientable if there does not exist an embedding of a Möbius strip into S. If such an embedding exists, then the surface is said to be nonorientable.

Definition 7.18. (Embedded Manifold) An embedded n-dimensional manifold is a subset M of $\mathbb{R}^k$ (with the subspace topology) such that for any $x \in M$, there exists a neighborhood $U \subseteq M$ of x and an open set $V \subseteq \mathbb{R}^n$ such that U and V are homeomorphic.

Clearly, an embedded manifold is also a manifold (every subset of $\mathbb{R}^k$ is Hausdorff and second countable). However, is every manifold homeomorphic to an embedded manifold? Indeed, this is called the Whitney Embedding Theorem, which we will not prove here.

Theorem 7.8. *(Whitney Embedding Theorem) Every compact n-dimensional manifold embeds in some Euclidean space* $\mathbb{R}^k$.

7.5. **Vector Fields.** Vector fields and winding numbers express the intimate connection between Calculus and Topology. Vector fields tell us about forces at varying points in some force field while winding numbers tell us important information about paths in those fields. We have already considered the fundamental group, which studies possible paths in the space. However, since we will not be considering just arbitrary groups but manifolds, we can do Calculus on these groups. Hence, vector fields yield more information about the space.

Definition 7.19. (Vector Field) If U is an open subset of $\mathbb{R}^n$, then a vector field on U is a function $V : U\mathbb{R}^n$ that assigns to each point $x \in U$ a vector $V(x)$.

Moreover, a function between open sets in $\mathbb{R}^n$ is said to be continuously differentiable if it has continuous partial derivatives. For our purposes, all vector fields will considered to be continuously differentiable. Furthermore, recall that a contour integral is independent of the parametrization of the path γ. So if $\gamma : I \to U$ is a continuously differentiable path in U, then the integral of the vector field V along γ is defined

$$\int_\gamma V = \int_0^1 V(\gamma(t)) \cdot \gamma'(t)\,\mathrm{d}t$$

Because of the independence of $\int$ on γ's parametrization, choose $\|\gamma\| = 1$. Then $V(\gamma(t)) \cdot \gamma'(t)$ is the projection of V onto γ', or the amount of component of V in the direction of movement in γ. Therefore, $\int_\gamma V$ is the total amount of the vector field that points along γ, assuming increasing t.

Theorem 7.9. *(Differentiable Homotopy) Let $U \subset \mathbb{R}^n$ be an open set and γ_0 and γ_1 are differentiable paths in U with common endpoints. Let V be a vector field on U with $\nabla \times V = 0$. If there exists a differentiable homotopy from γ_0 to γ_1, then*

$$\int_{\gamma_0} V = \int_{\gamma_1} V$$

Proof: We prove this only in the case where H is injective. The general case follows similarly but with far more calculation. Let $\gamma_0 - \gamma_1$ be the path consisting of traversing γ_0 and then γ_1 in the negative direction. Then

$$\int_{\gamma_0} V - \int_{\gamma_1} V = \gamma_{\gamma_0 - \gamma_1} V$$

In this case, $\gamma_0 - \gamma_1$ forms a loop based at $\gamma_0(0)$. Moreover, this path bounds the region $D = \mathrm{Im}(H)$. Using Green's Theorem, we obtain

$$\int_{\gamma_0 - \gamma_1} V = \int_D \nabla \times V$$

But then we cheaply obtain

$$\int_{\gamma_0} V - \int_{\gamma_1} V = \int_D \nabla \times V = \int_D 0 = 0$$

We obtain then the following trivial corollary.

Corollary 7.3. *If $n \neq m$ then the loops γ_n and γ_m are defined as in Theorem 7.9 are not homotopic in $\mathbb{R}^2 - \{0\}$ and γ_0 is the only one that is homotopic to the constant map.*

Then we can define a very useful topological property of a curve in relation to the space in which it sits.

Definition 7.20. (Winding Number) Given a point $P = (a, b) \in \mathbb{R}^2$, define a vector field $V_{\theta,P}$ by

$$V_{\theta,P}(x, y) = V_\theta(x - a, y - b) = \left(\frac{-(y-b)}{(x-a)^2 + (y-b)^2} \ , \ \frac{(x-a)}{(x-a)^2 + (y-b)^2} \right)$$

Given a differentiable loop $\gamma : I \to \mathbb{R}^2 - P$, then the winding number around P is

$$W(\gamma, P) = \frac{1}{2\pi} \oint_\gamma V_{\theta,P}$$

The winding number is simply the amount of times a curve completes a circle around a point. The sign of the winding number tells which direction the curve has rotated around the point, positive being counterclockwise. Which then allows us to make the following statement

Theorem 7.10. *Two differentiable paths in $\mathbb{R}^2 - \{0\}$ are differentiably homotopic if and only if they have the same winding number.*

In fact, this concept is crucial in the method developed by William Thurston in his method of turning a sphere inside out (something that cannot be done with a circle, specifically, because doing so would result in a change in winding number).

7.6. **Condensed Matter Physics and Order Fields.** We finally have all the necessary tools and vocabulary to tackle the issue of topology in condensed matter physics. The key idea here will be symmetry. However, the notion of symmetry as we will be utilizing it is not the common perception of symmetry. For example, does the cube have "a lot" of symmetry. Many people would say, "yes". After all, there are 24 possible rigid motions of the cube alone, not including reflection symmetries.

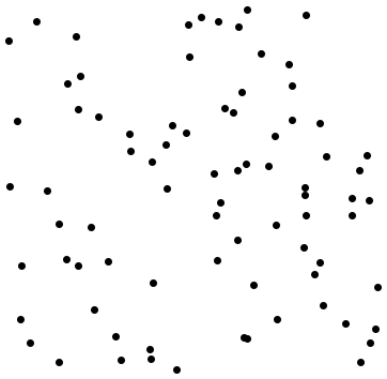
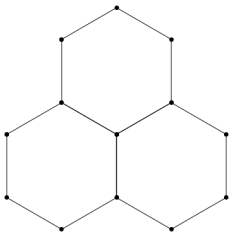
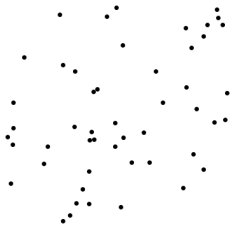


FIGURE 7. A collection of random points.

What are the symmetries of the collection of points in Figure 7? Are there any? A symmetry is essentially an action on an object which makes the object "look" like it did before. Notice for the points in Figure 7, any possible rotation or reflection will return the points to a state which resembles their state in Figure 7. However, there are objects one tends to think of as being very "symmetrical" that are not so random. Take the sphere, any possible 3-dimensional rotation of the sphere returns the sphere to a state which resembles the previous state, as does any reflection. In essence, the objects one tends to think of as being very "symmetrical" actually have very few possible symmetries. For example,



(A) A ice molecular arrangement.



(B) A collection of water molecules.

FIGURE 8. Notice the difference in the symmetries of water and its solid form - ice.

Notice in Figure 8, ice only has rotation symmetries of $\frac{2\pi}{3}$ and $\frac{4\pi}{3}$ radians (or any multiple of $\frac{pi}{3}$ radians for a symmetry with some shift). Therefore, the ice breaks rotational

invariance as it can only be shifted by multiples of $\frac{\pi}{3}$ radians. Moreover, the lattice can only be shifted by certain amounts, that is multiplies of the lattice unit. So ice breaks translational invariance. However, the water will appear the "same" under all possible symmetries.[10]

As evident from our examples and discussion thus far, our viewpoint will not be to focus on individual particles but rather focus on how large collections of particles come together to form a global behavior for the object. A perfect example of this is iron. Metals, such as iron, are full of free elections, that is electrons which are free to "move about" (that is exchange with other electrons' nuclei which they are orbiting). These movements allow the material to be a good conductor as well as allow the substance to become magnetic. The electrons moving about in the iron create a local direction of magnetization. If all these directions are aligned, then the iron will act as a magnet. However, if the local directions of magnetization vary largely over the iron, it will not act as a magnet. Though no single electron determines this, large groups of electrons and the directions of magnetization attached to them come together to determine the properties of the iron. Moreover, the iron itself does not act exactly as a solid. Its electrons move as those of a liquids but its nuclei are rigidly fixed in place as those of a crystal. To describe these global behaviors that are induced by smaller local behaviors (or orders), we look at the symmetries of the object in question.

Definition 7.21. (Order Field) An order field is a map from an open set $X \subseteq \mathbb{R}^n$ minus a singular set Z to a manifold R.

In many cases, R is a topological group of symmetries that is quotiented by a finite group of symmetries of the molecules involved. Then the order can be thought of as determining the direction the molecules are facing.

Example 7.13. If we imagine the molecules being examined as small cubes in $\mathbb{R}^3$ and assume that the cubes will prefer to orient themselves in the same direction, then R is $SO(3)/S_4$, where $SO(3)$ are all possible rotations in $\mathbb{R}^3$ and S_4 is the symmetries of the cube (obviously this is not exactly so but the rigid symmetries of the cube are isomorphic to S_4, as one can show).

However, it is important when working with R to make ones assumptions carefully as R could easily experience odd forces - such as quantum effects - that alter the "traditional" order field in ways one would not expect. The safest thing is to always only assume that R is a manifold.

Definition 7.22. (Order Parameter Space) Let R denote a manifold that we refer to as the order parameter space. Let X be an open subset of $\mathbb{R}$ and let Z be a subset of X. An order field (or the order parameter field) on X with singular set Z is a map

$$\Phi : X - Z \to R$$

The set Z is called the singular set of Φ and the points in Z are called singular points, singularities, or defects of Φ.

Definition 7.23. (Index) Let z be a point in Z and let γ be a loop in X whose winding number around z is 1 and whose winding number around any other point of Z is zero. Then the index of z in the order field Φ is

$$\text{index}_z\Phi = [\Phi \circ \gamma] \in \pi_1(R, z)$$

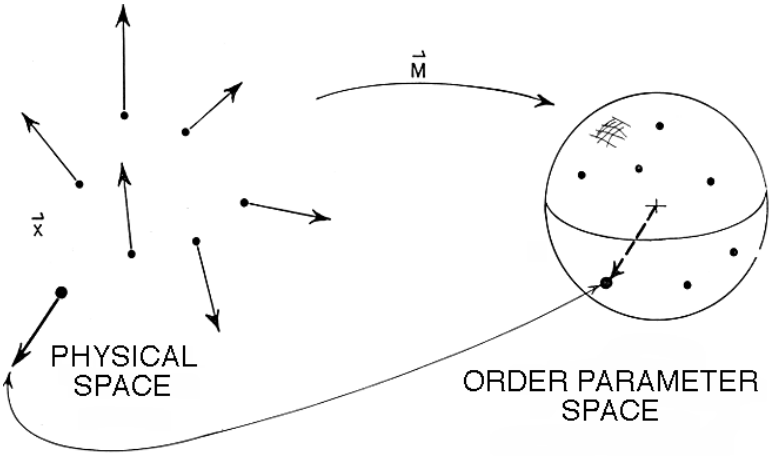


FIGURE 9. A collection of particles with their direction of local magnetization $M(x)$ on the left. On the right, we see the order parameter space for this collection local magnetizations.

Example 7.14. Examine Figure 9. On the left, at each point $\mathbf{x} = (x, y, z)$, we have a direction for the local magnetization, $M(x)$. The length of M is primarily fixed by the material in question. However, notice the direction of magnetization is undetermined. In becoming a magnet, the material has broken symmetry directions (as then only certain rotations suffice to return the material to a similar previous state). The order parameter M labels the various broken symmetry directions chosen by the materials. Specifically, take M to be the order parameter for a magnet. At a given temperature, the amount of magnetization $|M| = M_0$ will remain essentially constant. However, the energy is often independent of the direction $\hat{M} = \frac{M}{M_0}$ of magnetization. Throughout the material, the direction of magnetization changes essentially smoothly. We can think of the order parameter field as an arrow at each point in space giving a direction for the local magnetization. Moreover, we can think of it as a function taking points in the space $\mathbf{x}$ to the points on the sphere $|M| = M_0$. The sphere $\mathbb{S}^2$ is the order parameter space for the magnet (though the dimension we place $\mathbb{S}^2$ in does not matter).[10]

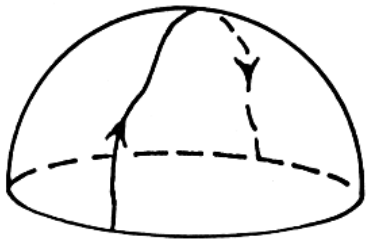


FIGURE 10. Here we represent the order parameter space for a nematic liquid crystal.

Example 7.15. Consider the space in Figure 10. A nematic liquid crystal (like those found in LCD displays in digital watches) is made up of long thing molecules that often align with each other. The molecules do not care which direction is regarded as up, their order parameter isn't exactly the vector $\mathbf{n}$ along the axis of molecules. Instead, it is a unit vector up to the equivalence $\mathbf{n} \equiv -\mathbf{n}$. The order parameter is then a half-sphere with the antipodal points equated. Then a loop over the entirety of the half-sphere is a closed loop and its intersections with the equator must correspond to the same orientations of the nematic molecules in the space.

Often, the material will have the lowest energy when the direction of magnetization is uniform. However, this does not often occur in nature - for example most iron pieces are not magnetic. This is because the material does not break symmetry uniformly. The individual parts of the material have a magnetization direction, but since they are not globally aligned, the material is not magnetic.

We have already stated that we would be interested in the symmetries of an object. So we will shift our focus to the symmetry groups of a molecule.

Definition 7.24. (Symmetry Group) Let G be a subgroup of $SO(2)$, called the symmetry group of a molecule. Then G consists of all rotations which are isometries of the molecule. G must have the form

$$G = G_n = \{0, \frac{1}{n}, \frac{2}{n}, \cdots, \frac{n-1}{n}\}$$

for some $n \in \mathbb{Z}_+$, where $\frac{m}{n}$ represents a rotation by $\frac{2\pi m}{n}$ - the only subgroups of $SO(2)$.

Definition 7.25. (n-Symmetry Field) Let X be an open subset of $\mathbb{R}$ and Z be a finite set of points in X. A n-symmetry field on X on X with a singular set Z is a map

$$\Phi : X - Z \to SO(2)/G_n$$

While every symmetry field are an order field, not every order field is a symmetry field. Furthermore, the order at a point $x \in X - Z$ is the rotation from some standard molecule to the molecule located at the point x. The order field defines, for each point $x \in X - Z$, some regular n-gon. Or in general,

$$SO(2)/G_n = [0, \frac{1}{n}] \quad \text{with } (0 \sim \frac{1}{n})$$

Then the index of z in the order field Φ is given as in definition of index above. Moreover, the index at a point z can be identified with a rational number $\frac{m}{n}$, with $m, n \in \mathbb{Z}$ and $n \neq 0$. Take a covering map $\mathbb{R} \to SO(2) \to SO(2)/G_n$, given by the quotient of $\mathbb{R}$ with the subgroup $\frac{1}{n}\mathbb{Z}$. Let $\overline{\Phi \circ \gamma} : [0,1] \to \mathbb{R}$ denote the lift of $\Phi \circ \gamma : [0,1] \to SO(2)/G_n$ beginning at $0 \in \mathbb{R}$. Then we have $\frac{m}{n} = \overline{\Phi \circ \gamma(1)}$. Then the index of a singularity in an order field is the amount that the molecules rotates traveling around the singularity.

Example 7.16. A X-Y ferromagnet is a thin film of material that has a lattice structure with some spin at each point in the lattice. When the material is in the ground state, the spins are confined to the plane. The spin acts as a direction vector giving a local direction of magnetization. The manifold R is then a circle

$$R = SO(2) \quad \text{with } \pi_1(SO(2)) = \mathbb{Z}$$

Then the index of any defect is some integer $n \in \mathbb{Z}$.

Example 7.17. The order parameter of superfluid helium-4 is defined by a complex number (the wave function) with a fixed magnitude. Then the manifold R is given by

$$R = SO(2) \quad \text{with } \pi_1(SO(2)) = \mathbb{Z}$$

which is similar to planar spins. Here, the magnitude is determined by several factors: temperature and pressure. However, these disappear at any temperature which is at least at the point where the phase changes from a superfluid to a normal fluid.

Example 7.18. A uniaxial nematic phase is a liquid crystalline phase of matter where the molecules appear like very narrow ellipses, whose symmetries are obviously like line segments. In a thin film, molecules will have their center of mass in some plane and will tilt at an angle of $\phi \in [0, \frac{\pi}{2}]$ with the normal vector defining the plane. Suppose that $\phi = \frac{\pi}{2}$, then the order is a 2-symmetry field with

$$R = SO(2)/G_2 \quad \text{with } \pi_1(SO(2)/G_2) = \frac{1}{2}\mathbb{Z} = \{\cdots, -1, \frac{-1}{2}, 0, \frac{1}{2}, 1, \cdots\}$$

The index of any defect is $\frac{n}{2}$ for some $n \in \mathbb{Z}$. Now if $\phi \in (0, \frac{\pi}{2})$, then the ends of the molecule are then indistinguishable by the direction, either above or below, in which they point from the plane. The order then looks like a direction field. The order parameter space is then

$$R = SO(2) \quad \text{with } \pi_1(SO(2)) = \mathbb{Z}$$

and the index of defect is also an integer. Finally, if $\phi = 0$, it must be the case that $R = SO(2)/SO(2) = 0$. Meaning, there are no stable defects.

Example 7.19. Suppose we have a 2-dimensional crystal, whose molecules lie in an integer lattice, then the crystal has an order field. The order of each molecule is measured as the vector in $\mathbb{R}^2$ from the position of the molecule to some position in an ideal lattice. Moreover, the order parameter is then periodic in both the x and y directions with period 1(because it is measured from the location of the molecule to any location in the lattice). Then it is the case that R is given by

$$R = \mathbb{T}^2 \quad \text{with } \pi_1(\mathbb{T}^2) = \mathbb{Z} \times \mathbb{Z}$$

The index of a defect is then a pair of integers $(m, n) \in \mathbb{Z} \times \mathbb{Z}$.

Now that we have sufficiently covered examples which are planar, we will now consider 3-dimensional pieces of materials whose order varies continuously with respect to position. Our definition of the order parameter need not change. However, for our purposes, we assume that Z is a finite collection of manifolds with dimension 0, 1, or 2 and unless stated $R = SO(3)/G$ for some subgroup G of $SO(3)$. But this need not always be the case for any given material. The defects now can be any point, curve, or surface and then their index is far more complicated than in the 2-dimensional case. So in the case that the defects are curves, that is with 1-dimensional components of Z called singular curves, Suppose that l is a singular curve for an order parameter Φ, the index of l in an order field Φ is

$$\text{Index}_l \Phi = [\Phi \circ \gamma] \in \pi_1(SO(3)/H)$$

where γ is any loop that goes only once around l, i.e. there is a homeomorphism from some open ball $B \subset X$ to a ball in $\mathbb{R}^3$ with radius 2 centered at the origin. Then take $l \cap B$ to the z-axis and the image of γ to the circle with radius 1 in the x, y-plane. The situation for a 0-dimensional singularity is more complicated and requires the second homotopy group, using 2-spheres in place of our previous circles.

Definition 7.26. (Second Homotopy Group) Let X be a topological space and b be a point in X. Let f be a map from the square $I^2T \times I$ to X such that $f(x, y) = b$ if (x, y) is in the boundary of I^2. We denote the equivalence class of all such maps that are homotopic to f, via a homotopy that always sends points in the boundary of T^2 to b, by $[f]$. The group of all such classes, where the product operation is defined similar to that for loops, is denoted $\pi_2(X, b)$ and called the second homotopy group of X based at b.

Since the sphere is of great importance in our analysis, it is important we know about their homotopy groups.

Theorem 7.11. *(n-Sphere Homotopy Groups)*
1. $\pi_n(\mathbb{S}^1) = 0$ *for* $n > 1$.
2. $\pi_n(\mathbb{S}^m) = 0$ *for* $n < m$.
3. $\pi_n(\mathbb{S}^n) = \mathbb{Z}$.

Moreover, the homotopy groups $\pi_n(\mathbb{S}^m)$ are all nontrivial for $m > n$ and often difficult to compute. Also, the reader should note that if X is a contractible space then $\pi_n(X, b)$ is trivial for all n. So in general, we can define the index of an order field $\Phi : X - Z \to R$ at some singularity $z \in X$ by taking a sphere f surrounding z only once but not surrounding any other point of Z. Then we have

$$\text{Index}_z \Phi = [\Phi \circ f] \in \pi_n(R)$$

A singularity z that has a nonzero index is called a hedgehog because these defects look like hedgehogs under a microscope. Now as with the two dimensional case, a symmetry field on X with a singular set Z is a map

$$\Phi : X - Z \to SO(3)/G$$

Of course, then the index of a singular curve l in the order field Φ is given by

$$\text{Index}_l\Phi = [\Phi \circ \gamma] \in \pi_1(SO(3)/G)$$

where γ is any loop which goes around l only once. In other words, the index of l is the continuous 1-parameter family of possible rotations that the molecules undergo as you travel around the curve l.

Example 7.20. Take the uniaxial nematic as given in Example 7.18 before but in 3-dimensions. The order is determined by the direction of $\mathbf{n}$. Then $R = SO(3)/(SO(2)/\mathbb{Z}_2) = \mathbb{S}^2/\mathbb{Z}_2 = \mathbb{RP}^2$. Of course,

$$R = \mathbb{RP}^2 \qquad \pi_1(\mathbb{RP}^2) = \mathbb{Z}_2 \qquad \pi_2(\mathbb{RP}^2) = \mathbb{Z}$$

But then of course uniaxial nematics can have singular curves and hedgehogs.

Example 7.21. A biaxial nematic has two axes with the symmetry of a pair of elongated ellipses making an "x". The symmetries are rotations by π about any of the axes. This is the symmetry of a prism whose cross sections are line segments. The symmetry group is then D_2 with

$$R = SO(3)/D_2 \qquad \pi_1(SO(3)/D_2) = \{e, R_x, R_y, R_z\} \qquad \pi_2(\mathbb{RP}^2) = 0$$

Where R_i is a reflection across the i-axis. The fundamental group here actually tells us about the space. Specifically, the nonabelian property of π_1 tells us the way the defects tangle and combine. Moreover, notice that biaxial nematics can have singular curves but not hedgehogs.

Example 7.22. A Heisenberg isotropic ferromagnet is given by the direction of the spin of the substance. It is then a direction field. Then

$$R = \mathbb{S}^2 \qquad \pi_1(\mathbb{S}^2) = 0 \qquad \pi_2(\mathbb{S}^2) = \mathbb{Z}$$

Example 7.23. Superfluid helium-3 in a dipole locked A-phase has the symmetry of a pair of orthogonal vectors $\mathbf{n}$ and $\mathbf{e} \in \mathbb{S}^3$. The only symmetry is the identity; therefore,

$$R = SO(3) \qquad \pi_1(SO(3)) = \mathbb{Z} \times \mathbb{Z} \qquad \pi_2(SO(3)) = 0$$

Then the substance can have singular curves but not hedgehogs.

Example 7.24. Consider superfluid helium-3 again but in the dipole free A-phase. The symmetries are determined by a pair of vectors that are free to rotate, $\mathbf{n} \in \mathbb{S}^2$ and $\mathbf{e} \in \mathbb{RP}^3$. The order is then determined by the function $O(x) = \mathbf{n}(x)\mathbf{e}(x)$, where x is a point in the medium. Then

$$R = (\mathbb{S}^2 \times \mathbb{RP}^3)/(\ (x, y) = (-x, -y)\ )$$

This yields $\pi_1() = \mathbb{Z}_4$ and $\pi_2(R) = \mathbb{Z}$. Then superfluid helium-3 in this state can have singular curves as well as hedgehogs.

Conclusion

It was our goal to see how the various pieces of Mathematics meet and interplay. Specifically, we examined the property of the symmetric group and its uses in the fields of Mathematics. Without doubt the theories developed are not trivial and are of great length and deep ideas. However, without great work one cannot expect to answer the great questions. Moreover, though the ideas may be vastly different and as a theoretical whole be none trivial, each individual step is simple and the idea quite trite. Indeed, we see how the roles of the symmetric group thread Galois Theory, Representation Theory, Lie Algebra Representations, Combinatorics, and Topology together. In this way, we see how one takes the big ideas and breaks them down into such simple concepts as permutations. All of Mathematics works in this way, taking the big ideas, breaking them down, and then generalizing them. We hope in this paper the reader has learned a new appreciation for this process.

8. Appendix

8.1. **Fundamental Theorem of Algebra.** Here we give two different proofs, an algebraic argument and a topological one, that says every polynomial $f(x) \in \mathbb{C}$ of degree n must have n roots in $\mathbb{C}$.

Theorem 8.1. *(Fundamental Theorem of Algebra - Proof 1)*[4] *Every polynomial $f(x) \in \mathbb{C}[x]$ of degree n has precisely n roots in $\mathbb{C}$ (counted with multiplicity). Equivalently, $\mathbb{C}$ is algebraically closed.*

Proof: It suffices to prove that every polynomial $f(x) \in \mathbb{C}[x]$ has a root in $\mathbb{C}$. Let τ denote the automorphism complex conjugation. If $f(x)$ has no root in $\mathbb{C}$ then neither does the conjugate polynomial $\overline{f}(x) = \tau f(x)$ (as conjugate roots come in pairs in $\mathbb{C}$) obtained by applying τ to the coefficients of $f(x)$ (since its roots are the conjugates of the roots of $f(x)$). The product $f(x)\overline{f}(x)$ has coefficients which are invariant under complex conjugation, hence has real coefficients. It suffices then to prove that a polynomial with real coefficients has a root in $\mathbb{C}$.

Supoose that $f(x)$ is a polynomial of degree n with real coefficients and write $n = 2^k m$, where m is odd. We prove that $f(x)$ has a root in $\mathbb{C}$ by induction on k. For $k = 0$, $f(x)$ has odd degree and polynomials of odd degree necessarily have a root in $\mathbb{R}$ so we are done. Now suppose that $k \geq 1$. Let $\alpha_1., \alpha_2, \cdots, \alpha_n$ be the roots of $f(x)$ and set $K = \mathbb{R}(\alpha_1, \alpha_2, \cdots, \alpha_n, i)$. Then K is a Galois extension of $\mathbb{R}$ containing $\mathbb{C}$ and the roots of $f(x)$. For any $t \in \mathbb{R}$, consider the polynomial

$$L_t = \prod_{1 \leq i < j \leq n} [x - (\alpha_i + \alpha_j + t\alpha_i\alpha_j)]$$

Any automorphism of $K/\mathbb{R}$ permutes the terms in this product so the coefficients of L_t are invariant under all the elements of $G(K/\mathbb{R})$. Hence, L_t is a polynomial with real coefficients. The degree of L_t is

$$\frac{n(n-1)}{2} = 2^{k-1}m(2^k m - 1) = 2^{k-1}m'$$

where m' is odd (since $k \geq 1$). The power of 2 in this degree is therefore less than k, so by induction the polynomial L_t has a root in $\mathbb{C}$. Hence, for each $t \in \mathbb{R}$, one of the elements $\alpha_i + \alpha_j + t\alpha_i\alpha_j$ for some i, j $(1 \leq i < j \leq n)$ is an element of $\mathbb{C}$. Since there are infinitely many choices for t and only finitely many values of i and j, we see that for some i and j (say, $i = 1$ and $j = 2$) there are distinct real numbers s and t with

$$\alpha_1 + \alpha_2 + s\alpha_1\alpha_2 \in \mathbb{C} \qquad \alpha_1 + \alpha_2 + t\alpha_1\alpha_2 \in \mathbb{C}$$

Since $s \neq t$, it follows that $a = \alpha_1 + \alpha_2 \in \mathbb{C}$ and $b = \alpha_1\alpha_2 \in \mathbb{C}$. But then α_1 and α_2 are the roots of the quadratic equation $ax^2 - ax + b$ with coefficients in $\mathbb{C}$, hence are elements of $\mathbb{C}$ because quadratic polynomials with coefficients in $\mathbb{C}$ have roots in $\mathbb{C}$ (i.e., there are no quadratic extension of $\mathbb{C}$). $\square$

Theorem 8.2. *(Fundamental Theorem of Algebra - Proof 2)*[9] *A polynomial equation*

$$x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0 = 0$$

of degree $n > 0$ with real or complex coefficients has at least one (real or complex) root.

Proof: We will show this in four steps:

1. We show that an induced homomorphism f_* of fundamental groups is injective.
2. We show that if $g : \mathbb{S}^1 \to \mathbb{R}^2 - \mathbf{0}$ is the map $g(z) = z^n$, then g is not nulhomotopic.
3. We prove a special case.
4. We then show the general case.

1. Consider the map $f : \mathbb{S}^1 \to \mathbb{S}^1$ given by $f(z) = z^n$, where $z \in \mathbb{C}$. We show the induced homomorphism f_* of fundamental groups is injective. Let $p_0 : I \to \mathbb{S}^1$ be a standard loop in $\mathbb{S}^1$.

   $$p_0(s) = e^{2\pi is} = (\cos 2\pi s, \sin 2\pi s)$$

   The image of I under f_* is the lop

   $$f(p_0(s)) = (e^{2\pi s})^n = (\cos 2\pi ns, \sin 2\pi ns)$$

   This loop lifts to the path $s \to ns$ in the covering space $\mathbb{R}$. Therefore, the loop $f \circ p_0$ corresponds to the integer n under the standard isomorphism of $\pi_1(\mathbb{S}^1, b_0)$ with the integers, whereas p_0 corresponds to the number 1. Thus, f_* is "multiplication by n" in the fundamental group of $\mathbb{S}^1$, so in particular, f_* is injective.
2. Now we show that if $g : \mathbb{S}^1 \to \mathbb{R}^2 - \mathbf{0}$ is the map $g(z) = z^n$, then g is not nulhomotopic. The map g equals the map f of step 1 followed by the inclusion map $j : \mathbb{S}^1 \to \mathbb{R}^2 - \mathbf{0}$. Now f_* is injective and j_* is injective because $\mathbb{S}^1$ is a retract of $\mathbb{R}^2 - \mathbf{0}$. Therefore, $g_* = j_* \circ f_*$ is injective. Therefore, g cannot be nulhomotopic.
3. Now we prove a special case based on a condition on the coefficients. Given a polynomial equation

   $$x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0 = 0$$

   we shall assume that

   $$|a_{n-1}| + \cdots + |a_1| + |a_0| < 1$$

   and show that the equation must have a root lying the the unit ball $\mathbb{B}^2$.

   Suppose there does not exist such a root. Then define a map $k : \mathbb{B}^2 \to \mathbb{R}^2 - \mathbf{0}$ by the equation

   $$k(z) = z^n + a_{n-1}z^{n-1} + \cdots + a_1z + a_0$$

   Let h be the restriction of k to $\mathbb{S}^1$. Because h extends to a map of the unit ball into $\mathbb{R}^2 - \mathbf{0}$ by the equation

   $$F(z, t) = z^n + t(a_{n-1}z^{n-1} + \cdots + a_0)$$

$F(z,t)$ is never $\mathbf{0}$ because

$$\begin{aligned}|F(z,t)| \geq & |z^n| - |t(a_{n-1}z^{n-1} + \cdots + a_0| \\ \geq & 1 - t(|a_{n-1}z^{n-1}| + \cdots + |a_0|) \\ = & 1 - t(|a_{n-1} + \cdots + |a_0|) > 0\end{aligned}$$

4. Now we can show the general case. Given a polynomial equation

$$x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 = 0$$

let us choose a real number $c > 0$ and substitute $x = cy$. Then we obtain the equation

$$(cy)^n + a_{n-1}(cy)^{n-1} + \cdots + a_1 cy + a_0 = 0$$

Or equivalently,

$$y^n + \frac{a_{n-1}}{c}y^{n-1} + \cdots + \frac{a_1}{c^{n-1}}y + \frac{a_0}{c^n} = 0$$

If this equation has the root $y = y_0$, then the original equation has the root $x_0 = cy_0$. We need merely choose c large enough that

$$\left|\frac{a_{n-1}}{c}\right| + \left|\frac{a_{n-2}}{c^2}\right| + \cdots + \left|\frac{a_1}{c^{n-1}}\right| + \left|\frac{a_0}{c^n}\right| < 1$$

But then the theorem reduces down to the case considered in Step 3. □

8.2. **Solving the Linear, Quadratic, Cubic, and Quartic Polynomials[4].** Here we show how to solve all possible polynomials of degree n for $0 \leq n \leq 5$.

1. $n = 0$: This is the constant polynomial and is only zero if it is the zero function, in which case it is zero infinitely many times. Thus, we discard it as a possibility as a polynomial of degree n should have n roots.

2. $n = 1$: This is the linear equation

$$a_1 x + a_0 = 0$$

for $a_0, a_1 \in \mathbb{C}$ with $a_1 \neq 0$. This polynomial has the trivial solution

$$x = -\frac{a_0}{a_1}$$

Here, the Galois group is obviously trivial.

3. $n = 2$: We solve this by a method called completing the square (a method known even to the Babylonians). Start with

$$ax^2 + bx + c = 0$$

with $a, b, c \in \mathbb{C}$ and $a \neq 0$. Division by a yields,

$$x^2 + \frac{b}{a}x + \frac{c}{a} = 0$$

We proceed as follows:

$$\begin{aligned}
x^2 + \frac{b}{a}x + \frac{c}{a} &= 0 \\
x^2 + \frac{b}{a}x + \frac{c}{a} + \left(\frac{b}{2a}\right)^2 - \left(\frac{b}{2a}\right)^2 &= 0 \\
x^2 + \frac{b}{a}x + \left(\frac{b}{2a}\right)^2 &= \left(\frac{b}{2a}\right)^2 - \frac{c}{a} \\
\left(x + \frac{b}{2a}\right)^2 &= \frac{b^2}{4a^2} - \frac{c}{a} \\
\left(x + \frac{b}{2a}\right)^2 &= \frac{b^2 - 4ac}{4a^2} \\
x + \frac{b}{2a} &= \sqrt{\frac{b^2 - 4ac}{4a^2}} \\
x + \frac{b}{2a} &= \frac{\pm\sqrt{b^2 - 4ac}}{2a} \\
x &= \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}
\end{aligned}$$

We can also calculate the Galois group of the quadratic. If we are given a polynomial $x^2 + ax + b$ with roots α, β, the discriminant D for this polynomial is $(\alpha - \beta)^2$ can be written as a polynomial in the elementary symmetric functions of the roots.

$$D = s_1^2 - 4s_2 = a^2 - 4b$$

The polynomial is separable if and only if $a^2 - 4b \neq 0$. The Galois group is a subgroup of S_2, the cyclic group of order 2 and is trivial (as A_2 is trivial) if and only if $a^2 - 4b$ is a rational square, which completely determines the possible Galois groups. If the polynomial is reducible (i.e. E is a square in F), then the Galois group is trivial (the splitting field is just F), while if the polynomial is irreducible then the Galois group is isomorphic to $\mathbb{Z}/2\mathbb{Z}$ since the splitting field is the quadratic extension $F(\sqrt{D})$.

4. Suppose the cubic polynomial is

$$f(x) = x^3 + ax^2 + bx + c$$

If we make the substitution $x = y - \frac{a}{3}$, the polynomial becomes

$$g(y) = y^3 + py + q$$

where

$$p = \frac{1}{3}(3b - a^2) \qquad q = \frac{1}{27}(2a^3 - 9ab + 27c)$$

The splitting fields for these two polynomials are the same since their roots differ by the constant $\frac{a}{3} \in F$ and since the formula for the discriminant involves the difference of roots, we see that these polynomials have the same discriminant.

Let the roots of the polynomial be α, β, and γ. We compute the discriminant of this polynomial in terms of p and q. Note that

$$g(y) = (y-\alpha)(y-\beta)(y-\gamma)$$

so that if we differentiate, we have

$$D_y g(y) = (y-\alpha)(y-\beta) + (y-\alpha)(y-\gamma) + (y-\beta)(y-\gamma)$$

Then

$$\begin{aligned} D_y g(\alpha) =& (\alpha-\beta)(\alpha-\gamma) \\ D_y g(\beta) =& (\beta-\alpha)(\beta-\gamma) \\ D_y g(\gamma) =& (\gamma-\alpha)(\gamma-\beta) \end{aligned}$$

Taking the product we see that

$$D = [(\alpha-\beta)(\alpha-\gamma)(\beta-\gamma)]^2 = -D_y g(\alpha) D_y g(\beta) D_y g(\gamma)$$

Since $D_y g(y) = 3y^2 + p$, we have

$$\begin{aligned} -D =& (3\alpha^2+p)(3\beta^2+p)(3\gamma^2+p) \\ =& 27\alpha^2\beta^2\gamma^2 + 9p(\alpha^2\beta^2 + \alpha^2\gamma^2 + \beta^2\gamma^2) + 3p^2(\alpha^2+\beta^2+\gamma^2) + p^3 \end{aligned}$$

The corresponding expression in the elementary symmetric functions of the roots are $s_1^2 - 4s_2$ and $s_1^2 - 2s_2$. Note that $s_1 = 0, s_2 = p, s_3 = -q$. We obtain

$$-D = 27(-q)^2 + 9p(p^2) + 3p^2(-2p) + p^3$$

so that

$$D = -4q^3 - 27q^2$$

The same as the discriminant of our original cubic. Expressing D in terms of a, b, c we obtain

$$D = a^2b^2 - 4b^3 - 4a^3c - 27c^2 + 18abc$$

Now we compute its Galois group. We need consider two cases. Case one is where the cubic $f(x)$ is reducible. Then it splits either into three linear factors or into a linear factor and an irreducible quadratic. In the first case the Galois group is trivial and in the second the Galois group is of order 2 (from above). Case 2 is when the polynomial $f(x)$ is irreducible. Then a root of $f(x)$ generates an extension of degree 3 over F, so the degree of the splitting field over F is divisible by 3. Since the Galois group is a subgroup of S_3, there are only two possibilities, namely A_3 or S_3. The Galois group is A_3 if and only if the discriminate D is a square.

Explicitly, if D is the square of an element of F, then the splitting field of the irreducible cubic $f(x)$ is obtained by adjoining any single root of $f(x)$ to F. The resulting field is Galois over F of degree 3 with a cyclic group of order 3 as a Galois group. If D is not the square of an element of F then the splitting field of $f(x)$ is of degree 6 over F, hence is the field $F(\theta, \sqrt{D})$ for any one of the roots θ of $f(x)$. This extension is Galois over F with Galois group S_3 (generators given by σ, which takes θ to one of the other roots of $f(x)$ and fixed $\sqrt{D}$ and τ, which takes $\sqrt{D}$ to $-\sqrt{D}$ and fixes θ).

In either case, the splitting field for the irreducible cubic $f(x)$ is obtained by adjoining $\sqrt{D}$ and root of $f(x)$ to F.

5. $n = 4$: Let the quartic polynomial be

$$f(x) = x^4 + ax^3 + bx^2 + cx + d$$

which under the substitution $x = y - \frac{a}{4}$ becomes the quartic

$$g(y) = y^4 + py^2 + qy + r$$

with

$$\begin{aligned} p =& \frac{1}{8}(-3a^2 + 8b) \\ q =& \frac{1}{8}(a^3 - 4ab + 8c) \\ r =& \frac{1}{256}(-3a^4 + 16a^2b - 64ac + 256d) \end{aligned}$$

Let the roots of $g(y)$ be $\alpha_1, \alpha_2, \alpha_3$, and α_4 and let G denote the Galois group for the splitting field of $g(y)$ (or of $f(x)$).

Suppose first that $g(y)$ is reducible. If $g(y)$ splits into a linear and a cubic, then G is the Galois group of the cubic, determined able. Suppose then that $g(y)$ splits into two irreducible quadratics. Then the splitting field is the extension $F(\sqrt{D_1}, \sqrt{D_2})$, where D_1 and D_2 are the discriminates of the two quadratics. If D_1 and D_2 do not differ by a square factor then this extension is a biquadratic extension and G is isomorphic to the Klein 4-subgroup of S_4. If D_1 is a square times D_2 then this extension is a quadratic extension and G is isomorphic to $\mathbb{Z}/2\mathbb{Z}$.

We are reduced to the situation where $g(y)$ is irreducible. In this case recall that the Galois group is transitive on the roots, i.e. it is possible to get from a given root to any other root by applying some automorphism of the Galois group. Examining the possibilities we see that the only transitive subgroups of S_4, hence the only possibilities are S_4,A_4, D_8 and its conjugates, $V = \{1, (1\ 2)(3\ 4), (1\ 3)(2\ 4), (1\ 4), (2\ 3)\}$, or $C = \{1, (1\ 2\ 3\ 4), (1\ 3)(2\ 4), (1\ 4\ 3\ 2)\}$ and its conjugates.

Consider the elements

$$\begin{aligned} \theta_1 =& (\alpha_1 + \alpha_2)(\alpha_3 + \alpha_4) \\ \theta_2 =& (\alpha_1 + \alpha_3)(\alpha_2 + \alpha_4) \\ \theta_3 =& (\alpha_1 + \alpha_4)(\alpha_2 + \alpha_3) \end{aligned}$$

in the splitting field for $g(y)$. These elements are permuted amongst themselves by the permutations in S_4. The stabilizer of θ_1 in S_4 is the dihedral group D_8. The stabilizers in S_4 of θ_2 and θ_3 are the conjugate dihedral subgroups of order 8. The subgroup of S_4 which stabilizes all three of these elements is the intersection of these subgroups, namely the Klein 4-group V.

Since S_4 merely permutes $\theta_1, \theta_2, \theta_3$, it follows that the elementary symmetric functions in the θ's are fixed by all the elements of S_4, hence are in F. An elementary computation

in symmetric functions shows that these elementary symmetric functions are $2p$, p^2-4r, and $-q^2$, which shows that $\theta_1, \theta_2, \theta_3$ are the roots of

$$h(x) = x^3 - 2px^2 + (p^2 - 4r)x + q^2$$

called the resolvent cubic for the quartic $g(y)$. Since

$$\begin{aligned}\theta_1 - \theta_2 =& \alpha_1\alpha_3 + \alpha_2\alpha_4 - \alpha_1\alpha_2 - \alpha_3\alpha_4 \\ =& -(\alpha_1 - \alpha_4)(\alpha_2 - \alpha_3)\end{aligned}$$

and similarly

$$\begin{aligned}\theta_1 - \theta_3 =& -(\alpha_1 - \alpha_3)(\alpha_2 - \alpha_4) \\ \theta_2 - \alpha_3 =& -(\alpha_1 - \alpha_2)(\alpha_3 - \alpha_4)\end{aligned}$$

we see that the discriminant of the resolvent cubic is the same as the discriminant of the quartic $g(y)$, hence also the discriminant of the quartic $f(x)$. Using our formulation for the discriminant of the cubic, we can easily compute the discrimination in terms of p, q, r:

$$D = 16p^4r - 4p^3q^2 - 128p^2r^2 + 144pq^2r - 27q^4 + 256r^3$$

from which one can give the formula for D in terms of a, b, c, d:

$$\begin{aligned}D =& -128b^2d^2 - 4a^3c^3 + 16b^4d - 4b^3c^2 - 27a^4d^2 + 18abc^3 \\ & 144a^2bd^2 - 192acd^2 + a^2b^2c^2 - 4a^2b^3d - 6a^2c^2d \\ & 144c^2d + 256d^3 - 27c^4 - 80ab^2cd + 18a^3bcd\end{aligned}$$

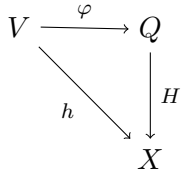
The splitting field for the resolvent cubic is a subfield of the splitting field of the quartic, so the Galois group of the resolvent cubic is a quotient of G. Hence, knowing the action of the Galois group on the roots of the resolvent cubic $h(x)$ gives information about the Galois group of $g(y)$ as below.

First, suppose that the resolvent cubic is irreducible. If D is not a square, then G is not contained in A_4 and the Galois group of the resolvent cubic in S_3, which implies that the degree of the splitting field for $g(y)$ is divisible by 6. The only possibility then is $G = S_4$. If the resolvent cubic is irreducible and D is a square, then G is a subgroup of A_4 and 3 divides the order of G (the Galois group of the resolvent cubic is A_3). The only possibility is $G = A_4$.

We are left in the case where the resolvent cubic is reducible. The first possibility is that $h(x)$ has 3 roots in F. Sine each of the elements $\theta_1, \theta_2, \theta_3$ is in F, every element of G fixes all three of these elements, which means $G \subseteq V$. The only possibility is that $G = V$. Now if $h(x)$ splits into a linear and a quadratic, then precisely one of $\theta_1, \theta_2, \theta_3$ is in F, say θ_1. The G stabilizes θ_1 but not θ_2 and θ_3, so we have $G \subseteq D_8$ and $G \nsubseteq V$. This leaves two possibilities: $G + D_8$ or $G = C$. One way to distinguish between these is to observe that $F(\sqrt{D})$ is the fixed field of the elements G in A_4. For the two cases being considered, we have $D_8 \cap A_4 = V$, $C \cap A_4 = \{1, (1\ 3)(2\ 4)\}$. The first group is transitive on the roots of $g(y)$, the second is not. It follows that the first case occurs if

and only if $g(y)$ is irreducible over $F(\sqrt{D})$. We may therefore determine G completely by factoring $g(y)$ in $F(\sqrt{D})$ and so completely determine the Galois group in all cases.

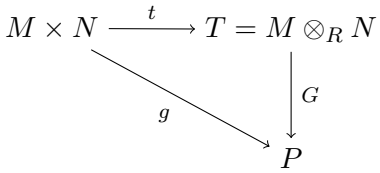
8.3. **Tensor Products.** Here we will discuss briefly what the tensor product is and how it relates to extension fields. We assume the reader is familiar with linear maps between vector spaces. Suppose that we want to extend this idea into a broader "new idea". Meaning that if V is a vector space with $v_0, w_0 \in V$, we want to find a vector space Q such that all linear transformations $\varphi : V \to Q$ given by $v \mapsto \overline{v}$, we have $\overline{v}_0 = \overline{w}_0$ and that this space is somehow the "best one". What we mean by "best one" is that it has all the universal mapping properties one would ever want in that create create the following commutative diagram



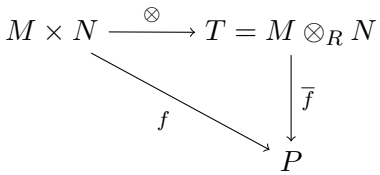
where h is a linear map to another vector space X and H is a unique function such that $H \circ \varphi = h$, i.e that $H(\overline{v}) = h(v)$ for all $v \in V$. So in a sense, the tensor product somehow captures all the information about linear maps from a vector space. However, here we construct the tensor product in a more general setting over a commutative ring.

Accordingly, let R be a commutative ring with unity $1 \neq 0$. Now let M, N, P be R-modules. The question we wish to solve is following:

Create a mapping t from $M \times N$ such that given any bilinear function $g : M \times N \to P$, there is a unique G such that the following diagram commutes.



That is, $G \circ t = g$ and t is the "best" linear map we discussed above. If we wanted to be truly rigorous, we say that given a commutative ring with unity $1 \neq 0$ R, a right R-module M, a left R-module N, and P any abelian group (though often this will too be an R-module), then the tensor product over R, $M \otimes_R N$ is an abelian group with a bilinear map $\otimes : M \times N \to M \otimes_R N$ that is universal in that for every abelian group P and bilinear map $f : M \times N \to P$, there is a unique group homomorphism $\overline{f} : M \otimes_R N \to P$ such that the following diagram commutes



that is $\overline{f} \circ \otimes = f$. How do we know such a space even exists? Well, we show that it exists.

Theorem 8.3. *The tensor product exists and is unique up to isomorphism.*

Proof: Uniqueness is quite trivial to show and we will not bother to demonstrate it here. To prove existence, let $W \subseteq V$ and

$$\begin{aligned} W =& \text{span}(v_0 - w_0) \\ =& \{\alpha(v_0 - w_0) \,|\, \alpha \in R\} \end{aligned}$$

Let $Q = V/W$, that is $p : V \to V/W$ where p is the projection of V onto W given by $v \mapsto v + W$. Now p is the "new idea" that we spoke of earlier. We have

$$\begin{aligned} \overline{v}_0 =& v_0 + W \\ \overline{w}_0 =& w_0 + W \\ v_0 - w_0 \in& W \end{aligned}$$

So now let $t : M \times N \to T$ be a bilinear function $t(m, n) = m \otimes n$ such that

1. $(m + m') \otimes n = m \otimes n + m' \otimes n$
2. $(rm) \otimes n = r(m \otimes n)$
3. $m \otimes (n + n') = m \otimes n + m \otimes n'$
4. $m \otimes (rn) = r(m \otimes n)$

where T is the space we shall construct. So consider the free module

$$R^{(S)} = F_S \to T$$

where $R^{(S)}$ is the set of δ_s, where $s \in S$, S is the set of generators, and the mapping is onto. Then $R^{(M \times N)}$ has the basis

$$\{\delta_s \,|\, s \in S\}$$

but this is a free module on (m, n)

$$\sum_{i=1}^{k} r_i(m_i, n_i) \; ; \; r_i \in R$$

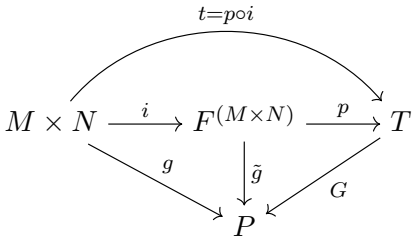
So consider the onto mapping $R^{(M \times N)} \to T$ given by $(m, n) \mapsto m \otimes n$. But we still don't have our space T! But notice in the 4 equations above for $t(m, n)$, if we were to set those to be 0 then they would all be in the ker of the mapping. So we can let R be the "relations", that is to say equations we want to be true. That is, let R be the submodule generated by our "equations"

1. $(m + m', n) - (m, n) - (m', n)$
2. $(m, n + n') - (m, n) - (m, n')$
3. $(rm, n) - r(m, n)$
4. $(m, rn) - r(m, n)$

Then we can let T be the space $R^{(M\times N)}/\ker$. But then we finally have

$$T = M \otimes N \stackrel{\text{def}}{=} F^{(M\times N)}/R$$

It only remains to show that we have the desired universal mapping properties.



Where i and p have the universal mapping property, g is bilinear, and G is the unique function from the theorem statement. We need to show that $g(m,n) = \tilde{g}(m,n)$. It is obvious that $\tilde{g}$ send all our "equations" to 0 if g sends them to 0. But does g send them to 0? Yes, because it sends things in the ker to 0. So g and $\tilde{g}$ are bilinear and equivalent. $\square$

Let us make a few comments. First, the free R-module we created in the theorem is *extremely* large! Say M, N, and R were the ring $\mathbb{R}$, then the free R-module contains $\mathbb{R}^2$ copies of $\mathbb{R}$! This size comes from the fact that the the construction was equivalent to

$$\bigoplus_{(m,n)\in M\times N} R\delta_{(m,n)}$$

so the sum is running over all possible pairs of vectors, not just those in the basis of M and N. We will not go further into the properties of tensor products. However, we can use them to construct fields with the roots of polynomials similarly to extension fields. Really what you do is extend scalars. Consider the polynomial $x^2 + 1 \in \mathbb{R}[x]$. We need to extend to $\mathbb{C}$ in some way. But then we can use

$$V_{\mathbb{C}} = \mathbb{C} \otimes_{\mathbb{R}} V$$

where V is a vector space over $\mathbb{R}$. This space has dimension 2dim V over $\mathbb{R}$ but dimension dim V over $\mathbb{C}$. In this vein, we can construct fields with the necessary roots of the polynomials as scalars.

8.4. **Lie Algebras.** Here we will construct many examples of the most commonly seen Lie algebras, following the method used by Hazewinkel, Gubareni, and Kirichenko[7].

1. Any vector space L with Lie bracket given by $[u, v] = 0$ for all $u, v \in L$ is an abelian Lie algebra. Moreover, all abelian Lie algebras look like this.
2. Suppose A is an associate algebra with operations $+, \cdot$ called addition and multiplication, respectively. Then let the Lie bracket $[\cdot, \cdot,]$ be given by

$$[x, y] = x \cdot y - y \cdot x$$

often called the bracket or commutator (less commonly the difference) of x and y for all $x, y \in A$. If A is not a commutative algebra (if A is commutative then we are in the case of the first Lie algebra we constructed), then the bracket satisfies:
   1. $[x, x] = 0$
   2. $[[x, y], z] + [[y, z], x] + [[z, x], y] = 0$

   for all $x, y, z \in A$. Then A becomes a Lie algebra with respect to the bracket operation $[\cdot, \cdot]$. So any associative algebra A with product $x \cdot y$ (often simply written xy) can be made into a Lie algebra.
3. Let $A = \mathrm{End}_k(V)$ be the endomorphism algebra of a k-vector space V. Then define the operation $[x, y]$ by $[x, y] = xy - yx$ for all $x, y \in A$. Then with respect to the operation $[\cdot, \cdot]$, A becomes a Lie algebra over the field k. To differentiate the Lie algebra from the "composition of endomorphisms" of the original algebra, one uses the notation $\mathfrak{gl}(V)$ for the Lie algebra. This is called the general linear Lie algebra. Any subalgebra of $\mathfrak{gl}(V)$ is called a linear Lie algebra.
4. Let $A = M_n(k)$ be the set of $n \times n$ matrices over a field k. Then define the operation $[\mathbf{X}, \mathbf{Y}] = \mathbf{XY} - \mathbf{YX}$, where $\mathbf{XY}$ denotes the usual matrix product. Then with respect to $[\cdot, \cdot]$, A becomes a Lie algebra that is denoted $\mathfrak{gl}(n, k)$. Now as a vector space, $\mathfrak{gl}(n, k)$ has basis of matrix units $e_{i,j}$ for $1 \leq i$ and $j \leq n$, where $e_{i,j}$ denotes the matrix with a 1 in the i, jth position and 0's elsewhere. It is a simple exercise to check that

$$[e_{i,j}, e_{r,s}] = \delta_{j,r} e_{i,s} - \delta_{i,s} e_{r,j}$$

where δ is the standard Kronecker delta function. According to Ado's Theorem, any finite dimensional Lie algebra over a field k of characteristic 0 is a linear Lie algebra. But then it must be isomorphic to some subalgebra of $\mathfrak{gl}(n, k)$.
5. Let $A = \{\mathbf{X} \in M_n(k) \,|\, \mathrm{trace}\ \mathbf{X} = 0\}$ be the set of all $n \times n$ over a field k of trace 0. Then A is a Lie algebra under the same bracket operation as $\mathfrak{gl}(n, k)$. This algebra is denoted $\mathfrak{sl}(n, k)$ and is called the special linear Lie algebra. Of course, $\mathfrak{sl}(n, k)$ has basis consisting of $\{e_{i,j} \,|\, 1 \leq i, j \leq n \wedge i \neq j\} \bigcup \{e_{i,i} - e_{i+1,i+1} \,|\, 1 \leq i \leq n - 1\}$.
6. Let $\mathfrak{b}(, nk) = \{\mathbf{X} = (x_{i,j}) \in M_n(k) \,|\, x_{i,j} = 0 \text{ when } i > j\}$ be the set of all upper triangular matrices. Then if we let $[\cdot, \cdot]$ be the same operation as in $\mathfrak{gl}(n, k)$. Then $\mathfrak{b}(n, k)$ is a Lie algebra over the field k.
7. Let $\mathfrak{n}(n, k) = \{\mathbf{X} = (x_{i,j}) \,|\, x_{i,j} = 0 \text{ when } i \geq j\}$ be the algebra of all strictly upper triangular matrices with the Lie bracket given by that in $\mathfrak{gl}(n, k)$.

8. Let $A = \{\mathbf{X} \in M_n(k) \,|\, \mathbf{X}^T = -\mathbf{X}\}$ be the set of all skew symmetric matrices over the field k. Then A is a Lie algebra with Lie bracket given by that in the matrix commutator. This algebra is often denoted $\mathfrak{O}(n,k)$.
9. Let $A = \{\mathbf{X} \in M_2(\mathbb{R}) \,|\, \begin{pmatrix} x & y \\ 0 & 0 \end{pmatrix}$ with $x, y \in \mathbb{R}\}$ be the subset of real 2×2 matrices with a zero second row. Then $A = \mathfrak{aft}(1)$ is the affine Lie algebra of the line with basis given by
$$\mathbf{X} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad , \quad \mathbf{Y} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$
and Lie bracket given by
$$[\mathbf{X}, \mathbf{Y}] = \mathbf{Y}$$
10. Let $\mathbb{R}^4$ be the real 4-dimensional Euclidean space with vectors $v = (x, z, y, t)$. Then considering the Lorentz inner product
$$\langle v, v \rangle_L = x^2 + y^2 + z^2 - t^2$$
If
$$I_{3,1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$
then we have
$$\langle v, v \rangle_L = v^T I_{3,1} v$$
Then the subset $\mathfrak{l}_{3,1}$ of all real 4×4 matrices M with
$$M^T I_{3,1} + I_{3,1} M = 0$$
is a Lie algebra called the Lorentz Lie algebra.
11. Let u, v be real-valued differentiable functions of $2n$ variables: $p = (p_1, \cdots, p_n)$ and $q = (q_1, \cdots, q_n)$. These variables could also possibly be dependent on time, t. Then we define the Poisson bracket by
$$\{p, q\} = \sum_{i=1}^{n} \left(\frac{\partial u}{\partial q_i} \frac{\partial v}{\partial p_i} - \frac{\partial u}{\partial p_i} \frac{\partial v}{\partial q_i} \right)$$
Then the space formed by the variables u, v under the Poisson bracket forms a Lie algebra.
12. Let A be an algebra, not necessarily associative, over a field k. Then a k-derivation of A is a k-linear mapping $D : A \to A$ such that
$$D(ab) = D(a)b + aD(b)$$
for all $a, b \in A$. The k-module of derivations carries a Lie algebra structure given by the commutator difference composition
$$[D_1, D_2] = D_1 D_2 - D_2 D_1$$

with $(D_1 D_2)(a) = D_1(D_2(a))$. The derivation is the algebraic counterpart to vector fields in topology, analysis, and differential geometry.

13. Let V be a vector space over k with basis L_n, c for $n \in \mathbb{Z}$. Then take

$$[L_n, c] = 0 \text{ for all } n$$

$$[L_m, L_n] = (m-n)L_{m+n} + \delta_{m+n,0}\frac{m^3 - m}{12}c$$

This action defines a Lie algebra called the Virasoro algebra.

14. If we quotient the center spanned by c in the Virasoro algebra again yields a quotient Lie algebra with basis L_n, $n \in \mathbb{Z}$ and bracket

$$[L_m, L_n] = (m-n)L_{m+n}$$

We can view this as a polynomial vector field on the circle

$$L_m = -z^{m+1}\frac{d}{dz} \text{ for } n \in \mathbb{Z}$$

This Lie algebra is known as the Witt algebra, though they were first defined by Cartan.

References

[1] Grove, Larry C. "Counting with Groups." *Groups and Characters*. New York: Wiley, 1997. 61-75. Print.

[2] Gallian, Joseph A. *Contemporary Abstract Algebra*. Independence, KY: Brooks Cole, 2009. Print.

[3] Fraleigh, John B. *A First Course in Abstract Algebra*. Reading, MA: Addison-Wesley Pub., 1967. Print.

[4] Dummit, David Steven., and Richard M. Foote. *Abstract Algebra*. 3rd ed. Hoboken, NJ: Wiley, 2004. Print.

[5] Sagan, Bruce Eli. *The Symmetric Group: Representations, Combinatorial Algorithms, and Symmetric Functions*. 2nd ed. New York: Springer, 2001. 1-90. Print.

[6] Stillwell, John. *Naive Lie Theory*. New York: Springer, 2008. Print.

[7] Nadiya Gubareni, Michiel Hazewinkel, and Vladimir V. Kiričenko. "Lie Algebras and Dynkin Diagrams." *Algebras, Rings and Modules*. Dordrecht [u.a.: Kluwer, 2010. 1-35. Print.

[8] Humphreys, James E. *Introduction to Lie Algebras and Representation Theory*. New York: Springer-Verlag, 1972. 1-72. Print.

[9] Munkres, James R. *Topology*. Upper Saddle River, NJ: Prentice Hall, 2000. Print.

[10] Sethna, James P. "Order Parameters, Broken Symmetry, and Topology." *Lectures in the Sciences of Complexity*. Ed. Daniel L. . Stein and Lynn Nadel. Redwood City, CA: Addison-Wesley, 1989. 243-88. Print.